Topics in word complexity

Steven Widmer

Thèse de L'Université de Lyon
Délivrée par
L'Université Claude Bernard Lyon 1
Ecole Doctorale Informatique et Mathématiques
Institut Camille Jordan
Diplôme de Doctorat
30 novembre, 2010

Topics in Word Complexity

par
Steven Widmer

Directeur de thèse : Dr. Luca Zamboni and Dr. Amy Glen

Jury : Dr. Gwénaël Richomme
Dr. Sébastien Ferenczi
Dr. Juhani Karhumäki
Dr. Anna Frid
Dr. Luca Zamboni

Contents

List of Symbols

In the following table, letters that are not otherwise defined shall be:

n, m	non-negative integer
k	positive integer
s, t	integers
β	real number
α	positive irrational
u, v	finite word
x	finite or infinite word over $\mathcal{A}$
y	infinite word over $\mathcal{A}$
w	aperiodic word over $\mathcal{A}$

Symbol	Definition
$\mathbb{Z}$	the integers
$\mathbb{N}$	the natural numbers, non-negative integers
$\mathbb{Q}$	the rational numbers
$\mathbb{R}$	the real numbers
$\lvert\beta\rvert$	absolute value of β
$\lfloor\beta\rfloor$	greatest integer $\leq \beta$
$\lceil\beta\rceil$	least integer $\geq \beta$
$s \equiv (t \mod k)$	s is congruent to t modulo k
∎	end of a proof
$\lvert A\rvert$	cardinality of the finite set A
$\mathcal{A}$	finite alphabet
$\mathcal{A}^*$	free monoid generated by $\mathcal{A}$, the set of finite words over $\mathcal{A}$

Symbol	Definition
ε	identity element of $\mathcal{A}^*$, empty word
$\mathcal{A}^+$	free semigroup $\mathcal{A}^* \setminus \varepsilon$
$\mathcal{A}^{\mathbb{N}}$	set of (right) infinite words over $\mathcal{A}$
$\mathcal{A}^{\mathbb{Z}}$	set of biinfinite words over $\mathcal{A}$
$\mathcal{A}^{\infty}$	the set $\mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}$
$Alph(x)$	letters in $\mathcal{A}$ used to form x
$\sim$	reversal operation on $\mathcal{A}^*$
$-$	complement operation on $\mathcal{A}^*$, only if $\|\mathcal{A}\| = 2$
$\|u\|$	the number of letters in u, length of u
$\|u\|_a$	the number of occurrences of the letter a in u
$\mathcal{F}(x)$	set of all factors of x
$x[n]$	the n shift of x, $x_n x_{n+1} x_{n+2} \cdots$
$x[n,m]$	factor $x_n x_{n+1} \cdots x_m$ of x, $n \leq m$
$\mathcal{F}_x(n)$	set of factors of x of length n, where $\|x\| \geq n$
$\rho_x(n)$	size of $\mathcal{F}_x(n)$, where $\|x\| \geq n$
$\Psi(u)$	Parikh vector of u
$\Psi_x(n)$	set of all distinct Parikh vectors of factors of x of length n, where $\|x\| \geq n$
$u \sim_{ab} v$	u and v are abelian equivalent
$\mathcal{F}_x^{ab}(n)$	set of non-abelian equivalent factors of x of length n,, where $\|x\| \geq n$
$\rho_x^{ab}(n)$	size of $\mathcal{F}_x^{ab}(n)$, where $\|x\| \geq n$
T	Thue-Morse word
μ_T	Thue-Morse morphism
F_k	Fraenkel word over k letters
$F_y(P)$	set of pattern words in y from the k-pattern P
$p_y^*(k)$	maximal pattern complexity function
$p_{ab}^*(k)$	maximal abelian pattern complexity function
$(p_0\ p_1\ \cdots\ p_{n-1})$	permutation of the numbers 1, 2, ..., n

Symbol	Definition
π_w	infinite permutation associated with w
$\pi_w[n,m]$	subpermutation of π_w from n to m, only defined when $n \leq m$
$\mathrm{Perm}(w)$	set of subpermutations of π_w
$\mathrm{Perm}^w(n)$	set of subpermutations of π_w of length n
$\mathrm{Perm}^w_{ev}(n)$	length n subpermutations with even starting index
$\mathrm{Perm}^w_{odd}(n)$	length n subpermutations with odd starting index
$\tau_w(n)$	size of $\mathrm{Perm}^w(n)$
$[\alpha_0, \alpha_1, \alpha_2, \ldots]$	continued fraction expansion of α
$[\overline{\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n}]$	periodic continued fraction expansion
$[\alpha_0, \alpha_1, \ldots, \alpha_n, \overline{\alpha_{n+1}, \ldots, \alpha_{n+m}}]$	eventually periodic continued fraction expansion
$n!$	n factorial, $n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1$
$\binom{n}{m}$	$\frac{n!}{m!(n-m)!}$ if $m \leq n$, and 0 otherwise
$\binom{n}{m}^{[k+1]}$	coefficient of x^m in the expansion of $(1+x+x^2+\cdots+x^k)^n$
$l_k(n)$	number of compositions of n into k parts

Abstract - Résumé

The main topics of interest in this thesis will be two types of complexity, abelian complexity and permutation complexity. Abelian complexity has been investigated over the past decades. Permutation complexity is a relatively new type of word complexity which investigates lexicographical ordering of shifts of an aperiodic word.

We will investigate two topics in the area of abelian complexity. Firstly we will consider an abelian variation of maximal pattern complexity. Secondly we consider an upper bound for words with the C-balance property. In the area of permutation complexity, we compute the permutation complexity function for a number of words. A formula for the complexity of Thue-Morse word is established by studying patterns in subpermutations and the action of the Thue-Morse morphism on the subpermutations. We then give a method to calculate the complexity of the image of certain words under the doubling map. The permutation complexity function of the image of the Thue-Morse word under the doubling map and the image of a Sturmian word under the doubling map are established.

Les principaux sujets d'intérêt de cette thèse concerneront deux notions de la complexité d'un mot infini : la complexité abélienne et la complexité de permutation. La complexité abélienne a été étudiée durant les dernières décennies. La complexité de permutation est, elle, une forme de complexité des mots relativement nouvelle qui associe à chaque mot apériodique de manière naturelle une permutation infinie.

Nous nous pencherons sur deux sujets dans le domaine de la complexité abélienne. Dans un premier temps, nous nous intéresserons à une notion abélienne de la maximal pattern complexity définie par T. Kamae. Deuxièmement, nous analyserons une limite supérieure de cette complexité pour les mots $C-$équilibré.

Dans le domaine de la complexité de permutation des mots apériodiques binaires, nous établissons une formule pour la complexité de permutation du mot de Thue-Morse, conjecturée par Makarov, en étudiant la combinatoire des sous-permutations sous l'action du morphisme de Thue-Morse. Par la suite, nous donnons une méthode générale pour calculer la complexité de permutation de l'image de certains mots sous l'application du morphisme du doublement des lettres. Finalement, nous déterminons la complexité de

permutation de l'image du mot de Thue-Morse et d'un mot Sturmien sous l'application du morphisme du doublement des lettres.

Sommaire

Les thèmes principaux de recherche de cette thèse concernent diverses notions de la complexité des mots infinis : la complexité abélienne, une variation abélienne de la complexité de Kamae, et la complexité de permutation.

Nous commençons par un étude combinatoire des mots $C-$équilibrés. Il sera démontré que les mots épisturmiens équilibrés ayant des fréquences de lettre distinctes obéissent à la conjecture de Fraenkel. Il sera aussi démontré qu'un mot récurrent équilibré w faiblement riche possédant au moins 3 lettres est en fait un mot périodique épisturmien, et si les fréquences des lettres sont distinctes, alors w obéira à la conjecture de Fraenkel.

Dans le domaine de la complexité abélienne, nous nous intéresserons à une variation abélienne de la maximal pattern complexity définie par T. Kamae. Il sera démontré que cette notion de complexité donne une classification des mots binaires apériodiques récurrents. Nous analyserons ensuite le lien entre la complexité abélienne et les mots $C-$équilibrés. Nous établissons une méthode générale permettant de calculer la valeur maximale de la complexité abélienne d'un mot récurrent infini équilibré. Nous trouvons ensuite une limite supérieure pour la complexité abélienne d'un mot $C-$équilibré et nous donnons des exemples de mots atteignant cette limite supérieure.

La majeure partie de la recherche présentée dans cette thèse concerne la complexité de permutation. La complexité de permutation prendra en compte l'ordre lexicographique des shifts d'un mot infini. D'abord, nous établissons une formule pour la complexité de permutation du mot de Thue-Morse, conjecturée par Makarov. Puis, nous donnons une méthode générale pour calculer la complexité de permutation d'un mot binaire apériodique sous l'application du morphisme du doublement des lettres. Finalement, nous calculons la fonction de la complexité de permutation de l'image d'un mot sturmien et du mot de Thue-Morse sous l'application du morphisme du doublement de lettre.

Acknowledgments

First and foremost, I would like to thank my wife, Alisa. Your love and support through this trying time has been unwavering, and for this I can never thank you enough. The antics of Baxter have also brought many smiles over these years. Together we make a great team.

I would also like to thank my thesis advisers, Luca Zamboni and Amy Glen. They introduced me to the study of words and gave me the ideas for this thesis. Thank you also for all of the time you spent reading my work and keeping me moving forward.

I wish to extend thanks to Mark Dukes for helping me keep things in perspective during all of our discussions in Iceland. Thanks also to Einar Steingrímsson for making me feel welcome in Iceland, and for all of his help at the end. A warm thanks also to all of my friends and family for being so supportive.

I would finally like to thank the mathematics faculty and staff of Texas State University, Baylor University, and the University of North Texas. Their teaching and support over the years inspired me to pursue this life of mathematics.

Financial support for this thesis was provided by grant number 090038012 from the Icelandic Research Fund.

Chapter 1

Introduction

Combinatorics on words is a relatively new field of study in discrete mathematics. The study of combinatorics on words has grown independently in different areas of mathematics; such as number theory, group theory, differential geometry, and probability. The applications of combinatorics on words have extended to various fields such as theoretical computer science, dynamical systems, biology, and linguistics. Many books have been written on the subject of combinatorics, but the best known may be the collected works written by a group of researchers under the nom de plume of Lothaire [27, 28, 29]. For a survey of the beginnings of combinatorics on words and the inter-working of the early discoveries see [9].

The original study of words has been attributed to Axel Thue (1863-1922) with the investigation of repetitions in words. Thue is credited with the discovery of the first square-free infinite word in [43]. In Thue's second paper on the subject, [44], he introduced what is now generally referred to as the Thue-Morse word, which he proved was overlap-free. After the work of Thue there were relatively few researchers working with words in the early twentieth century. A greater number of researchers gave interest to words in the 1950's in connection with areas of discrete mathematics. Interest in combinatorics on words grew rapidly after the publication of the publication of Lothaire's book ([27]).

Possibly the most studied class of words is the *Sturmian words*. The history of Sturmian words dates back to the 1700's, with the first formal investigation by Morse and Hedlund [37] in 1940. There have been many surveys written on the topic of Sturmian words, with [8, 10] as examples.

A natural extension of studying words is word complexity. For example, the Sturmian words are the class of aperiodic binary words with minimal factor complexity. Many different notions of word complexity have been introduced over the years, such as factor complexity, abelian complexity, palindromic complexity, pattern complexity, and recently permutation complexity. The main goal of this thesis is to investigate topics related to different types of complexity, namely abelian complexity and permutation complexity, as well as maximal abelian pattern complexity.

1.1 Thesis Outline

In Chapter 2 we give some background information necessary for the later chapters. Preliminary notation and terminology are defined, and some classic results in combinatorics on words are given.

We start by looking at a class of balanced words. In [38] it is shown that balanced episturmian words with distinct letter frequencies obey Fraenkel's conjecture. In Chapter 3, we show that a recurrent balanced weakly rich word w with at least 3 letters is in fact a periodic episturmian word, and if w is a word with distinct letter frequencies then w will obey Fraenkel's conjecture.

In Chapter 4, we investigate abelian complexity of words. First we will consider a new type of complexity related to abelian complexity, maximal abelian pattern complexity. It will be shown that maximal abelian pattern complexity will classify the recurrent aperiodic binary words. We then investigate how abelian complexity and the C-balance property are related. We first develop a method to calculate the maximal value for abelian complexity of an infinite balanced recurrent word. We then find an upper bound for the abelian complexity of an infinite C-balanced word. It is unknown if this upper bound is the least upper bound, but we give some examples of words over a relatively small alphabet which achieve the given upper bound.

In Chapter 5, we investigate a relatively new type of word complexity called permutation complexity. Permutation complexity will consider the lexicographic ordering of shifts of an infinite word. For this reason permutation complexity is only defined for aperiodic words, because no two shifts of an aperiodic word will be identical. We will give the permutation complexity function of the Thue-Morse word. We will also give a method to calculate the permutation complexity of an infinite word which is the image of an uniformly recurrent aperiodic binary word under the doubling map. We then give a formula for the permutation complexity of the image of a Sturmian word under the doubling map, and give the permutation complexity function of the image of the Thue-Morse word under the doubling map.

1.2 Future Research

From the different areas of research in this thesis, there seem to be many possibilities for future research topics. For example, one avenue of further research would be to find bounds for maximal abelian pattern complexity of binary aperiodic words that are not necessarily recurrent. Another direction is to find bounds for maximal abelian pattern complexity of words on an alphabet of size $k \geq 3$.

Early work in this thesis was focused on the link between abelian complexity and the notion of C-balance. The direction of this thesis changed before many questions could be answered. The main question related to abelian complexity is if the upper bound given in Section 4.2.2 is the least upper bound for abelian complexity of C-balanced words. Some examples of balanced recurrent words have been found which attain the upper bound for abelian complexity, but no method has been developed to construct balanced recurrent words over an alphabet with 7 or more letters.

The majority of the later research in this thesis was in the area of permutation complexity. The permutation complexity has been calculated for some well-known words, but many other classes of words have not been investigated. One avenue of future research would be to calculate the permutation complexity of some other well-known words. Additionally, early research has been focused on binary words. Another research direction is to investigate the permutation complexity of words over a k-letter alphabet, where $k \geq 3$.

The action of the Thue-Morse morphism on subpermutations played a key role in the investigation of the permutation complexity of the Thue-Morse word. The action of the doubling map on subpermutations played a key role as well in the investigation of the permutation complexity of words which are the image of an uniformly recurrent aperiodic binary word under the doubling map. One direction of future research is to generalize permutation complexity results to aperiodic words which are the image of an aperiodic binary word under a morphism. This could be an aperiodic word which is the fixed points of a morphism, or maybe the image of any word under some fixed morphism. An early thought leads us to believe that the investigation of the action of morphisms on subpermutations will answer many open questions about permutation complexity.

A method has been developed by Makarov to find the permutations of length n generated by binary words, but no method has been developed to determine if a given permutation of length n can be attained by a binary word. A natural question to ask is, are there permutations which are not attainable from a binary word? The answer to this question is yes, for example the permutation $(2\,1\,3\,4)$ will never occur in a binary word. Some additional research topics are

- Is there a classification of the unattainable permutations?
- What are the necessary and sufficient conditions for a permutation to be attainable by a binary word, or a word over a general k-letter alphabet?
- Are there permutations that are not attainable from a word on a k-letter alphabet?

- How long are unattainable permutations on a k-letter alphabet?

Chapter 2

Background

In this chapter, we present some preliminary definitions and results about words and morphisms. We will also introduce the notion of word complexity and give some examples and results. We will finish the chapter by introducing some well-known words and results which will be used in this writing.

2.1 Words

A *word* is a finite, (right) infinite, or biinfinite sequence of symbols taken from a finite non-empty set, $\mathcal{A}$, called an *alphabet*. The elements of $\mathcal{A}$ are called *letters*. For any word w over the alphabet $\mathcal{A}$, denote $Alph(w) = \mathcal{B} \subseteq \mathcal{A}$ to be the subset of letters in $\mathcal{A}$ that are used to form the word w. In what follows, if w is a word over $\mathcal{A}$, then $Alph(w) = \mathcal{A}$ unless otherwise noted.

2.1.1 Finite Words

A *finite word* over $\mathcal{A}$ is an element of the free monoid $\mathcal{A}^*$, generated from $\mathcal{A}$ by concatenation of the letters in $\mathcal{A}$ and is represented by juxtaposition of letters and words. For example, if $u = \text{POWER}$ and $v = \text{NAP}$ the concatenation of u and v is $uv = \text{POWERNAP}$. It should be noted that this operation is not commutative, because $vu = \text{NAPPOWER}$. Concatenation is an associative operation because $(uv)w = u(vw) = uvw$ for all $u, v, w \in \mathcal{A}^*$. The identity element ε of $\mathcal{A}^*$ is called the *empty word*, and the free semigroup over $\mathcal{A}$ is defined by $\mathcal{A}^+ = \mathcal{A}^* \setminus \varepsilon$.

A finite word $u \in \mathcal{A}^*$ has the form $u = a_1 a_2 \ldots a_n$ with each $a_i \in \mathcal{A}$ and $n \geq 0$, and the *length* of u is the number of symbols in the sequence and is denoted $|u| = n$ (if $n = 0$, then $u = \varepsilon$ and $|u| = |\varepsilon| = 0$).

Denote $\sim$ to be the *reversal operation* on $\mathcal{A}^*$. Let $\tilde{\varepsilon} = \varepsilon$ and for $u = u_1 u_2 \ldots u_n \in$

$\mathcal{A}^+$ the *reversal* of u is denoted by $\tilde{u} = u_n \ldots u_2 u_1$. For example if $u =$ PEANUT and $v =$ RACECAR, then $\tilde{u} =$ TUNAEP and $\tilde{v} =$ RACECAR. A word $u \in \mathcal{A}^*$ is said to be a *palindrome* if $u = \tilde{u}$, by definition the empty-word ε is a palindrome. The word v in the previous example is a palindrome, as are the English words KAYAK, ROTATOR, and AIBOHPHOBIA (which is the fear of palindromes).

2.1.2 Infinite Words

A *(right) infinite word* over $\mathcal{A}$ is a list of letters in $\mathcal{A}$ indexed by $\mathbb{N}$. An infinite word has the form $\omega = \omega_0\omega_1\omega_2\ldots$ with each $\omega_i \in \mathcal{A}$. The set of all infinite words over $\mathcal{A}$ is denoted $\mathcal{A}^{\mathbb{N}}$, and let $\mathcal{A}^\infty = \mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}$. A *biinfinite word* is a list of letters in $\mathcal{A}$ indexed by $\mathbb{Z}$ and the set of all biinfinite words is denoted $\mathcal{A}^{\mathbb{Z}}$. In what follows, all infinite words will be considered to be right infinite words unless otherwise noted.

A word $\omega \in \mathcal{A}^{\mathbb{N}}$ is said to be *periodic* of period p if p is the least integer so that $\omega_i = \omega_{i+p}$ for each $i \in \mathbb{N}$. A word $\omega \in \mathcal{A}^{\mathbb{N}}$ is said to be *eventually periodic* of period p if p is the least integer so that for some $N \in \mathbb{N}$, $\omega_i = \omega_{i+p}$ for each $i > N$. It should be noted that all periodic words are eventually periodic. A word $\omega \in \mathcal{A}^{\mathbb{N}}$ is said to be *aperiodic* if it is not periodic or eventually periodic.

2.1.3 Factors

A finite word x is a *factor* of $w \in \mathcal{A}^\infty$ if $w = uxv$ for some $u \in \mathcal{A}^*$ and $v \in \mathcal{A}^\infty$. The word x is a called a *prefix* of ω if $u = \varepsilon$, and is called a *proper prefix* of ω if $v \neq \varepsilon$. The word x is a called a *suffix* of ω if $v = \varepsilon$, and is called a *proper suffix* of ω if $u \neq \varepsilon$.

For any word $\omega \in \mathcal{A}^\infty$, define $\mathcal{F}(\omega)$ to be the set of all factors of ω and $\mathcal{F}_\omega(n)$ to be the set of all factors of ω of length n where $|\omega| \geq n$.

A factor u of a word ω is said to be *right special* (resp. *left special*) in ω if there are at least two distinct letters a, b so that ua and ub (resp. au and bu) are also factors of ω. A factor that is both right and left special is called *bispecial*.

The infinite word $\omega \in \mathcal{A}^{\mathbb{N}}$ is said to be *recurrent* if for any prefix p of ω there exists a prefix q of ω so that $q = pvp$ for some $v \in \mathcal{A}^*$. Equivalently, a word ω is recurrent if each factor of ω occurs infinitely often in ω. The word $\omega \in \mathcal{A}^{\mathbb{N}}$ is *uniformly recurrent* if each factor occurs infinitely often with bounded gaps. Thus if ω is uniformly recurrent, for each integer $n > 0$ there is a positive integer N so that for each factor v of ω with $|v| = N$, $\mathcal{F}_\omega(n) \subset \mathcal{F}(v)$.

2.1.4 Lexicographic Order

Suppose the letters of $\mathcal{A}$ are ordered with a linear order $<$. Then all the elements of $\mathcal{A}^*$ can be linearly ordered by the *lexicographic order*, $<$. For words $u, v \in \mathcal{A}^*$, we say $u < v$ if and only if either u is a proper prefix of v or $u = xau'$ and $v = xbv'$, for $x, u', v' \in \mathcal{A}^*$ and $a, b \in \mathcal{A}$ with $a < b$. To see an example of this ordering, the lexicographic order is the ordering used in a dictionary or phone book with the intuitive ordering $a < b < \cdots < y < z$ on the letters. For a more explicit example consider the English words

$$\text{RACE} < \text{RACECAR} < \text{RACING},$$

where the larger letters show where one word is greater than the previous word in the list.

This ordering can be naturally extended to elements of $\mathcal{A}^{\mathbb{N}}$. Let $u, v \in \mathcal{A}^{\mathbb{N}}$, where $u = u_1u_2\cdots$ and $v = v_1v_2\cdots$, we say $u < v$ if and only if there is some $i \geq 0$ so that $u_i < v_i$ and $u_j = v_j$ for each $0 \leq j < i$.

2.1.5 Complement of a Word

Supposing $|\mathcal{A}| = 2$, or $\mathcal{A} = \{0, 1\}$ is a *binary alphabet*, we can define another operation on $\mathcal{A}^\infty$. Denote $-$ to be the *complement operation* on $\mathcal{A}^\infty$. Let $\overline{a}$ denote the *complement* of $a \in \mathcal{A}$, that is $\overline{0} = 1$ and $\overline{1} = 0$. For example, if $u \in \mathcal{A}^*$ with $u = 101001$ then $\overline{u} = 010110$. If $\omega = \omega_1\omega_2\omega_3\cdots \in \mathcal{A}^\infty$, the *complement* of ω is defined to be the word composed of the complement of the letters in ω, that is $\overline{\omega} = \overline{(\omega_1\omega_2\omega_3\cdots)} = \overline{\omega}_1\overline{\omega}_2\overline{\omega}_3\cdots$. For a word $\omega \in \mathcal{A}^\infty$ we say the set of factors of ω is *closed under complementation* if for each $u \in \mathcal{F}(\omega)$ then $\overline{u} \in \mathcal{F}(\omega)$.

2.2 Morphisms

Let $\mathcal{A}$ and $\mathcal{B}$ be two finite alphabets. A map $\varphi : \mathcal{A}^* \mapsto \mathcal{B}^*$ so that $\varphi(uv) = \varphi(u)\varphi(v)$ for any $u, v \in \mathcal{A}^*$ is called a *morphism* of $\mathcal{A}^*$ into $\mathcal{B}^*$, and φ is defined by the image of each letter in $\mathcal{A}$. Note that $\varphi(\varepsilon) = \varepsilon$.

Example Let $\mathcal{A} = \{a, f, s, t\}$, $\mathcal{B} = \{a, e, f, i, r\}$, and $\varphi : \mathcal{A}^* \mapsto \mathcal{B}^*$ be defined by

$$\varphi : \begin{cases} a \mapsto a \\ f \mapsto fer \\ s \mapsto rr \\ t \mapsto i. \end{cases}$$

Then we have $\varphi(\text{FAST}) = \text{FERARRI}$.

A *morphism on* $\mathcal{A}$ is a morphism from $\mathcal{A}^*$ into $\mathcal{A}^*$, also called an *endomorphism* of $\mathcal{A}$. A morphism φ is said to be *non-erasing* if the image of any non-empty word is not empty. All morphisms contained in this paper will be considered to be non-erasing unless otherwise noted.

A morphism on $\mathcal{A}$ worth noting here is the *doubling map*, d, defined by

$$d : a \mapsto aa$$

for each $a \in \mathcal{A}$. The doubling map will be discussed more in Section 5.3.

If there is a positive integer N so that $|\varphi(a)| = N$ for each $a \in \mathcal{A}$, then φ is called an N-*uniform morphism*. The morphism μ_T on $\mathcal{A} = \{0, 1\}$ defined by

$$\mu_T : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

is a 2-uniform morphism.

The action of a morphism φ on $\mathcal{A}$ can naturally be extended from $\mathcal{A}^*$ to $\mathcal{A}^{\mathbb{N}}$. For any $\omega = \omega_0\omega_1\omega_2 \ldots \in \mathcal{A}^{\mathbb{N}}$, we define $\varphi(\omega) = \varphi(\omega_0)\varphi(\omega_1)\varphi(w_2)\ldots$ as in the case for words in $\mathcal{A}^*$. We say that a word ω is a *fixed point* of the morphism φ if $\varphi(\omega) = \omega$.

The n-*iteration* of a morphism φ on some $a \in \mathcal{A}$ is denoted $\varphi^n(a)$ and is defined by

$$\varphi^0(a) = a, \quad \varphi^n(a) = \varphi(\varphi^{n-1}(a)) \quad \text{for } n \geq 1.$$

If φ is a morphism on $\mathcal{A}$ and if $\varphi(a) = au$ for some $a \in \mathcal{A}$ and $u \in \mathcal{A}^+$, φ is said to be *prolongable* on a. If φ is a morphism on $\mathcal{A}$ that is prolongable on some $a \in \mathcal{A}$, then $\varphi^n(a)$ is a proper prefix of $\varphi^{n+1}(a)$ for each $n \in \mathbb{N}$. The limit of the sequence $\{\varphi^n(a)\}_{n\in\mathbb{N}}$ will be the unique infinite word

$$\omega = \lim_{n\to\infty} \varphi^n(a) = \varphi^\infty(a) = au\varphi(u)\varphi^2(u)\cdots$$

where ω is a fixed point of φ, and we say that ω is *generated* by φ. A morphism φ on $\mathcal{A}$ is said to be *primitive* if there is a positive integer k so that for each $a \in \mathcal{A}$, $\varphi^k(a)$ contains all the letters of $\mathcal{A}$.

2.3 Word Complexity

The study of combinatorics on words eventually led to the investigation of how complex a word can be. There have been different ways to define the complexity of a word. A natural way to define the complexity of a word ω is to count the number of distinct factors of ω of each length, which is known as the factor complexity of a word. Recently a generalization of factor complexity led to counting the number of pairwise non-abelian equivalent factors, formally compiled in [40], known as the abelian complexity of a word. Other ways to measure the complexity of a word have been developed over the years, and many well-known classes of words have been classified using different definitions of complexity. The Sturmian words have been classified as the words having minimal factor complexity without being periodic, but they also have minimal abelian complexity.

A common question for any notion of complexity is to find the complexity of some well-known words. For example a classification of Sturmian words has been given by palindromic complexity, introduced by Droubay and Pirillo in [16]. For an infinite word ω, let $h_\omega(n)$ be the number of palindromic factors of length n of ω.

Theorem 2.3.1 ([16]) *Let ω be an infinite word. Then ω is Sturmian if and only if*

$$h_\omega(n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 2 & \text{otherwise} \end{cases}$$

In this section we are concerned with some different definitions of word complexity. In Section 2.3.1 we present some fundamental results of studies in factor complexity. Some of the research of this author has been concerned with abelian complexity and pattern complexity, so these types of complexity are considered in Sections 2.3.2 and 2.3.3, respectively. Permutation complexity is left out of this section, but is investigated in Chapter 5.

2.3.1 Factor Complexity

For a word $\omega \in \mathcal{A}^\infty$, recall from Section 2.1.3, $\mathcal{F}(\omega)$ is the set of all factors of a word ω, and $\mathcal{F}_\omega(n)$ is the set of all factors of ω of length n, where $|\omega| \geq n$. We are now ready to define the notion of factor complexity. Define the function ρ_ω by

$$\rho_\omega(n) = |\mathcal{F}_\omega(n)| .$$

The function $\rho_\omega : \mathbb{N} \mapsto \mathbb{N}$ is called the *factor complexity function*, or *subword complexity function*, of ω and it counts the number of distinct factors of ω for each length.

A natural investigation led to the possible bounds of factor complexity for words. Early work in the area of factor complexity (see [13, 36]) characterized the biinfinite (resp. right infinite) words with bounded factor complexity as the periodic (resp. eventually periodic) words.

Proposition 2.3.2 [13] *An infinite word $x \in \mathcal{A}^{\mathbb{N}}$ is eventually periodic if and only if $\rho_x(n) \leq n$ for some $n \geq 1$. A biinfinite word $y \in \mathcal{A}^{\mathbb{Z}}$ is periodic if and only if $\rho_y(n) \leq n$ for some $n \geq 1$.*

Therefore periodic and eventually periodic words have bounded factor complexity. For a word $\omega \in \mathcal{A}^{\mathbb{N}}$, where $|\mathcal{A}| = k$, the maximal value for the factor complexity for a length n will be k^n. Thus for an aperiodic infinite word ω, the bounds for factor complexity are

$$n + 1 \leq \rho_\omega(n) \leq k^n$$

for all $n \geq 1$.

One consideration to factor complexity are words with maximal factor complexity. Suppose $|\mathcal{A}| = k$, $k \geq 2$, and fix an integer $n \geq 1$. A trivial way to construct a finite word $u \in \mathcal{A}^*$ with maximal factor complexity $\rho_u(n) = k^n$ is to concatenate all k^n distinct words of length n. By this construction, the word u has maximal factor complexity and $|u| = n \cdot k^n$, but many words of length n appear as factors of u multiple times. The minimal length cyclic words to contain all k^n distinct words of length n are the de Bruijn cycles ([47, 42]), $B(n, k)$. The length of a de Bruijn cycle $B(k, n)$ is k^n, and each word of length n appears as a factor in $B(k, n)$ exactly once. There are a total of $k^{-n}(k!)^{n^{k-1}}$ such cycles, so $B(k, n)$ is not unique. The de Bruijn cycles will be discussed further in Section 2.4.1. An infinite word over $\mathcal{A}$ with maximal factor complexity can be created by concatenating all words of length n, for each $n \geq 1$. For example, the binary word

$$C = 01\,00011011\,000001010011100101110111\cdots$$

contains all binary words of length n as a factor, and thus $\rho_C(n) = 2^n$ for each $n \geq 1$.

Aperiodic words with minimal factor complexity are the *Sturmian words*, first studied in [36], having $n + 1$ distinct factors of length n, for each $n \geq 1$. First we note that Sturmian words are binary words since they have 2 distinct factors of length 1. Then for each length $n \geq 1$, there is exactly one factor of length n that can be followed by more than one letter, or else the number of factors of length $n + 1$ will not be $n + 2$. Thus for each length, $n \geq 1$, Sturmian words have exactly one right special factor (as well as exactly one left special factor) of length n. Sturmian words have been studied extensively and will be discussed further in Section 2.4.2.

Extending the idea of words with minimal factor complexity to a general k-letter alphabet can be done by limiting the number of left and right special factors. The first natural generalization in this direction was given by Arnoux and Rauzy [5] for the case of $k = 3$. An *Arnoux-Rauzy word* has exactly one left special factor and one right special factor of each length, and for each right (resp. left) special factor u of ω, ua (resp. au) is a factor of ω for each $a \in \mathcal{A}$. Thus, an Arnoux-Rauzy word over $\mathcal{A}$, with $|\mathcal{A}| = k$, has k factors of length 1, and then an additional $k-1$ factors for each additional length. The Sturmian words are precisely the Arnoux-Rauzy words over a 2-letter alphabet. Arnoux-Rauzy words have a factor complexity function with linear growth with respect to factor length, and for an Arnoux-Rauzy word $x \in \mathcal{A}^{\mathbb{N}}$, $\rho_x(n) = (k-1)n + 1$. Another generalization of words with linear factor complexity are the episturmian words which will be discussed further in Section 2.4.3.

A nice property of Sturmian words, Arnoux-Rauzy words, and episturmian words is that their sets of factors are closed under reversal. That is, if $\omega \in \mathcal{A}^{\mathbb{N}}$ is such a word and $u \in \mathcal{F}(\omega)$, then $\tilde{u} \in \mathcal{F}(\omega)$. Factor complexity of other well-known words has also been investigated. For example, the factor complexity of the Thue-Morse word ([11, 31]) and generalized Thue-Morse words ([46]) have been found to increase with linear bounds, with respect to factor length.

2.3.2 Abelian Complexity

A natural extension of factor complexity is abelian complexity. In the sense that abelian group is a group with a commutative operation, abelian complexity will count the number of occurrences of letters in factors. Much of the content in this section is taken from [40]. We define abelian complexity as follows.

For any finite word $u \in \mathcal{A}^*$ and for each $a \in \mathcal{A}$, let $|u|_a$ denote the number of occurrences of the letter a in the word u. Any two words u and v in $\mathcal{A}^*$ are said to be *abelian equivalent*, denoted $u \sim_{ab} v$, if $|u|_a = |v|_a$ for each $a \in \mathcal{A}$, and it is readily verified that $\sim_{ab}$ defines an equivalence relation on $\mathcal{A}^*$. For example,

ELEVEN PLUS TWO $\sim_{ab}$ TWELVE PLUS ONE.

The *frequency* of the letter $a \in \mathcal{A}$ in the word $u \in \mathcal{A}^*$ is defined to be $|u|_a / |u|$. The *frequency* of the letter $a \in \mathcal{A}$ in an infinite word $\omega \in \mathcal{A}^{\mathbb{N}}$ is defined to be

$$\lim_{n\to\infty} \frac{|p_n|_a}{n},$$

if the limit exists, where p_n is the prefix of ω of length n.

We can then expand on the definition of factor complexity and define

$$\mathcal{F}_\omega^{ab}(n) = \mathcal{F}_\omega(n) / \sim_{ab}$$

to be the set of non-abelian equivalent factors of ω of length n, and let

$$\rho_\omega^{ab}(n) = \left|\mathcal{F}_\omega^{ab}(n)\right|.$$

The function $\rho_\omega^{ab} : \mathbb{N} \mapsto \mathbb{N}$ is called the *abelian complexity function* of ω and it counts the number of non-abelian equivalent factors of ω of length n.

The alphabet $\mathcal{A}$ will generally consist of the numbers $\{0, 1, 2, \ldots k-1\}$. Then for each $u \in \mathcal{A}^*$ we can naturally define the *Parikh vector* associated to u by

$$\Psi(u) = (|u|_0, |u|_1, \ldots, |u|_{k-1}).$$

For example, if $\mathcal{A} = \{0, 1, 2\}$ and $u = 0102010 \in \mathcal{A}^*$ then $\Psi(u) = (4, 2, 1)$.

Extending this notion to an infinite word $\omega \in \mathcal{A}^{\mathbb{N}}$, we define

$$\Psi_\omega(n) = \{\ \Psi(u)\ \mid\ u \in \mathcal{F}_\omega(n)\ \}$$

to be the set of all distinct Parikh vectors of factors of ω of length n. It is also useful to note that $\rho^{ab}_\omega(n) = |\Psi_\omega(n)|$.

Let a and b be letters in $\mathcal{A} = \{0, 1, \ldots, k-1\}$ and let $u \in \mathcal{A}^*$. If $a = b$ then $\Psi(au) = \Psi(ub)$. When $a \neq b$, the vector $V = \Psi(au) - \Psi(ub)$ will be the vector so that $V_{a+1} = 1$, $V_{b+1} = -1$ and for each other $i \in \mathcal{A}$, $V_i = 0$. This shows how Parikh vectors change when considering two successive factors of the same length of a word ω. These observations directly imply the following fact.

Fact 2.3.3 ([40]) *If an infinite word ω has two factors u and v of same length n for which the i-th entry of the Parikh vector are p and $p+c$ respectively for some p and $c > 0$, then for all $l = 0, \ldots, c$, there exist factors u_l of ω whose i-th entry is $p + l$.*

Using the notation from Fact 2.3.3, it should be clear that $\rho^{ab}_\omega(n) \geq c + 1$. For an infinite word $\omega \in \mathcal{A}^N$ it is said that ω is *C-balanced* (where C is a positive integer) if for all factors u and v with $|u| = |v|$, $||u|_a - |v|_a| \leq C$ for each $a \in \mathcal{A}$. It should be clear that if a word is C-balanced then it is $(C+1)$-balanced. If a word ω is 1-balanced, we say that the word ω is *balanced*. An initial link between C-balance and abelian complexity can be seen in the next lemma.

Lemma 2.3.4 ([40]) *For a word $\omega \in \mathcal{A}^{\mathbb{Z}} \cup \mathcal{A}^{\mathbb{N}}$, the abelian complexity function ρ^{ab}_ω is bounded if and only if ω is C-balanced for some positive integer C.*

Proof If ρ^{ab}_ω is bounded by some K, then it is easy to see ω is $(K-1)$-balanced. Conversely, if ω is C-balanced then for any positive integer n the Parikh vectors of factors of length n can take on at most $C+1$ distinct values. Thus $\rho^{ab}_\omega(n) \leq (C+1)^{|\mathcal{A}|}$. ∎

A natural question with regards to abelian complexity of a word ω is, what are the maximal and minimal values for the abelian complexity of ω? Minimal values for abelian complexity are achieved by periodic words.

Lemma 2.3.5 ([13], Remark 4.07) *Let $\omega \in \mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\mathbb{Z}}$. Then ω is periodic of period p if and only if $\rho^{ab}_\omega(p) = 1$.*

The maximum number of elements of $\Psi_\omega(n)$ can be calculated with relative ease. Each element in $\Psi_\omega(n)$ is a k-tuple $\Psi(u) = (i_0, i_1, \ldots, i_{k-1})$ where $i_0 + i_1 + \ldots + i_{k-1} = n$. Thus, the maximal size of $\Psi_\omega(n)$ is the same as the number of ways to write n as the sum of k non-negative integers. This value (see [49]), $l_k(n)$, is called the number of compositions of n into k parts and is given by the binomial coefficient

$$l_k(n) = \binom{n+k-1}{k-1}.$$

Thus, for an infinite word ω over a k-letter alphabet $\mathcal{A}$

$$\rho^{ab}_\omega(n) \leq l_k(n),$$

for each positive integer n. A word with maximal factor complexity will contain all words of length n as factors, and thus a word with maximal factor complexity will also have maximal abelian complexity. One topic of this thesis will be to investigate the link between the notions of abelian complexity and C-balance, namely an optimal upper bound for abelian complexity for a word that is C-balanced. This topic will be covered in more detail in Section 4.2.

As introduced in the previous section (2.3.1), the Sturmian words are the class of aperiodic words with minimal factor complexity. An equivalent definition of Sturmian words is that they are also the class of aperiodic balanced binary words (see Theorem 2.4.2). The following theorem gives another classification of the Sturmian words.

Theorem 2.3.6 ([13]) *Let $\omega \in \{0,1\}^{\mathbb{N}}$. Then ω is Sturmian if and only if $\rho^{ab}_\omega(n) = 2$ for all $n \geq 1$.*

Thus the Sturmian words are the aperiodic words with minimal abelian complexity.

The abelian complexity of the Thue-Morse word has also been calculated. Moreover, the class of words having the same abelian complexity as the Thue-Morse word has been characterized. This class of words is described as the image of any aperiodic binary word under the Thue-Morse morphism, μ_T, defined as

$$\mu_T : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10. \end{cases}$$

Theorem 2.3.7 ([40]) *The abelian complexity of an aperiodic binary word ω is*

$$\rho^{ab}_\omega(n) = \begin{cases} 2 & \text{for } n \text{ odd} \\ 3 & \text{for } n \neq 0 \text{ even} \end{cases}$$

if and only if there exists a word ω' so that $\omega = \mu_T(\omega')$, $\omega = 0\mu_T(\omega')$, or $\omega = 1\mu_T(\omega')$.

Since the Thue-Morse word, T, clearly satisfies the conditions of Theorem 2.3.7, we have a formula to calculate the abelian complexity of the Thue-Morse word.

So far we have seen some examples of words with linear factor complexity functions having low abelian complexity. This is not always the case. It is possible to construct words with an exponential factor complexity function, yet have bounded abelian complexity. Take for example the word C defined previously,

$$C = 01\,00011011\,000001010011100101110111\cdots$$

and consider the word $C' = \mu_T(C)$. Theorem 2.3.7 says that this word will have bounded abelian complexity ($\rho^{ab}_{C'}(n) \leq 3$), yet $\rho_{C'}(2n) = 2^n$ has exponential growth.

It is also possible to construct a word with maximal abelian complexity and linear factor complexity. This example is given in [40]. Let f and g be morphisms defined by $f(a) = abc$, $f(b) = bbb$, $f(c) = ccc$, $g(a) = 0 = g(c)$, and $g(b) = 1$. It should be clear that f is prolongable on a, and thus the word $u = f^{\infty}(a)$ is a fixed point of f. The word $\omega = g(u) = g(f^{\infty}(a))$ is the word

$$\omega = 0 \prod_{i \geq 0} 1^{3^i} 0^{3^i} = 010111000\cdots .$$

It is readily verified that $\rho^{ab}_{\omega}(n)$ is maximal for each $n \geq 1$, and thus not bounded. It turns out that this word ω is an automatic sequence, and thus it has linear factor complexity (see [3], Theorems 6.3.2 and 10.3.1). An *automatic sequence* is defined as follows. Starting with a morphism f on the alphabet $\mathcal{B}$, where f is prolongable on some $b \in \mathcal{B}$, and a projection map $g : \mathcal{B} \mapsto \mathcal{A}$. If u is a fixed point of f, then $g(u)$ is an automatic sequence over $\mathcal{A}$. An automatic sequence can also be seen as an word generated by a finite automaton.

Inspired by the result of Sturmian words having constant abelian complexity, G. Rauzy asked if there existed an infinite word ω so that $\rho^{ab}_{\omega}(n) = 3$ for all $n \geq 0$ ([39]). In general, it is possible to create a word with constant abelian complexity. For example (see [40]), let $k \geq 3$, and let s be a Sturmian word on the alphabet $\{0, 1\}$. Then the word $\omega' = (k-1)(k-2)\ldots 2s$ will have $\rho^{ab}_{\omega'}(n) = k$ for all $n \geq 1$. The case where $k = 3$ gives an answer to Rauzy's question, but it is a trivial answer since ω' is not recurrent. An example of an infinite recurrent word ω with $\rho^{ab}_{\omega} = 3$ is a word ω, where ω is the image of a Sturmian word s on $\{a, b\}$ under the morphism φ where $\varphi(a) = 012$ and $\varphi(b) = 021$, ([40]). It has recently been shown (see [14]) that for $k \geq 4$, there is no recurrent word over a k letter alphabet with constant abelian complexity.

2.3.3 Pattern Complexity

The idea of maximal pattern complexity was introduced by T. Kamae ([26]). The pattern complexity function is defined as follows. Let k be a positive integer. A *k-pattern* P is a sequence of k integers with

$$0 = P_0 < P_1 < \cdots < P_{k-1}$$

A $(k+1)$-pattern P' is called an *immediate extension* of a k-pattern P if $P'(i) = P(i)$ for $0 \le i \le (k-1)$, and we call this P an *immediate restriction* of P'.

Now let ω be an infinite word over the finite alphabet $\mathcal{A}$. For each k-pattern P define

$$F_\omega(P) = \left\{ \ \omega_{n+P_0}\omega_{n+P_1}\cdots\omega_{n+P_{k-1}} \ \mid n \ge 0 \ \right\}$$

and let p^*_ω be defined as

$$p^*_\omega(k) = \sup_P |F_\omega(P)|$$

where the supremum is taken over all k-patterns P. The function $p^*_\omega : \mathbb{N} \mapsto \mathbb{N}$ is called the *maximal pattern complexity function*. A k-pattern P will *attain* $p^*_\omega(k)$ if $|F_\omega(P)| = p^*_\omega(k)$. The maximal value for the pattern complexity of a k-pattern will be $|\mathcal{A}|^k$, since for each of the k windows there are $|\mathcal{A}|$ possible letters to choose from. Thus for an infinite word ω over $\mathcal{A}$, we have $p^*_\omega(k) \le |\mathcal{A}|^k$

Maximal pattern complexity can be used to characterize eventually periodic infinite words.

Theorem 2.3.8 ([26]) *An infinite word ω is eventually periodic if and only if for some n, $p^*_\omega(n) \le 2n - 1$.*

The class of infinite words having maximal pattern complexity $p^*_\omega(k) = 2k$ are called *pattern Sturmian words*. Note that a pattern Sturmian word ω must be binary, since $p^*_\omega(1) = 2$.

In [26], it is shown that for any irrational $0 < \alpha < 1$, and interval $I \subset [0,1]$ with $0 < |I| < 1$ and any $x \in [0,1)$, the word $\omega = \mathcal{R}(\alpha, I, x, \mathbb{Z})$ defined by

$$\omega(n) = \begin{cases} 0 & x + n\alpha \in I (\text{mod } \mathbb{Z}) \\ 1 & \text{otherwise} \end{cases}$$

is a pattern Sturmian word. Thus every Sturmian word is pattern Sturmian ([26]), but if $|I| \notin \{\alpha, 1-\alpha\}$ then ω is not Sturmian.

In [25] an interesting property is shown if the set I defined above is not an interval but is a particular closed set.

Theorem 2.3.9 ([25]) *For any irrational rotation α, there exists a closed set S in $[0,1)$ so that $p^*_\omega(k) = 2^k$, $k \geq 1$, for almost all $x \in [0,1)$ with $\omega = \mathcal{R}(\alpha, S, x, \mathbb{Z})$*

The notion of pattern complexity can be naturally extended to abelian pattern complexity. For a k-pattern P and an infinite word ω we consider the set

$$\left\{ \ \Psi(\omega_{n+P_0}\omega_{n+P_1}\cdots\omega_{n+P_{k-1}}) \ \mid n \geq 0 \ \right\},$$

the set of Parikh vectors of the pattern words. Then counting the size of this set will give the *maximal abelian pattern complexity*. Maximal abelian pattern complexity will be discussed further in Section 4.1.

2.4 Some Well-Known Words

Many well-known words have been studied extensively. When a new notion or idea of combinatorics on words is introduced, a first step is to see what happens with some of the well-known words. For example if a new type of complexity is introduced, does it classify some of the well-known words, or define a new class of words? Maybe a new family of morphisms does something interesting to a class of words.

Expanding some of the notions of minimal factor complexity led to the discovery of other classes of words, namely the Arnoux-Rauzy words (or strict episturmian) and the episturmian words. These classes of words will be described in more detail in the following sections. First we will look at some well-known finite words and then consider infinite words.

2.4.1 Finite Words

The first type of finite words to discuss are the *Lyndon words*. This notation and terminology can be found in [27]

A word is *primitive* if it is not a power of another word. Thus if x is primitive and $x = z^n$ for $z \in \mathcal{A}^*$, then $n = 1$ and $x = z$. The finite words $x, y \in \mathcal{A}^*$ are said to be *conjugate* if there exist words $u, v \in \mathcal{A}^*$ so that

$$x = uv \qquad y = vu$$

A *cyclic permutation* is a map σ so that if $a \in \mathcal{A}$ and $v \in \mathcal{A}^*$, then $\sigma(av) = va$. Thus two words x and y are conjugate if there is some n so that $\sigma^n(x) = y$. Conjugacy defines an equivalence relation on $\mathcal{A}^*$, where two words are equivalent if they are conjugate.

If we define an order on the alphabet $\mathcal{A} = \{0, 1, \ldots, k-1\}$, namely $0 < 1 < \cdots < k-1$, we can use lexicographical ordering on the elements of $\mathcal{A}^*$. The *Lyndon words* are precisely the set of words which are minimal in their conjugacy class, or l is a Lyndon word if for each $1 \leq n \leq |l| - 1$ then $l < \sigma^n(l)$. An equivalent way to define Lyndon words is that a word l is a Lyndon word if and only if for any factorization $l = uv$, with both u and v non-empty, then $l < v$. Thus l is less than any of its proper suffixes. By this definition Lyndon words must be primitive words, or else if $l = xx$ then lexicographically $x < l$ and l is not a Lyndon word. The first few Lyndon words on the alphabet $\mathcal{A} = \{0, 1\}$ with $0 < 1$ are: $0, 1, 01, 001, 011, 0001, 0011, 0111, \cdots$. The general formula for $L_k(n)$, the number of Lyndon words of length n on an alphabet with k letters, is

$$L_k(n) = \frac{1}{n} \sum_{d|n} \mu\left(\frac{n}{d}\right) k^d.$$

The sum is taken over all d which divide n and $\mu(m)$ is the Möbius function, which is defined to be 1 if m is a square-free positive integer with an even number of distinct prime factors, 0 if m is not square-free, and -1 otherwise.

Another well-known class of finite words are the *de Bruijn cycles*. The de Bruijn cycles, $B(k, n)$, are the minimal length cyclic words to contain all k^n distinct words of length n ([47, 42]). The length of a de Bruijn cycle $B(k, n)$ is k^n and each word of length n appears once as a factor. There are a total of $k^{-n}(k!)^{n^{k-1}}$ such cycles, so $B(k, n)$ is not unique. The construction and enumeration of such words over a binary alphabet was first completed by Sainte-Marie [42], and the extension to a general k-letter alphabet is stated and solved by van Aardenne-Ehrenfest and de Bruijn in [47]. The factors of length n of a cyclic word $B(k, n) = b_1 b_2 \ldots b_{k^n}$ are of the form $b_i b_{i+1} \ldots b_{i+n-1}$ for $i \leq k^n - n + 1$ or $b_i b_{i+1} \ldots b_{k^n} b_1 \ldots b_{i-k^n+n-1}$ for $k^n - n + 1 < i$. For example, it is readily verified that the words

$$B(3, 3) = 000111222012022110021210102$$
$$B(2, 5) = 00000100011001010011101011011111$$

contain all words, of their respective lengths, as factors exactly once.

The next set of finite words to consider are the *Fraenkel words*, which are used to construct the so-called *Fraenkel sequences*, in connection to a conjecture by Fraenkel (see [19]). A combinatorial version of the conjecture (from [38]) can be read as:

Conjecture 2.4.1 (Fraenkel) *For a k letter alphabet $\{1, 2, \ldots, k\}$ with $k \geq 3$, there is a unique balanced word, up to letter permutations and shifts, that has distinct letter frequencies.*

This word is called *Fraenkel's sequence* and is written as $(F_k)^\infty$, where the *Fraenkel words* are the F_k and are defined recursively as $F_1 = 1$ and $F_n = F_{n-1} n F_{n-1}$ for $n \geq 2$. Fraenkel's conjecture has been verified for an alphabet of size up to 7 ([4, 7, 22]).

The final well-known class of finite words to be discussed here are the finite Sturmian words. The set of finite Sturmian words is the set of all balanced finite binary words. For example the word 001001 is a finite Sturmian word, while the word 000101 is not a finite Sturmian word due to an imbalance in the factors 000 and 101. Another definition for the finite Sturmian words is, a word u is a finite Sturmian word if and only if it is a factor of some (infinite) Sturmian word s.

2.4.2 Sturmian Words

An infinite word s is a *Sturmian word* if for each $n \geq 0$, s has exactly $n+1$ distinct factors of length n, or $\rho_s(n) = n + 1$ (the only factor of length $n = 0$ being the empty-word). Thus since $\rho_s(1) = 2$, it should be clear that Sturmian words are binary words.

Example An example of a Sturmian word is the *Fibonacci word*, t, where

$$t = 010010100100101001 01\ldots$$

The Fibonacci word can be constructed in a number of ways. One way to construct t is by iteration of the morphism $\tau : 0 \mapsto 01; 1 \mapsto 0$. Thus $t = \tau^{\infty}(0)$, and t is a fixed point of τ. Is is good to note that for each $n \geq 0$ we have $\tau^n(0)$ as a prefix of t. These finite iterations of τ are called the *finite Fibonacci words*, $f_n = \tau^n(0)$. The morphism τ is called the Fibonacci morphism because for each $n \geq 0$ the length of $f_n = \tau^n(0)$ is the n-th Fibonacci number. This word can also be defined recursively, $f_{n+2} = f_{n+1}f_n$ where $f_{n-1} = 1$ and $f_0 = 0$. Yet another way to view the Fibonacci word is as a mechanical word (defined below) with irrational slope $\alpha = 1/\phi^2$, where $\phi = (1 + \sqrt{5})/2$ is the golden ratio.

As stated above, there are many equivalent definitions for the class of Sturmian words. Before some equivalent definitions are given, let's look at words created geometrically by an irrational slope. These descriptions have been given many times in the literature (see, e.g. [28], Chapter 2).

Given two real numbers α and ρ with $0 \leq \alpha \leq 1$, define the words $s_{\alpha,\rho}$ and $s'_{\alpha,\rho}$ by

$$s_{\alpha,\rho}(n) = \lfloor \alpha(n+1) + \rho \rfloor - \lfloor \alpha n + \rho \rfloor$$
$$s'_{\alpha,\rho}(n) = \lceil \alpha(n+1) + \rho \rceil - \lceil \alpha n + \rho \rceil$$

for each $n \geq 0$. For simplicity we can assume $0 \leq \rho < 1$ or $0 < \rho \leq 1$. It is readily verified that the words $s_{\alpha,\rho}$ and $s'_{\alpha,\rho}$ are over the alphabet $\mathcal{A} = \{0, 1\}$. The word $s_{\alpha,\rho}$ is called the *lower mechanical word*, $s'_{\alpha,\rho}$ is the *upper mechanical word*, with slope α and intercept ρ. It should be clear that $s_{0,\rho} = s'_{0,\rho} = 0^{\infty}$ and $s_{1,\rho} = s'_{1,\rho} = 1^{\infty}$, so we will assume that $0 < \alpha < 1$ unless otherwise noted. The condition on $0 < \alpha < 1$ is more of a simplification. If we let $\alpha > 0$ then we would have $\lfloor \alpha \rfloor \le s_{\alpha,\rho}(n) \le 1 + \lfloor \alpha \rfloor$, and $s_{\alpha,\rho}$ would be a binary word over the alphabet $\mathcal{A} = \{\lfloor \alpha \rfloor, 1 + \lfloor \alpha \rfloor\}$.

A mechanical word is said to be *irrational* or *rational* according to if its slope is irrational or rational. When $\alpha n + \rho$ is not an integer we have $1 + \lfloor \alpha n + \rho \rfloor = \lceil \alpha n + \rho \rceil$, so $s_{\alpha,\rho} = s'_{\alpha,\rho}$ except when $\alpha n + \rho$ is an integer. Thus if α is irrational, $\alpha n + \rho$ can be an integer at most once so $s_{\alpha,\rho}$ and $s'_{\alpha,\rho}$ can differ by at most one factor of length 2. If α is rational, say $\alpha = q/p$ for $q, p \in \mathbb{N}$ with $p \neq 0$, it is easy to see $s_{\alpha,\rho}$ is periodic of period p. So for each n we have

$$\begin{aligned} s_{\frac{q}{p},\rho}(n+p) &= \left\lfloor \frac{q}{p}(n+1+p) + \rho \right\rfloor - \left\lfloor \frac{q}{p}(n+p) + \rho \right\rfloor \\ &= \left\lfloor q + \frac{q}{p}(n+1) + \rho \right\rfloor - \left\lfloor q + \frac{q}{p}(n) + \rho \right\rfloor \\ &= \left\lfloor \frac{q}{p}(n+1) + \rho \right\rfloor - \left\lfloor \frac{q}{p}n + \rho \right\rfloor \\ &= s_{\frac{q}{p},\rho}(n) \end{aligned}$$

The case where $0 < \alpha < 1$ and $\rho = 0$ is going to come up again, so let's note that case now. In this case we have $s_{\alpha,0}(0) = \lfloor \alpha \rfloor = 0$ and $s'_{\alpha,0}(0) = \lceil \alpha \rceil = 1$, and if α is irrational (thus αn is irrational for all $n \ge 1$) we get

$$s_{\alpha,0} = 0c_\alpha, \qquad s'_{\alpha,0} = 1c_\alpha$$

It is also helpful to notice

$$s_{\alpha,\alpha} = c_\alpha, \qquad s'_{\alpha,\alpha} = c_\alpha$$

where c_α is called the *characteristic Sturmian word* with slope α. Characteristic words will be discussed more later.

Now we will see some of the equivalent definitions of infinite Sturmian words, proven by Morse and Hedlund.

Theorem 2.4.2 ([37]) *Let s be an infinite word. The following are equivalent:*

1. *s is Sturmian.*
2. *s is balanced and aperiodic.*
3. *s is irrational mechanical.*

Thus the class of Sturmian words corresponds to the words generated by irrational mechanical words. An interesting point is that there are uncountably many distinct Sturmian words due to uncountably many irrational numbers in the interval $[0, 1]$. A word is said to be a *finite Sturmian word* if it is a factor of some Sturmian word. Thus, as stated at the end of Section 2.4.1, the set of finite Sturmian words is the set

$$\{u \in \{0,1\}^* | \quad u \text{ is balanced}\}$$

All Sturmian words considered in this writing will be infinite Sturmian words, unless otherwise noted.

There also exists a relation between the slope of a Sturmian word and the frequency of the letters in the word. Recall from Section 2.3.2 that the frequency of the letter a in a finite word is the number of occurrences of a divided by the length of the word. In the case of an infinite word ω, the frequency of the letter a is the limit (if the limit exists) of the frequency of a in the prefixes of ω. Let s be a lower mechanical word over $\{0,1\}$ with irrational slope α and intercept ρ. Then for each $m \geq 1$ there exists an $n_m \in \mathbb{N}$ where $0 \leq n_m \leq m-1$ so that

$$\frac{n_m}{m} < \alpha < \frac{n_m+1}{m}$$

Thus for each m, any factor u of s where $|u| = m$ we know

$$\left\lfloor \frac{n_m}{m} m + \rho \right\rfloor \leq |u|_1 \leq \left\lfloor \frac{n_m+1}{m} m + \rho \right\rfloor$$

$$n_m \leq |u|_1 \leq n_m + 1$$

Thus the frequency of the letter 1 in s will be bounded by $\frac{n_m}{m}$ and $\frac{n_m+1}{m}$. It is readily verified that the sequences $(\frac{n_m}{m})_{m\in\mathbb{N}}$ and $(\frac{n_m+1}{m})_{m\in\mathbb{N}}$ will both limit to α, and thus the frequency of the letter 1 in s will be α. The frequency of the letter 0 in s will then be $1-\alpha$.

Mechanical words can be interpreted other ways. For different applications of words, different interpretations may be more useful. This first alternative interpretation will be mentioned again in Section 4.2.1, and can be found in [28]. For this we will consider the line $y = \beta x + \rho$ with $\beta > 0$ and any $\rho \in \mathbb{R}$, and with no other restrictions on β and ρ. This line will then define a sequence of points $Q_0, Q_1, \ldots$ where the line intersects the lines of the grid of non-negative integer points. Thus if we let $Q_n = (x_n, y_n)$ for each $n \geq 0$ we have $x_0 < x_1 < \cdots$, $y_0 < y_1 < \cdots$, and at least one of x_n or y_n an integer. We call Q_n *horizontal* if y_n is an integer, and *vertical* if x_n is an integer. If both x_n and y_n are integers we create an additional point Q_{n-1} before Q_n and say Q_{n-1} is horizontal and Q_n is vertical (we could say Q_{n-1} is vertical and Q_n is horizontal, but would always use the same choice), and then repeat in this manner when ever both x_n and y_n are both integers.

If we associate a 0 to each vertical point and a 1 to each horizontal point we get the word $K_{\beta,\rho}$ called the *(lower) cutting sequence* with slope β and intercept ρ. More formally, to each Q_n associate the point $I_n = (u_n, v_n)$ where:

$$(u_n, v_n) = \begin{cases} (\lceil x_n \rceil, y_n - 1) & \text{if } Q_n \text{ is horizontal} \\ (x_n, \lfloor y_n \rfloor) & \text{if } Q_n \text{ is vertical} \end{cases}$$

The I_n points are below (resp. below and to the right) of Q_n if Q_n is vertical (resp. horizontal). Similar points J_n can be defined to the left (resp. above and to the left) of Q_n if Q_n is horizontal (resp. vertical) to define the upper cutting sequence $K'_{\beta,\rho}$. It is readily verified that $u_n + v_n = n$ for each $n \geq 0$, and the cutting sequence $K_{\beta,\rho}$ is defined by:

$$K_{\beta,\rho}(n) = v_{n+1} - v_n = 1 + u_n - u_{n+1}$$

As with the mechanical words, the upper and lower cutting words will only differ where $\beta n + \rho$ is an integer. Let's look at the special case where $\rho = 0$ and β is irrational, and we get infinite words where the first letter is different. Thus we have the word C_β where

$$K_{\beta,0} = 0C_\beta, \qquad K'_{\beta,0} = 1C_\beta$$

and C_β is similar to a characteristic word for mechanical words. Cutting sequences are related to mechanical words $(s_{\alpha,\rho})$ by the following identity:

$$K_{\beta,\rho} = s_{\frac{\beta}{1+\beta},\frac{\rho}{1+\beta}}$$

Some other interpretations of mechanical words can be seen. In [37], Morse and Hedlund show that mechanical words can be realized by coding the orbit of a point on the circle of circumference equal to one under a shift of angle $0 < \alpha < 1$, where the circle is partitioned into two intervals of size α and $1-\alpha$. Sturmian words have also been shown to be represented by coding square billiard words (see [48]).

Now let's consider some properties of the factors of Sturmian words.

Proposition 2.4.3 ([28]) *The set $\mathcal{F}(s)$ of factors of a Sturmian word s is closed under reversal.*

Proposition 2.4.4 ([28]) *Let s and t be Sturmian words.*

1. *If s and t have the same slope, then $\mathcal{F}(s) = \mathcal{F}(t)$.*
2. *If s and t have distinct slopes, then $\mathcal{F}(s) \cap \mathcal{F}(t)$ is finite.*

Thus for two Sturmian words $s_{\alpha,\rho}$ and $s_{\alpha,\rho'}$, we know $\mathcal{F}(s_{\alpha,\rho}) = \mathcal{F}(s_{\alpha,\rho'})$. Moreover, we know for any Sturmian word $s_{\alpha,\rho}$, $\mathcal{F}(s_{\alpha,\rho}) = \mathcal{F}(s_{\alpha,0}) = \mathcal{F}(c_\alpha) = \mathcal{F}(C_{\alpha/(1-\alpha)})$. Thus when talking about the set of factors of a Sturmian word with slope α, it is sufficient to talk about the set of factors of the characteristic Sturmian word with the same slope α.

Since the words $s_{\alpha,0} = 0c_\alpha$ and $s'_{\alpha,0} = 1c_\alpha$ have the same slope, they have the same set of factors. Therefore, for each $n \geq 1$ the prefix p_n of length n of c_α is a left special factor. That is, both $0p_n$ and $1p_n$ are factors of c_α. More formally,

Proposition 2.4.5 ([28]) *For every Sturmian word s, either $0s$ or $1s$ is Sturmian. A Sturmian word s is characteristic if and only if $0s$ and $1s$ are both Sturmian.*

This, along with Proposition 2.4.3, implies that the set of right special factors of a Sturmian word s with slope α are the reversal of the prefixes of the characteristic Sturmian word c_α.

Proposition 2.4.3 was used to generalize the idea of minimal factor complexity from aperiodic binary words to aperiodic words over an alphabet with k letters, for $k \geq 3$. These classes of words are the Arnoux-Rauzy words (or strict episturmian words) and the episturmian words, which will be discussed more in Section 2.4.3.

Now let's look at another property relating an irrational number α and the characteristic Sturmian word with slope α. This explanation also comes from [28]. Every irrational number α has a unique expansion as a continued fraction, so

$$\alpha = d_0 + \cfrac{1}{(1+d_1) + \frac{1}{d_2 + \frac{1}{\cdots}}}$$

where each d_i an integer, $d_0 \geq 0$, $d_1 \geq 0$, and $d_i > 0$ for each $i \geq 2$. We then write

$$\alpha = [d_0, (1+d_1), d_2, \ldots]$$

If the sequence $(d_i)_{i\in\mathbb{N}}$ is eventually periodic where $d_i = d_{i+p}$ for each $i > N$, we write

$$\alpha = \left[d_0, (1+d_1), \ldots, d_N, \overline{d_{N+1}, d_{N+2}, \ldots, d_{N+p}}\right]$$

For the case of characteristic Sturmian words with irrational slope α, we have $0 < \alpha < 1$, and thus $\alpha = [0, (1+d_1), d_2, \ldots]$. To the sequence $(d_i)_{i\geq 1}$ we associate the sequence of words $(s_n)_{n\geq -1}$ where

$$s_{-1} = 1, \qquad s_0 = 0, \qquad s_n = s_{n-1}^{d_n} s_{n-2} \quad (n \geq 1)$$

The sequence $(s_n)_{n\geq -1}$ is a *standard sequence*, and the sequence $(d_1, d_2, \ldots)$ is called its *directive sequence*. It should be clear that each s_n is a prefix of s_{n+1} for each $n \geq 1$.

Example The directive sequence $(1, 1, 1, \ldots) = (\overline{1})$ gives the standard sequence defined by $s_1 = 01$, $s_2 = 010$, $s_3 = 01001$, $\cdots$, $s_n = s_{n-1}s_{n-2}$, which are exactly the finite Fibonacci words.

The property we have for the directive sequence is that the limit of the sequence $(s_n)_{n\geq -1}$ gives the characteristic Sturmian word c_α. Formally,

Proposition 2.4.6 ([28]) *Let* $\alpha = [0, (1+d_1), d_2, \ldots]$ *be the continued fraction expansion of some irrational* α*, with* $0 < \alpha < 1$*, and let* $(s_n)_{n \geq -1}$ *be the standard sequence associated to the directive sequence* $(d_1, d_2, \ldots)$*. Then every* s_n *is a prefix of* c_α *and*

$$c_\alpha = \lim_{n \to \infty} s_n.$$

Thus, the continued fraction expansion of an irrational α can be used to construct the characteristic Sturmian word of slope α. Likewise, if a word ω is a characteristic Sturmian word we can use the form of ω to construct the continued fraction expansion of the irrational value of its slope.

Another characterization of characteristic Sturmian words deals with palindromic prefixes. In Theorem 2.4.7, the equivalence of (1), (2), and (3) is given in [15], while the equivalence between (3) and (4) is given in [30]. The *right palindromic closure* of the finite word u, denoted $(u)^+$, is the shortest palindrome that has u as a prefix. For example, $(01011)^+ = 01011010$ and $(11001)^+ = 110011$. A palindrome u is a *central factor* of the palindrome w if $w = vu\tilde{v}$ for some $v \in \mathcal{A}^*$.

Theorem 2.4.7 ([15, 30]) *For an aperiodic word* ω *on the binary alphabet* $\mathcal{A} = \{0, 1\}$*, the following are equivalent:*

1. *For any prefix* v *of* ω*,* $(v)^+$ *is a prefix of* ω*.*
2. *The leftmost occurrence of a palindromic factor of* ω *is a central factor of a palindromic prefix of* ω*.*
3. *There exist an infinite sequence of palindromes* $u_1 = \varepsilon$*,* $u_2, u_3, \ldots$ *and an infinite word* $\Delta(s) = x_1 x_2 \cdots$*, each* $x_i \in \mathcal{A}$*, and* $\Delta(s) \in \mathcal{A}^{\mathbb{N}} \backslash (\mathcal{A}^* 0^{\mathbb{N}} \cup \mathcal{A}^* 1^{\mathbb{N}})$ *so that the* u_i *are prefixes of* ω *and* $u_{i+1} = (u_i x_i)^+$ *for all* $i \geq 1$*.*
4. *The word* ω *is a characteristic Sturmian word.*

The word $\Delta(s)$ in (3) of Theorem 2.4.7 is called the *directive word* (not to be confused with the directive sequence dealing with the continued fraction expansion, the directive sequence was a general sequence of integers rather than over the alphabet $\mathcal{A}$). Other notation in the literature also use the PAL operator. The comparison between PAL and the notation in Theorem 2.4.7 is as follows: $x = x_1 x_2 \ldots x_n$, each $x_i \in \mathcal{A}$, $u_1 = \varepsilon$, $\ldots$, $u_{n+1} = (u_n x_n)^+$, then $PAL(x) = u_{n+1}$.

It is also helpful to point out that in [15], any infinite word on a general alphabet $\mathcal{A}$ that satisfies properties (1) and (2) in Theorem 2.4.7 is uniformly recurrent (specifically cases (1), (2), and (3) in Theorem 2.4.12, which Sturmian words satisfy). Therefore characteristic Sturmian words are uniformly recurrent. Since a Sturmian word has the same set of factors as the characteristic word of the same slope, Sturmian words are uniformly recurrent as well.

The idea of morphisms related to Sturmian words is another topic that has been of interest in the area of combinatorics on words (see [10, 28]). A morphism f is called a *Sturmian morphism* if $f(s)$ is a Sturmian word for every Sturmian word s. The set of all Sturmian morphisms form a monoid with the operation of composition. As we will see, the building blocks of the Sturmian morphisms are the following three morphisms:

$$\begin{aligned} \tau &: \ \tau(0) = 01, \ \tau(1) = 0 \\ \tilde{\tau} &: \ \tilde{\tau}(0) = 10, \ \tilde{\tau}(1) = 0 \\ \theta &: \ \theta(0) = 1, \ \theta(1) = 0 \end{aligned}$$

The Fibonacci morphism used in the example at the beginning of this section is exactly the morphism τ listed here.

Proposition 2.4.8 ([10]) *The morphisms τ and $\tilde{\tau}$ are Sturmian.*

To test if a morphism φ is a Sturmian morphism could be a difficult task if we really needed to test if φ preserved all Sturmian words. Fortunately, that much work is not necessary thanks to following result.

Theorem 2.4.9 ([10]) *A morphism φ is Sturmian if and only if the word*

$$\varphi(10^2 10^2 1010^2 101)$$

is primitive and balanced. Moreover, it is decidable whether a morphism is Sturmian.

Thus to test if a morphism is Sturmian, is suffices to apply the morphism to a word of length 14 and see of it is primitive and balanced. For example, $\tau(10^2 10^2 1010^2 101) =$ 0010100101001001010010 which is readily verified to be primitive and balanced.

A characterization of Sturmian morphisms can be seen in the following theorem.

Theorem 2.4.10 ([10]) *A morphism φ is Sturmian if and only if φ is a composition of the morphisms θ, τ, and $\tilde{\tau}$.*

Then classifying the morphisms that preserve characteristic Sturmian words we have the following result.

Theorem 2.4.11 ([10]) *Let φ be a morphism, and let $0 < \alpha, \beta < 1$ be two irrational numbers so that $c_\alpha = \varphi(c_\beta)$. Then φ is a composition of the morphisms θ and τ.*

Thus any composition of the morphisms θ, τ, and $\tilde{\tau}$ will give a Sturmian morphism, but only a composition of only the θ and τ morphisms will preserve the characteristic Sturmian words. Therefore the set of Sturmian morphisms are a monoid under composition with generators θ, τ, and $\tilde{\tau}$, while the morphisms that preserve characteristic Sturmian words are a submonoid of the monoid of Sturmian morphisms generated by θ and τ.

2.4.3 Episturmian Words

As stated before the Sturmian words are the words with exactly $n+1$ distinct factors of length n. Therefore, Sturmian words have exactly one left special and one right special factor for each length. Since the set of factors of a Sturmian word are closed under reversal we know if u is a right special factor of the Sturmian word s, then $\tilde{u}$ is a left special factor of s. Extending the singular left and right special factor property to an alphabet $\mathcal{A}$, where $|\mathcal{A}| = k \geq 3$, we have the following definition using terminology from [20], and introduced in [15].

Definition An infinite word s on the finite alphabet $\mathcal{A}$ is *episturmian* if the set of factors of s are closed under reversal and s has at most one left special factor (equivalently right special factor) of each length. An episturmian word s is *standard episturmian* if all the left special factors s are prefixes of s.

The standard episturmian words are a generalization of the characteristic Sturmian words to an alphabet of size $k \geq 3$. We also have the property that if a word ω is episturmian then $\mathcal{F}(\omega) = \mathcal{F}(s)$ for some standard episturmian word s ([15]). The definition of episturmian words does allow for periodic words, and thus we say that an episturmian word is aperiodic (resp. periodic) if its associated standard episturmian word is aperiodic (resp. periodic). All episturmian words mentioned here will be aperiodic episturmian words, unless otherwise noted.

Another class of words that are contained in the class of episturmian words are the Arnoux-Rauzy words (or *strict episturmian words*).

Definition A word ω is an *Arnoux-Rauzy word* if the set of factors of ω is closed under reversal and ω has exactly one right (resp. left) special factor of each length, and if u is a right (resp. left) special of ω then ua (resp. au) is a factor of ω for each $a \in Alph(\omega)$.

In the case where ω is a strict episturmian word and $Alph(\omega) = \mathcal{B} \subseteq \mathcal{A}$, then ω is said to be $\mathcal{B}$-strict, and if $\mathcal{B} = \mathcal{A}$ then ω is said to be strict or $\mathcal{A}$-strict.

A characterization of standard episturmian words dealing with palindromic prefixes exists, similar to Theorem 2.4.7.

Theorem 2.4.12 ([15]) *For an infinite word $\omega \in \mathcal{A}^{\mathbb{N}}$, the following are equivalent:*

1. *For any prefix v of ω, $(v)^+$ is a prefix of ω.*

2. *The leftmost occurrence of a palindromic factor of ω is a central factor of a palindromic prefix of ω.*

3. *There exist an infinite sequence of palindromes $u_1 = \varepsilon$, $u_2, u_3, \ldots$ and an infinite word $\Delta(s) = x_1 x_2 \cdots$, each $x_i \in \mathcal{A}$, so that $u_{i+1} = (u_i x_i)^+$ for all $i \geq 1$ and all the u_i are prefixes of ω.*

4. *The word ω is a standard episturmian word.*

The word $\Delta(s)$ in (3) of Theorem 2.4.12 is called the *directive word* of the standard episturmian word ω. The notation $\Delta = x_1 x_2 \cdots$ will be the directive word of a standard episturmian word, unless otherwise noted. As stated after Theorem 2.4.7, words satisfying cases (1), (2), and (3) in Theorem 2.4.12 are uniformly recurrent. Therefore episturmian words are uniformly recurrent. Thus, eventually periodic episturmian words are purely periodic.

Example The standard episturmian word with directed word $\Delta = (012)^\infty$ is known as the *Tribonacci word* (or *Rauzy word*, [39]). The word begins as follows:

$$r = \tau_3^\infty(0) = \mathbf{01}0\mathbf{2}010\mathbf{0}102010\mathbf{1}020100102010\mathbf{2}01001\ldots$$

where the bold letters follow the palindromic prefixes u_i, or for $\Delta = x_1 x_2 \cdots$ it is written $u_i \mathbf{x_i}$. Another way to view the Tribonacci word r is as the unique fixed point of the morphism τ_3,

$$\tau_3: \quad 0 \mapsto 01; \qquad 1 \mapsto 02; \qquad 2 \mapsto 0$$

To generalize this, for an alphabet $\mathcal{A} = \{1, 2 \cdots, k\}$, for $k \geq 2$, the *k-bonacci* word has directed word $(12 \ldots k)^\infty$ and is the unique fixed point $\tau_k^\infty(1)$ of the morphism τ_k where $1 \mapsto 12$, $2 \mapsto 13$, $\cdots$, $k \mapsto 1$. If $k = 2$ we get the Fibonacci word.

As stated in the previous section, the set of Sturmian morphisms is the monoid of morphisms that preserve Sturmian words. The morphisms that generate the Sturmian morphisms are easily generalized to a general alphabet of size $k \geq 3$. The monoid of all *episturmian morphisms* $\mathcal{E}$ (see [15, 20, 24]) is generated, using composition, by the following three morphisms:

$$\psi_a: \ \psi_a(a) = a, \ \psi_a(b) = ab \text{ for any letter } b \neq a$$
$$\overline{\psi}_a: \ \overline{\psi}_a(a) = a, \ \overline{\psi}_a(b) = ba \text{ for any letter } b \neq a$$
$$\theta_{ab}: \ \theta_{ab}(a) = b, \ \theta_{ab}(b) = a, \ \theta_{ab}(x) = x, \text{ for any letter } x \notin \{a, b\}$$

For any morphism $f \in \mathcal{E}$, the decomposition of f need not be unique. In [15] it is shown that these morphisms satisfy the relation $\theta_{ab}\psi_a = \psi_b\theta_{ab}$, and thus $\psi_a = \theta_{ab}\psi_b\theta_{ab}$. The monoid of *standard episturmian morphisms* $\mathcal{S}$ is the submonoid of $\mathcal{E}$ generated by the ψ_a and θ_{ab}. The monoid of *pure episturmian morphisms* $\mathcal{E}_p$ is the submonoid of $\mathcal{E}$ generated by the ψ_a and $\overline{\psi}_a$. Characterizing the episturmian and standard episturmian morphisms, we have the following theorem.

Theorem 2.4.13 ([24]) *A morphism φ is episturmian (resp. standard episturmian) if there exist strict episturmian (resp. standard episturmian) words $\boldsymbol{m}$ and $\boldsymbol{t}$ so that $\boldsymbol{m} = \varphi(\boldsymbol{t})$.*

Moreover, the episturmian (resp. standard episturmian) morphisms preserve the set of episturmian (resp. standard episturmian) words.

The following theorem shows a nice property of the standard episturmian words, namely the distribution of the most frequent letter.

Theorem 2.4.14 ([15]) *An infinite word $s \in \mathcal{A}^{\mathbb{N}}$ is a standard episturmian word if and only if there exist a standard episturmian word t and $a \in \mathcal{A}$ so that $s = \psi_a(t)$. Moreover, the first letter of s is a, t is unique and the directive words satisfy $\Delta(s) = a\Delta(t)$.*

Thus, for a standard episturmian word s over $\mathcal{A}$ there is a letter $a \in \mathcal{A}$ so that for each factor u of s with $|u| = 2$, then $|u|_a \geq 1$. Therefore, balanced standard episturmian words have a letter that satisfies the conditions of Lemma 3.2.2 (see later).

When generalizing Sturmian words to an arbitrary finite alphabet by preserving the singular (or at most one) special factor it is possible to lose balance. Thus there are balanced episturmian words as well as unbalanced episturmian words. For example, the Tribonacci word $t = 0102010010201\cdots$ is not balanced, but is 2-balanced (see [41]). In [38], Paquin and Vuillon classified the balanced standard episturmian words.

Theorem 2.4.15 ([38]) *Any balanced standard episturmian word s over an alphabet $\mathcal{A} = \{1, 2, \ldots, k\}$, where $k \geq 3$, has a directive word, up to a letter permutation, in one of the three following families:*

1. $\Delta(s) = 1^n \left(\prod_{i=2}^{k-1} i\right)(k)^\infty = 1^n 23\ldots(k-1)(k)^\infty$, *with* $n \geq 1$;
2. $\Delta(s) = \left(\prod_{i=1}^{j-1} i\right) 1 \left(\prod_{i=j}^{k-1} i\right)(k)^\infty = 12\ldots(j-1)1j\ldots(k-1)(k)^\infty$, *with* $3 \leq j \leq k-1$;
3. $\Delta(s) = \left(\prod_{i=1}^{k} i\right)(1)^\infty = 123\ldots k(1)^\infty$;

Thus the balanced episturmian words with distinct letter frequencies satisfy Fraenkel's conjecture (from [38], and stated in Proposition 3.1.1 below). Then considering this theorem with a result from [15], we have Corollary 2.4.17.

Theorem 2.4.16 ([15]) *A standard episturmian word s is eventually periodic if and only if its directive word Δ has the form ga^∞, with $g \in \mathcal{A}^*$ and $a \in \mathcal{A}$. Moreover, s is periodic.*

Thus, the following Corollary comes rather naturally.

Corollary 2.4.17 ([38]) *Every balanced standard episturmian word over 3 or more letters is periodic.*

For a periodic word s, there will be an $n \in \mathbb{N}$ so that no factor of length n can be right special. Thus, none of the Arnoux-Rauzy words are balanced. In fact, it has been shown in [12] that an Arnoux-Rauzy word can be constructed that is not C-balanced for any positive integer C.

Chapter 3

Weakly Rich Words

In this chapter we will investigate weakly rich words. One topic in [21] involved showing that recurrent balanced weakly rich words are necessarily balanced periodic episturmian words. Thus, recurrent balanced weakly rich words with distinct letter frequencies obey Fraenkel's conjecture.

Note: Results of this chapter appear in [21].

3.1 Preliminaries

Given a (finite or infinite) word x over $\mathcal{A}$ and any $a \in \mathcal{A}$, a *complete return to* a in x is a factor of x of the form aya, where $|y|_a = 0$. We say that an infinite word ω over $\mathcal{A}$ is *weakly rich* if for each $a \in \mathcal{A}$, all complete returns to a in ω are palindromes. For example, the infinite periodic word $(acbcaacbc)^\infty$ can be verified to be weakly rich. It should be clear that each binary word is weakly rich, since complete returns in binary words would look like either 01^k0 or 10^l1 for integers $k, l \geq 0$.

As mentioned in Section 2.4.3, Theorem 2.4.15, balanced episturmian words are characterized into three classes. One of these classes involved words with distinct letter frequencies and they, up to letter permutation and shifts, correspond to the Fraenkel sequences.

Proposition 3.1.1 ([38]) *Suppose* $\boldsymbol{t}$ *is a balanced episturmian word with* $Alph(\boldsymbol{t}) = \{1, 2, \ldots, k\}$*,* $k \geq 3$*. If* $\boldsymbol{t}$ *has mutually distinct frequencies, then up to a letter permutation,* $\boldsymbol{t}$ *is a shift of* $(F_k)^\infty$

Also recall from Section 2.4.3, the episturmian morphism ψ_a, defined as:

$$\psi_a : \begin{cases} a \mapsto a \\ b \mapsto ab \quad \text{for } b \neq a \end{cases}$$

This morphism will play a fundamental role in the construction of balanced weakly rich words.

3.2 Main Result

To show the main result of this section, we need the following notation.

Definition Let $x = x_1x_2x_3\ldots \in \mathcal{A}^{\mathbb{N}}$ and let a be a new symbol not in $\mathcal{A}$. Then define $\sigma_a : \mathcal{A}^{\mathbb{N}} \mapsto (A \cup \{a\})^{\mathbb{N}}$ by

$$\sigma_a(x) = ax_1a^{\epsilon_1}x_2a^{\epsilon_2}x_3a^{\epsilon_3}\cdots$$

where $\epsilon_i \in \{1, 2\}$, with $\epsilon_i = 2$ if and only if $x_i = x_{i+1}$.

The theorem to be shown is then stated as follows.

Theorem 3.2.1 ([21]) *Suppose ω is a recurrent balanced weakly rich word over $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geq 3$. Then, up to a letter permutation, ω is either:*

(1) a shift of the periodic word

$$\psi_1^n \circ \psi_2 \circ \cdots \circ \psi_{k-1}(k^\infty), \text{ for some } n \geq 1;$$

(2) or a shift of the periodic word

$$\sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_j \circ \psi_{j+1}^2 \circ \cdots \circ \psi_{k-1}(k^\infty), \text{ for some } 1 \leq j \leq k-2.$$

The proof of Theorem 3.2.1 will require several lemmas. For the following lemmas, we will assume that $\omega \in \mathcal{A}^{\mathbb{N}}$ and $|\mathcal{A}| \geq 3$. For each $a \in \mathcal{A}$ denote $g_a = \sup |u|$, where the supremum is taken over all factors u of ω not containing a and observe that each g_a is finite.

Lemma 3.2.2 ([45]) *Suppose $\omega \in \mathcal{A}^{\mathbb{N}}$ is balanced, and let $a \in \mathcal{A}$ be such that the frequency of a in ω is at least $1/3$. Then the word $\omega' \in (A\backslash\{a\})^{\mathbb{N}}$ obtained from ω by deleting all occurrences of the letter a in ω is also balanced.*

The following (Lemmas 3.2.3, 3.2.4, 3.2.5 and Corollaries 3.2.6, 3.2.7) are stated and proven in [21] (numbered 5.9, 5.10, 5.11, 5.12, 5.13 respectively).

Lemma 3.2.3 *Suppose $\omega \in \mathcal{A}^{\mathbb{N}}$ is a recurrent balanced weakly rich word, and let $a \in \mathcal{A}$ be such that $g_a \leq g_x$ for all $x \in \mathcal{A}$. Then the word $\omega' \in (A\backslash\{a\})^{\mathbb{N}}$ obtained from ω by deleting all occurrences of the letter a in ω is also a recurrent balanced weakly rich word.*

Lemma 3.2.4 *Suppose ω and ω' are as in Lemma 3.2.3. Suppose ω' contains the factor bb for some $b \in A\backslash\{a\}$. Then ω is a shift of $\sigma_a(\omega')$. In particular, the complete returns to b in ω are of the form baab or baxab for some $x \in A\backslash\{a,b\}$*

Lemma 3.2.5 *Suppose ω and ω' are as in Lemma 3.2.3 and let $b \in A\backslash\{a\}$. Then bbb is not a factor of ω'.*

Proof [Theorem 3.2.1] The proof of Theorem 3.2.1 is done by induction on k, the number of letters in the alphabet. Suppose ω is a recurrent balanced weakly rich word on the alphabet $\mathcal{A}_3 = \{1,2,3\}$ and without loss of generality we can assume $g_1 \leq g_2 \leq g_3$. Let $\omega' \in \{2,3\}^{\mathbb{N}}$ be obtained from ω by deleting all occurrences of the letter 1 in ω. First suppose that ω' does not contain the factor 22. Then $\omega' = (23)^\infty = \psi_2(3^\infty)$. Thus, the only complete return to 2 in ω' is 232. Then because ω is weakly rich, there must be some $m \geq 0$ so that the only complete return to 2 in ω is 21^m31^m2 and thus $\omega = \psi_1^m(\omega') = \psi_1^m \circ \psi_2(3^\infty)$. Next, suppose that the factor 22 does occur in ω'. If the first returns to 3 in ω' were of the form 323 and 3223 Lemma 3.2.4 implies that 31213 and 31211213 will be complete returns to 3 in ω, contradicting the fact that ω is balanced. Thus ω' is a shift of the periodic word $(223)^\infty = \psi_2^2(3^\infty)$. Then ω is a shift of $\sigma_1(\omega') = \sigma_1 \circ \psi_2^2(3^\infty)$. Therefore Theorem 3.2.1 holds for $k = 3$.

Next, let $k > 3$ and suppose ω is a recurrent balanced weakly rich word over $\mathcal{A}_k = \{1,2,\ldots,k\}$. Then by the induction hypothesis, assume that Theorem 3.2.1 holds for any recurrent balanced weakly rich words over a $k-1$ letter alphabet. Without loss of generality we again assume that $g_1 \leq g_2 \leq \ldots \leq g_k$. Let $\omega' \in \{2,3,\ldots,k\}^{\mathbb{N}}$ be obtained from ω by deleting all occurrences of the letter 1 in ω. By Lemma 3.2.3, ω' is a recurrent balanced weakly rich word. By the induction hypothesis, ω' is a shift of either $\psi_2^n \circ \psi_3 \circ \cdots \circ \psi_{k-1}(k^\infty)$ for some $n \geq 1$, or $\sigma_2 \circ \sigma_3 \circ \cdots \circ \sigma_j \circ \psi_{j+1}^2 \circ \cdots \circ \psi_{k-1}(k^\infty)$, for some $2 \leq j \leq k-2$.

First suppose that ω' does not contain the factor 22. Then ω' must be a shift of $\psi_2 \circ \psi_3 \circ \cdots \circ \psi_{k-1}(k^\infty)$. Thus the complete returns to 2 in ω' are of the form $2a2$ for some $a \in \{3,4,\ldots,k\}$ and there exists some $m \geq 1$ so that the complete returns to 2 in ω are of the form 21^ma1^m2. Therefore in this case $\omega = \psi_1^m(\omega') = \psi_1^m \circ \psi_2 \circ \psi_3 \circ \cdots \circ \psi_{k-1}(k^\infty)$.

Next suppose that ω' does contain the factor 22. Then ω' is a shift of either $\psi_2^2 \circ \psi_3 \circ \cdots \circ \psi_{k-1}(k^\infty)$ or $\sigma_2 \circ \sigma_3 \circ \cdots \circ \sigma_j \circ \psi_{j+1}^2 \circ \cdots \circ \psi_{k-1}(k^\infty)$, for some $2 \leq j \leq k-2$. Lemma 3.2.4 implies that ω is a shift of either $\sigma_1 \circ \psi_2^2 \circ \psi_3 \circ \cdots \circ \psi_{k-1}(k^\infty)$ or $\sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_j \circ \psi_{j+1}^2 \circ \cdots \circ \psi_{k-1}(k^\infty)$, for some $2 \leq j \leq k-2$. Therefore by the induction hypothesis, Theorem 3.2.1 is true. ∎

Thus once we have Theorem 3.2.1, we have the following corollaries.

Corollary 3.2.6 *Suppose that ω is a recurrent balanced weakly rich word over $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geq 3$. Then ω is a balanced periodic episturmian word.*

This is true by showing

$$\sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_j \circ \psi_{j+1}^2 \circ \psi_{j+2} \circ \cdots \circ \psi_{k-1}(k^\infty) =$$
$$\psi_1 \circ \psi_2 \circ \cdots \circ \psi_j \circ \psi_{j+1} \circ \psi_1 \circ \psi_{j+2} \circ \cdots \circ \psi_{k-1}(k^\infty)$$

Corollary 3.2.7 *Suppose that ω is a recurrent balanced weakly rich word over $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geq 3$. If the letters in ω have distinct frequencies, then up to a letter permutation, ω is a shift of $(F_k)^\infty$.*

This is then true by Proposition 3.1.1 and Corollary 3.2.6.

Chapter 4

Topics In Abelian Complexity

In this chapter, we will consider two problems related to abelian complexity. First, in Section 4.1 we will investigate an abelian variation of maximal pattern complexity. The main result will be to classify the set of binary recurrent aperiodic words using maximal abelian pattern complexity.

In Section 4.2, we will investigate the link between abelian complexity and the C-balance property. First we will develop a method to calculate the maximal value for abelian complexity of infinite balanced words. Next an upper bound for the abelian complexity of infinite recurrent C-balanced words is given. We end this chapter with a conjecture about words which achieve the maximal abelian complexity value.

4.1 Maximal Abelian Pattern Complexity

The goal of this section is to classify eventually periodic words using maximal abelian pattern complexity. Recall from Section 2.3.3 the notion of pattern complexity. A k-pattern P is a sequence of k integers with $0 = P_0 < P_1 < \cdots < P_{k-1}$, and denote by $|P| = k$ the number of terms in P. Then set $F_\omega(P)$ to be the set of pattern words over a k-pattern, and $p^*_\omega(k) = \sup_P |F_\omega(P)|$. The *maximal abelian pattern complexity function* is defined by

$$p^*_{ab}(k) = \sup_P \left|\left\{ \ \Psi(\omega_{n+P_0}\omega_{n+P_1}\cdots\omega_{n+P_{k-1}}) \ \mid \ n \geq 0 \ \right\}\right|$$

where the supremum is taken over all k-patterns P. The maximal value for $p^*_{ab}(k)$ on an infinite binary word will be $k+1$.

An *infinite pattern* $\mathcal{P}$ is an infinite sequence of integers $(\mathcal{P}_i)_{i\geq 0}$ so that

$$0 = \mathcal{P}_0 < \mathcal{P}_1 < \mathcal{P}_2 < \cdots$$

A finite or infinite pattern $\mathcal{P}' = N_0 < N_1 < N_2 < \cdots$ is a *sub-pattern* of $\mathcal{P}$ if the sequence $(N_i)_{i\geq 1}$ is a subsequence of the sequence $(\mathcal{P}_i)_{i\geq 0}$, and we write $\mathcal{P}' \subset \mathcal{P}$.

Let ω be an infinite binary word. For an infinite pattern $\mathcal{P}$ and some positive integer k, we say

$$\omega[\mathcal{P}](k) = \{ \ \omega_{n+\mathcal{P}_0}\omega_{n+\mathcal{P}_1}\cdots\omega_{n+\mathcal{P}_{k-1}} \ \mid n \geq 0 \ \}$$

and

$$\omega[\mathcal{P}]^* = \bigcup_{k\geq 0} \omega[\mathcal{P}](k)$$

Given a finite sub-pattern $P = (N_i)_{0\leq i\leq k} \subset \mathcal{P} = (\mathcal{P}_i)_{i\geq 0}$ say

$$\omega[P] = \{ \ \omega_{n+N_0}\omega_{n+N_1}\cdots\omega_{n+N_k} \ \mid n \geq 0 \ \}$$

Lemma 4.1.1 *Let ω be an infinite recurrent binary word. There exists an infinite pattern $\mathcal{P}$ so that:*

$$(*) \qquad u_1u_2\ldots u_n \in \omega[\mathcal{P}]^* \Rightarrow u_1u_2\ldots(u_n)^m \in \omega[\mathcal{P}]^* \ \ \textit{for } m \geq 1$$

Proof Let ω be an infinite recurrent binary word. The infinite pattern $\mathcal{P}$ will be the limit of the patterns ${}_nP$, defined as follows for each n.

Let ${}_nP$ be an n-pattern. Since $\omega[{}_nP]$ is finite there is a prefix of ω which contains all pattern words found by ${}_nP$ in ω, so let U_n be a prefix of ω so that $\omega[{}_nP] = U_n[{}_nP]$. Since ω is recurrent there will be another occurrence of U_n, infinitely many to be precise, so that $\omega = U_nvU_n\cdots$ with $z = |U_nv| \geq |U_n|$. Note that for each $0 \leq i \leq |U_n|$, $\omega_i = \omega_{i+z}$.

We now define ${}_{n+1}P$ as the immediate extension of ${}_nP$, so ${}_{n+1}P_i = {}_nP_i$ for each $0 \leq i \leq n-1$, where ${}_{n+1}P_n = {}_{n+1}P_{n-1} + z$. Then for any $u_1u_2\ldots u_n \in \omega[{}_nP]$, it is true that $u_1u_2\ldots u_n \in U_n[{}_nP]$ so there is some $0 \leq i \leq |U_n| - {}_nP_{n-1}$ so

$$u_1u_2\ldots u_n = u_{i+{}_nP_0}u_{i+{}_nP_1}\ldots u_{i+{}_nP_{n-1}}.$$

For each $0 \leq j \leq n-1$, $i+{}_nP_j = i+{}_{n+1}P_j$, and ${}_{n+1}P_n = {}_{n+1}P_{n-1} + z$ so

$$u_{i+{}_{n+1}P_{n-1}} = u_{i+{}_{n+1}P_n}.$$

Thus $u_1u_2\ldots u_nu_n \in \omega[{}_{n+1}P]$.

So $\mathcal{P} = \lim {}_nP$, and if for some n, $u_1u_2\ldots u_n \in \omega[\mathcal{P}]^*$ it follows that $u_1u_2\ldots(u_n)^m \in \omega[\mathcal{P}]^*$ for each $m \geq 1$. ∎

Lemma 4.1.2 *Let $\mathcal{P}$ be an infinite pattern satisfying the above condition $(*)$, and let $\mathcal{P}'$ be any infinite sub-pattern of $\mathcal{P}$. Then $\mathcal{P}'$ also satisfies $(*)$.*

Proof Let $u_1u_2\dots u_n \in \omega[\mathcal{P}']^*$ and let $m' \geq 0$. Then since $\mathcal{P}' \subset \mathcal{P}$, there is a word $v_1v_2\dots v_l \in \omega[\mathcal{P}]^*$ so that $\mathcal{P}'_n = \mathcal{P}_l$ and for each $1 \leq i < n$ there is some j so that $\mathcal{P}'_i = \mathcal{P}_j$. Then for each $m \geq 0$ $v_1v_2\dots v_l^m \in \omega[\mathcal{P}]^*$. Thus there is some m so that $\mathcal{P}'_{n+m'} = \mathcal{P}_{l+m}$ and so that for some $x \in \mathbb{N}$, $v_1v_2\dots v_l^m = \omega_{x+\mathcal{P}_0}\omega_{x+\mathcal{P}_a}\dots\omega^m_{x+\mathcal{P}_{l-1}}$, and thus $\omega_{x+\mathcal{P}'_0}\omega_{x+\mathcal{P}'_1}\dots\omega^{m'}_{x+\mathcal{P}'_{n-1}} = u_1u_2\dots u_n^{m'} \in \omega[\mathcal{P}']^*$. ∎

Proposition 4.1.3 *Let ω be an infinite binary word and let $\mathcal{P}$ be an infinite pattern. Then for each positive integer K, there exists an infinite sub-pattern $\mathcal{P}' = \mathcal{P}_K \subset \mathcal{P}$ so that for any two finite sub-patterns $P, Q \subset \mathcal{P}'$ with $|P| = |Q|$ and $1 \leq |P|, |Q| \leq K$, then*

$$\omega[P] = \omega[Q].$$

Proof This will be done by induction, creating a series of nested infinite patterns

$$\mathcal{P}' = \mathcal{P}_K \subset \mathcal{P}_{K-1} \subset \dots \subset \mathcal{P}_1 = \mathcal{P}$$

so that for each $1 \leq i \leq K$ we have $\omega[P] = \omega[Q]$ for all finite sub-patterns P and Q of $\mathcal{P}_i$ with $|P| = |Q|$ and $1 \leq |P|, |Q| \leq i$.

Begin by constructing $\mathcal{P}_2$. For two sub-patterns $P, Q \subset \mathcal{P}_1$ with $|P| = |Q| = 2$ say that

$$P \sim_2 Q \iff \omega[P] = \omega[Q].$$

Thus $\sim_2$ forms an equivalence relation on all sub-patterns of $\mathcal{P}_1$ of length 2, and thus naturally defines a finite coloring (since there are finitely many sets of words of a finite length) on the set of size 2 sub-patterns of $\mathcal{P}_1$. Equivalently we can say we have a finite coloring on the set of all 2-element subsets of $\mathbb{N}$; where two patterns P and Q are monochromatic if and only if $P \sim_2 Q$. Recall the following well-known theorem of Ramsey:

Theorem 4.1.4 (Ramsey) *Let k be a positive integer. Then given any finite coloring of the set of all k-element subset of $\mathbb{N}$; there exists an infinite set $\mathcal{A} \subset \mathbb{N}$ such that any two k-element subsets of $\mathcal{A}$ are monochromatic.*

Thus by the above theorem, there exists an infinite sub-pattern $\mathcal{P}_2 \subset \mathcal{P}_1$ so that any two sub-patterns of $\mathcal{P}_2$ are $\sim_2$ equivalent. Then, having constructed $\mathcal{P}_n \subset \mathcal{P}_{n-1} \subset \dots \subset \mathcal{P}_1$ we construct $\mathcal{P}_{n+1}$ as follows. For sub-patterns $P, Q \subset \mathcal{P}_n$ with $|P| = |Q| = n+1$ say that

$$P \sim_{n+1} Q \iff \omega[P] = \omega[Q].$$

This then defines a finite coloring of the set of size $n+1$ sub-patterns. Thus, again by Ramsey's Theorem, there exists an infinite sub-pattern $\mathcal{P}_{n+1} \subset \mathcal{P}_n$ so that any two sub-patterns of $\mathcal{P}_{n+1}$ are $\sim_{n+1}$ equivalent. Also, for any $1 \leq i \leq n+1$ any sub-patterns of length i are $\sim_i$ equivalent since $\mathcal{P}_{n+1} \subset \mathcal{P}_i$. ∎

As a corollary to Proposition 4.1.3, we have the following

Corollary 4.1.5 *Let ω, $\mathcal{P}$, K, and $\mathcal{P}'$ be as in Proposition* 4.1.3*. If $u \in \omega[\mathcal{P}']$ and $|u| < K$, then $au \in \omega[\mathcal{P}']$ for some $a \in Alph(\omega)$.*

Now as the main result the classification of eventually periodic words by maximal abelian pattern complexity, namely the words with maximal abelian pattern complexity $p^*_{ab}(n) \leq n$ for some $n \geq 1$.

Theorem 4.1.6 *Let $\omega \in \{0,1\}^{\mathbb{N}}$ be a recurrent aperiodic word. Then the maximal abelian pattern complexity is $p^*_{ab}(n) = n + 1$ for all $n \geq 1$.*

Proof By Lemma 4.1.1 there exists an infinite pattern $\mathcal{P}$ satisfying $(*)$. Let K be a positive integer, and then by Proposition 4.1.3 there exists a sub-pattern $\mathcal{P}' \subset \mathcal{P}$ so that for two sub-patterns $P, Q \subset \mathcal{P}'$ with $|P| = |Q|$ and $1 \leq |P|, |Q| \leq K$ we know $\omega[P] = \omega[Q]$. Moreover, by Lemma 4.1.2, $\mathcal{P}'$ also satisfies $(*)$.

Now show that for each pair of non-negative integers (k,l), where $1 \leq k + l \leq K$, there exists a word $u \in \omega[\mathcal{P}']$ so that $\Psi(u) = (|u|_0, |u|_1) = (k,l)$. This will be done by induction on $k + l$. Since ω is aperiodic, $\omega[\mathcal{P}']$ will contain both 0 and 1, so the result is true for $k + l = 1$.

Next, suppose the result is proved for all pairs (k,l) where $k + l = n$, and consider the pair $(k, l+1)$ where $k + l + 1 = n + 1$, $k \geq 0$, and $l \geq -1$. Consider the following three cases.

Case 1: $l = -1$. Since $0 \in \omega[\mathcal{P}']$, and $\mathcal{P}'$ has (*), $0^k \in \omega[\mathcal{P}']$, and $\Psi(0^k) = (k, 0)$. Thus the result is true for $l = -1$.

Case 2: $l = 0$. Since both $0, 1 \in \omega[\mathcal{P}']$, and from $(*)$ both $00, 11 \in \omega[\mathcal{P}']$. Then since ω is aperiodic, both $01, 10 \in \omega[\mathcal{P}']$. By $(*)$, $10^k \in \omega[\mathcal{P}']$ and $\Psi(10^k) = (k, 1)$. Thus the result is true for $l = 0$

Case 3: $l > 0$. By the induction hypothesis, there exists a word $u \in \omega[\mathcal{P}']$ so that $\Psi(u) = (k,l)$, and there is at least one occurrence of 1 in u since $l > 0$. If u ends in a 1, then by $(*)$ $u1 \in \omega[\mathcal{P}']$, and $\Psi(u1) = (k, l+1)$ and we are done. Thus suppose that u ends in a 0. By Corollary 4.1.5 either $0u$ or $1u$ is in $\omega[\mathcal{P}']$. If $1u \in \omega[\mathcal{P}']$ then $\Psi(1u) = (k, l+1)$ and we are done. Thus suppose $\omega[\mathcal{P}']$ does not contain $1u$, so $0u \in \omega[\mathcal{P}']$

Let $u' = 0u0^{-1}$. Then $u' \in \omega[\mathcal{P}']$, $\Psi(u') = \Psi(u)$, and u' is a cyclic shift of u. If u' ends in a 1 then $u'1 \in \omega[\mathcal{P}']$ and we are done, otherwise u' ends with a 0. If $1u' \in \omega[\mathcal{P}']$ then we are done, otherwise by Corollary 4.1.5 $0u' \in \omega[\mathcal{P}']$. In this case, set $u'' = 0u'0^{-1}$ and repeat as above. Since there is at least one occurrence of 1 in u there will eventually be a conjugate of u which ends in a 1, say $v1 \in \omega[\mathcal{P}']$ with $\Psi(v1) = (k,l)$. Thus $v11 \in \omega[\mathcal{P}']$ and $\Psi(v11) = (k, l+1)$. Thus the result is true for $l > 0$. ∎

Therefore, if ω is a binary recurrent aperiodic word, then it has maximal abelian pattern complexity $p^*_{ab}(n) = n + 1$ for all $n \geq 1$.

4.2 Abelian Complexity of C-Balanced Words

In this section we will investigate the link between the notions of abelian complexity and C-balance. To be more specific, we will investigate the upper bound for abelian complexity of C-balanced words.

As stated in Section 2.3.2, the upper bound for abelian complexity of a word ω over a k-letter alphabet is

$$\rho^{ab}_{\omega}(n) \leq l_k(n) = \binom{n+k-1}{k-1}.$$

A natural question to ask is, are there words which achieve this upper bound? The answer to this question is yes, as any word with maximal factor complexity will have maximal abelian complexity.

Recall Lemma 2.3.4 ([40]): For a word $\omega \in \mathcal{A}^{\mathbb{Z}} \cup \mathcal{A}^{\mathbb{N}}$, the function ρ^{ab}_{ω} is bounded if and only if ω is C-balanced for some positive integer C, namely the abelian complexity is bounded by $(C+1)^{|\mathcal{A}|}$. Thus an initial link is seen between abelian complexity and the C-balance property. The first question to ask is if the bound $\rho^{ab}_{\omega}(n) \leq (C+1)^{|\mathcal{A}|}$ is optimal? The answer to this question is no because this much variation of the values in the Parikh vectors does not take into account that the sum of the values in the Parikh vector must be equal to n.

In Section 4.2.1 we develop a way to calculate the abelian complexity of aperiodic balanced words. This calculation will involve the connection between aperiodic balanced binary words and constant gap sequences. Thus given a balanced word ω over a k-letter alphabet, we can find the maximal value for the abelian complexity of ω. Then in Section 4.2.2 we will calculate an upper bound for abelian complexity of C-balanced words. In Section 4.2.3 we calculate the maximal abelian complexity of balanced aperiodic words as well as some balanced periodic words. We also give a conjecture about words which achieve the abelian complexity upper bound found in Section 4.2.2.

4.2.1 Abelian Complexity of Balanced Aperiodic Words

The goal of this section is to determine the abelian complexity of balanced aperiodic words over an alphabet $\mathcal{A}$, where $|\mathcal{A}| \geq 3$. For this we will need a new definition. A word $G \in \mathcal{A}^{\mathbb{N}}$ is of *constant gap* if for each $a \in \mathcal{A}$ there is a period p_a so that if $G_i = a$ then $G_{i+p_a} = a$. The following proposition should be clear.

Proposition 4.2.1 ([4]) *A constant gap sequence $G \in \mathcal{A}^{\mathbb{N}}$ is periodic and balanced.*

Proof Let p be the least common multiple of the set of integers $\{ p_a \mid a \in A \}$, where p_a is the period of $a \in \mathcal{A}$. Then for each $i \geq 0$ we have $G_i = G_{i+p}$, so G is periodic.

Then for any factor u of G, with $|u| = n$, we have for each $a \in \mathcal{A}$

$$\left\lfloor \frac{n}{p_a} \right\rfloor \leq |u|_a \leq \left\lfloor \frac{n}{p_a} \right\rfloor + 1$$

Thus for any factors u, v of G with $|u| = |v|$, we have $||u|_a - |v|_a| \leq 1$, and G is balanced. ∎

In [4], the authors describe the possible frequencies for the letters in a constant gap sequence. For each $k \geq 1$, there will be finitely many constant gap sequences over k letters. Here is a list of the constant gap sequences, up to a permutation or shift, on k letters for $k = 1, 2, 3, 4$:

$$\begin{array}{ll} k = 1 & (1)^\infty \\ k = 2 & (12)^\infty \\ k = 3 & (123)^\infty; (1213)^\infty \\ k = 4 & (1234)^\infty; (121314)^\infty; (12131214)^\infty; (123124)^\infty \end{array}$$

We suspect that there are 10 constant gap sequences over a 5 letter alphabet, and at least 24 over a 6 letter alphabet. The number of constant gap sequences that exist over an alphabet of size k, for an arbitrary k, is still an open problem, see [4].

For a constant gap sequence $G = (g)^\infty = (g_1 g_2 \cdots g_p)^\infty$, call g the *base* of G. Let u be a factor of G so that $|u| < p$. Call the factor v of G the *compliment* of u in G if uv is a cyclic shift of the base g.

Another note to make is about the abelian complexity of a constant gap sequence G. For any $n \geq 1$ there exists numbers $m \geq 0$ and $b \equiv (n \mod p)$ so that $n = mp + b$, and for any factor u of length n of G, u will contain m copies of a cyclic shift of g, and then a cyclic factor of length b of g. Thus $\rho^{ab}_G(n) = \rho^{ab}_G(b)$. Thus the sequence of abelian complexity values $(\rho^{ab}_G(i))_{i \geq 1}$ is periodic, since $\rho^{ab}_G(i) = \rho^{ab}_G(i + p)$ for each $i \geq 1$. Moreover, the sequence $(\rho^{ab}_G(i))_{1 \leq i < p}$ is a palindrome. Suppose that u_b is a factor of G of length $b < p$, and let $v_b = \Psi(u_b)$. Let u'_b be the compliment of u, so $|u'_b| = p - b$. Thus $\Psi(g) = \Psi(u_b u'_b) = \Psi(u_b) + \Psi(u'_b)$, so $\Psi(u'_b) = \Psi(g) - \Psi(u_b)$. Thus, a Parikh vector of a factor of length b will determine a Parikh vector of a factor of length $p - b$. Therefore, $\rho^{ab}_G(b) = \rho^{ab}_G(p - b)$.

In [22] and [23] the following characterization of balanced words involving constant gap sequences and balanced binary words is given, also listed and proven in [4].

Theorem 4.2.2 *Let s be a balanced word over $\{0, 1\}$. Construct a new word ω by replacing in s, the subsequence of 0's by a constant gap sequence G on alphabet $\mathcal{A}_1$, and* *the subsequence of 1's by a constant gap sequence H on a disjoint alphabet $\mathcal{A}_2$. Then ω is balanced on the alphabet $\mathcal{A}_1 \cup \mathcal{A}_2$.*

Theorem 4.2.3 *Let $\omega \in \mathcal{A}^{\mathbb{N}}$ be a balanced aperiodic word. Then there exists a partition of $\mathcal{A}$ into two sets $\mathcal{A}_0$ and $\mathcal{A}_1$ so that the word s defined by:*

$$u_n = 0 \quad \text{if } \omega_n \in \mathcal{A}_0$$
$$u_n = 1 \quad \text{if } \omega_n \in \mathcal{A}_1$$

is balanced. Moreover, the words z_0 and z_1 constructed from ω by keeping only the letters from $\mathcal{A}_0$ and $\mathcal{A}_1$ respectively have constant gaps.

Thus, a balanced aperiodic word is generated from a balanced aperiodic binary word on $\{a, b\}$ by replacing the a's with a constant gap sequence G and the b's by a constant gap sequence H, where $Alph(G) \cap Alph(H) = \emptyset$. Since the balanced aperiodic binary words are precisely the Sturmian words, we can use properties of Sturmian words to help calculate the abelian complexity of balanced aperiodic words on an alphabet of size $k \geq 3$.

We now recall some properties of Sturmian words from Section 2.4.2. First, for any Sturmian word s and for each $n \geq 1$, $\rho_s^{ab}(n) = 2$. Thus for each $n \geq 1$ there exist unique values r_n and t_n so that $\Psi_s(n) = \{(r_n, t_n), (r_n - 1, t_n + 1)\}$. Secondly, for any Sturmian word s, there a mechanical word $s_{\alpha,\rho}$ so that $s_{\alpha,\rho}(i) = s(i)$ for each $i \geq 0$. Recall that the prefixes of $c_\alpha = s_{\alpha,\alpha}$ are left special factors, and $\mathcal{F}(c_\alpha) = \mathcal{F}(s_{\alpha,\rho})$ for any $0 \leq \rho$, and $s_{\alpha,0} = 0c_\alpha$.

Sturmian words can also be generated by cutting words, $K_{\beta,\rho}$. For the cutting word $K_{\beta,0}$, where $\beta = \alpha/(1-\alpha)$, we have $K_{\beta,0} = s_{\alpha,0}$ and $\mathcal{F}(K_{\beta,0}) = \mathcal{F}(s_{\alpha,0})$. Thus the prefix of length $n+1$, for each $n \geq 0$, of $K_{\beta,0}$ is of the form $0p_n$ where p_n is the left special factor of length n of $K_{\beta,0}$, and thus $1p_n$ is also a factor of $K_{\beta,0}$.

The word $K_{\beta,0}$ is generated by considering the line $y = \beta x$ and how it travels through the plane. For positive integers a and b, if the line passes through the square $[a, a+1] \times [b, b+1]$, then there is some n so that $I_n = (a+1, b)$ (where I_n is defined for cutting words in Section 2.4.2) and $a + b + 1 = n$. The next proposition will, for each $n \geq 1$, connect the point I_n with the unique values r_n and t_n.

Lemma 4.2.4 *Given a Sturmian word w with slope α, define the cutting word $K_{\beta,0}$, with $\beta = \alpha/(1-\alpha)$, and $I_n = (u_n, v_n)$ as above. Then*

$$\Psi_w(n) = \{I_n, I_n + (-1, 1)\} = \{(u_n, v_n), (u_n - 1, v_n + 1)\}$$

for each $n > 0$.

Proof For $\beta = \alpha/(1-\alpha)$, $K_{\beta,0} = s_{\alpha,0}$. Since w has slope α,

$$\mathcal{F}(w) = \mathcal{F}(s_{\alpha,0}) = \mathcal{F}(K_{\beta,0}).$$

If

$$K_{\beta,0}(0)K_{\beta,0}(1)\dots K_{\beta,0}(n) = 0p_n$$

is the prefix of $K_{\beta,0}$, and also of $s_{\alpha,0}$, of length $n+1$ then $\{(|p_n|_0+1, |p_n|_1), (|p_n|_0, |p_n|_1+1)\}$ are the possible Parikh vectors for factors of length $n+1$ of $K_{\beta,0}$, because of Theorem 2.3.6, and thus for w as well. For each $n \geq 0$, $I_n = (u_n, v_n)$ and $u_n + v_n = n$ by definition.

Induction on the prefix length n will be used to show the claim is true. For $n = 1$, $I_0 = (0,0)$ and $I_1 = (1,0)$. Then $K_{\beta,0}(0) = 0$, $p_0 = \varepsilon$ so $\Psi_w(1) = \Psi_{K_{\beta,0}}(1) = \{I_1, I_1 + (-1,1)\} = \{(1,0),(0,1)\}$.

Then suppose $n \geq 1$ and the claim is true for all lengths less than or equal to n, so $I_n = (|p_{n-1}|_0 + 1, |p_{n-1}|_1) = (u_n, v_n)$ and $\Psi_w(n) = \Psi_{K_{\beta,0}}(n) = \{I_n, I_n + (-1,1)\}$.

Case 1: $K_{\beta,0}(n+1) = 0$. Then $p_n = p_{n-1}0$, so $(|p_n|_0, |p_n|_1) = (|p_{n-1}|_0 + 1, |p_{n-1}|_1)$ and $\Psi_{K_{\beta,0}}(n+1) = \{(u_n+1, v_n), (u_n, v_n+1)\}$. Thus

$$\begin{aligned} K_{\beta,0}(n+1) = 0 = v_{n+1} - v_n \quad &\Rightarrow \quad v_{n+1} = v_n \\ n = u_{n+1} + v_{n+1} \quad &\Rightarrow \quad u_{n+1} = u_n + 1 \end{aligned}$$

Thus $I_{n+1} = (u_{n+1}, v_{n+1}) = (u_n + 1, v_n)$.

Case 2: $K_{\beta,0}(n+1) = 1$. Then $p_n = p_{n-1}1$, so $(|p_n|_0, |p_n|_1) = (|p_{n-1}|_0, |p_{n-1}|_1 + 1)$ and $\Psi_{K_{\beta,0}}(n+1) = \{(u_n, v_n+1), (u_n - 1, v_n + 2)\}$. Thus

$$\begin{aligned} K_{\beta,0}(n+1) = 1 = v_{n+1} - v_n \quad &\Rightarrow \quad v_{n+1} = v_n + 1 \\ n = u_{n+1} + v_{n+1} \quad &\Rightarrow \quad u_{n+1} = u_n \end{aligned}$$

Thus $I_{n+1} = (u_{n+1}, v_{n+1}) = (u_n, v_n + 1)$.

In either case, $\Psi_w(n+1) = \Psi_{K_{\beta,0}}(n+1) = \{I_{n+1}, I_{n+1} + (-1,1)\}$ ∎

Lemma 4.2.5 *Let w be a Sturmian word and $K_{\beta,0}$ be as in Lemma 4.2.4. For all positive integers k, l, a, and b with $0 \leq a < k$ and $0 \leq b < l$; there is an $n > 0$ so that $I_n = (r_n, t_n)$ with $r_n \equiv (a \mod k)$ and $t_n \equiv (b \mod l)$. Moreover, there are infinitely many n for which this holds.*

Proof Let k, l, a, and b be as in the hypothesis. Any line with an irrational slope will have a dense orbit in the torus generated by $[0,1] \times [0,1]$. Thus, the line $y = \beta x$ will also have a dense orbit in the torus generated by $[0,k] \times [0,l]$, and for each $n > 0$ the points $I_n = (r_n, t_n)$ will correspond to the points $I'_n = ((r_n \mod k), (t_n \mod l))$ on the torus.

Thus since the line that generates $K_{\beta,0}$ has a dense orbit in the torus, there is an $n > 0$ so that $I'_n = (a,b)$, and thus $r_n \equiv (a \mod k)$ and $t_n \equiv (b \mod l)$. This happens infinitely many times due to the density of the path. ∎

Thus for any factor u of length n of a Sturmian word w over $\{0,1\}$ where $u = w_i w_{i+1} \ldots w_{i+n-1}$ for some i, $I_i = (r_i, t_i)$, so there are r_i occurrences of 0 to the left of u and t_i occurrences of 1 to the left of u. Since $|u| = n$ then $\Psi(u) \in \{(r_n, t_n), (r_n - 1, t_n + 1)\}$ because $\Psi_w(n) = \{(r_n, t_n), (r_n - 1, t_n + 1)\}$.

Lemma 4.2.6 *Let w be a Sturmian word and $K_{\beta,0}$, k, l, a, and b be as in Lemma 4.2.5. For any $n > 0$, there exist n_1, n_2 so that $I'_{n_1} = (a, b) = I'_{n_2}$, and $\Psi(u_{n_1}) = (r_n, t_n)$ and $\Psi(u_{n_2}) = (r_n - 1, t_n + 1)$, where $u_{n_1} = w_{n_1} w_{n_1+1} \ldots w_{n_1+n-1}$ and $u_{n_2} = w_{n_2} w_{n_2+1} \ldots w_{n_2+n-1}$ are factors of w of length n.*

Proof There are factors U and V of length n of w so that $\Psi(U) = (r_n, t_n)$ and $\Psi(V) = (r_n - 1, t_n + 1)$. Thus let $I'_{n_U} = (a_U, b_U)$ and $I'_{n_V} = (a_V, b_V)$ be so that $U = w_{n_U} \ldots w_{n_U+n-1}$ and $V = w_{n_V} \ldots w_{n_V+n-1}$. Then the line that generates $K_{\beta,0}$ enters the square $[a_U, a_U + 1] \times [b_U, b_U + 1]$ and there is a small interval, J_U , so that any time the line passes through this interval, the word U will be the next n letters. A similar interval, J_V will exist for the square $[a_V, a_V + 1] \times [b_V, b_V + 1]$ where the word V will be the next n letters.

Then considering the interval J_U in the same relative location on the square $[a, a + 1] \times [b, b + 1]$, when the line passes through J_U the resulting n letters will have (r_n, t_n) as their Parikh vector. Likewise, since the path of the line that generates $K_{\beta,0}$ is dense in the orbit, when the line passes through the interval J_V on the square $[a, a+1] \times [b, b+1]$, the resulting n letters will have $(r_n - 1, t_n + 1)$ as their Parikh vector. Since the path of the line that generates $K_{\beta,0}$ is dense in the orbit, there is an n_1 so that $I'_{n_1} = (a, b)$ and the line enters the square $[a, a + 1] \times [b, b + 1]$ in J_U and an n_2 so that $I'_{n_2} = (a, b)$ and the line enters the square $[a, a+1] \times [b, b+1]$ in J_V. Thus for $u_{n_1} = w_{n_1} w_{n_1+1} \ldots w_{n_1+n-1}$ and $u_{n_2} = w_{n_2} w_{n_2+1} \ldots w_{n_2+n-1}$ we have $\Psi(u_{n_1}) = (r_n, t_n)$ and $\Psi(u_{n_2}) = (r_n - 1 t_n + 1)$. ∎

Thus for w, k, l, a, and b as in the previous lemmas, there exist factors u and v of length n of w so that $\Psi(u) = (r_n, t_n)$, $\Psi(v) = (r_n - 1, t_n + 1)$, and there are $(a \mod k)$ occurrences of 0 and $(b \mod l)$ occurrences of 1 to the left of both u and v.

Now that we have these Lemmas, let's look at the main result of this section dealing with the abelian complexity of balanced aperiodic words on an alphabet of size at least 3.

Theorem 4.2.7 *Let $\omega \in \mathcal{A}^{\mathbb{N}}$ be a balanced aperiodic word, with $|\mathcal{A}| \geq 3$. Thus ω is the image of a Sturmian word s over $\{0,1\}$ obtained by replacing the subsequence of 0's by a constant gap sequence G_0 on alphabet $\mathcal{A}_0$, and the subsequence of 1's by a constant gap sequence G_1 on a disjoint alphabet $\mathcal{A}_1$, where $\mathcal{A}_0 \cup \mathcal{A}_1 = \mathcal{A}$. For each $n > 0$, let r_n and t_n be unique so that $\Psi_s(n) = \{(r_n, t_n), (r_n - 1, t_n + 1)\}$. Then:*

$$\rho^{ab}_{\omega}(n) = \rho^{ab}_{G_0}(r_n) \cdot \rho^{ab}_{G_1}(t_n) + \rho^{ab}_{G_0}(r_n - 1) \cdot \rho^{ab}_{G_1}(t_n + 1)$$

Proof By Theorems 4.2.2 and 4.2.3, ω is the image of a Sturmian word s over $\{0,1\}$ obtained by replacing the subsequence of 0's by a constant gap sequence G_0 on alphabet $\mathcal{A}_0$, and the subsequence of 1's by a constant gap sequence G_1 on a disjoint alphabet $\mathcal{A}_1$, where $\mathcal{A}_0 \cup \mathcal{A}_1 = \mathcal{A}$.

Consider the factors of length n of ω. Each factor of length n of ω corresponds to a factor of length n of s where the 0's have been replaced by letters in G_0 and the 1's have been replaced by letters in G_1. Thus to find $\rho^{ab}_{\omega}(n)$ we need to consider the number of ways to map factors of length n of s by replacing the 0's by consecutive letters from G_0 and the 1's by consecutive letters from G_1.

So for a factor u of length n of s, $\Psi(u) \in \{(r_n, t_n), (r_n - 1, t_n + 1)\}$. By Lemma 4.2.5, the first 0 in u can be replaced by any letter in the base of G_0, and the first 1 can be replaced by any letter in the base of G_1. By Lemma 4.2.6 the image of u in ω can have either r_n or $r_n - 1$ letters from G_0 and either t_n or $t_n + 1$ letters from G_1 respectively.

Thus there are $\rho^{ab}_{G_0}(r_n)$ ways to select r_n distinct consecutive letters from G_0, and for each of those ways there are $\rho^{ab}_{G_1}(t_n)$ ways to select t_n distinct consecutive letters from G_1, and all possible ways can occur. Thus if u is a factor of s with $\Psi(u) = (r_n, t_n)$, there are $\rho^{ab}_{G_0}(r_n) \cdot \rho^{ab}_{G_1}(t_n)$ possible ways to select distinct (not having the same Parikh vector) consecutive letters from G_0 and G_1. Likewise, if u is a factor of s with $\Psi(u) = (r_n - 1, t_n + 1)$ there are $\rho^{ab}_{G_0}(r_n - 1) \cdot \rho^{ab}_{G_1}(t_n + 1)$ possible ways to select distinct consecutive letters from G_0 and G_1. Since these methods are independent we have $\rho^{ab}_{G_0}(r_n) \cdot \rho^{ab}_{G_1}(t_n) + \rho^{ab}_{G_0}(r_n - 1) \cdot \rho^{ab}_{G_1}(t_n + 1)$ ways to select distinct letters from G_0 and G_1 to replace the 0's and 1's in a factor of length n of s.

Therefore, any factor of length n of ω will be constructed from either r_n consecutive letters from G_0 and t_n consecutive letters from G_1, or $r_n - 1$ consecutive letters from G_0 and $t_n + 1$ consecutive letters from G_1. Thus we have $\rho^{ab}_{\omega}(n) = \rho^{ab}_{G_0}(r_n) \cdot \rho^{ab}_{G_1}(t_n) + \rho^{ab}_{G_0}(r_n - 1) \cdot \rho^{ab}_{G_1}(t_n + 1)$. ∎

Using this formula to find the abelian complexity of a balanced word over a 4 letter alphabet, we would consider what constant gap sequences are used to create it. We could use a constant gap sequence over three letters with another constant gap sequence over one letter, or two constant gap sequences over two letters.

Example Suppose $\omega \in \mathcal{A}^{\mathbb{N}}$ is a balanced word, $|\mathcal{A}| = 5$, and ω is created from a Sturmian word s over $\{a, b\}$ by replacing the subsequence of a's by a constant gap sequence $G_a = (123124)^{\infty}$, and the subsequence of b's by a constant gap sequence $G_b = (5)^{\infty}$. Then $(\rho^{ab}_{G_a}(i))_{i \geq 1} = (4, 5, 2, 5, 4, 1)^{\infty}$ and $(\rho^{ab}_{G_b}(i))_{i \geq 1} = (1)^{\infty}$. Thus there exists an n so that $I_n = (u_n, v_n)$ and $u_n \equiv (2 \mod 6)$, note that for any n $(v_n \mod 1) \equiv 0$, and so

$$\rho^{ab}_{\omega}(n) = \rho^{ab}_{G_a}((u_n \mod 6)) \cdot \rho^{ab}_{G_b}(v_n) + \rho^{ab}_{G_a}((u_n - 1 \mod 6)) \cdot \rho^{ab}_{G_b}(v_n + 1) = 5 * 1 + 4 * 1 = 9$$

We now look at the upper bound for abelian complexity of a C-balanced word.

4.2.2 Upper Bound For Abelian Complexity

As stated in the introduction, if we consider finite words of length n over a k-letter alphabet $\mathcal{A}$, $k \geq 2$, there will be at most $l_k(n) = \binom{n+k-1}{k-1}$ possible distinct Parikh vectors. Now we will consider looking at the set of possible Parikh vectors in a different way. Let $\mathcal{A} = \{0, 1, \ldots, k-1\}$ and $u \in \mathcal{A}^+$ with $|u| = n$. Thus,

$$\sum_{i=0}^{k-1} |u|_i = n \quad \Longleftrightarrow \quad |u|_0 = n - \sum_{i=1}^{k-1} |u|_i,$$

so we see the Parikh vector for u only depends on $k-1$ values since we know the word has length n. Now consider the set $\Lambda_{k,n}$, which is the subset of $\mathbb{Z}^{k-1}$, the set of $(k-1)$-tuples of integers, where

$$\Lambda_{k,n} = \left\{ (x_1, x_2, \ldots, x_{k-1}) \in \mathbb{Z}^{k-1} \;\middle|\; \sum_{i=1}^{k-1} x_i \leq n; \quad \forall i \;\; x_i \in \mathbb{Z}, \;\; x_i \geq 0 \right\}.$$

Then to see the correspondence between $\Lambda_{k,n}$ and the collection of all possible Parikh vectors for words of length n over a k-letter alphabet, consider the collection of k-tuples of non-negative integers $(i_0, i_1, \ldots, i_{k-1})$ where $i_0 + i_1 + \cdots i_{k-1} = n$, and also note $i_0 = n - (i_1 + \cdots i_{k-1})$. Map such a k-tuple to $\Lambda_{k,n}$ as follows:

$$(i_0, i_1, i_2, \ldots, i_{k-1}) \mapsto (i_1, i_2, \ldots, i_{k-1})$$

This mapping gives a bijective relation, which can be seen as follows. By definition, each $i_j \geq 0$, and since $i_0 + i_1 + \cdots i_{k-1} = n$ we know $i_1 + \cdots i_{k-1} \leq n$ so the mapping is into $\Lambda_{k,n}$. If two k-tuples u and v map to the same $\lambda \in \Lambda_{k,n}$, then for each i we know $u_i = v_i$ and thus $u_0 = n - (u_1 + \cdots + u_{k-1}) = n - (v_1 + \cdots + v_{k-1}) = v_0$, so $u = v$. Then for some $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{k-1}) \in \Lambda_{k,n}$, the k-tuple $(n - (\lambda_1 + \lambda_2 + \cdots + \lambda_{k-1}), \lambda_1, \lambda_2, \ldots, \lambda_{k-1})$, which satisfies the above conditions, would map to λ.

The set $\Lambda_{k,n}$ can be equipped with a metric, giving a discrete metric space. The most useful metric on the space in the sense of C-balance is defined as follows, for $\lambda, \kappa \in \Lambda_{k,n}$:

$$\eta(\lambda, \kappa) = \sup \left\{ |\lambda_1 - \kappa_1|, |\lambda_2 - \kappa_2|, \ldots, |\lambda_{k-1} - \kappa_{k-1}|, \left| \sum_{i=1}^{k-1} \lambda_i - \sum_{i=1}^{k-1} \kappa_i \right| \right\}.$$

In the case of the metric, the difference $|\lambda_i - \kappa_i|$ corresponds to the difference in the number of occurrences of the letter i, and the term $|\sum \lambda_i - \sum \kappa_i|$ corresponds to the difference in the number of occurrences of the letter 0.

Thus for an infinite word $\omega \in \mathcal{A}^{\mathbb{N}}$, the set $\Psi_\omega(n)$ can be embedded as a subset of $\Lambda_{k,n}$. Then if the word ω is C-balanced, we know if $\lambda, \kappa \in \Lambda_{k,n}$ correspond to Parikh vectors of factors of ω of length n then $\eta(\lambda,\kappa) \leq C$. Therefore, if a word ω is C-balanced the embedding of $\Psi_\omega(n)$ would be bounded by a sphere, in the sense of the metric η, of diameter C and thus the abelian complexity, ρ^{ab}_ω, would be bounded above by the cardinality of such a sphere.

Theorem 4.2.8 *Let $|\mathcal{A}| = k \geq 2$, and $\omega \in \mathcal{A}^{\mathbb{N}}$ be C-balanced. Then for each n,*

$$\rho^{ab}_\omega(n) \leq \binom{k}{\lfloor \frac{Ck}{2} \rfloor}^{[C+1]}.$$

Proof Let $\mathcal{A} = \{0, 1, \ldots, k-1\}$, and $\omega \in \mathcal{A}^{\mathbb{N}}$ be C-balanced. For any $u, v \in \mathcal{F}_\omega(n)$ we know $\left||u|_i - |v|_i\right| \leq C$, so for sufficiently large n there exist integers t_i so that $t_i \leq |u|_i \leq t_i + C$ for each $i = 1, 2, \ldots, k-1$. Let $B_{\omega,n}$ be the subset of $\Lambda_{k,n}$ so that:

$$B_{\omega,n} = \{\ \lambda \in \Lambda_{k,n}\ \mid t_i \leq \lambda_i \leq t_i + C \text{ for each } i = 1, 2, \ldots, k-1\ \}.$$

Another way to view $B_{\omega,n}$ is as a hypercube with side length C in $\mathbb{Z}^{k-1}$.

For the sake of finding the sphere of diameter C with the largest possible cardinality we can assume that each $t_i \geq 0$ and $\sum (t_i + C) \leq n$, i.e. $B_{\omega,n}$ is totally contained in $\Lambda_{k,n}$. Then it should be clear that the image of $\Psi_\omega(n)$ in $\Lambda_{k,n}$ is a subset of $B_{\omega,n}$. For each $m = 0, 1, \ldots, C(k-1)$ define

$$D_m = \left\{\ \lambda \in B_{\omega,n}\ \middle|\ \sum_{i=1}^{k-1} (\lambda_i - t_i) = m\ \right\}$$

to be the set of parallel hyperplanes in $B_{\omega,n}$, where all factors of ω corresponding to points in D_m would have the same number of occurrences of the letter 0. Thus $B_{\omega,n}$ is the disjoint union of all the D_m, and the image of $\Psi_\omega(n)$ will be a subset of $D_m \cup D_{m+1} \cup \cdots \cup D_{m+C}$ for some $m = 0, 1, \ldots, C(k-2)$. Thus for the largest $(C+1)$ of the D_m sets,

$$\rho^{ab}_\omega(n) = |\Psi_\omega(n)| \leq \sum_{i=0}^{C} |D_{m+i}|.$$

Thus we need to find the cardinality of the sets D_m, and then we can find an upper bound for the abelian complexity.

For each m, the cardinality of D_m is the number of ways to write m as the sum of $k-1$ non-negative integers less than or equal to C, or $m = i_1 + i_2 + \cdots i_{k-1}$ where $0 \leq i_j \leq C$ for each $j = 1, 2, \ldots, k-1$. For each m, this value is known. This is the same value as the coefficient of x^m in the expansion of $(1 + x + \cdots + x^C)^{k-1}$. Thus if $(1 + x + \cdots + x^C)^{k-1} = a_0 + a_1 x + a_2 x^2 + \cdots + a_{C(k-1)} x^{C(k-1)}$, we know

$$a_0 < a_1 < \cdots < a_{\lfloor \frac{C(k-1)}{2} \rfloor} = a_{\lceil \frac{C(k-1)}{2} \rceil} > \cdots > a_{C(k-1)},$$

and thus $|D_m| = a_m$. Then considering the coefficients of $(1+x+\cdots+x^C)^k = b_0+b_1x+\cdots b_{Ck}x^{Ck}$, we have $b_{\lfloor \frac{Ck}{2} \rfloor} = b_{\lceil \frac{Ck}{2} \rceil}$ are the greatest coefficients, and they are the sum of the greatest $C+1$ coefficients of $(1+x+\cdots+x^C)^{k-1}$.

To find the coefficients of $(1+x+\cdots+x^C)^k = b_0+b_1x+\cdots b_{Ck}x^{Ck}$, we will need to define some new notation. We first define

$$\binom{k}{m}^{[2]} = \binom{k}{m},$$

and for $C \geq 2$ we define recursively

$$\binom{k}{m}^{[C+1]} = \sum_{i=0}^{m} \binom{k}{m-i}\binom{m-i}{i}^{[C]}.$$

From [17], the coefficient b_m of x^m in the expansion of $(1+x+x^2+\cdots+x^C)^k$ is

$$\binom{k}{m}^{[C+1]} = \binom{k}{m}\binom{m}{0}^{[C]} + \binom{k}{m-1}\binom{m-1}{1}^{[C]} + \binom{k}{m-2}\binom{m-2}{2}^{[C]} + \cdots$$

where $\binom{s}{t} = 0$ if $s \leq 0$, $t \leq 0$, or $s < t$. Thus using this notation from Euler we have a nice way to write the value for the upper bound of the abelian complexity of a infinite word ω on a k-letter alphabet that is C-balanced. We have:

$$\rho_\omega^{ab}(n) \leq \binom{k}{\lfloor \frac{Ck}{2} \rfloor}^{[C+1]}.$$

■

In the case of a balanced word ω, we have $\rho_\omega^{ab}(n) \leq \binom{k}{\lfloor \frac{k}{2} \rfloor}$, and for $k = 1,2,3,4,5,6,7$ we have $\rho_\omega^{ab}(n)$ is bounded by $1,2,3,6,10,20,35$ respectively. As discussed in the previous section, the abelian complexity for an aperiodic balanced word ω on k letters can be determined by the abelian complexity of the constant gap sequences that are used to create it.

For example, if an alphabet is composed of 5 letters then a recurrent aperiodic balanced word over that alphabet will not achieve the maximum values of 10.

An aperiodic balanced word ω on a 5 letter alphabet has $\rho_\omega^{ab}(n) \leq 9$.

Thus if ω achieves the maximal abelian complexity, it must be (eventually) periodic. In the following section we will investigate the maximum value the abelian complexity can achieve. At the time of this writing, only balanced words have been considered. This is because there exist nice ways to generalize recurrent balanced words, namely the method described in Section 4.2.1.

4.2.3 Calculating Abelian Complexity

The previous two sections dealt with some theorems and lemmas, but this section will look at some of the numbers. Initially work has been done with a k-letter alphabet, with $k \leq 6$. This initial work leads to a conjecture about C-balanced words with maximal abelian complexity.

In Section 4.2.1 the constant gap sequences on k letters for $k = 1, 2, 3, 4$ are listed. These sequences are listed again below, but listed with them is the periodic abelian complexity sequence.

$k=1$	CGS	$(1)^\infty$			
	ρ^{ab}	$(1)^\infty$			
$k=2$	CGS	$(12)^\infty$			
	ρ^{ab}	$(2,1)^\infty$			
$k=3$	CGS	$(123)^\infty$	$(1213)^\infty$		
	ρ^{ab}	$(3,3,1)^\infty$	$(3,2,3,1)^\infty$		
$k=4$	CGS	$(1234)^\infty$	$(121314)^\infty$	$(12131214)^\infty$	$(123124)^\infty$
	ρ^{ab}	$(4,4,4,1)^\infty$	$(4,3,6,3,4,1)^\infty$	$(4,3,5,2,5,3,4,1)^\infty$	$(4,5,2,5,4,1)^\infty$

Any balanced aperiodic word over an alphabet of 5 or less letters will be created from a Sturmian word over any two of the above constant gap sequences. For example, to create a balanced word over a 5 letter alphabet we could use a constant gap sequence over 4 letters and another over 1 letter, or a sequence over 3 letters and another over 2 letters. While we can use two constant gap sequences over 4 letters to create a balanced aperiodic word over 8 letters, we can not find all balanced aperiodic words over 8 letters with the above sequences.

Let ω be a balanced aperiodic word over a k-letter alphabet, where $2 < k \leq 5$. Then ω can be created from a Sturmian word s and constant gap sequences g and h, as described in Section 4.2.1. To see the possible values for $\rho^{ab}_\omega(n)$, consider the set

$$\mathcal{S}_\omega = \left\{ \rho^{ab}_g(r) \cdot \rho^{ab}_h(t) + \rho^{ab}_g(r-1) \cdot \rho^{ab}_g(t+1) \mid r \geq 1,\ \ t \geq 1 \right\},$$

and thus $\rho^{ab}_\omega(n) \in \mathcal{S}_\omega$. Since the abelian complexity sequence for both g and h are periodic, $|\mathcal{S}_\omega|$ will be finite. Thus the maximal value for $\rho^{ab}_\omega(n)$ will be $\max \mathcal{S}_\omega$.

Example (1) Suppose ω is a balanced aperiodic word constructed from a Sturmian word s and constant gap sequences $g = (123124)^\infty$ and $h = (5)^\infty$. Then

$$\mathcal{S}_\omega = \{5+4, 2+5, 5+2, 4+5, 1+4\} = \{5, 7, 9\}$$

and the maximal value for $\rho^{ab}_\omega(n)$ will be 9.

(2) Suppose ω is a balanced aperiodic word constructed from a Sturmian word s and constant gap sequences $g = (1213)^\infty$ and $h = (45)^\infty$. Then

$$\mathcal{S}_\omega = \{2 \cdot 2 + 3 \cdot 1, 3 \cdot 2 + 2 \cdot 1, 1 \cdot 2 + 3 \cdot 1, 2 \cdot 1 + 3 \cdot 2, 3 \cdot 1 + 2 \cdot 2, 1 \cdot 1 + 3 \cdot 2\} = \{5, 7, 8\}$$

and the maximal value for $\rho_\omega^{ab}(n)$ will be 8.

It turns out that for a recurrent aperiodic balanced word ω over a 5 letter alphabet, $\rho_\omega^{ab}(n) \leq 9$. This comes from considering the abelian complexity sequence of constant gap sequences and the possible combinations of them to create a recurrent aperiodic balanced word over a 5 letter alphabet. Thus no balanced aperiodic word over a 5 letter alphabet will achieve the maximum value of $\binom{5}{\lfloor \frac{5}{2} \rfloor} = 10$. Thus if a balanced word over a 5 letter alphabet does have abelian complexity of 10, for some factor length, it is not an aperiodic word.

As stated before, we believe there are 10 constant gap sequences over 5 letters. Each of these 10 sequences (up to permutation or shift) are listed with its periodic abelian complexity sequence.

$(12345)^\infty$ $(5,5,5,5,1)^\infty$	$(12341235)^\infty$ $(5,5,7,2,7,5,5,1)^\infty$
$(123125)^\infty$ $(5,6,3,6,5,1)^\infty$	$(123124125)^\infty$ $(5,7,3,9,9,3,7,5,1)^\infty$
$(12131415)^\infty$ $(5,4,8,4,8,4,5,1)^\infty$	$(1213121412131215)^\infty$ $(5,4,7,3,8,5,7,2,7,5,8,3,7,4,5,1)^\infty$
$(121314121315)^\infty$ $(5,4,9,5,7,2,7,5,9,4,5,1)^\infty$	$(132415231425)^\infty$ $(5,6,9,3,7,2,7,3,9,6,2,1)^\infty$
$(123124123125)^\infty$ $(5,6,3,8,7,2,7,8,3,6,5,1)^\infty$	$(121312141215)^\infty$ $(5,4,7,3,9,6,9,3,7,4,5,1)^\infty$

If the method used in the previous example is used with the constant gap sequences above, we find that a balanced aperiodic word over a 6 letter alphabet will have abelian complexity bounded above by 18.

Conjecture 4.2.9 *Suppose $|\mathcal{A}| = k$, $k \geq 5$, and $\omega \in \mathcal{A}^{\mathbb{N}}$ is a recurrent C-balanced word. If there exists an n so that*

$$\rho_\omega^{ab}(n) = \binom{k}{\lfloor \frac{Ck}{2} \rfloor}^{[C+1]}.$$

then ω is eventually periodic.

Work has been done to find words which attain the upper bound found in Theorem 4.2.8. For example, the following words

$$\alpha_5 = (12312412512312412)^\infty$$
$$\alpha_6 = (12312412512312412612312412)^\infty$$

are balanced words which achieve maximal abelian complexity, and it is readily verified that $\rho^{ab}_{\alpha_5}(4) = 10$ and $\rho^{ab}_{\alpha_6}(13) = 20$. At the time of this writing, no balanced word over an alphabet of size at least 7 has been found to achieve the value found through Theorem 4.2.8.

For each $k \geq 1$, there will be finitely many constant gap sequences over k letters, [4], but the number of constant gap sequences that exist over an alphabet of size k, for an arbitrary k, is still an open problem. There seems to be many unanswered questions in finding the optimal upper bound for the abelian complexity of C-balanced words.

Chapter 5

Permutation Complexity

Permutation complexity of aperiodic words is a relatively new notion of word complexity which was first introduced and studied by Makarov [32] based on ideas of S.V. Avgustinovich (see the acknowledgments in [18]), and is based on the idea of an infinite permutation associated to an aperiodic word. For an aperiodic word ω, no two shifts of ω are identical. Thus, given an order on the symbols used to compose ω, no two shifts of ω are equal lexicographically. The infinite permutation associated with ω is the linear order on $\mathbb{N}$ induced by the lexicographic order of the shifts of ω. The permutation complexity of the word ω will be the number of distinct subpermutations of a given length of the infinite permutation associated with ω.

This chapter will have a few different permutation complexity results. Section 5.1 has some preliminary information about infinite permutations induced by an aperiodic word, as well as some basic properties about these infinite permutations. In Section 5.2 we calculate the permutation complexity of the Thue-Morse word. Section 5.3 deals with the permutation complexity of infinite words which are the image of an aperiodic uniformly recurrent word under the doubling map, d. We also give the permutation complexity of the image of Sturmian words under d, and the image of the Thue-Morse word under d.

5.1 Preliminaries

Infinite permutations associated with infinite aperiodic words over a binary alphabet act fairly well-behaved, but many of the arguments used for binary words break down when used with words over more than two symbols. Given a subpermutation of length n of an infinite permutation associated with a binary word, a portion of length $n-1$ of the word can be recovered from the subpermutation. This is not always the case for subpermutations associated with words over 3 or more symbols. For example, consider the permutation $(1\,2\,3)$. If this permutation is associated with a binary word over $\{0, 1\}$, with

$0 < 1$, it could only correspond to the word 00. On the other hand, if this permutation is associated with a word over 3 symbols, suppose $\{0, 1, 2\}$ with $0 < 1 < 2$, then the permutation could be associated with any of 00, 01, or 11.

For binary words the subpermutations depend on the order on the symbols used to compose ω, but the permutation complexity does not depend on the order. For words over 3 or more symbols, not only do the subpermutations depend on the order on the alphabet but so does the permutation complexity. For example, consider the Fibonacci word

$$t = 0100101001001010010100100101\ldots,$$

defined by iterating the morphism $0 \mapsto 01, 1 \mapsto 0$ on the letter 0, and suppose the 1s are replaced by alternating a's and b's to create the word:

$$\hat{t} = 0a00b0a00b00a0b00a0b00a00b0a\ldots.$$

If the symbols in $\hat{t}$ are ordered $0 < a < b$ there will be 5 distinct subpermutations of length 3, and if the symbols are ordered $a < 0 < b$ there will be only 4 distinct subpermutations of length 3.

In this section we will give some basic definitions which will be used in this chapter.

5.1.1 Infinite Permutation

The idea of an infinite permutation that will be here used was introduced in [18]. Since we will be investigating the permutation complexity of infinite words, the set used in the following definition will be $\mathbb{N}$ rather than an arbitrary countable set. To define an *infinite permutation* π, start with a linear order $\prec_\pi$ on $\mathbb{N}$, together with the usual order $<$ on $\mathbb{N}$. To be more specific, an infinite permutation is the ordered triple $\pi = \langle \mathbb{N}, \prec_\pi, < \rangle$, where $\prec_\pi$ and $<$ are linear orders on $\mathbb{N}$. The notation to be used here will be $\pi(i) < \pi(j)$ rather than $i \prec_\pi j$.

5.1.2 Permutations Induced by Words

Given an aperiodic word $\omega = \omega_0\omega_1\omega_2\ldots$ on an alphabet $\mathcal{A}$, fix a linear order on $\mathcal{A}$. We will use the binary alphabet $\mathcal{A} = \{0, 1\}$ with the natural ordering $0 < 1$. Once a linear order is set on the alphabet, we can then define an order on the natural numbers based on the lexicographic order of shifts of ω. Considering two shifts of ω with $a \neq b$, $\omega[a] = \omega_a\omega_{a+1}\omega_{a+2}\ldots$ and $\omega[b] = \omega_b\omega_{b+1}\omega_{b+2}\ldots$, we know $\omega[a] \neq \omega[b]$ because ω is aperiodic. Thus there exists some minimal number $c \geq 0$ so that $\omega_{a+c} \neq \omega_{b+c}$ and $\omega_{a+i} = \omega_{b+i}$ for each $0 \leq i < c$. We call π_ω the infinite permutation associated with ω and say that $\pi_\omega(a) < \pi_\omega(b)$ if $\omega_{a+c} < \omega_{b+c}$, else we say $\pi_\omega(b) < \pi_\omega(a)$.

For natural numbers $a \leq b$ consider the factor $\omega[a, b] = \omega_a \omega_{a+1} \ldots \omega_b$ of ω of length $b - a + 1$. Denote the finite permutation of $\{1, 2, \ldots, b - a + 1\}$ corresponding to the linear order by $\pi_\omega[a, b]$. That is, $\pi_\omega[a, b]$ is the permutation of $\{1, 2, \ldots, b - a + 1\}$ so that for each $0 \leq i, j \leq (b - a)$, $\pi_\omega[a, b](i) < \pi_\omega[a, b](j)$ if and only if $\pi_\omega(a + i) < \pi_\omega(a + j)$. Say that $p = p_0 p_1 \cdots p_n$ is a *(finite) subpermutation* of π_ω if $p = \pi_\omega[a, a + n]$ for some $a, n \geq 0$. For the subpermutation $p = \pi_\omega[a, a + n]$ of $\{1, 2, \cdots, n + 1\}$, we say the *length* of p is $n + 1$.

Denote the set of all subpermutations of π_ω by Perm^ω, and for each positive integer n let

$$\mathrm{Perm}^\omega(n) = \{ \ \pi_\omega[i, i + n - 1] \ \mid i \geq 0 \ \}$$

denote the set of distinct finite subpermutations of π_ω of length n. The *permutation complexity function* of ω is defined as the total number of distinct subpermutations of π_ω of a length n, denoted $\tau_\omega(n) = |\mathrm{Perm}^\omega(n)|$.

Example Let's consider the well-known Fibonacci word,

$$t = 0100101001001010010100100101\ldots,$$

with the alphabet $\mathcal{A} = \{0, 1\}$ ordered as $0 < 1$. We can see $t[2] = 001010\ldots$ is lexicographically less than $t[1] = 100101\ldots$, and thus $\pi_t(2) < \pi_t(1)$.

Then for a subpermutation, consider the factor $t[3, 5] = 010$. We see $\pi_t[3, 5] = (231)$ because in lexicographic order if we have $\pi_t(5) < \pi_t(3) < \pi_t(4)$.

We will also be interested in the form of the subpermutations of π_ω.

Definition For a binary word u of length $n - 1$, say that p has *form* u if

$$p_i < p_{i+1} \iff u_i = 0$$

for each $i = 0, 1, \ldots, n - 2$. Two permutations p and q of $\{1, 2, \ldots, n\}$ have the *same form* if for each $i = 0, 1, \ldots, n - 1$,

$$p_i < p_{i+1} \iff q_i < q_{i+1}.$$

Then given a subpermutation p of π_ω we define the following restrictions of p.

Definition Let $p = \pi_\omega[a, a + n]$ be a subpermutation of the infinite permutation π_ω. The *left restriction of* p, denoted by $L(p)$, is the subpermutation of p so that $L(p) = \pi[a, a + n - 1]$. The *right restriction of* p, denoted by $R(p)$, is the subpermutation of p so that $R(p) = \pi[a + 1, a + n]$. The *middle restriction of* p, denoted by $M(p)$, is the subpermutation of p so that $M(p) = R(L(p)) = L(R(p)) = \pi[a + 1, a + n - 1]$.

For each i, there are $p_i - 1$ terms in p that are less than p_i and there are $n - p_i$ terms that are greater than p_i. Thus consider $i \in \{0, 1, \ldots, n-1\}$ and the values of $L(p)_i$ and $R(p)_i$. If $p_0 < p_{i+1}$ there will be $p_{i+1} - 2$ terms in $R(p)$ less than $R(p)_i$ so we have $R(p)_i = p_{i+1} - 1$. In a similar sense, if $p_n < p_i$ we have $L(p)_i = p_i - 1$. If $p_0 > p_{i+1}$ there will be $p_{i+1} - 1$ terms in $R(p)$ less than $R(p)_i$ so we have $R(p)_i = p_{i+1}$. In a similar sense, if $p_n > p_i$ we have $L(p)_i = p_i$.

The values in $M(p)$ can be found by finding the values in $R(L(p))$ or $L(R(p))$. Since $R(L(p))$ or $L(R(p))$ correspond to the same subpermutation of p, $R(L(p))_i < R(L(p))_j$ if and only if $L(R(p))_i < L(R(p))_j$. Therefore $R(L(p)) = L(R(p))$.

It should also be clear that if there are two subpermutations $p = \pi_\omega[a, a+n]$ and $q = \pi_\omega[b, b+n]$, $a \neq b$, so that $p = q$ then $L(p) = L(q)$, $R(p) = R(q)$, and $M(p) = M(q)$ since if $p = q$ then $p_i < p_j$ if and only if $q_i < q_j$.

5.1.3 General Permutation Complexity Properties

Initially work has been done with infinite binary words (see [6, 18, 32, 33, 34]). Suppose $\mathcal{A} = \{0, 1\}$ and $\omega = \omega_0\omega_1\omega_2 \ldots \in \mathcal{A}^{\mathbb{N}}$ is an aperiodic word. First let's look at some remarks about permutations generated by binary words where we use the natural order, $0 < 1$, on $\mathcal{A}$.

Claim 5.1.1 ([32]) *For an aperiodic word ω over $\mathcal{A} = \{0, 1\}$ with the natural ordering we have:*

1. $\pi_\omega(i) < \pi_\omega(i+1)$ *if and only if* $\omega_i = 0$.
2. $\pi_\omega(i) > \pi_\omega(i+1)$ *if and only if* $\omega_i = 1$.
3. *If* $\omega_i = \omega_j$, *then* $\pi_\omega(i) < \pi_\omega(j)$ *if and only if* $\pi_\omega(i+1) < \pi_\omega(j+1)$.

Proof (1) Suppose $\pi_\omega(i) < \pi_\omega(i+1)$ and assume $\omega_i = 1$. Then if $\omega_{i+1} = 0$ we have a contradiction, so $\omega_{i+1} = 1$ must be true. Thus there is some $m \geq 2$ so

$$\omega[i] = 1^m 0 \cdots, \quad \omega[i+1] = 1^{m-1} 0 \cdots$$

Thus $\pi_\omega(i) > \pi_\omega(i+1)$, which is a contradiction. Therefore $\omega_i = 0$.

Conversely suppose that $\omega_i = 0$. If $\omega_{i+1} = 1$ then we are done so suppose $\omega_{i+1} = 0$ as well. Thus there is some $m \geq 2$ so

$$\omega[i] = 0^m 1 \cdots, \quad \omega[i+1] = 0^{m-1} 1 \cdots$$

Thus $\pi_\omega(i) < \pi_\omega(i+1)$.

(2) This is true, as it is the contrapositive of (1).

(3) Let $\omega_i = \omega_j$. If $\pi_\omega(i) < \pi_\omega(j)$, then there is some $a \in \mathcal{A}$ and $u \in \mathcal{F}(\omega)$ so that

$$\omega[i] = au0\cdots, \quad \omega[j] = au1\cdots,$$

and thus

$$\omega[i+1] = u0\cdots, \quad \omega[j+1] = u1\cdots,$$

so $\pi_\omega(i+1) < \pi_\omega(j+1)$.

If $\pi_\omega(i+1) < \pi_\omega(j+1)$, we see $\pi_\omega(i) < \pi_\omega(j)$ by a similar argument. ∎

Lemma 5.1.2 ([32]) *Given two infinite binary words $u = u_0u_1\cdots$ and $v = v_0v_1\cdots$ with $\pi_u[0, n+1] = \pi_v[0, n+1]$, it follows that $u[0, n] = v[0, n]$.*

Proof Let $\mathcal{A} = \{0, 1\}$ be equipped with the natural ordering, $0 < 1$, and $u, v \in \mathcal{A}^{\mathbb{N}}$ be aperiodic words with $\pi_u[0, n+1] = \pi_v[0, n+1]$. For $0 \le i \le n$, $\pi_u(i) < \pi_u(i+1)$ if and only if $\pi_v(i) < \pi_v(i+1)$, so $u_i = v_i$ by Claim 5.1.1. Therefore $u[0, n] = v[0, n]$. ∎

We do have a trivial upper bound for $\tau_\omega(n)$ being the number of permutations of length n, which is $n!$. Lemma 5.1.2 directly implies a lower bound for the permutation complexity for a binary aperiodic word ω, namely the factor complexity of ω. Thus, initial bounds on the permutation complexity can be seen to be:

$$\rho_\omega(n-1) \le \tau_\omega(n) \le n!.$$

Recall from Section 2.1.5 the complement operator on $\mathcal{A}^\infty$. For $\mathcal{A} = \{0, 1\}$, $\overline{0} = 1$ and $\overline{1} = 0$. If $u \in \mathcal{A}^\infty$, $\overline{u} = \overline{u}_1\overline{u}_2\overline{u}_3\cdots$. Also recall that ω the set of factors of ω is closed under complementation if for each $u \in \mathcal{F}(\omega)$ then $\overline{u} \in \mathcal{F}(\omega)$. The following lemma shows an interesting property of the subpermutations of the infinite permutation π_ω.

Lemma 5.1.3 *Let $\omega = \omega_0\omega_1\omega_2\cdots$ be an aperiodic binary word with factors closed under complementation. If p is a subpermutation of π_ω of length n, then the subpermutation q defined by $q_i = n - p_i + 1$ for each i, is also a subpermutation of π_ω of length n.*

Proof Let p be a subpermutation of π_ω. There is an $a \in \mathbb{N}$ so that $p = \pi_\omega[a, a+n-1]$. For each $i, j \in \{0, 1, \ldots, n-1\}$, if $p_i < p_j$ then $\omega[a+i] < \omega[a+j]$ and there is some finite word $u_{i,j}$ so that

$$\omega[a+i] = u_{i,j}0\cdots, \quad \omega[a+j] = u_{i,j}1\cdots.$$

Let v be the prefix of $\omega[a]$ so that for each $i, j \in \{0, 1, \ldots, n-1\}$, v contains both $u_{i,j}0$ and $u_{i,j}1$. Since the set of factors of ω is closed under complementation, $\overline{v}$ is a factor of ω. There is a b so that $\overline{v}$ is a prefix of $\omega[b]$, and let $q = \pi_\omega[b, b+n-1]$. For each $i, j \in \{0, 1, \ldots, n-1\}$, if $p_i < p_j$

$$\omega[b+i] = \overline{u}_{i,j}1\cdots, \quad \omega[b+j] = \overline{u}_{i,j}0\cdots$$

and thus, $q_i > q_j$.

For any $i \in \{0, 1, \ldots, n-1\}$ there are $p_i - 1$ many j so that $p_j < p_i$ and there are $n - p_i$ many j so that $p_j > p_i$. Therefore there are $n - p_i$ many j so that $q_j < q_i$, so $q_i = n - p_i + 1$. ∎

For an aperiodic word $\omega \in \mathcal{A}^{\mathbb{N}}$, the following lemma shows the relationship of the permutation complexity of ω and $\overline{\omega}$.

Lemma 5.1.4 *Let $\omega = \omega_0\omega_1\omega_2\cdots$ be an aperiodic binary word, and let $\overline{\omega}$ be the complement of ω. For each $n \geq 1$,*

$$\tau_\omega(n) = \tau_{\overline{\omega}}(n).$$

Proof For some $a \neq b$, suppose $\omega[a] < \omega[b]$. Thus there is some (possibly empty) factor u of ω so that $\omega[a] = u0\cdots$ and $\omega[b] = u1\cdots$. Thus $\overline{\omega}[a] = \overline{u}1\cdots$ and $\overline{\omega}[b] = \overline{u}0\cdots$, so $\overline{\omega}[a] > \overline{\omega}[b]$.

For both ω and $\overline{\omega}$ it should be clear $\tau_\omega(1) = \tau_{\overline{\omega}}(1) = 1$, namely the subpermutation (1). Let $n \geq 2$. For a permutation p of $\{1, 2, \ldots, n\}$, define the permutation $\tilde{p}$ of $\{1, 2, \ldots, n\}$ by

$$\tilde{p}_i = n - \tilde{p}_i + 1$$

for each i.

Let $p = \pi_\omega[a, a+n-1]$ be a subpermutation of π_ω, so p is a permutation of $\{1, 2, \ldots, n\}$. Let $q = \pi_{\overline{\omega}}[a, a+n-1]$ be a subpermutation of $\pi_{\overline{\omega}}$. For each $0 \leq i, j \leq n-1$, $i \neq j$, if $p_i < p_j$ then $q_i > q_j$.

Let $0 \leq i \leq n-1$. There are $p_i - 1$ many j so that $p_j < p_i$ and there are $n - p_i$ many j so that $p_j > p_i$. Therefore there are exactly $n - p_i$ many j so that $q_j < q_i$, so $q_i = n - p_i + 1$. Thus $q = \tilde{p}$ and for any $p \in \mathrm{Perm}^\omega(n)$ we have $\tilde{p} \in \mathrm{Perm}^{\overline{\omega}}(n)$, so

$$\left|\mathrm{Perm}^\omega(n)\right| \leq \left|\mathrm{Perm}^{\overline{\omega}}(n)\right|.$$

By a similar argument we can see $p = \tilde{q}$ and for $q \in \mathrm{Perm}^{\overline{\omega}}(n)$ we have $\tilde{q} \in \mathrm{Perm}^\omega(n)$, so

$$\left|\mathrm{Perm}^{\overline{\omega}}(n)\right| \leq \left|\mathrm{Perm}^\omega(n)\right|.$$

Therefore $\left|\mathrm{Perm}^\omega(n)\right| = \left|\mathrm{Perm}^{\overline{\omega}}(n)\right|$ and $\tau_\omega(n) = \tau_{\overline{\omega}}(n)$. ∎

5.1.4 Permutation Complexity of Sturmian Words

Consider now a case where we start with a Sturmian word. An interesting property of characteristic Sturmian words can be seen in the next proposition, listed in [8].

Proposition 5.1.5 ([1, 8, 28]) *Let $0 < \alpha < 1$ be an irrational and s be Sturmian word with slope α over $\{0, 1\}$ with $0 < 1$. Then*

$$\pi_{0c_\alpha}(0) \leq \pi_s(0) \leq \pi_{1c_\alpha}(0).$$

Therefore the characteristic Sturmian word with slope α can be used to generate the least and greatest infinite permutations associated to a Sturmian word with slope α. If a word s is Sturmian then we know $\rho_s(n) = n + 1$ for each n. We have the following lemma.

Lemma 5.1.6 ([33]) *Let s be a Sturmian word. For natural numbers a_1 and a_2 we have $\pi_s[a_1, a_1 + n + 1] = \pi_s[a_2, a_2 + n + 1]$ if and only if $s[a_1, a_1 + n] = s[a_2, a_2 + n]$.*

Thus for a Sturmian word s we have $\tau_s(n) = \rho_s(n - 1)$. As a result of the above lemma we have the following.

Theorem 5.1.7 ([33]) *An aperiodic binary word ω is Sturmian if and only if $\tau_\omega(n) = n$ for each $n \geq 1$.*

Proof If ω is Sturmian it follows directly from Lemma 5.1.6 that $\tau_\omega(n) = \rho_\omega(n-1) = n$.

If $\tau_\omega(n) = n$ for all n, then we have:

$$n + 1 = \tau_\omega(n + 1) \geq \rho_\omega(n) > n$$

and thus $\rho_\omega(n) = n + 1$ for all n, and therefore ω is Sturmian. ∎

Therefore, permutation complexity can be used to classify the Sturmian words.

5.2 Permutation Complexity of the Thue-Morse Word

The Thue-Morse word, $T = T_0T_1T_2\cdots$, is:

$$T = 0110100110010110100101100110100 1\cdots,$$

which can be generated by the morphism:

$$\mu_T : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

by iterating on the letter 0. Axel Thue introduced this word in his studies of repetitions in words, and proved that the word T is overlap-free ([44]). A word ω is said to be *overlap-free* if it does not contain a factor of the form $vuvuv$ for words u and v, with v non-empty.

The Thue-Morse word was again discovered independently by Marston Morse in 1921 [35] through his study of differential geometry, and used in the foundations of symbolic dynamics. For a more in depth look at further properties, independent discoveries, and applications of the Thue-Morse word see [2].

The factor complexity of the Thue-Morse word was computed independently by two groups in 1989, Brlek [11] and de Luca and Varricchio [31]. Our proof of the permutation complexity of the Thue-Morse word does not use the factor complexity function.

The permutation complexity of the Thue-Morse word can be found as follows. For any $n \geq 2$, we can write n as $n = 2^r + p$, with $0 < p \leq 2^r$. Using this notation, it will be shown that the formula for the permutation complexity of T, initially conjectured by M. Makarov, is

$$\tau_T(n) = 2(2^{r+1} + p - 2).$$

We give a a non-trivial proof of this formula here. The method of this proof relies upon special properties of the Thue-Morse word and the morphism that generates it. It is currently not clear how to generalize this method to a wider class of words.

The infinite permutation associated with the Thue-Morse word, π_T, is introduced in Section 5.2.1. Patterns found in the subpermutations of π_T are studied in Section 5.2.3, while Section 5.2.4 investigates when a specific pattern occurs. The formula for the permutation complexity is established in Section 5.2.5. Low order subpermutations are listed in Appendix A to be used as a base case for induction arguments.

Note: Results of this section (5.2) appear in [50].

5.2.1 The Thue-Morse Permutation

In this section the action of the Thue-Morse morphism on the subpermutations of π_T will be investigated. This action will induce a well-defined map on the subpermutations of π_T and lead to an initial upper-bound on the permutation complexity of T.

It can readily be verified that if a is a natural number then

$$\mu_T(T[a]) = T[2a]$$

because for any letter $x \in \{0, 1\}$, $|\mu_T(x)| = 2$. A nice property of the factors of T is that any factor of length 5 or greater contains either 00 or 11. Another interesting property is that for any $i \in \mathbb{N}$, $T[2i, 2i+1]$ will be either 01 or 10. Thus any occurrence of 00 or 11 must be a factor of the form $T[2i+1, 2i+2]$ for some $i \in \mathbb{N}$. Therefore any factors $T[2i, 2i+n]$ and $T[2j+1, 2j+1+n]$ where $n \geq 4$ cannot be equal based on the location of the factors 00 or 11.

Let π_T be the infinite permutation associated to the Thue-Morse word, T. Let a and n be natural numbers and suppose we want to determine if $T[a] < T[a+n]$. There will be some (possibly empty) factor u of T, and suffixes x and y of T so that $T[a] = u\lambda x$ and $T[a+n] = u\bar{\lambda}y$, for $\lambda \in \{0, 1\}$. If $|u| \geq n+1$ we would have $T_{a+i} = T_{a+n+i}$ for each $i = 0, 1, \ldots, n$, and thus $T[a, a+n] = T[a+n, a+2n]$, and $T[a, a+2n]$ would violate the fact that T is overlap-free. Thus $|u| \leq n$, and if $|u| = n$ we have $T[a, a+n-1] = T[a+n, a+2n-1]$ and $T_{2n} = \overline{T_a}$. Therefore the subpermutation $\pi_T[a, a+n]$ can be determined within the factor $T[a, a+2n]$ of length $2n+1$. Thus the trivial bounds for the permutation complexity of the Thue-Morse word T are

$$\rho_T(n-1) \leq \tau_T(n) \leq \rho_T(2n-1).$$

Since the factor complexity of the Thue-Morse word is known (see [11, 31]) we can find all factors of a given length. Thus for any natural number n, all factors of T of length $2n-1$ can be identified and thus the set of all subpermutations of π_T of length n, $\mathrm{Perm}^T(n)$, can be identified as well. The subpermutations of $\{1, 2, \ldots, n\}$ have been identified for relatively low n (up to $n = 65$) and in these cases no more than two subpermutations of any length were identified to have the same form. In other words, for any factor u of T of length $n \leq 64$ there are at most two subpermutations of length $n+1$ having form u.

This section will deal with some properties of π_T. Something to note about the Thue-Morse morphism is that it is an order preserving morphism, as shown by the following lemma.

Lemma 5.2.1 *For natural numbers a and b, $T[a] < T[b]$ if and only if $\mu_T(T[a]) < \mu_T(T[b])$.*

Proof If $T[a] < T[b]$, then there exists a finite factor u of T, and suffixes x and y of T so that

$$T[a] = u0x \quad \text{and} \quad T[b] = u1y.$$

Thus we can see

$$\mu_T(T[a]) = \mu_T(u)01\mu_T(x) \quad \text{and} \quad \mu_T(T[b]) = \mu_T(u)10\mu_T(y)$$

and therefore $\mu_T(T[a]) < \mu_T(T[b])$.

Then suppose $T[a] > T[b]$, and we see $\mu_T(T[a]) > \mu_T(T[b])$ by a similar argument. Thus $\mu_T(T[a]) < \mu_T(T[b])$ will imply $T[a] < T[b]$. ∎

Lemma 5.2.2 *If u and v are shifts of T so that for some a and b $u = 0T[a]$ and $v = 1T[b]$, and hence $u < v$, $\mu_T(u) = 01\mu_T(T[a])$, and $\mu_T(v) = 10\mu_T(T[b])$. Thus $0\mu_T(T[b]) < 01\mu_T(T[a]) < 10\mu_T(T[b]) < 1\mu_T(T[a])$.*

Proof The first letters in $T[a]$ will be either 01 or 1, thus $\mu_T(T[a])$ will start with either 0110 or 10. The first letters in $T[b]$ will be either 10 or 0, thus $\mu_T(T[b])$ will start with either 1001 or 01, respectively.

Then $0\mu_T(T[b])$ will start with 01001 or 001 and $01\mu_T(T[a])$ will start with 010110 or 0110. Thus $001 < 01001 < 010110 < 0110$, and

$$0\mu_T(T[b]) < 01\mu_T(T[a]).$$

Then $10\mu_T(T[b])$ will start with 101001 or 1001 and $1\mu_T(T[a])$ will start with 10110 or 110. Thus $1001 < 101001 < 10110 < 110$, and

$$10\mu_T(T[b]) < 1\mu_T(T[a]).$$

Therefore

$$0\mu_T(T[b]) < 01\mu_T(T[a]) < 10\mu_T(T[b]) < 1\mu_T(T[a]).$$

∎

Let u be a factor of T of length n. There is an $a \in \mathbb{N}$ so that $u = T[a, a+n-1]$. Also recall that $|u|_1$ is the number of occurrences of the letter 1 in u, and that $|u|_1 = n - |u|_0$. Let $p = \pi_T[a, a+n]$ be a subpermutation of π_T with form u. Then $\mu_T(u) = T[2a, 2a + 2n - 1]$, and let p' be the subpermutation $p' = \pi_T[2a, 2a + 2n]$ with form $\mu_T(u)$. When Lemma 5.2.2 is used with this notation, for $0 \leq i, j \leq n-1$, where $T_{a+i} = 0$ and $T_{a+j} = 1$, we have $p_i < p_j$ and $p'_{2j+1} < p'_{2i} < p'_{2j} < p'_{2i+1}$. The following lemma describes the values of p' in terms of the values of p.

Proposition 5.2.3 *Let u, p, and p' be as described above. For any $i \in \{0, 1, \ldots, n\}$:*

$$p'_{2i} = p_i + |u|_1$$

and for any $i \in \{0, 1, \ldots, n-1\}$:

$$p'_{2i+1} = \begin{cases} p_i + |u|_1 + (n+1) & \text{if } p_i < p_{i+1} \text{ and } p_i < p_n \\ p_i + |u|_1 + n & \text{if } p_i < p_{i+1} \text{ and } p_i > p_n \\ p_i + |u|_1 - n & \text{if } p_i > p_{i+1} \text{ and } p_i < p_n \\ p_i + |u|_1 - (n+1) & \text{if } p_i > p_{i+1} \text{ and } p_i > p_n \end{cases}$$

Proof To take care of the p'_{2i} terms, let $i \in \{0, 1, \ldots, n\}$. There will be $p_i - 1$ many j so that $p_i > p_j$, so there are $p_i - 1$ many j so that $p'_{2i} > p'_{2j}$. Clearly, if $p_i < p_j$ then $p'_{2i} < p'_{2j}$. So there are exactly $p_i - 1$ many even j so that $p'_{2i} > p'_j$. There are $|u|_1$ many j so that $T_{a+j} = 1$, so there are $|u|_1$ many j so that $p'_{2i} > p'_{2j+1}$ and $|u|_0$ many j so that $T_{a+j} = 0$, so $p'_{2i} < p'_{2j+1}$. So there are exactly $|u|_1$ many odd j so that $p'_{2i} > p'_j$. Thus there are exactly $p_i - 1 + |u|_1$ many j so that $p'_{2i} > p'_j$, and therefore $p'_{2i} = (p_i - 1 + |u|_1) + 1 = p_i + |u|_1$.

The p'_{2i+1} terms will be done in two cases. First when $p_i < p_{i+1}$ and then when $p_i > p_{i+1}$.

Case a: Suppose that $p_i < p_{i+1}$, so $T_{a+i} = 0$ by Claim 5.1.1. For each $j = 0, 1, \ldots, n$ we must have $p'_{2i+1} > p'_{2j}$, so for each even j (there are $n + 1$ many such j) $p'_{2i+1} > p'_j$. There are $|u|_1$ many j so that $T_{a+j} = 1$, so there are $|u|_1$ many j so that $p'_{2i+1} > p'_{2j+1}$. Thus the only other j where p'_{2j+1} can be less than p'_{2i+1} are $j \in \{0, 1, \ldots, n-1\}$ where $T_{a+j} = 0$ and $p_i > p_j$.

Subcase a.1: If $p_i < p_n$ then there are $p_i - 1$ many j so that $T_{a+j} = 0$ and $p_i > p_j$, and then $n - p_i - |u|_1 = |u|_0 - p_i$ many j so that $T_{a+j} = 0$ and $p_i < p_j$. Thus there can only be $(n + 1) + |u|_1 + p_i - 1$ many j so that $p'_{2i+1} > p'_j$, and therefore $p'_{2i+1} = (n + 1) + |u|_1 + p_i - 1 + 1 = p_i + |u|_1 + (n + 1)$.

Subcase a.2: If $p_i > p_n$ then there are $p_i - 2$ many j so that $T_{a+j} = 0$ and $p_i > p_j$ (since T_{a+n} is not in $u = T[a, a + n - 1]$), and then $n - (p_i - 1) - |u|_1 = |u|_0 - (p_i - 1)$ many j so that $T_{a+j} = 0$ and $p_i < p_j$. Thus there can only be $(n+1) + |u|_1 + p_i - 2$ many j so that $p'_{2i+1} > p'_j$, and therefore $p'_{2i+1} = (n + 1) + |u|_1 + p_i - 2 + 1 = p_i + |u|_1 + n$.

Case b: Suppose that $p_i > p_{i+1}$, so $T_{a+i} = 1$ by Claim 5.1.1. For each $j = 0, 1, \ldots, n$ we must have $p'_{2i+1} < p'_{2j}$, so for each even j (there are $n + 1$ many such j) $p'_{2i+1} < p'_j$. There are $|u|_0$ many j so that $T_{a+j} = 0$, so there are $|u|_0$ many j so that $p'_{2i+1} < p'_{2j+1}$. Thus the only other j where p'_{2j+1} can be less than p'_{2i+1} are $j \in \{0, 1, \ldots, n-1\}$ where $T_{a+j} = 1$ and $p_i > p_j$.

Subcase b.1: If $p_i < p_n$ then there are $(p_i - 1) - |u|_0$ many j so that $T_{a+j} = 1$ and $p_i > p_j$, and there can only be $|u|_1 - (p_i - 1 - |u|_0) - 1 = n - p_i$ many j so that $T_{a+j} = 1$ and $p_i < p_j$ (since T_{a+n} is not in $u = T[a, a + n - 1]$). Thus there can only be $(p_i - 1) - |u|_0 = p_i - 1 - (n - |u|_1) = p_i + |u|_1 - n - 1$ many j so that $p'_{2i+1} > p'_j$, and therefore $p'_{2i+1} = p_i + |u|_1 - n - 1 + 1 = p_i + |u|_1 - n$.

Subcase b.2: If $p_i > p_n$ then there are $(p_i - 2) - |u|_0$ many j so that $T_{a+j} = 1$ and $p_i > p_j$ (since T_{a+n} is not in $u = T[a, a + n - 1]$), and there can only be $|u|_1 - (p_i - 2 - |u|_0) - 1 = (n + 1) - p_i$ many j so that $T_{a+j} = 1$ and $p_i < p_j$. Thus there can only be $(p_i - 2) - |u|_0 = p_i - 2 - (n - |u|_1) = p_i + |u|_1 - n - 2$ many j so that $p'_{2i+1} > p'_j$, and therefore $p'_{2i+1} = p_i + |u|_1 - n - 2 + 1 = p_i + |u|_1 - (n + 1)$. ∎

Fix a subpermutation $p = \pi_T[a, a+n]$, and then let $p' = \pi_T[2a, 2a+2n]$. So the terms of p' can be defined using the method defined in Proposition 5.2.3. Let $q = \pi_T[b, b+n]$, $b \neq a$, be a subpermutation of π_T and let $q' = \pi_T[2b, 2b+2n]$ as in Proposition 5.2.3. The following lemma concerns the relationship of p and q to p' and q'. Therefore the idea of p' can be used to define a map on the subpermutations of π_T, and the map will be well-defined by Proposition 5.2.3.

Lemma 5.2.4 *$p \neq q$ if and only if $p' \neq q'$.*

Proof Supposing that $p \neq q$, there are $i, j \in \{0, 1, \ldots, n\}$ so that $p_i < p_j$ and $q_i > q_j$. Since the Thue-Morse morphism is order preserving we have $p'_{2i} < p'_{2j}$ and $q'_{2i} > q'_{2j}$, so $p' \neq q'$.

Now to show by contrapositive, suppose that $p = q$, so $p_i = q_i$ for each $i \in \{0, 1, \ldots, n\}$. Since $p = q$, p and q have the same form, because $p_i < p_{i+1}$ if and only if $q_i < q_{i+1}$, so $T[a, a+n-1] = T[b, b+n-1]$ and thus $T[2a, 2a+2n-1] = T[2b, 2b+2n-1]$. Then by Proposition 5.2.3 it should be clear that for each $j \in \{0, 1, \ldots, 2n\}$ we have $p'_j = q'_j$, and thus $p' = q'$.

Therefore if $p' \neq q'$ then $p \neq q$. ∎

The next corollary follows directly from Lemma 5.2.4.

Corollary 5.2.5 *If $p = \pi_T[a, a+n] = \pi_T[b, b+n]$ for some $a \neq b$, then $\pi_T[2a, 2a+2n] = \pi_t[2b, 2b+2n]$.*

Thus there is a well-defined function on the subpermutations of π_T. Let $p = \pi_T[a, a+n]$, and define $\phi(p) = p' = \pi_T[2a, 2a+2n]$ using the formula in Proposition 5.2.3. Thus we have the map

$$\phi : \mathrm{Perm}^T(n+1) \mapsto \mathrm{Perm}^T(2n+1)$$

which is injective by Lemma 5.2.4.

Not all subpermutations of π_T will be the image under ϕ of another subpermutation. Let $n \geq 5$ and a be natural numbers. Then n and a can be either even or odd, and for the subpermutation $\pi_T[a, a+n]$, there exist natural numbers b and m so that one of 4 cases hold:

1. $\pi_T[a, a+n] = \pi_T[2b, 2b+2m]$, even starting position with odd length.
2. $\pi_T[a, a+n] = \pi_T[2b, 2b+2m-1]$, even starting position with even length.
3. $\pi_T[a, a+n] = \pi_T[2b+1, 2b+2m]$, odd starting position with even length.
4. $\pi_T[a, a+n] = \pi_T[2b+1, 2b+2m+1]$, odd starting position with odd length.

Consider two subpermutations of length $n > 5$, $\pi_T[2c, 2c+n]$ and $\pi_T[2d+1, 2d+n+1]$. The subpermutations $\pi_T[2c, 2c+n]$ will have form $T[2c, 2c+n-1]$, and $\pi_T[2d+1, 2d+n+1]$ will have form $T[2d+1, 2d+n]$. Since the length of these factors is at least 5, we know $T[2c, 2c+n-1] \neq T[2d+1, 2d+n]$, and thus $\pi_T[2c, 2c+n] \neq \pi_T[2d+1, 2d+n+1]$ because they do not have the same form. Thus we can break up the set $\mathrm{Perm}^T(n)$ into two classes of subpermutations, namely the subpermutations that start at an even position or an odd position. So say that $\mathrm{Perm}^T_{ev}(n)$ is the set of subpermutations p of length n so that $p = \pi_T[2b, 2b+n-1]$ for some b, and that $\mathrm{Perm}^T_{odd}(n)$ is the set of subpermutations p of length n so that $p = \pi_T[2b+1, 2b+n]$ for some b. Thus

$$\mathrm{Perm}^T(n) = \mathrm{Perm}^T_{ev}(n) \cup \mathrm{Perm}^T_{odd}(n),$$

where we have

$$\mathrm{Perm}^T_{ev}(n) \cap \mathrm{Perm}^T_{odd}(n) = \emptyset.$$

Thus for $n \geq 3$, $\mathrm{Perm}^T_{ev}(2n+1)$ is the set of all subpermutations of length $2n+1$ starting at an even position. So for $\pi_T[2a, 2a+2n]$, we know there is a subpermutation $p = \pi_T[a, a+n]$ so that $\phi(p) = p' = \pi_T[2a, 2a+2n]$. Thus the map

$$\phi : \mathrm{Perm}^T(n+1) \mapsto \mathrm{Perm}^T_{ev}(2n+1)$$

is also a surjective map, and is thus a bijection.

Recall the left, right, and middle restrictions of a subpermutation. If $p = \pi_T[a, a+n]$ then $L(p) = \pi[a, a+n-1]$, $R(p) = \pi[a+1, a+n]$, and $M(p) = R(L(p)) = L(R(p)) = \pi[a+1, a+n-1]$. These restrictions will be helpful to count the size of the sets $\mathrm{Perm}^T_{odd}(2n)$, $\mathrm{Perm}^T_{ev}(2n)$, and $\mathrm{Perm}^T_{odd}(2n+1)$.

For $p = \pi_T[a, a+n]$, we can then define three additional maps by looking at the left, right, and middle restrictions of $\phi(p) = p'$. These maps are

$$\begin{aligned}
\phi_L &: \mathrm{Perm}^T(n+1) \mapsto \mathrm{Perm}^T_{ev}(2n) \\
\phi_R &: \mathrm{Perm}^T(n+1) \mapsto \mathrm{Perm}^T_{odd}(2n) \\
\phi_M &: \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^T_{odd}(2n+1)
\end{aligned}$$

and are defined by

$$\begin{aligned}
\phi_L(p) &= L(\phi(p)) = L(p') \\
\phi_R(p) &= R(\phi(p)) = R(p') \\
\phi_M(p) &= M(\phi(p)) = M(p')
\end{aligned}$$

It can be readily verified that these three maps are surjective. To see an example of this, consider the map ϕ_L, and let $\pi_T[2b, 2b+2n-1]$ be a subpermutation in $\mathrm{Perm}^T_{ev}(2n)$. Then for the subpermutation $p = \pi_T[b, b+n]$, $\phi_L(p) = L(p') = \pi_T[2b, 2b+2n-1]$ so ϕ_L is surjective. A similar argument will show that ϕ_R and ϕ_M are also surjective.

Lemma 5.2.6 *For $n \geq 2$:*

$$\tau_T(2n) \leq 2(\tau_T(n+1))$$
$$\tau_T(2n+1) \leq \tau_T(n+1) + \tau_T(n+2)$$

Proof Let $n \geq 2$. We have:

$$\left|\text{Perm}^T_{ev}(2n)\right| \leq \left|\text{Perm}^T(n+1)\right|$$
$$\left|\text{Perm}^T_{odd}(2n)\right| \leq \left|\text{Perm}^T(n+1)\right|$$
$$\left|\text{Perm}^T_{ev}(2n+1)\right| = \left|\text{Perm}^T(n+1)\right|$$
$$\left|\text{Perm}^T_{odd}(2n+1)\right| \leq \left|\text{Perm}^T(n+2)\right|$$

since ϕ is a bijection, and the 3 maps ϕ_L, ϕ_R, and ϕ_M are all surjective. Thus we have the following inequalities:

$$\begin{aligned}
\tau_T(2n) &= \left|\text{Perm}^T(2n)\right| = \left|\text{Perm}^T_{ev}(2n)\right| + \left|\text{Perm}^T_{odd}(2n)\right| \\
&\leq \left|\text{Perm}^T(n+1)\right| + \left|\text{Perm}^T(n+1)\right| = 2(\tau_T(n+1)) \\
\tau_T(2n+1) &= \left|\text{Perm}^T(2n+1)\right| = \left|\text{Perm}^T_{ev}(2n+1)\right| + \left|\text{Perm}^T_{odd}(2n+1)\right| \\
&\leq \left|\text{Perm}^T(n+1)\right| + \left|\text{Perm}^T(n+2)\right| = \tau_T(n+1) + \tau_T(n+2)
\end{aligned}$$

∎

The three maps ϕ_L, ϕ_R, and ϕ_M are not injective maps. To see this, consider the subpermutations

$$p = \pi_T[5,9] = (2\,3\,5\,4\,1)$$
$$q = \pi_T[23,27] = (1\,3\,5\,4\,2).$$

Both of these subpermutations have form $T[5,8] = T[23,26] = 0011$. Then applying the maps we see:

$$p' = \phi(p) = \pi_T[10,18] = (4\,8\,5\,9\,7\,2\,6\,1\,3)$$
$$q' = \phi(q) = \pi_T[46,54] = (3\,8\,5\,9\,7\,2\,6\,1\,4)$$

$$\phi_L(p) = \pi_T[10,17] = (3\,7\,4\,8\,6\,2\,5\,1) = \pi_T[46,53] = \phi_L(q)$$

$$\phi_R(p) = \pi_T[11,18] = (7\,4\,8\,6\,2\,5\,1\,3) = \pi_T[47,54] = \phi_R(q)$$

$$\phi_M(p) = \pi_T[11,17] = (6\,3\,7\,5\,2\,4\,1) = \pi_T[47,53] = \phi_M(q)$$

So $p' \neq q'$ but $\phi_L(p) = \phi_L(q)$, $\phi_R(p) = \phi_R(q)$, and $\phi_M(p) = \phi_M(q)$, and these maps are not injective in general. Hence the values in Lemma 5.2.6 are only upper bounds. The next goal is to determine when these maps are not injective.

5.2.2 Creating $\phi(p)$ From a Table

Given a subpermutation $p = \pi_T[a, a+n]$, we can use a table to calculate the subpermutation $p' = \phi(p) = \pi_T[2a, 2a+2n]$. To show how this is done we will use an example. Let

$$p = \pi_T[10, 18] = (4\,8\,5\,9\,7\,2\,6\,1\,3)$$

$$u = T[10, 17] = 01011010,$$

so p has form u and $|u|_1 = 4$. Let $p' = \phi(p) = \pi_T[20, 36]$, so by Proposition 5.2.3

$$p' = (8\ 16\ 12\ 3\ 9\ 17\ 13\ 4\ 11\ 2\ 6\ 15\ 10\ 1\ 5\ 14\ 7).$$

The table we create will have 3 columns. In the first column of the table we list all the values in the subpermutation p from least to greatest. We will add a horizontal line just above the least value which corresponds to a 1 in u and add an empty box in the third column of the row corresponding to the last element in p. In this example the least value of p which corresponds to a 1 in u is the number 6 so the horizontal line will be between $p_2 = 5$ and $p_6 = 6$. The last element of p is $p_8 = 3$, so we place the empty box in the last column of row 3. Since p is a permutation of $\{1, 2, \ldots, 9\}$ we have the table in Figure 1. The empty box will act as a place holder. In all the steps listed below we will ignore the location with the empty box and write nothing in that spot. In the second column of the table we will write the value we get from adding $|u|_1$ to the number in the first column, see Figure 2. Note $|u|_1 = 4$ and there are 4 rows below the horizontal line.

1		
2		
3		□
4		
5		
6		
7		
8		
9		

Fig. 1

$\longrightarrow$

1	5	
2	6	
3	7	□
4	8	
5	9	
6	10	
7	11	
8	12	
9	13	

Fig. 2

In the third column we start by writing a 1 in the first row below the horizontal line and then increasing by 1 for the following columns, see Figure 3. To finish the table we will add 1 to the last value in the second column (a 13 in this example) and write the result in the first row of the third column, and then increase by 1 in the following columns until we reach the horizontal line, see Figure 4.

1	5	
2	6	
3	7	$\square$
4	8	
5	9	
6	10	1
7	11	2
8	12	3
9	13	4

Fig. 3

$\longrightarrow$

1	5	14
2	6	15
3	7	$\square$
4	8	16
5	9	17
6	10	1
7	11	2
8	12	3
9	13	4

Fig. 4

Now that we have the completed table in Figure 4 we can see how ϕ will alter the permutation p. We have $p_2 = 5$, so $p'_4 = 9$ and $p'_5 = 17$. From the table in Figure 4, when we find the row with a 5 in the first column we see the next elements in the row are 9 and 17. This is exactly the behavior described in Lemma 5.2.2 and Proposition 5.2.3. Thus to create p' from the table we find the row which has the value p_i in the first column, and then p'_{2i} is the value in the second column and p'_{2i+1} is the value in the third column. So

$$\begin{aligned}
p_0 = 4 &\implies p'_0 = 8 \quad p'_1 = 16 \\
p_1 = 8 &\implies p'_2 = 12 \quad p'_3 = 3 \\
p_2 = 5 &\implies p'_4 = 9 \quad p'_5 = 17 \\
&\vdots \\
p_7 = 1 &\implies p'_{14} = 5 \quad p'_{15} = 14 \\
p_8 = 3 &\implies p'_{16} = 7
\end{aligned}$$

which is readily verified to be the same as p' listed above.

One note about this example. It should be fairly clear for the subpermutation $p = \pi_T[10, 18]$ that $T_{18} = 0$. In the table above $p_8 = 3$ is in the portion of the table which corresponds to the 0's. Consider the subpermutation $q = (1\ 3\ 6\ 5\ 2\ 4)$ of π_T which has form $v = 00110$, and $\phi(q) = (3\ 9\ 5\ 11\ 8\ 2\ 7\ 1\ 4\ 10\ 6)$. Each of q_0, q_1, and q_4 correspond to a 0, and q_2 and q_3 each correspond to a 1. Because v is a right special factor of T it is not clear if q_6 will correspond to a 0 or a 1. It turns out that when we construct the table to create $\phi(q)$, it does not matter. In Figure 5 we construct the table as above, and in

Figure 6 we construct the table assuming q_6 corresponds to a 1, but note that $|v|_1 = 2$.

1	3	9
2	4	10
3	5	11
4	6	□
5	7	1
6	8	2

Fig. 5

or

1	3	9
2	4	10
3	5	11
4	6	□
5	7	1
6	8	2

Fig. 6

In either case we will construct the same subpermutation q' from the table. A note about this example is that it is a contrived example. It turns out that only 001101 will correspond to the subpermutation q, and 001100 will correspond to the subpermutation (246513). This example was chosen to show that if there is some ambiguity with what letter corresponds with the last value in a subpermutation, the table construction will still work. The reason the above definition used the value $|u|_1$ to place the horizontal line was to make the steps in the table creation explicit.

5.2.3 Type k and Complementary Pairs

An interesting pattern occurs in some subpermutations of π_T. The subpermutations that follow this pattern are said to be subpermutations of type k which is described in the next definition. Proposition 5.2.3 will be used inductively to show the maps ϕ, ϕ_L, ϕ_R, and ϕ_M preserve subpermutations of type k. An induction argument with this fact will be used to show that two subpermutations have the same form if and only if they are a complementary pair of type k, defined below. A corollary of this will determine when the maps ϕ_L, ϕ_R, and ϕ_M are bijective.

Definition A subpermutation $p = \pi_T[a, a+n]$ is of *type* k, for $k \geq 1$, if p can be decomposed as

$$p = (\alpha_1 \cdots \alpha_k \lambda_1 \cdots \lambda_l \beta_1 \cdots \beta_k)$$

where $\alpha_i = \beta_i + \epsilon$ for each $i = 1, 2, \ldots, k$ and an $\epsilon \in \{-1, 1\}$.

Some examples of subpermutations of type 1, 2, and 3 (resp.) are:

$$\pi_T[5, 9] = (2\,3\,5\,4\,1)$$
$$\pi_T[20, 25] = (2\,5\,4\,1\,3\,6)$$
$$\pi_T[6, 12] = (3\,7\,5\,1\,2\,6\,4)$$

Definition Suppose that the subpermutation $p = \pi_T[a, a+n]$ is of type k so that for some $\epsilon \in \{-1, 1\}$, $\alpha_i = \beta_i + \epsilon$ for each $i = 1, 2, \ldots, k$. If there exists a subpermutation $q = \pi_T[b, b+n]$ of type k so that p and q can be decomposed as:

$$p = \pi_T[a, a+n] = (\alpha_1 \cdots \alpha_k \lambda_1 \cdots \lambda_l \beta_1 \cdots \beta_k)$$
$$q = \pi_T[b, b+n] = (\beta_1 \cdots \beta_k \lambda_1 \cdots \lambda_l \alpha_1 \cdots \alpha_k)$$

then p and q are said to be a *complementary pair of type k*. If p and q are a *complementary pair of type $k \leq 0$* then $p = q$.

The subpermutations

$$\pi_T[5, 9] = (2\,3\,5\,4\,1)$$
$$\pi_T[23, 27] = (1\,3\,5\,4\,2)$$

are a complementary pair of type 1. The following subpermutation of type 1

$$\pi_T[0, 3] = (2\,4\,3\,1)$$

does not have a complementary pair, since $(1\,4\,3\,2)$ is not a subpermutation of π_T.

The following proposition considers subpermutations of type k, and complementary pairs of type k.

Proposition 5.2.7 *Suppose $p = \pi_T[a, a+n]$ and $q = \pi_T[b, b+n]$ are a complementary pair of type k, with $k \geq 1$.*

(a) $\phi(p)$ is of type $2k-1$, and if $k \geq 2$ then $\phi_L(p)$ and $\phi_R(p)$ are of type $2k-2$ and $\phi_M(p)$ is of type $2k-3$.

(b) $\phi(p)$ and $\phi(q)$ are a complementary pair of type $2k-1$.

(c) $\phi_L(p)$ and $\phi_L(q)$ are a complementary pair of type $2k-2$.

(d) $\phi_R(p)$ and $\phi_R(q)$ are a complementary pair of type $2k-2$.

(e) $\phi_M(p)$ and $\phi_M(q)$ are a complementary pair of type $2k-3$.

Proof Since p and q are a complementary pair of type k they can be decomposed as

$$p = \pi_T[a, a+n] = (\alpha_1 \cdots \alpha_k \lambda_1 \cdots \lambda_l \beta_1 \cdots \beta_k)$$
$$q = \pi_T[b, b+n] = (\beta_1 \cdots \beta_k \lambda_1 \cdots \lambda_l \alpha_1 \cdots \alpha_k)$$

and for $\epsilon \in \{-1, 1\}$, $\alpha_i = \beta_i + \epsilon$ for each $i = 1, 2, \ldots, k$. Since p and q are a complementary pair they have the same form, which is shown relatively quickly at the beginning of the proof of Theorem $SameFormIFFCompPair$. Let $u = T[a, a+n-1] = T[b, b+n-1]$ be the form of p and q. For the values of k and l, $2k+l = n+1$ and $4k+2l-1 = 2n+1$.

(a) The first thing to show is that $\phi(p)$ is of type $2k-1$.

For $i \in \{0, 1, \ldots, k-1\}$ we have $p_i = p_{n-(k-1)+i} + \epsilon$, so by Proposition 5.2.3:

$$p'_{2i} = p'_{2(n-(k-1)+i)} + \epsilon$$

For $i \in \{0, 1, \ldots, k-2\}$, $p_i < p_{i+1}$ if and only if $p_{n-(k-1)+i} < p_{n-(k-1)+i+1}$, and $p_i < p_n$ if and only if $p_{n-(k-1)+i} < p_n$ since p_i and $p_{n-(k-1)+i}$ are consecutive values. By Proposition 5.2.3:

$$p'_{2i+1} = p'_{2(n-(k-1)+i)+1} + \epsilon$$

So for each $i \in \{0, 1, \ldots, 2k-2\}$: $p'_i = p'_{2n-2k+2+i} + \epsilon$, and $\phi(p)$ can be decomposed as

$$\phi(p) = \pi_T[2a, 2a+2n] = (\alpha'_1 \cdots \alpha'_{2k-1} \lambda'_1 \cdots \lambda'_{2l+1} \beta'_1 \cdots \beta'_{2k-1}),$$

where $\alpha'_i = \beta'_i + \epsilon$, so $\phi(p) = p'$ is of type $2k-1$.

Next, suppose that $k \geq 2$ so $2k-1 \geq 3$, we show that $\phi_L(p) = L(p')$ and $\phi_R(p) = R(p')$ are of type $2k-2$ and $\phi_M(p)$ is of type $2k-3$.

Let $i \in \{0, 1, \ldots, 2k-3\}$, and consider $\phi_L(p) = L(p')$. Since p'_i and $p'_{2n-2k+2+i}$ are consecutive values, $p'_i < p'_{2n}$ if and only if $p'_{2n-2k+2+i} < p'_{2n}$. So if $L(p')_i = p'_i$ then $L(p')_{2n-2k+2+i} = p'_{2n-2k+2+i}$, and if $L(p')_i = p'_i - 1$ then $L(p')_{2n-2k+2+i} = p'_{2n-2k+2+i} - 1$. In either case, $L(p')_i = L(p')_{2n-2k+2+i} + \epsilon$ and there is a decomposition

$$\phi_L(p) = \pi_T[2a, 2a+2n-1] = (\alpha''_1 \cdots \alpha''_{2k-2} \lambda''_1 \cdots \lambda''_{2l+2} \beta''_1 \cdots \beta''_{2k-2}),$$

and $\phi_L(p)$ is of type $2k-2$. A similar argument will show $\phi_R(p)$ is of type $2k-2$ and $\phi_M(p)$ is of type $2k-3$.

(b) From (a), $\phi(q) = q'$ is of type $2k-1$. Since p and q are a complementary pair of type k, $p_i = p_{n-k+1+i} + \epsilon = q_i + \epsilon = q_{n-k+1+i}$ for each $i \in \{0, 1, \ldots, k-1\}$, and $p_{k+i} = q_{k+i}$ for each $i \in \{0, 1, \ldots, l-1\}$. We can assume that $\epsilon = 1$ by exchanging the role of p and q. Thus for $i \in \{0, 1, \ldots, k-1\}$:

$$\begin{aligned} p'_{2i} &= p'_{2(n-k+1+i)} + \epsilon \\ p'_{2i} &= q'_{2(n-k+1+i)} \\ q'_{2(n-k+1+i)} &= q'_{2i} + \epsilon \end{aligned}$$

For $i \in \{0, 1, \ldots, k-2\}$:

$$\begin{aligned} p'_{2i+1} &= p'_{2(n-k+1+i)+1} + \epsilon \\ p'_{2i+1} &= q'_{2(n-k+1+i)+1} \\ q'_{2(n-k+1+i)+1} &= q'_{2i+1} + \epsilon \end{aligned}$$

We know $p_{k-1} = p_n + \epsilon = q_{k-1} + \epsilon = q_n$, so $p_{k-1} > p_n$ and $q_{k-1} < q_n$. Thus if $p_{k-1} < p_k$

$$p'_{2k-1} = p_{k-1} + |u|_1 + n = q_{k-1} + 1 + |u|_1 + n = q_{k-1} + |u|_1 + (n+1) = q'_{2k-1}$$

and if $p_{k-1} > p_k$

$$p'_{2k-1} = p_{k-1} + |u|_1 - (n+1) = q_{k-1} + 1 + |u|_1 - (n+1) = q_{k-1} + |u|_1 - n = q'_{2k-1}.$$

By Proposition 5.2.3, since $p_{k+i} = q_{k+i}$ for each $i \in \{0, 1, \ldots, l-1\}$,

$$\begin{aligned} p'_{2(k+i)} &= q'_{2(k+i)} \\ p'_{2(k+i)+1} &= q'_{2(k+i)+1} \end{aligned}$$

Thus there are decompositions of $\phi(p) = p'$ and $\phi(q) = q'$ so that

$$\phi(p) = \pi_T[2a, 2a+2n] = (\alpha'_1 \cdots \alpha'_{2k-1} \lambda'_1 \cdots \lambda'_{2l+1} \beta'_1 \cdots \beta'_{2k-1}),$$

$$\phi(q) = \pi_T[2b, 2b+2n] = (\beta'_1 \cdots \beta'_{2k-1} \lambda'_1 \cdots \lambda'_{2l+1} \alpha'_1 \cdots \alpha'_{2k-1}),$$

where $\alpha'_i = \beta'_i + \epsilon$. Therefore $\phi(p) = p'$ and $\phi(q) = q'$ are a complementary pair of type $2k-1$.

(c) From (b), $\phi(p) = p'$ and $\phi(q) = q'$ are a complementary pair of type $2k-1$. Suppose $k \geq 2$ and so $2k-3 \geq 1$, and let $i \in \{0, 1, \ldots, 2k-3\}$, then $p'_i = q'_i + \epsilon = p'_{2n-2k+2+i} + \epsilon = q'_{2n-2k+2+i}$. Thus p'_i and $p'_{2n-2k+2+i}$ are consecutive values, as are q'_i and $q'_{2n-2k+2+i}$, also $p'_{2n} < p'_i$ if and only if $p'_{2n} < p'_{2n-2k+2+i}$, and

$$p'_{2n} < p'_i \text{ and } p'_{2n} < p'_{2n-2k+2+i} \iff q'_{2n} < q'_i \text{ and } q'_{2n} < q'_{2n-2k+2+i}.$$

If $L(p')_i = p'_i - 1$ or $L(p')_i = p'_i$, we have $L(q')_i = q'_i - 1$ or $L(q')_i = q'_i$ (resp.), and $L(p')_i = L(q')_i + \epsilon = L(p')_{2n-2k+2+i} + \epsilon = L(q')_{2n-2k+2+i}$.

Now let $i \in \{0, 1, \ldots, 2l\}$, so $p'_{2k-1+i} = q'_{2k-1+i}$. Thus $p'_{2n} < p'_{2k-1+i}$ if and only if $q'_{2n} < q'_{2k-1+i}$, and so we have $L(p')_{2k-1+i} = L(q')_{2k-1+i}$.

Then $p'_{2k-2} = q'_{2k-2} + \epsilon = p'_{2n} + \epsilon = q'_{2n}$, so $p'_{2k-2} > p'_{2n}$ if and only if $q'_{2k-2} < q'_{2n}$. If $p'_{2k-2} > p'_{2n}$ and $q'_{2k-2} < q'_{2n}$, then $p'_{2k-2} = q'_{2k-2} + 1 = p'_{2n} + 1 = q'_{2n}$ so

$$L(p')_{2k-2} = p'_{2k-2} - 1 = q'_{2k-2} = L(q')_{2k-2}.,$$

If $p'_{2k-2} < p'_{2n}$ and $q'_{2k-2} > q'_{2n}$, then $p'_{2k-2} = q'_{2k-2} - 1 = p'_{2n} - 1 = q'_{2n}$ so

$$L(p')_{2k-2} = p'_{2k-2} = q'_{2k-2} - 1 = L(q')_{2k-2}.$$

In either case, $L(p')_{2k-2} = L(q')_{2k-2}$. Thus there are decompositions of $\phi_L(p) = L(p')$ and $\phi_L(q) = L(q')$ so that

$$\phi_L(p) = \pi_T[2a, 2a+2n-1] = (\alpha'_1 \cdots \alpha'_{2k-2}\lambda'_1 \cdots \lambda'_{2l+2}\beta'_1 \cdots \beta'_{2k-2}),$$

$$\phi_L(q) = \pi_T[2b, 2b+2n-1] = (\beta'_1 \cdots \beta'_{2k-2}\lambda'_1 \cdots \lambda'_{2l+2}\alpha'_1 \cdots \alpha'_{2k-2}),$$

where $\alpha'_i = \beta'_i + \epsilon$. Therefore $\phi_L(p)$ and $\phi_L(q)$ are a complementary pair of type $2k-2$.

Now suppose that $k = 1$ and so $2k - 1 = 1$. Then $p'_0 = q'_0 + \epsilon = p'_{2n} + \epsilon = q'_{2n}$ and $p'_i = q'_i$ for $i = 1, 2, \ldots, 2n-1$. If $p'_0 > p'_{2n}$ and $q'_0 < q'_{2n}$, then $p'_0 = q'_0 + 1 = p'_{2n} + 1 = q'_{2n}$ so

$$L(p')_0 = p'_0 - 1 = q'_0 = L(q')_0.$$

If $p'_0 < p'_{2n}$ and $q'_0 > q'_{2n}$, then $p'_0 = q'_0 - 1 = p'_{2n} - 1 = q'_{2n}$ so

$$L(p')_0 = p'_0 = q'_0 - 1 = L(q')_0.$$

In either case, $L(p')_0 = L(q')_0$. Then for each $i \in \{1, 2, \ldots, 2n-1\}$, $p'_i = q'_i$, and $p'_{2n} < p'_i$ if and only if $q'_{2n} < q'_i$ so $L(p')_i = L(q')_i$. Therefore, if $k = 1$ then $\phi_L(p) = \phi_L(q)$.

(d) A similar argument from part (c) will show $\phi_R(p)$ and $\phi_R(q)$ are a complementary pair of type $2k-2$.

(e) If $k \geq 2$, from (d) $\phi_R(p)$ and $\phi_R(q)$ are a complementary pair of type $2k-2$. A similar argument from part (c) will show $L(\phi_R(p)) = \phi_M(p)$ and $L(\phi_R(q)) = \phi_M(q)$ are a complementary pair of type $2k-3$. If $k = 1$, then $\phi_R(p) = \phi_R(q)$ and $\phi_L(p) = \phi_L(q)$ so $\phi_M(p) = \phi_M(q)$. ∎

Theorem 5.2.8 *Let p and q be distinct subpermutations of π_T. Then p and q have the same form if and only if p and q are a complementary pair of type k, for some $k \geq 1$.*

Proof First, suppose that p and q are a complementary pair of type k, for some $k \geq 1$. So there are decompositions:

$$p = \pi_T[a, a+n] = (\alpha_1 \cdots \alpha_k \lambda_1 \cdots \lambda_l \beta_1 \cdots \beta_k)$$
$$q = \pi_T[b, b+n] = (\beta_1 \cdots \beta_k \lambda_1 \cdots \lambda_l \alpha_1 \cdots \alpha_k)$$

so that for $\epsilon \in \{-1, 1\}$, $\alpha_i = \beta_i + \epsilon$ for each $i \in \{1, 2, \ldots, k\}$.

For each $i \in \{0, 1, \ldots, k-2\}$, p_i and $p_{n-k+1+i}$ are consecutive values, as are q_i and $q_{n-k+1+i}$, so

$$p_i < p_{i+1} \text{ and } p_{n-k+1+i} < p_{n-k+1+i+1} \iff q_i < q_{i+1} \text{ and } q_{n-k+1+i} < q_{n-k+1+i+1}.$$

Since $p_{k-1} = q_{k-1} + \epsilon$, $p_{k+l} + \epsilon = q_{k+l}$, $p_k = q_k$, and $p_{k+l-1} = q_{k+l-1}$:

$$\begin{aligned} p_{k-1} < p_k &\iff q_{k-1} < q_k \\ p_{k+l-1} < p_{k+l} &\iff q_{k+l-1} < q_{k+l}. \end{aligned}$$

For each $i \in \{0, 1, \ldots, l-2\}$, $p_{k+i} = q_{k+i}$, so

$$p_{k+i} < p_{k+i+1} \iff q_{k+i} < q_{k+i+1}.$$

Therefore $p_i < p_{i+1}$ if and only if $q_i < q_{i+1}$ for each $i \in \{0, 1, \ldots, n-1\}$, so p and q have the same form.

To show that distinct subpermutations with the same form are a complementary pair of type k, for some $k \geq 1$, an induction argument will be used. The subpermutations of lengths 2 through 9 are listed in Appendix A, along with the form of the subpermutations. It can be seen that distinct subpermutations with the same form are a complementary pair of type k, for some $k \geq 1$.

Assume that $n \geq 9$ and that the theorem is true for all subpermutations of length at most n. Let p' and q' be distinct subpermutations of length $n+1$ with the same form, so $p'_i < p'_{i+1}$ if and only if $q'_i < q'_{i+1}$ for each $i = 0, 1, \ldots, n-1$.

Then

$$p', q' \in \text{Perm}^T_{ev}(n+1) \quad \text{or} \quad p', q' \in \text{Perm}^T_{odd}(n+1).$$

If, without loss of generality, $p' \in \text{Perm}^T_{ev}(n+1)$ and $q' \in \text{Perm}^T_{odd}(n+1)$, then $p' = \pi_T[2a, 2a+n]$ and $q' = \pi_T[2b+1, 2b+n+1]$, so $T[2a, 2a+n-1] = T[2b+1, 2b+n]$. Since $n \geq 9$, $T[2a, 2a+n-1]$ will contain either 00 or 11, so there is some c so that $T[2a+2c+1, 2a+2c+2]$ is 00 or 11. Then also, $T[2b+1+2c+1, 2b+1+2c+2] = T[2b+2c+2, 2b+2c+3]$ must be the same as $T[2a+2c+1, 2a+2c+2]$, but $T[2b+2c+2, 2b+2c+3]$ is either $\mu_T(0) = 01$ or $\mu_T(1) = 10$, so $T[2b+2c+2, 2b+2c+3] \neq T[2a+2c+1, 2a+2c+2]$. Therefore, either $p', q' \in \text{Perm}^T_{ev}(n+1)$ or $p', q' \in \text{Perm}^T_{odd}(n+1)$

Thus one of the 4 following cases must hold:

1. $p', q' \in \text{Perm}^T_{ev}(n+1)$ and $n+1$ is odd.
2. $p', q' \in \text{Perm}^T_{ev}(n+1)$ and $n+1$ is even.
3. $p', q' \in \text{Perm}^T_{odd}(n+1)$and $n+1$ is even.

4. $p', q' \in \text{Perm}^T_{odd}(n+1)$and $n+1$ is odd.

Case 1 Suppose $p', q' \in \text{Perm}^T_{ev}(n+1)$ and $n+1 = 2m+1$, so there are numbers a and b so that $p' = \pi_T[2a, 2a+2m]$ and $q' = \pi_T[2b, 2b+2m]$, and

$$p = \pi_T[a, a+m] \qquad q = \pi_T[b, b+m],$$

$$p' = \phi(p) \qquad q' = \phi(q).$$

If $T[a, a+m-1] \neq T[b, b+m-1]$ then $T[2a, 2a+2m-1] \neq T[2b, 2b+2m-1]$. Hence

$$T[a, a+m-1] = T[b, b+m-1]$$

and p and q have the same form. Since $p = q$ would imply $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type k, for some $k \geq 1$. Therefore, by Proposition 5.2.7, $\phi(p) = p'$ and $\phi(q) = q'$ are a complementary pair of type $2k-1$.

Case 2 Suppose $p', q' \in \text{Perm}^T_{ev}(n+1)$ and $n+1 = 2m$, so there are numbers a and b so that $p' = \pi_T[2a, 2a+2m-1]$ and $q' = \pi_T[2b, 2b+2m-1]$, and

$$p = \pi_T[a, a+m] \qquad q = \pi_T[b, b+m],$$

$$p' = \phi_L(p) \qquad q' = \phi_L(q).$$

Since p' and q' have the same form, $T[2a, 2a+2m-2] = T[2b, 2b+2m-2]$. Thus $T_{2a+2m-2} = T_{2b+2m-2}$ implies $T_{a+m-1} = T_{b+m-1}$, so $T[2a+2m-2, 2a+2m-1] = T[2b+2m-2, 2b+2m-1]$ and

$$T[2a, 2a+2m-1] = T[2b, 2b+2m-1].$$

If $T[a, a+m-1] \neq T[b, b+m-1]$ then $T[2a, 2a+2m-1] \neq T[2b, 2b+2m-1]$. Hence

$$T[a, a+m-1] = T[b, b+m-1]$$

and p and q have the same form. Since $p = q$ would imply $\phi(p) = \phi(q)$, and thus $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type k, for some $k \geq 1$. If $k = 1$, then $\phi_L(p)$ and $\phi(q)_L$ are a complementary pair of type $2k-2 = 0$ and $p' = q'$, thus $k \geq 2$. Therefore, by Proposition 5.2.7, $\phi_L(p) = p'$ and $\phi_L(q) = q'$ are a complementary pair of type $2k-2 \geq 2$.

Case 3 Suppose $p', q' \in \mathrm{Perm}^T_{odd}(n+1)$ and $n+1 = 2m$, so there are numbers a and b so that $p' = \pi_T[2a+1, 2a+2m]$ and $q' = \pi_T[2b+1, 2b+2m]$, and

$$p = \pi_T[a, a+m] \qquad q = \pi_T[b, b+m],$$

$$p' = \phi_R(p) \qquad q' = \phi_R(q).$$

Since p' and q' have the same form, $T[2a+1, 2a+2m-1] = T[2b+1, 2b+2m-1]$. Thus $T_{2a+1} = T_{2b+1}$ implies $T_a = T_b$, so $T[2a, 2a+1] = T[2b, 2b+1]$ and

$$T[2a, 2a+2m-1] = T[2b, 2b+2m-1].$$

If $T[a, a+m-1] \neq T[b, b+m-1]$ then $T[2a, 2a+2m-1] \neq T[2b, 2b+2m-1]$. Hence

$$T[a, a+m-1] = T[b, b+m-1]$$

and p and q have the same form. Since $p = q$ would imply $\phi(p) = \phi(q)$, and thus $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type k, for some $k \geq 1$. If $k = 1$, then $\phi_R(p)$ and $\phi_R(q)$ are a complementary pair of type $2k-2 = 0$ and $p' = q'$, thus $k \geq 2$. Therefore, by Proposition 5.2.7, $\phi_R(p) = p'$ and $\phi(q)_R = q'$ are a complementary pair of type $2k-2 \geq 2$.

Case 4 Suppose $p', q' \in \mathrm{Perm}^T_{odd}(n+1)$ and $n+1 = 2m+1$, so there are numbers a and b so that $p' = \pi_T[2a+1, 2a+2m+1]$ and $q' = \pi_T[2b+1, 2b+2m+1]$, and

$$p = \pi_T[a, a+m+1] \qquad q = \pi_T[b, b+m+1],$$

$$p' = \phi_M(p) \qquad q' = \phi_M(q).$$

Since p' and q' have the same form, $T[2a+1, 2a+2m] = T[2b+1, 2b+2m]$. As in cases 2 and 3 we find $T[a, a+m] = T[b, b+m]$ and p and q have the same form. Since $p = q$ would imply $\phi(p) = \phi(q)$, and thus $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type k, for some $k \geq 1$. If $k = 1$, then $\phi_M(p)$ and $\phi_M(q)$ are a complementary pair of type $2k-3 = -1$ and $p' = q'$, thus $k \geq 2$. Therefore, by Proposition 5.2.7, $\phi_M(p) = p'$ and $\phi_M(q) = q'$ are a complementary pair of type $2k-3 \geq 1$.

Therefore subpermutations p and q have the same form if and only if p and q are a complementary pair of type k, for some $k \geq 1$. ∎

There are a number of useful corollaries of Theorem 5.2.8. These corollaries give the number of subpermutations that can have the same form and show when the maps ϕ_L, ϕ_R, and ϕ_M are not injective.

Corollary 5.2.9 *For a subpermutation p of π_T, there can be at most one subpermutation q of π_T so that p and q are a complementary pair.*

Proof Assume that p is a subpermutation of π_T so that p and q are a complementary pair of type s, and p and r are a complementary pair of type t. Moreover, $s \neq t$, and thus $q \neq r$. Then there are decompositions:

$$p = \pi_T[a, a+n] = (\alpha_1 \cdots \alpha_s \lambda_1 \cdots \lambda_x \beta_1 \cdots \beta_s)$$
$$q = \pi_T[b, b+n] = (\beta_1 \cdots \beta_s \lambda_1 \cdots \lambda_x \alpha_1 \cdots \alpha_s)$$

so that for $\epsilon_s \in \{-1, 1\}$, $\alpha_i = \beta_i + \epsilon_s$ for each $i = 1, 2, \ldots, s$, and

$$p = \pi_T[a, a+n] = (\alpha'_1 \cdots \alpha'_t \lambda'_1 \cdots \lambda'_y \beta'_1 \cdots \beta'_t)$$
$$r = \pi_T[b, b+n] = (\beta'_1 \cdots \beta'_t \lambda'_1 \cdots \lambda'_y \alpha'_1 \cdots \alpha'_t)$$

so that for $\epsilon_t \in \{-1, 1\}$, $\alpha'_i = \beta'_i + \epsilon_t$ for each $i = 1, 2, \ldots, t$.

Since p and q are a complementary pair they have the same form, as do p and r. Thus q and r are distinct subpermutations with the same form, so by Theorem 5.2.8 q and r are a complementary pair of type k, for some k.

If $\beta_1 = \beta'_1$ then $p_{n-s+1} = p_{n-t+1}$, but since $s \neq t$ this cannot happen. Thus $\beta_1 \neq \beta'_1$ and $\epsilon_s \neq \epsilon_t$, so $\epsilon_s = -\epsilon_t$. Hence

$$\begin{array}{ccccc} & & \alpha_1 = \beta_1 + \epsilon_s & \Rightarrow & \beta_1 = \alpha_1 - \epsilon_s \\ \alpha'_1 = \beta'_1 + \epsilon_t & \Rightarrow & \beta'_1 = \alpha'_1 - \epsilon_t & \Rightarrow & \beta'_1 = \alpha_1 + \epsilon_s. \end{array}$$

Therefore $q_0 \neq r_0 \pm 1$, and q and r are not a complementary pair, contradicting the assumption. ∎

The next corollary follows directly from Theorem 5.2.8 and Corollary 5.2.9

Corollary 5.2.10 *For a factor u of T, there are at most two subpermutations of π_T with form u.*

The next corollary shows when the maps $\phi_L(p)$, $\phi_R(p)$, and $\phi_M(p)$ are not injective.

Corollary 5.2.11 *For subpermutations $p = \pi_T[a, a+n]$ and $q = \pi_T[b, b+n]$, where $p \neq q$:*

(a) $\phi_L(p) = \phi_L(q)$ if and only if p and q are a complementary pair of type 1.

(b) $\phi_R(p) = \phi_R(q)$ if and only if p and q are a complementary pair of type 1.

(c) $\phi_M(p) = \phi_M(q)$ if and only if p and q are a complementary pair of type 1.

Proof It should be clear for all three cases that if p and q are a complementary pair of type 1 then

$$\phi_L(p) = \phi_L(q) \qquad \phi_R(p) = \phi_R(q) \qquad \phi_M(p) = \phi_M(q)$$

by Proposition 5.2.7. For the three cases, let $p = \pi_T[a, a+n]$ and $q = \pi_T[b, b+n]$ and $p \neq q$.

(a) Suppose $\phi_L(p) = \phi_L(q)$, so $\pi_T[2a, 2a+2n-1] = \pi_T[2b, 2b+2n-1]$ and $T[2a, 2a+2n-2] = T[2b, 2b+2n-2]$. As before we find $T[2a, 2a+2n-1] = T[2b, 2b+2n-1]$ and $T[a, a+n-1] = T[b, b+n-1]$, so p and q have the same form. By Theorem 5.2.8, p and q are a complementary pair of type $k \geq 1$. If $k > 1$, then $\phi_L(p)$ and $\phi_L(q)$ are a complementary pair of type $2k-2 > 1$, so $\phi_L(p) \neq \phi_L(q)$. Therefore p and q are a complementary pair of type 1.

(b) Suppose $\phi_R(p) = \phi_R(q)$. This again implies $T[a, a+n-1] = T[b, b+n-1]$ so p and q have the same form. As in case (a) we then find p and q are a complementary pair of type 1.

(c) Suppose $\phi_M(p) = \phi_M(q)$. This again implies $T[a, a+n-1] = T[b, b+n-1]$ so p and q have the same form. As in case (a) we then find p and q are a complementary pair of type 1. ∎

So when there are complementary pairs of type 1 none of the maps ϕ_L, ϕ_R, and ϕ_M are injective, and thus they are not bijective. In cases where there are no complementary pairs of type 1 the maps ϕ_L, ϕ_R, and ϕ_M are injective and the inequalities in Lemma 5.2.6 become equalities. So we need to know when complementary pairs of type 1 will occur, and how many complementary pairs there are.

5.2.4 Type 1 Pairs

This section investigates when complementary pairs of type 1 arise and the number of pairs that occur. To show when the maps ϕ_L, ϕ_R, and ϕ_M are bijections we need to consider when complementary pairs of type 1 occur. The following lemma shows when there are complementary pairs of type k, for each $k \geq 0$. An induction argument will be used with Proposition 5.2.7 and Theorem 5.2.8 to show that all complementary pairs of a given length are of same type.

Proposition 5.2.12 *Let $n > 4$ be a natural number and let p and q be subpermutations of π_T of length $n+1$ with the same form. There exist r and c so that $n = 2^r + c$, where $0 \leq c < 2^r$.*

(a) If $0 \leq c < 2^{r-1}+1$, then either $p = q$ or p and q are a complementary pair of type $c+1$.

(b) If $2^{r-1}+1 \leq c < 2^r$, then $p = q$.

Proof This will be proved using an induction argument on r. By looking at the subpermutations in Appendix A it can be readily verified that the lemma is true for $r = 2$ and $c = 0, 1, 2, 3$, so for $n = 4, 5, 6, 7$. Suppose that $r > 2$ and that the statement of the lemma is true when $n < 2^r$. It will be shown that it is true for all $n = 2^r + c$ where $0 \leq c < 2^r$.

(a) Let $n = 2^r + c$ with $0 \leq c < 2^{r-1}+1$. If $p' = q'$ the proposition is satisfied, so assume that $p' \neq q'$. As it was stated in the proof of Theorem 5.2.8, if $p' \in \mathrm{Perm}^T_{ev}(n+1)$ and $q' \in \mathrm{Perm}^T_{odd}(n+1)$, then p' and q' cannot have the same form. We must also consider when $n+1$ is both even and odd. So there will be four subcases to consider, when $p', q' \in \mathrm{Perm}^T_{ev}(n+1)$ or when $p', q' \in \mathrm{Perm}^T_{odd}(n+1)$ and when $n+1$ is even or odd.

Case a.1: Suppose $p', q' \in \mathrm{Perm}^T_{ev}(n+1)$ and $n+1$ is odd, so c is even. There is a d so that $c = 2d$, with $0 \leq d < 2^{r-2}+1$, and there are numbers a and b so that $p' = \pi_T[2a, 2a+2^r+2d]$ and $q' = \pi_T[2b, 2b+2^r+2d]$, and

$$p = \pi_T[a, a+2^{r-1}+d], \qquad q = \pi_T[b, b+2^{r-1}+d],$$

$$p' = \phi(p), \qquad q' = \phi(q).$$

We again find $T[a, a+2^{r-1}+d-1] = T[b, b+2^{r-1}+d-1]$, so p and q have the same form. Since $p = q$ would imply $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type $d+1$. Therefore, by Proposition 5.2.7, $\phi(p) = p'$ and $\phi(q) = q'$ are a complementary pair of type $2(d+1)-1 = 2d+1 = c+1$.

Case a.2: Suppose $p', q' \in \mathrm{Perm}^T_{ev}(n+1)$ and $n+1$ is even, so c is odd. There is a d so that $c = 2d+1$, with $0 \leq d < 2^{r-2}+1$, and there are numbers a and b so that $p' = \pi_T[2a, 2a+2^r+2d+1]$ and $q' = \pi_T[2b, 2b+2^r+2d+1]$, and

$$p = \pi_T[a, a+2^{r-1}+d+1], \qquad q = \pi_T[b, b+2^{r-1}+d+1],$$

$$p' = \phi_L(p), \qquad q' = \phi_L(q).$$

Since p' and q' have the same form, $T[2a, 2a+2^r+2d] = T[2b, 2b+2^r+2d]$. We again find $T[a, a+2^{r-1}+d] = T[b, b+2^{r-1}+d]$, so p and q have the same form. Since $p = q$ would imply $\phi(p) = \phi(q)$, and thus $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type $d+2$. Therefore, by Proposition 5.2.7, $\phi_L(p) = p'$ and $\phi_L(q) = q'$ are a complementary pair of type $2(d+2) - 2 = 2d+2 = c+1$.

Case a.3: Suppose $p', q' \in \text{Perm}^T_{odd}(n+1)$ and $n+1$ is even, so c is odd. There is a d so that $c = 2d+1$, with $0 \leq d < 2^{r-2}+1$, and there are numbers a and b so that $p' = \pi_T[2a+1, 2a+2^r+2d+2]$ and $q' = \pi_T[2b+1, 2b+2^r+2d+2]$, and

$$p = \pi_T[a, a+2^{r-1}+d+1], \qquad q = \pi_T[b, b+2^{r-1}+d+1],$$

$$p' = \phi_R(p), \qquad q' = \phi_R(q).$$

Since p' and q' have the same form, $T[2a+1, 2a+2^r+2d+1] = T[2b+1, 2b+2^r+2d+1]$. We again find $T[a, a+2^{r-1}+d] = T[b, b+2^{r-1}+d]$, so p and q have the same form. Since $p = q$ would imply $\phi(p) = \phi(q)$, and thus $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type $d+2$. Therefore, by Proposition 5.2.7, $\phi_R(p) = p'$ and $\phi(q)_R = q'$ are a complementary pair of type $2(d+2) - 2 = 2d+2 = c+1$.

Case a.4: Suppose $p', q' \in \text{Perm}^T_{odd}(n+1)$ and $n+1$ is odd, so c is even. There is a d so that $c = 2d$, with $0 \leq d < 2^{r-2}+1$, and there are numbers a and b so that $p' = \pi_T[2a+1, 2a+2^r+2d+1]$ and $q' = \pi_T[2b+1, 2b+2^r+2d+1]$, and

$$p = \pi_T[a, a+2^{r-1}+d+1], \qquad q = \pi_T[b, b+2^{r-1}+d+1],$$

$$p' = \phi_M(p), \qquad q' = \phi_M(q).$$

Since p' and q' have the same form, $T[2a+1, 2a+2^r+2d] = T[2b+1, 2b+2^r+2d]$. We again find $T[a, a+2^{r-1}+d] = T[b, b+2^{r-1}+d]$, so p and q have the same form. Since $p = q$ would imply $\phi(p) = \phi(q)$, and thus $p' = q'$, it must be that $p \neq q$. By the induction hypothesis, p and q are a complementary pair of type $d+2$. Therefore, by Proposition 5.2.7, $\phi_M(p) = p'$ and $\phi_M(q) = q'$ are a complementary pair of type $2(d+2) - 3 = 2d+1 = c+1$.

(b) Let $n = 2^r + c$ with $2^{r-1}+1 \leq c < 2^r$. There will again be the four subcases from part (a) when $2^{r-1}+1 \leq c < 2^r - 2$, when $p', q' \in \text{Perm}^T_{ev}(n+1)$ or when $p', q' \in \text{Perm}^T_{odd}(n+1)$ and when $n+1$ is even or odd. There will also be 2 additional special cases to consider, which are when $c = 2^r - 2$ and $c = 2^r - 1$.

Case b.1: Suppose $p', q' \in \text{Perm}_{ev}^T(n+1)$ and $n+1$ is odd, so c is even. There is a d so that $c = 2d$, with $2^{r-2}+1 \leq d < 2^{r-1}$, and there are numbers a and b so that $p' = \pi_T[2a, 2a+2^r+2d]$ and $q' = \pi_T[2b, 2b+2^r+2d]$, and

$$p = \pi_T[a, a+2^{r-1}+d], \qquad q = \pi_T[b, b+2^{r-1}+d],$$

$$p' = \phi(p), \qquad q' = \phi(q).$$

As in case **a.1**, $T[a, a+2^{r-1}+d-1] = T[b, b+2^{r-1}+d-1]$, so p and q have the same form. By the induction hypothesis $p = q$, so by Corollary 5.2.5, $p' = \phi(p) = \phi(q) = q'$.

Case b.2: Suppose $p', q' \in \text{Perm}_{odd}^T(n+1)$ and $n+1$ is odd, so c is even. There is a d so that $c = 2d$, with $2^{r-2}+1 \leq d < 2^{r-1}$, and there are numbers a and b so that $p' = \pi_T[2a+1, 2a+2^r+2d+1]$ and $q' = \pi_T[2b+1, 2b+2^r+2d+1]$, and

$$p = \pi_T[a, a+2^{r-1}+d+1], \qquad q = \pi_T[b, b+2^{r-1}+d+1],$$

$$p' = \phi_M(p), \qquad q' = \phi_M(q).$$

As in case **a.4**, $T[a, a+2^{r-1}+d] = T[b, b+2^{r-1}+d]$, so p and q have the same form. By the induction hypothesis $p = q$, so by Corollary 5.2.5, $\phi(p) = \phi(q)$ and therefore $p' = \phi_M(p) = \phi_M(q) = q'$.

Case b.3: Suppose $p', q' \in \text{Perm}_{ev}^T(n+1)$ and $n+1$ is even, so c is odd. There is a d so that $c = 2d+1$, with $2^{r-2}+1 \leq d < 2^{r-1}$, and there are numbers a and b so that $p' = \pi_T[2a, 2a+2^r+2d+1]$ and $q' = \pi_T[2b, 2b+2^r+2d+1]$, and

$$p = \pi_T[a, a+2^{r-1}+d+1], \qquad q = \pi_T[b, b+2^{r-1}+d+1],$$

$$p' = \phi_L(p), \qquad q' = \phi_L(q).$$

As in case **a.2**, $T[a, a+2^{r-1}+d] = T[b, b+2^{r-1}+d]$, so p and q have the same form. By the induction hypothesis $p = q$, so by Corollary 5.2.5, $\phi(p) = \phi(q)$ and therefore $p' = \phi_L(p) = \phi_L(q) = q'$.

Case b.4: Suppose $p', q' \in \text{Perm}_{odd}^T(n+1)$ and $n+1$ is even, so c is odd. There is a d so that $c = 2d+1$, with $2^{r-2}+1 \leq d < 2^{r-1}$, and there are numbers a and b so that $p' = \pi_T[2a+1, 2a+2^r+2d+2]$ and $q' = \pi_T[2b+1, 2b+2^r+2d+2]$, and

$$p = \pi_T[a, a+2^{r-1}+d+1], \qquad q = \pi_T[b, b+2^{r-1}+d+1],$$

$$p' = \phi_R(p), \qquad q' = \phi_R(q).$$

As in case **a.3**, $T[a, a + 2^{r-1} + d] = T[b, b + 2^{r-1} + d]$, so p and q have the same form. By the induction hypothesis $p = q$, so by Corollary 5.2.5, $\phi(p) = \phi(q)$ and therefore $p' = \phi_R(p) = \phi_R(q) = q'$.

Case b.5: Suppose $c = 2^r - 2$. Thus $n = 2^r + 2^r - 2 = 2^{r+1} - 2$, and the subpermutations p' and q' will have odd length. There will be two subcases, these being when $p', q' \in \mathrm{Perm}^T_{ev}(n+1)$ and when $p', q' \in \mathrm{Perm}^T_{odd}(n+1)$.

Case b.5.i: Suppose $p', q' \in \mathrm{Perm}^T_{ev}(n+1)$. There are numbers a and b so that $p' = \pi_T[2a, 2a + 2^{r+1} - 2]$ and $q' = \pi_T[2b, 2b + 2^{r+1} - 2]$, and

$$p = \pi_T[a, a + 2^r - 1], \qquad q = \pi_T[b, b + 2^r - 1],$$

$$p' = \phi(p), \qquad q' = \phi(q).$$

As in cases **a.1** and **b.1**, $T[a, a+2^r-2] = T[b, b+2^r-2]$, so p and q have the same form. By the induction hypothesis $p = q$, so by Corollary 5.2.5, $p' = \phi(p) = \phi(q) = q'$.

Case b.5.ii: Suppose $p', q' \in \mathrm{Perm}^T_{odd}(n+1)$. There are numbers a and b so that $p' = \pi_T[2a+1, 2a + 2^{r+1} - 1]$ and $q' = \pi_T[2b+1, 2b + 2^{r+1} - 1]$, and

$$p = \pi_T[a, a + 2^r], \qquad q = \pi_T[b, b + 2^r],$$

$$p' = \phi_M(p), \qquad q' = \phi_M(q).$$

As in cases **a.2** and **b.2**, $T[a, a + 2^r - 1] = T[b, b + 2^r - 1]$, so p and q have the same form. If $p = q$ then $\phi(p) = \phi(q)$, by Corollary 5.2.5, and $p' = M(\phi(p)) = M(\phi(q)) = q'$. If $p \neq q$ then by case **a.1**, p and q are a complementary pair of type 1. Therefore, by Proposition 5.2.7, $p' = \phi_M(p) = \phi_M(q) = q'$.

Case b.6: Suppose $c = 2^r - 1$. Thus $n = 2^r + 2^r - 1 = 2^{r+1} - 1$, and the subpermutations p' and q' will have even length. There will be two subcases, these being when $p', q' \in \mathrm{Perm}^T_{ev}(n+1)$ and when $p', q' \in \mathrm{Perm}^T_{odd}(n+1)$.

Case b.6.i: Suppose $p', q' \in \mathrm{Perm}^T_{ev}(n+1)$. There are numbers a and b so that $p' = \pi_T[2a, 2a + 2^{r+1} - 1]$ and $q' = \pi_T[2b, 2b + 2^{r+1} - 1]$, and

$$p = \pi_T[a, a + 2^r], \qquad q = \pi_T[b, b + 2^r],$$

$$p' = \phi_L(p), \qquad q' = \phi_L(q).$$

As in cases **a.2** and **b.3**, $T[a, a+2^r-1] = T[b, b+2^r-1]$, so p and q have the same form. If $p = q$ then $\phi(p) = \phi(q)$, by Corollary 5.2.5, and $p' = L(\phi(p)) = L(\phi(q)) = q'$. If $p \neq q$ then by case **a.1**, p and q are a complementary pair of type 1. Therefore, by Proposition 5.2.7, $p' = \phi_L(p) = \phi_L(q) = q'$.

Case b.6.ii: Suppose $p', q' \in \text{Perm}^T_{odd}(n+1)$. There are numbers a and b so that $p' = \pi_T[2a+1, 2a+2^{r+1}]$ and $q' = \pi_T[2b+1, 2b+2^{r+1}]$, and

$$p = \pi_T[a, a+2^r], \qquad q = \pi_T[b, b+2^r],$$

$$p' = \phi_R(p), \qquad q' = \phi_R(q).$$

As in cases **a.3** and **b.4**, $T[a, a+2^r-1] = T[b, b+2^r-1]$, so p and q have the same form. If $p = q$ then $\phi(p) = \phi(q)$, by Corollary 5.2.5, and $p' = R(\phi(p)) = R(\phi(q)) = q'$. If $p \neq q$ then by case **a.1**, p and q are a complementary pair of type 1. Therefore, by Proposition 5.2.7, $p' = \phi_R(p) = \phi_R(q) = q'$.

Therefore the lemma is true when $n = 2^r + c$ with $0 \leq c < 2^r$, and therefore for all n. ∎

Thus, only subpermutations of length $2^r + 1$ can be a complementary pair of type 1, and we have the following corollary.

Corollary 5.2.13 *If $n \neq 2^r$, for $r \geq 1$, then for any subpermutations $p = \pi_T[a, a+n]$ and $q = \pi_T[b, b+n]$*

(a) $\phi_L(p) = \phi_L(q)$ if and only if $p = q$.

(b) $\phi_R(p) = \phi_R(q)$ if and only if $p = q$.

(c) $\phi_M(p) = \phi_M(q)$ if and only if $p = q$.

Proof It should be clear in each case that if $p = q$ then

$$\phi_L(p) = \phi_L(q) \qquad \phi_R(p) = \phi_R(q) \qquad \phi_M(p) = \phi_M(q).$$

Suppose $\phi_L(p) = \phi_L(q)$. If $p \neq q$, by Corollary 5.2.11 p and q are a complementary pair of type 1. By Proposition 5.2.12, p and q are cannot be complementary pair of type 1, therefore $p = q$.

A similar argument will show if $\phi_R(p) = \phi_R(q)$ then $p = q$, and if $\phi_M(p) = \phi_M(q)$ then $p = q$. ∎

We now consider the number of factors u of T of length 2^r that have two subpermutations which form a complementary pair of type 1.

Lemma 5.2.14 *Let $n = 2^r$ or $2^r + 1$, with $r \geq 2$. Then there are exactly 2^r factors u of T of length n so that there exist subpermutations $p = \pi_T[a, a+n]$ and $q = \pi_T[b, b+n]$ with form u and $p \neq q$.*

Proof It can be readily verified by looking at the subpermutations in Appendix A that the lemma is true for $r = 2$. So there are 4 factors u of T of length 4 with two distinct subpermutations of length 5 with form u, and there are 4 factors v of T of length 5 with two distinct subpermutations of length 6 with form v.

Suppose $r \geq 2$ and that the lemma is true for r. We now show the lemma is true for $r+1$. Let Γ be the set of factors of length 2^r, $|\Gamma| = 2^r$, so that for $u \in \Gamma$ there are subpermutations p and q with form u so that $p \neq q$, hence, by Proposition 5.2.12, p and q are a complementary pair of type 1. Let Γ' be the set of factors of length 2^{r+1} so that if $u \in \Gamma'$ then there exist subpermutations p and q with form u so that $p \neq q$. Let Δ be the set of factors of length $2^r + 1$, $|\Delta| = 2^r$, so that for $v \in \Delta$ there are subpermutations p and q with form v so that $p \neq q$, hence, by Proposition 5.2.12, p and q are a complementary pair of type 2. Let Δ' be the set of factors of length $2^{r+1} + 1$ so that if $v \in \Delta'$ then there exist subpermutations p and q with form v so that $p \neq q$.

The sizes of Γ' and Δ' will be considered in two cases.

Case Γ': Any factor in Γ' will either start in an even position or an odd position, call these sets of factors Γ'_{ev} and Γ'_{odd} and hence

$$\Gamma' = \Gamma'_{ev} \cup \Gamma'_{odd}.$$

Since the factors are of length $2^{r+1} \geq 8$, for any factors $s \in \Gamma'_{ev}$ and $t \in \Gamma'_{odd}$, $s \neq t$, thus

$$\Gamma'_{ev} \cap \Gamma'_{odd} = \emptyset.$$

There will be two subcases to establish the size of Γ', first by showing the size of Γ'_{ev} and then the size of Γ'_{odd}.

Subcase Γ'_{ev}: For $u \in \Gamma$ there are subpermutations p and q of π_T of length $2^r + 1$, so that $p \neq q$. By Proposition 5.2.12, p and q are a complementary pair of type 1. By Proposition 5.2.7 $\phi(p)$ and $\phi(q)$ are a complementary pair of type 1, so $\phi(p) \neq \phi(q)$ and they both have form $\mu_T(u)$. Therefore for each $u \in \Gamma$, $\mu_T(u) \in \Gamma'_{ev}$. Hence

$$|\Gamma'_{ev}| \geq |\Gamma|.$$

Suppose that $u' \in \Gamma'_{ev}$, so there are subpermutations $p' = \pi_T[2a, 2a + 2^{r+1}]$ and $q' = \pi_T[2b, 2b + 2^{r+1}]$ with form $u' = T[2a, 2a + 2^{r+1} - 1] = T[2b, 2b + 2^{r+1} - 1]$, so that $p' \neq q'$. Hence there exist subpermutations p and q so that $\phi(p) = p'$ and $\phi(q) = q'$. As in case **a.1** of Proposition 5.2.12, p and q are a complementary pair of type 1 with form u where $\mu_T(u) = u'$. Thus for each $u' \in \Gamma'_{ev}$, there is some $u \in \Gamma$ so that $\mu_T(u) = u'$. Hence

$$|\Gamma'_{ev}| \leq |\Gamma|.$$

Therefore $|\Gamma'_{ev}| = |\Gamma|$.

Subcase Γ'_{odd}: For $u \in \Delta$, $u = T[a, a+2^r]$, there are subpermutations p and q of π_T of length $2^r + 2$, so that $p \neq q$. By Proposition 5.2.12, p and q are a complementary pair of type 2. By Proposition 5.2.7, $\phi(p)$ and $\phi(q)$ are a complementary pair of type 3 with form $\mu_T(u) = T[2a, 2a + 2^{r+1} + 1]$ and $\phi_M(p)$ and $\phi_M(q)$ are a complementary pair of type 1, so $\phi_M(p) \neq \phi_M(q)$ and they both have form $T[2a+1, 2a+2^{r+1}]$. Therefore for each $T[a, a+2^r] \in \Delta$, $T[2a+1, 2a+2^{r+1}] \in \Gamma'_{odd}$. Hence

$$|\Gamma'_{odd}| \geq |\Delta|\,.$$

Suppose that $u' \in \Gamma'_{odd}$, so there are subpermutations $p' = \pi_T[2a+1, 2a+2^{r+1}+1]$ and $q' = \pi_T[2b+1, 2b+2^{r+1}+1]$ with form $u' = T[2a+1, 2a+2^{r+1}] = T[2b+1, 2b+2^{r+1}]$, so that $p' \neq q'$. Hence there exist subpermutations p and q so that $\phi_M(p) = p'$ and $\phi_M(q) = q'$. As in case **a.4** of Proposition 5.2.12, p and q are a complementary pair of type 2 with form $T[a, a+2^r]$. Thus for each $u' \in \Gamma'_{odd}$, there is some $T[a, a+2^r] \in \Delta$ so that $u' = T[2a+1, 2a+2^{r+1}]$. Hence

$$|\Gamma'_{odd}| \leq |\Delta|\,.$$

Therefore $|\Gamma'_{odd}| = |\Delta|$.

Therefore

$$|\Gamma'| = |\Gamma'_{ev}| + |\Gamma'_{odd}| = |\Gamma| + |\Delta| = 2^r + 2^r = 2^{r+1}.$$

Case Δ': Any factor in Δ' will either start in an even position or an odd position, call these sets of factors Δ'_{ev} and Δ'_{odd} and hence

$$\Delta' = \Delta'_{ev} \cup \Delta'_{odd}.$$

Since the factors are of length $2^{r+1} + 1 \geq 8$, for any factors $s \in \Delta'_{ev}$ and $t \in \Delta'_{odd}$, $s \neq t$, thus

$$\Delta'_{ev} \cap \Delta'_{odd} = \emptyset.$$

There will be two subcases to establish the size of Δ', first by showing the size of Δ'_{ev} and then the size of Δ'_{odd}.

Subcase Δ'_{ev}: For $u \in \Delta$, $u = T[a, a+2^r]$, there are subpermutations p and q of π_T of length $2^r + 2$, so that $p \neq q$. By Proposition 5.2.12, p and q are a complementary pair of type 2. By Proposition 5.2.7, $\phi(p)$ and $\phi(q)$ are a complementary pair of type 3 with form $\mu_T(u) = T[2a, 2a + 2^{r+1} + 1]$ and $\phi_L(p)$ and $\phi_L(q)$ are a complementary pair of type 2, so $\phi_L(p) \neq \phi_L(q)$ and they both have form $T[2a, 2a+2^{r+1}]$. Therefore for each $T[a, a+2^r] \in \Delta$, $T[2a, 2a+2^{r+1}] \in \Delta'_{ev}$. Hence

$$|\Delta'_{ev}| \geq |\Delta|\,.$$

Suppose that $u' \in \Delta'_{ev}$, so there are subpermutations $p' = \pi_T[2a, 2a + 2^{r+1} + 1]$ and $q' = \pi_T[2b, 2b + 2^{r+1} + 1]$ with form $u' = T[2a, 2a + 2^{r+1}] = T[2b, 2b + 2^{r+1}]$, so that $p' \neq q'$. Hence there exist subpermutations p and q so that $\phi_L(p) = p'$ and $\phi_L(q) = q'$. As in case **a.3** of Proposition 5.2.12, p and q are a complementary pair of type 2 with form $u = T[a, a + 2^r]$. Thus for each $u' \in \Delta'_{ev}$, there is some $T[a, a + 2^r] \in \Delta$ so that $u' = T[2a, 2a + 2^{r+1}]$. Hence

$$|\Delta'_{ev}| \leq |\Delta| .$$

Therefore $|\Delta'_{ev}| = |\Delta|$.

Subcase Δ'_{odd}: A symmetric argument to the argument used in Subcase Δ'_{ev} will show $|\Delta'_{odd}| = |\Delta|$.

Therefore

$$|\Delta'| = |\Delta'_{ev}| + |\Delta'_{odd}| = |\Delta| + |\Delta| = 2^r + 2^r = 2^{r+1}.$$

∎

Now we know when there are complementary pairs of type 1, and how many pairs of type 1 there are in each case.

5.2.5 Permutation Complexity of T

We are now ready to give a recursive definition for the permutation complexity of T. To show this we consider when the maps ϕ, ϕ_L, ϕ_R, and ϕ_M are bijective. After the recursive definition is given, it will be shown that the recursive definition yields a formula for the permutation complexity.

Proposition 5.2.15 *Let $n \in \mathbb{N}$. When $2n + 1 = 2^r - 1$, for some $r \geq 3$:*

$$\tau_T(2n + 1) = \tau_T(n + 1) + \tau_T(n + 2) - 2^{r-1}.$$

When $2n = 2^r$, for some $r \geq 3$:

$$\tau_T(2n) = 2(\tau_T(n + 1) - 2^{r-1}).$$

For all other $n \geq 3$:

$$\tau_T(2n) = 2(\tau_T(n + 1))$$
$$\tau_T(2n + 1) = \tau_T(n + 1) + \tau_T(n + 2).$$

Proof For any n,

$$\tau_T(n) = \left|\mathrm{Perm}^T(n)\right| = \left|\mathrm{Perm}^T_{ev}(n)\right| + \left|\mathrm{Perm}^T_{odd}(n)\right| .$$

This proof will be done in three cases. The first is when $2n+1 = 2^r - 1$ for some $r \geq 3$, the second is when $2n = 2^r$ for some $r \geq 3$, and the third for all other n.

Case $2n+1 = 2^r - 1$**:** It can be readily verified by looking at the subpermutations in Appendix A that the proposition is true for $r = 3$. Suppose $r \geq 3$ and the lemma is true for r. We show that the lemma is true for $r+1$, so $2n+1 = 2^{r+1} - 1$.

Since the map

$$\phi : \mathrm{Perm}^T(n+1) \mapsto \mathrm{Perm}^T_{ev}(2n+1)$$

is a bijection, $\left|\mathrm{Perm}^T_{ev}(2n+1)\right| = \left|\mathrm{Perm}^T(n+1)\right| = \tau_T(n+1)$. The map

$$\phi_M : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^T_{odd}(2n+1)$$

is a surjective map , but it is not injective because $n+2 = 2^r + 1$. So there are 2^r factors u of length 2^r with a complementary pair of type 1 by Proposition 5.2.12 and Lemma 5.2.14. Thus there are exactly 2^r complementary pairs of type 1 in $\mathrm{Perm}^T(n+2)$. So 2^{r+1} subpermutations in $\mathrm{Perm}^T(n+2)$ will be mapped to 2^r subpermutations in $\mathrm{Perm}^T_{odd}(2n+1)$ under ϕ_M. The other $\mathrm{Perm}^T(n+2) - 2^{r+1}$ subpermutations in $\mathrm{Perm}^T(n+2)$ are pairwise distinct and not complementary pairs, and thus will be pairwise distinct under ϕ_M. Hence

$$\left|\mathrm{Perm}^T_{odd}(2n+1)\right| = \left(\left|\mathrm{Perm}^T(n+2)\right| - 2^{r+1}\right) + 2^r = \tau_T(n+2) - 2^r.$$

Therefore

$$\tau_T(2n+1) = \tau_T(n+1) + \tau_T(n+2) - 2^r.$$

Case $2n+1 = 2^r$**:** It can be readily verified by looking at the subpermutations in Appendix A that the proposition is true for $r = 3$. Suppose $r \geq 3$ and the lemma is true for r, and we show that the lemma is true for $r+1$, so $2n+1 = 2^{r+1}$.

The map

$$\phi_L : \mathrm{Perm}^T(n+1) \mapsto \mathrm{Perm}^T_{ev}(2n)$$

is a surjective map, but it is not injective because $n+1 = 2^r + 1$. So there are 2^r factors u of length 2^r with a complementary pair of type 1 by Proposition 5.2.12 and Lemma 5.2.14. Thus there are exactly 2^r complementary pairs of type 1 in $\mathrm{Perm}^T(n+1)$. So 2^{r+1} subpermutations in $\mathrm{Perm}^T(n+1)$ will be mapped to 2^r subpermutations in $\mathrm{Perm}^T_{ev}(2n)$ under ϕ_M. The other $\left|\mathrm{Perm}^T(n+1)\right| - 2^{r+1}$ subpermutations in $\mathrm{Perm}^T(n+1)$ are pairwise distinct and not complementary pairs, and thus will be pairwise distinct under ϕ_L. Hence

$$\left|\mathrm{Perm}^T_{ev}(2n)\right| = \left(\left|\mathrm{Perm}^T(n+1)\right| - 2^{r+1}\right) + 2^r = \left|\mathrm{Perm}^T(n+1)\right| - 2^r.$$

The map

$$\phi_R : \mathrm{Perm}^T(n+1) \mapsto \mathrm{Perm}^T_{odd}(2n)$$

is a surjective map, but it is not injective because $n+1 = 2^r + 1$. By a similar argument to above we can see

$$\left|\text{Perm}^T_{odd}(2n)\right| = \left|\text{Perm}^T(n+1)\right| - 2^r.$$

Therefore

$$\tau_T(2n) = (\left|\text{Perm}^T(n+1)\right| - 2^r) + (\left|\text{Perm}^T(n+1)\right| - 2^r) = 2(\tau_T(n+1) - 2^r).$$

Case $n \geq 3$: It can be readily verified by looking at the subpermutations in Appendix A that the proposition is true for $n = 3$. Suppose $n \geq 3$ and the lemma is true for n, and we show that the lemma is true for $n+1$. Since $2(n+1)+1, 2(n+1) \notin \{2^r - 1, 2^r | r \geq 2\}$ for any r, we have $n+2, n+3 \notin \{2^r + 1 | r \geq 2\}$. So for $2(n+1)$ and $2(n+1)+1$ each of the maps

$$\begin{aligned}
\phi &: \text{Perm}^T(n+2) \mapsto \text{Perm}^T_{ev}(2(n+1)+1) \\
\phi_L &: \text{Perm}^T(n+2) \mapsto \text{Perm}^T_{ev}(2(n+1)) \\
\phi_R &: \text{Perm}^T(n+2) \mapsto \text{Perm}^T_{odd}(2(n+1)) \\
\phi_M &: \text{Perm}^T(n+3) \mapsto \text{Perm}^T_{odd}(2(n+1)+1)
\end{aligned}$$

are all bijections. Therefore:

$$\begin{aligned}
\left|\text{Perm}^T_{ev}(2(n+1)+1)\right| &= \left|\text{Perm}^T(n+2)\right| = \tau_T(n+2) \\
\left|\text{Perm}^T_{ev}(2(n+1))\right| &= \left|\text{Perm}^T(n+2)\right| = \tau_T(n+2) \\
\left|\text{Perm}^T_{odd}(2(n+1))\right| &= \left|\text{Perm}^T(n+2)\right| = \tau_T(n+2) \\
\left|\text{Perm}^T_{odd}(2(n+1)+1)\right| &= \left|\text{Perm}^T(n+3)\right| = \tau_T(n+3).
\end{aligned}$$

So:

$$\begin{aligned}
\tau_T(2(n+1)) &= \left|\text{Perm}^T_{ev}(2(n+1))\right| + \left|\text{Perm}^T_{odd}(2(n+1))\right| = 2(\tau_T(n+2)) \\
\tau_T(2(n+1)+1) &= \left|\text{Perm}^T_{ev}(2(n+1)+1)\right| + \left|\text{Perm}^T_{odd}(2(n+1)+1)\right| \\
&= \tau_T(n+2) + \tau_T(n+3).
\end{aligned}$$

∎

Theorem 5.2.16 *For any $n \geq 6$, where $n = 2^r + p$ with $0 < p \leq 2^r$,*

$$\tau_T(n) = 2(2^{r+1} + p - 2).$$

Proof The proof will be done by induction on n. The above formula can be readily verified by looking at the subpermutations listed in Appendix A for $n \leq 9$. Suppose the theorem is true for all values less than or equal to $2n$.

Case $2n+1 = 2^r - 1$: Suppose $2n+1 = 2^r - 1 = 2^{r-1} + 2^{r-1} - 1$, then $n = 2^{r-1} - 1$, so $n+1 = 2^{r-1} = 2^{r-2} + 2^{r-2}$ and $n+2 = 2^{r-1} + 1$. Thus:

$$\tau_T(n+1) = 2(2^{r-2+1} + 2^{r-2} - 2) = 2(2^{r-1} + 2^{r-2} - 2) = 2(3(2^{r-2}) - 2)$$
$$\tau_T(n+2) = 2(2^{r-1+1} + 1 - 2) = 2(2^r - 1)$$

From Proposition 5.2.15:

$$\begin{aligned}\tau_T(2n+1) &= 2(3(2^{r-2}) - 2) + 2(2^r - 1) - 2^{r-1} = 2(3(2^{r-2}) - 2 + 2^r - 1 - 2^{r-2}) \\ &= 2(2(2^{r-2}) + 2^r - 3) = 2(2^r + (2^{r-1} - 1) - 2)\end{aligned}$$

Case $2n+2 = 2(n+1) = 2^r$: Suppose $2n+2 = 2(n+1) = 2^r = 2^{r-1} + 2^{r-1}$:

$$\begin{aligned}\tau_T(2(n+1)) &= 2(2(2^r - 1) - 2^{r-1}) = 2(2^{r+1} - 2^{r-1} - 2) = 2(3(2^{r-1}) - 2) \\ &= 2(2(2^{r-1}) + 2^{r-1} - 2) = 2(2^r + 2^{r-1} - 2)\end{aligned}$$

Case else: Suppose $2n+1 = 2^r + p$, $2n+2 = 2(n+1) = 2^r + p + 1$, and $0 < p < 2^r - 1$. Since $2n+1 = 2^r + p$ is odd, p is odd. So $n = 2^{r-1} + \frac{p-1}{2}$, $n+1 = 2^{r-1} + \frac{p+1}{2}$, and $n+2 = 2^{r-1} + \frac{p+3}{2}$. Thus:

$$\tau_T(n+1) = 2(2^r + \frac{p+1}{2} - 2)$$
$$\tau_T(n+2) = 2(2^r + \frac{p+3}{2} - 2).$$

From Proposition 5.2.15:

$$\begin{aligned}\tau_T(2n+1) &= 2(2^r + \frac{p+1}{2} - 2) + 2(2^r + \frac{p+3}{2} - 2) \\ &= 2(2^r + 2^r + \frac{p+1}{2} + \frac{p+3}{2} - 2 - 2) = 2(2^{r+1} + \frac{2p+4}{2} - 4) \\ &= 2(2^{r+1} + p - 2)\end{aligned}$$

$$\begin{aligned}\tau_T(2(n+1)) &= 2(2(2^r + \frac{p+3}{2} - 2)) = 2(2^{r+1} + p + 3 - 4) \\ &= 2(2^{r+1} + (p+1) - 2).\end{aligned}$$

Therefore, for all $n \geq 6$, where $n = 2^r + p$ with $0 < p \leq 2^r$, $\tau_T(n) = 2(2^{r+1} + p - 2)$ ∎

5.3 Permutation Complexity and the Doubling Map

Recall the doubling map, d, on $\mathcal{A}$ defined by $d : a \mapsto aa$ for each $a \in \mathcal{A}$. Let $\mathcal{A} = \{0, 1\}$, and $\omega \in \mathcal{A}^{\mathbb{N}}$ be an aperiodic uniformly recurrent word. In this section we will investigate permutation complexity of $d(\omega)$. We will give an upper bound for the permutation complexity of $d(\omega)$ by looking at how the doubling map alters the subpermutations of π_ω. After some general results are established we will give a method to calculate the permutation complexity of $d(s)$, where s is a Sturmian word, and give the permutation complexity function of $d(T)$, where T is the Thue-Morse word.

5.3.1 Uniformly Recurrent Words

Let ω be an aperiodic uniformly recurrent word over $\{0, 1\}$, and π_ω be the infinite permutation associated with ω using the natural order on the alphabet. We would like to describe the infinite permutation associated with $d(\omega)$, the image of ω under the doubling map. If $u = \omega[a, a + n - 1]$ is a factor of ω of length n, it is helpful to note $d(u) = d(\omega)[2a, 2a + 2n - 1]$ will be a factor of $d(\omega)$ of length $2n$.

Since ω is a uniformly recurrent word it can not have arbitrarily long strings of contiguous 0 or 1. Thus there are $k_0, k_1 \in \mathbb{N}$ so that

$$\begin{gathered} 10^{k_0}1 \\ 01^{k_1}0 \end{gathered}$$

are factors of ω, but 0^{k_0+1} and 1^{k_1+1} are not. We then define the following class of words:

$$\begin{aligned} C_0 &= 0^{k_0} \\ C_1 &= 0^{k_0-1}1 \\ C_2 &= 0^{k_0-2}1 \\ &\vdots \\ C_{k_0-1} &= 01 \\ C_{k_0} &= 10 \\ &\vdots \\ C_{k_0+k_1-2} &= 1^{k_1-1}0 \\ C_{k_0+k_1-1} &= 1^{k_1}. \end{aligned}$$

For each $i \in \mathbb{N}$, $\omega[i] = \omega_i\omega_{i+1}\cdots$ can have exactly one the above classes of words as a prefix. It should be clear $C_0 < C_1 < \cdots < C_{k_0+k_1-1}$, and so $d(C_i) < d(C_j)$ for $i < j$ since the doubling map d is order preserving, as shown in Lemma 5.3.1. The next lemma will not only show that the doubling map is an order preserving map, but also the order of the image of ω_i under the doubling map.

Lemma 5.3.1 *Let ω be as above. Suppose $\omega[a]$ and $\omega[b]$ are two shifts of ω for some $a \neq b$ so that $\omega[a] < \omega[b]$. Moreover, suppose C_i is a prefix of $\omega[a]$ and C_j is a prefix of $\omega[b]$ where $i \leq j$. Then $d(\omega[a]) < d(\omega[b])$, and*

(a) If $\omega_a = \omega_b = 0$ and $i < j$, then $d(\omega)[2a] < d(\omega)[2a+1] < d(\omega)[2b] < d(\omega)[2b+1]$.

(b) If $\omega_a = \omega_b = 0$ and $i = j$, then $d(\omega)[2a] < d(\omega)[2b] < d(\omega)[2a+1] < d(\omega)[2b+1]$.

(c) If $\omega_a = 0$ and $\omega_b = 1$, then $d(\omega)[2a] < d(\omega)[2a+1] < d(\omega)[2b+1] < d(\omega)[2b]$.

(d) If $\omega_a = \omega_b = 1$and $i < j$, then $d(\omega)[2a+1] < d(\omega)[2a] < d(\omega)[2b+1] < d(\omega)[2b]$.

(e) If $\omega_a = \omega_b = 1$and $i = j$, then $d(\omega)[2a+1] < d(\omega)[2b+1] < d(\omega)[2a] < d(\omega)[2b]$.

Proof Since $\omega[a] < \omega[b]$, there is some (possibly empty) factor u of ω so that

$$\omega[a] = u0\cdots$$

$$\omega[b] = u1\cdots$$

and thus

$$d(\omega[a]) = d(u)00\cdots$$

$$d(\omega[b]) = d(u)11\cdots$$

so $d(\omega[a]) < d(\omega[b])$ and d is an order preserving map.

Each of the cases will be looked at independently.

(a) Suppose $\omega_a = \omega_b = 0$ and $i < j$. Since both $\omega[a]$ and $\omega[b]$ start with 0, $\omega[a]$ has $0^{k_0-i}1$ as a prefix and $\omega[b]$ has $0^{k_0-j}1$ as a prefix. Thus $d(\omega)[2a]$ has $0^{2(k_0-i)}1$ as a prefix and $d(\omega)[2b]$ has $0^{2(k_0-j)}1$ as a prefix, and the result follows from

$$0^{2(k_0-i)}1 < 0^{2(k_0-i)-1}1 < 0^{2(k_0-j)}1 < 0^{2(k_0-j)-1}1.$$

(b) Suppose $\omega_a = \omega_b = 0$ and $i = j$. Since both $\omega[a]$ and $\omega[b]$ start with 0, $\omega[a]$ and $\omega[b]$ have $0^{k_0-i}1$ as a prefix. Thus $d(\omega)[2a]$ and $d(\omega)[2b]$ have $0^{2(k_0-i)}1$ as a prefix. Since $\omega[a] < \omega[b]$ is given and d is an order preserving map

$$d(\omega)[2a] < d(\omega)[2b]$$

$$d(\omega)[2a+1] < d(\omega)[2b+1]$$

$$0^{2(k_0-i)}1 < 0^{2(k_0-i)-1}1.$$

Thus $d(\omega)[2a] < d(\omega)[2b] < d(\omega)[2a+1] < d(\omega)[2b+1]$.

(c) Suppose $\omega_a = 0$ and $\omega_b = 1$, so $i < j$. Since $\omega[a]$ starts with 0 and $\omega[b]$ starts with 1, $\omega[a]$ has $0^{k_0-i}1$ as a prefix and $\omega[b]$ has $1^{j-k_0+1}0$ as a prefix. Thus $d(\omega)[2a]$ has $0^{2(k_0-i)}1$ as a prefix and $d(\omega)[2b]$ has $1^{2(j-k_0)+2}0$ as a prefix, and the result follows from

$$0^{2(k_0-i)}1 < 0^{2(k_0-i)-1}1 < 1^{2(j-k_0)+1}0 < 1^{2(j-k_0)+2}0.$$

(d) Suppose $\omega_a = \omega_b = 1$ and $i < j$. Since both $\omega[a]$ and $\omega[b]$ start with 1, $\omega[a]$ has $1^{i-k_0+1}0$ as a prefix and $\omega[b]$ has $1^{j-k_0+1}0$ as a prefix. Thus $d(\omega)[2a]$ has $1^{2(i-k_0)+2}0$ as a prefix and $d(\omega)[2b]$ has $1^{2(j-k_0)+2}0$ as a prefix, and the result follows from

$$1^{2(i-k_0)+1}0 < 1^{2(i-k_0)+2}0 < 1^{2(j-k_0)+1}0 < 1^{2(j-k_0)+2}0.$$

(e) Suppose $\omega_a = \omega_b = 1$ and $i = j$. Since both $\omega[a]$ and $\omega[b]$ start with 1, $\omega[a]$ and $\omega[b]$ have $1^{i-k_0+1}0$ as a prefix. Thus $d(\omega)[2a]$ and $d(\omega)[2b]$ have $1^{2(i-k_0)+2}0$ as a prefix. Since $\omega[a] < \omega[b]$ is given and d is an order preserving map

$$d(\omega)[2a] < d(\omega)[2b]$$
$$d(\omega)[2a+1] < d(\omega)[2b+1]$$
$$1^{2(i-k_0)+1}0 < 1^{2(i-k_0)+2}0.$$

Thus $d(\omega)[2a+1] < d(\omega)[2b+1] < d(\omega)[2a] < d(\omega)[2b]$. ∎

Because ω is uniformly recurrent, for $k = \sup\{k_0, k_1\}$, there is an N_k so any factor u of ω of length $n \geq N_k$ will contain all factors of ω of length k as a subword, and thus u will have C_j as a subword for each j.

One note about the factors of $d(\omega)$. For $n \geq N_k$ and two factors $u = d(\omega)[2x, 2x+2n]$ and $v = d(\omega)[2y+1, 2y+2n+1]$ of $d(\omega)$, then $u \neq v$. This is because a prefix of u will begin with an even number of one letter (either $0^{2m}1$ or $1^{2m}0$ for some m), and a prefix of v will begin with an odd number of one letter (either $0^{2m+1}1$ or $1^{2m+1}0$ for some m).

Let u be a factor of ω of length $n \geq N_k$. There is an a so that $u = \omega[a, a+n-1]$. For each $0 \leq i \leq n-1$ there is one j so that $\omega[a+i]$ has C_j as a prefix. In the factor $\omega[a, a+n+k-2]$ of length $n+k-1$, we will know explicitly which C_j is a prefix of the shift $\omega[a+i]$ for each $0 \leq i \leq n-1$. Let $p = \pi_\omega[a, a+n+k-1]$ be a subpermutation of π_ω of length $n+k$. The factor $\omega[a, a+n+k-2]$ of length $n+k-1$ is the form of p, and has u as a prefix.

For each $j \in \{0, 1, \ldots, k_0 + k_1 - 1\}$ define

$$\gamma_j = \{ \, i \mid 0 \leq i \leq n-1 \text{ and } C_j \text{ is a prefix of } \omega[a+i] \, \}.$$

So $|\gamma_0| + |\gamma_1| + \cdots + |\gamma_{k_0+k_1-1}| = n$ and $\gamma_i \cap \gamma_j = \emptyset$ for $i \neq j$. For each $j \in \{0, 1, \ldots, k_0 + k_1 - 1\}$ define

$$S_j = \sum_{i=0}^{j} |\gamma_i|$$

and say $S_{-1} = 0$. Since $|u| \geq N_k$, each γ_j is not empty, so $|\gamma_j| \geq 1$ for each j. We can see $d(u) = d(\omega)[2a, 2a+2n-1]$, and let p' be the subpermutation $p' = \pi_{d(\omega)}[2a, 2a+2n-1]$. Using Lemma 5.3.1 and the size of each of the γ_j sets we can determine the values of p' based on the values of $L^k(p)$, the k-left restriction of p.

Proposition 5.3.2 *Let ω, u, p, and p' be as above. For each $0 \leq i \leq n-1$, there is a j so $\omega[a+i]$ has C_j as a prefix.*

(a) If $p_i < p_{i+1}$ then $p'_{2i} = L^k(p)_i + S_{j-1}$ and $p'_{2i+1} = L^k(p)_i + S_j$

(b) If $p_i > p_{i+1}$ then $p'_{2i} = L^k(p)_i + S_j$ and $p'_{2i+1} = L^k(p)_i + S_{j-1}$

Proof Let $0 \leq i \leq n-1$ and suppose that C_j is a prefix of $\omega[a+i]$ for some $0 \leq j \leq k_0 + k_1 - 1$.

(a) Suppose $p_i < p_{i+1}$, and so $\omega_{a+i} = u_i = 0$.

For p'_{2i}, there are $L^k(p)_i - 1$ many $h \in \{0, 1, \ldots, n-1\}$ so that $p_i > p_h$, and thus $L^k(p)_i - 1$ many h so that $p'_{2i} > p'_{2h}$. Likewise there are $n - L^k(p)_i$ many h so that $p'_{2i} < p'_{2h}$. By Lemma 5.3.1 if $m < j$ and $h \in \gamma_m$ then $p'_{2i} > p'_{2h+1}$, and if $m \geq j$ and $h \in \gamma_m$ then $p'_{2i} < p'_{2h+1}$. Thus there are S_{j-1} many h so that $p'_{2i} > p'_{2h+1}$, and likewise there are $n - S_{j-1}$ many h so that $p'_{2i} < p'_{2h+1}$. Therefore there are exactly $L^k(p)_i - 1 + S_{j-1}$ many h so that $p'_{2i} > p'_h$, so

$$p'_{2i} = L^k(p)_i - 1 + S_{j-1} + 1 = L^k(p)_i + S_{j-1}.$$

For p'_{2i+1}, there are $L^k(p)_i - 1$ many h so that $p_i > p_h$, and thus $L^k(p)_i - 1$ many h so that $p'_{2i} > p'_{2h}$ and $p'_{2i+1} > p'_{2h+1}$. Likewise there are $n - L^k(p)_i$ many h so that $p'_{2i+1} < p'_{2h+1}$. By Lemma 5.3.1 if $m \leq j$ and $h \in \gamma_m$ then $p'_{2i+1} > p'_{2h}$, and if $m > j$ and $h \in \gamma_m$ then $p'_{2i+1} < p'_{2h}$. Thus there are S_j many h so that $p'_{2i+1} > p'_{2h}$, and likewise there are $n - S_j$ many h so that $p'_{2i+1} < p'_{2h}$. Therefore there are exactly $L^k(p)_i - 1 + S_j$ many h so that $p'_{2i} > p'_h$, so

$$p'_{2i+1} = L^k(p)_i - 1 + S_j + 1 = L^k(p)_i + S_j.$$

(b) Suppose $p_i > p_{i+1}$, and so $\omega_{a+i} = u_i = 1$.

For p'_{2i}, there are $L^k(p)_i - 1$ many h so that $p_i > p_h$, and thus $L^k(p)_i - 1$ many h so that $p'_{2i} > p'_{2h}$. Likewise there are $n - L^k(p)_i$ many h so that $p'_{2i} < p'_{2h}$. By Lemma 5.3.1 if $m \le j$ and $h \in \gamma_m$ then $p'_{2i} > p'_{2h+1}$, and if $m > j$ and $h \in \gamma_m$ then $p'_{2i} < p'_{2h+1}$. Thus there are S_j many h so that $p'_{2i} > p'_{2h+1}$, and likewise there are $n - S_j$ many h so that $p'_{2i} < p'_{2h+1}$. Therefore there are exactly $L^k(p)_i - 1 + S_j$ many h so that $p'_{2i} > p'_h$, so

$$p'_{2i} = L^k(p)_i - 1 + S_j + 1 = L^k(p)_i + S_j.$$

For p'_{2i+1}, there are $L^k(p)_i - 1$ many h so that $p_i > p_h$, and thus $L^k(p)_i - 1$ many h so that $p'_{2i} > p'_{2h}$ and $p'_{2i+1} > p'_{2h+1}$. Likewise there are $n - L^k(p)_i$ many h so that $p'_{2i+1} < p'_{2h+1}$. By Lemma 5.3.1 if $m < j$ and $h \in \gamma_m$ then $p'_{2i+1} > p'_{2h}$, and if $m \ge j$ and $h \in \gamma_m$ then $p'_{2i+1} < p'_{2h}$. Thus there are S_{j-1} many h so that $p'_{2i+1} > p'_{2h}$, and likewise there are $n - S_{j-1}$ many h so that $p'_{2i+1} < p'_{2h}$. Therefore there are exactly $L^k(p)_i - 1 + S_{j-1}$ many h so that $p'_{2i} > p'_h$, so

$$p'_{2i+1} = L^k(p)_i - 1 + S_{j-1} + 1 = L^k(p)_i + S_{j-1}.$$

■

The following corollaries show some nice properties that follow from Proposition 5.3.2. The first corollary (5.3.3) gives an example of when distinct subpermutations of π_ω will lead to the same subpermutation of $\pi_{d(\omega)}$. The next corollary (5.3.4) shows when two subpermutations of π_ω will definitely lead to distinct subpermutations of $\pi_{d(\omega)}$.

Corollary 5.3.3 *Let ω be as defined above. If $\pi_\omega[a, a+n+k-1]$ and $\pi_\omega[b, b+n+k-1]$, $a \neq b$, are subpermutations of π_ω where $\pi_\omega[a, a+n-1] = \pi_\omega[b, b+n-1]$ and for each $0 \le i \le n-1$, there is some j so that both $\omega[a+i]$ and $\omega[b+i]$ have C_j as a prefix. Then $\pi_{d(\omega)}[2a, 2a+2n-1] = \pi_{d(\omega)}[2b, 2b+2n-1]$.*

Proof Let $p = \pi_\omega[a, a+n+k-1]$ and $q = \pi_\omega[b, b+n+k-1]$, $a \neq b$, with p and q as in the hypothesis. For each $0 \le i \le n-1$, $L^k(p)_i = L^k(q)_i$ and $\omega[a+i]$ and $\omega[b+i]$ have C_j as a prefix for some j, so $p'_{2i} = q'_{2i}$ and $p'_{2i+1} = q'_{2i+1}$ by Proposition 5.3.2, and $p' = q'$. ■

Corollary 5.3.4 *Let ω be as defined above. If $p = \pi_\omega[a, a+n+k-1]$ and $q = \pi_\omega[a, a+n+k-1]$ are subpermutations of π_ω where one of the following conditions is true*

(a) $\omega[a, a+n-1] \neq \omega[b, b+n-1]$

(b) $L^k(p) \neq L^k(q)$

then $p' \neq q'$.

Proof (a) Since $\omega[a, a+n-1] \neq \omega[b, b+n-1]$, then there is an $0 \leq i \leq n-1$ so that, without loss of generality, $\omega_{a+i} = 0$ and $\omega_{b+i} = 1$. Thus $d(\omega)[2a+2i, 2a+2i+1] = 00$ and $d(\omega)[2b+2i, 2b+2i+1] = 11$ so $p'_{2i} < p'_{2i+1}$ and $q'_{2i} > q'_{2i+1}$, and $p' \neq q'$.

(b) Since $L^k(p) \neq L^k(q)$, then there are $0 \leq i, j \leq n-1$, $i \neq j$, so that, without loss of generality, $L^k(p)_i < L^k(p)_j$ and $L^k(q)_i > L^k(q)_j$, so $\omega[a+i] < \omega[a+j]$ and $\omega[b+i] > \omega[b+j]$. Thus $d(\omega)[2a+2i] < d(\omega)[2a+2j]$ and $d(\omega)[2b+2i] > d(\omega)[2b+2j]$ so $p'_{2i} < p'_{2j}$ and $q'_{2i} > q'_{2j}$, and $p' \neq q'$. ∎

Fix a subpermutation $p = \pi_\omega[a, a+n+k-1]$, and let $p' = \pi_{d(\omega)}[2a, 2a+2n-1]$. The terms of p' can be defined using the method given in Proposition 5.3.2. Let $q = \pi_\omega[b, b+n+k-1]$, $b \neq a$, be a subpermutation of π_ω and let $q' = \pi_{d(\omega)}[2b, 2b+2n-1]$ as in Proposition 5.3.2. The following lemma shows that if $p = q$ we know $p' = q'$, but the converse of this is not necessarily true. The objective here is using the idea of p' to define a map from the set of subpermutations of π_ω to the set of subpermutations of $\pi_{d(\omega)}$, and this map will be well-defined by Proposition 5.3.2.

Lemma 5.3.5 *If* $p = q$*, then* $p' = q'$.

Proof Suppose that $p = q$. So $p_i = q_i$ and thus $L^k(p)_i = L^k(q)_i$ for each $0 \leq i \leq n-1$. Since $p = q$, p and q have the same form, so $\omega[a, a+n+k-1] = \omega[b, b+n+k-1]$ and for each $0 \leq i \leq n-1$ if $\omega[a+i]$ has C_j as a prefix, for some j, then $\omega[b+i]$ has C_j as a prefix as well. Then by Proposition 5.3.2 it should be clear that for each $0 \leq i \leq 2n-1$ we have $p'_i = q'_i$, and thus $p' = q'$. ∎

Corollary 5.3.6 *Let* ω *be as defined above. If* $\pi_\omega[a, a+n+k-1] = \pi_\omega[b, b+n+k-1]$ *for some* $a \neq b$*, then* $\pi_{d(\omega)}[2a, 2a+2n-1] = \pi_{d(\omega)}[2b, 2b+2n-1]$.

Thus there is a well-defined function from the set of subpermutations of π_ω to the set of subpermutations of $\pi_{d(\omega)}$. Let $p = \pi_\omega[a, a+n+k-1]$, and define $\delta(p) = p' = \pi_{d(\omega)}[2a, 2a+2n-1]$ using the formula in Proposition 5.3.2. Thus we have the map

$$\delta : \text{Perm}^\omega(n+k) \mapsto \text{Perm}^{d(\omega)}(2n)$$

5.3.2 Creating $\delta(p)$ From a Table

As in Section 5.2.2, given a subpermutation p we can create $\delta(p)$ by using a table. The table created in Section 5.2.2 used properties of the Thue-Morse morphism, namely the behavior described in Lemma 5.2.2. In the table representing the behavior of δ, the properties of the doubling map d will be considered and the behavior described in Lemma 5.3.1 and Proposition 5.3.2 will be used.

We will again construct the table by using an example. In this example we will use a subpermutation p of π_T and find $\delta(p)$. When we calculate $\delta(p)$ we are interested in the C_j classes. In the case of the Thue-Morse word there will be 4 of these classes, namely $C_0 = 00$, $C_1 = 01$, $C_2 = 10$, $C_3 = 11$, with $k = 2$.

Let $u = T[9, 15] = 0010110$ which is a factor of length 7, we will consider the subpermutation

$$p = \pi_T[9, 17] = (2\,4\,8\,5\,9\,7\,3\,6\,1)$$

of length 9. The form of p is $T[9, 16] = 00101101$, and $L^2(p) = \pi_T[9, 17] = (1\,3\,6\,4\,7\,5\,2)$. From the form of p we see $\gamma_0 = \{0\}$, $\gamma_1 = \{1, 3, 6\}$, $\gamma_2 = \{2, 5\}$, and $\gamma_3 = \{4\}$. Thus $S_0 = 1$, $S_1 = 4$, $S_2 = 6$, and $S_3 = 7$. Let $p' = \delta(p) = \pi_T[18, 31]$, so by Proposition 5.2.3

$$p' = (1\ 2\ 4\ 7\ 12\ 10\ 5\ 8\ 14\ 13\ 11\ 9\ 3\ 6).$$

The table we will create will have 3 columns. In this example there will be an additional column to the left of the table listing the 4 C_j classes. In the first column of the table we list all the values in the subpermutation $L^2(p)$, because $k = 2$, from least to greatest.

We will add a double horizontal line just before the least value which corresponds to a 1 in u, and a single horizontal line between the values which correspond to different C_j classes. There will be S_0 many shifts within u which have C_0 as a prefix so the first horizontal line will be placed after the number S_0. Then there will be $S_1 - S_0$ many shifts within u which have C_1 as a prefix, so the next horizontal line will be placed after S_1. Continue in this fashion and place a horizontal line after S_j for each j. The double horizontal line will be placed after S_{k_0-1}, because if $L^2(p)_i \leq S_{k_0-1}$ then $u_i = 0$, and if $L^2(p)_i > S_{k_0-1}$ then $u_i = 1$. In this example, we place a horizontal line after $S_0 = 1$ and $S_2 = 6$, and the double horizontal line after $S_1 = 4$, see Figure 1.

We will fill in the table by the C_i classes. We start with a 1 in the top row of the second column, and then increasing by 1 for the following columns increase by 1 until we reach the first horizontal line. The next number will go in the top of the third column, and increase by 1 for the following rows until we reach the first horizontal line. In this example there is only one element in γ_0, see Figure 2.

00	1		
	2		
01	3		
	4		
10	5		
	6		
11	7		

Fig. 1

$\longrightarrow$

00	1	1	2
	2		
01	3		
	4		
10	5		
	6		
11	7		

Fig. 2

The next number will go in the top of the second column for the C_1 class, and increase by 1 for the following rows until we reach the next horizontal line. The next number will go in the top of the third column for the C_1 class and increase by 1 for the following rows until we reach the next horizontal line. We then continue in this fashion for the remaining C_j classes which begin with a 0, or for the classes above the double horizontal line. In this example there are 2 classes (namely C_0 and C_1) above the double horizontal line, see Figure 3.

The numbers below the double line correspond to classes which begin with a 1, so the values in the third row will be less than the values in the second row. For these classes we start in the top row of the third column and then move to the second row, which is a slight adjustment from the above steps. Continuing in this fashion we get the table in Figure 4.

00	1	1	2
	2	3	6
01	3	4	7
	4	5	8
10	5		
	6		
11	7		

Fig. 3

$\longrightarrow$

00	1	1	2
	2	3	6
01	3	4	7
	4	5	8
10	5	11	9
	6	12	10
11	7	14	13

Fig. 4

Now that we have the completed table in Figure 4 we can see how δ will alter the permutation p. We have $L^2(p)_1 = 3$, so $p'_2 = 4$ and $p'_3 = 7$. From the table in Figure 4, when we find the row with a 3 in the first column we see the next elements in the row are

4 and 7. This is exactly the behavior described in Lemma 5.3.1 and Proposition 5.3.2. Thus to create p' from the table we find the row which has the value $L^2(p)_i$ in the first column, and then p'_{2i} is the value in the second column and p'_{2i+1} is the value in the third column. So

$$\begin{aligned}
L^2(p)_0 = 1 &\implies p'_0 = 1 \quad p'_1 = 2 \\
L^2(p)_1 = 3 &\implies p'_2 = 4 \quad p'_3 = 7 \\
L^2(p)_2 = 6 &\implies p'_4 = 12 \quad p'_5 = 10 \\
&\vdots \\
L^2(p)_6 = 2 &\implies p'_{12} = 3 \quad p'_{13} = 6
\end{aligned}$$

which is readily verified to be the same as p' listed above.

5.3.3 Injective Restrictions

Not all subpermutations of $\pi_{d(\omega)}$ will be the image under δ of another subpermutation. Let $n > 2N_k$ and a be natural numbers. Then n and a can be either even or odd, and for the subpermutation $\pi_{d(\omega)}[a, a+n-1]$, there exist natural numbers b and m so that one of 4 cases hold:

1. $\pi_{d(\omega)}[a, a+n] = \pi_{d(\omega)}[2b, 2b+2m]$, even starting position with odd length.
2. $\pi_{d(\omega)}[a, a+n] = \pi_{d(\omega)}[2b, 2b+2m-1]$, even starting position with even length.
3. $\pi_{d(\omega)}[a, a+n] = \pi_{d(\omega)}[2b+1, 2b+2m]$, odd starting position with even length.
4. $\pi_{d(\omega)}[a, a+n] = \pi_{d(\omega)}[2b+1, 2b+2m-1]$, odd starting position with odd length.

Consider two subpermutations $\pi_{d(\omega)}[2c, 2c+n]$ and $\pi_{d(\omega)}[2d+1, 2d+n+1]$, with $n > 2N_k$. The subpermutation $\pi_{d(\omega)}[2c, 2c+n]$ will have form $d(\omega)[2c, 2c+n-1]$, and $\pi_{d(\omega)}[2d+1, 2d+n+1]$ will have form $d(\omega)[2d+1, 2d+n]$. Since the length of these factors is at least $2N_k$, these factors each contain both 0 and 1, so $d(\omega)[2c, 2c+n-1] \neq d(\omega)[2d+1, 2d+n]$, and thus $\pi_{d(\omega)}[2c, 2c+n] \neq \pi_{d(\omega)}[2d+1, 2d+n+1]$ because they do not have the same form. Thus we can break up the set $\mathrm{Perm}^{d(\omega)}(n)$ into two classes of subpermutations, namely the subpermutations that start at an even position or an odd position. So say that $\mathrm{Perm}^{d(\omega)}_{ev}(n)$ is the set of subpermutations p of length n so that $p = \pi_{d(\omega)}[2b, 2b+n-1]$ for some b, and that $\mathrm{Perm}^{d(\omega)}_{odd}(n)$ is the set of subpermutations p of length n so that $p = \pi_{d(\omega)}[2b+1, 2b+n]$ for some b. Thus

$$\mathrm{Perm}^{d(\omega)}(n) = \mathrm{Perm}^{d(\omega)}_{ev}(n) \cup \mathrm{Perm}^{d(\omega)}_{odd}(n),$$

where

$$\text{Perm}^{d(\omega)}_{ev}(n) \cap \text{Perm}^{d(\omega)}_{odd}(n) = \emptyset.$$

Thus for $n \geq N_k$, $\text{Perm}^{d(\omega)}_{ev}(2n)$ is the set of all subpermutations of length $2n$ starting at an even position. So for $\pi_{d(\omega)}[2a, 2a + 2n - 1]$, we have the subpermutation $p = \pi_\omega[a, a+n+k-1]$, and $\delta(p) = p' = \pi_{d(\omega)}[2a, 2a+2n-1]$. Thus the map

$$\delta : \text{Perm}^\omega(n+k) \mapsto \text{Perm}^{d(\omega)}_{ev}(2n)$$

is a surjective map.

For $p = \pi_\omega[a, a+n+k-1]$, we can then define three additional maps by looking at the left, right, and middle restrictions of $\delta(p) = p'$. These maps are

$$\begin{aligned}
\delta_L &: \text{Perm}^\omega(n+k) \mapsto \text{Perm}^{d(\omega)}_{ev}(2n-1) \\
\delta_R &: \text{Perm}^\omega(n+k) \mapsto \text{Perm}^{d(\omega)}_{odd}(2n-1) \\
\delta_M &: \text{Perm}^\omega(n+k) \mapsto \text{Perm}^{d(\omega)}_{odd}(2n-2)
\end{aligned}$$

and are defined by

$$\begin{aligned}
\delta_L(p) &= L(\delta(p)) = L(p') \\
\delta_R(p) &= R(\delta(p)) = R(p') \\
\delta_M(p) &= M(\delta(p)) = M(p')
\end{aligned}$$

It can be readily verified that these three maps are surjective. To see an example of this, consider the map δ_L, and let $\pi_{d(\omega)}[2b, 2b+2n-2]$ be a subpermutation of $\pi_{d(\omega)}$ in $\text{Perm}^{d(\omega)}_{ev}(2n-1)$. Then for the subpermutation $p = \pi_\omega[b, b+n+k-1]$, $\delta_L(p) = L(p') = \pi_{d(\omega)}[2b, 2b+2n-2]$ so δ_L is surjective. A similar argument will show that δ_R and δ_M are also surjective.

Lemma 5.3.7 *For $n \geq N_k$:*

$$\begin{aligned}
\tau_{d(\omega)}(2n-1) &\leq 2(\tau_\omega(n+k)) \\
\tau_{d(\omega)}(2n) &\leq \tau_\omega(n+k) + \tau_\omega(n+k+1)
\end{aligned}$$

Proof Let $n \geq N_k$. We have:

$$\begin{aligned}
\left|\text{Perm}^{d(\omega)}_{ev}(2n-1)\right| &\leq |\text{Perm}^\omega(n+k)| \\
\left|\text{Perm}^{d(\omega)}_{odd}(2n-1)\right| &\leq |\text{Perm}^\omega(n+k)| \\
\left|\text{Perm}^{d(\omega)}_{ev}(2n)\right| &\leq |\text{Perm}^\omega(n+k)| \\
\left|\text{Perm}^{d(\omega)}_{odd}(2n)\right| &\leq |\text{Perm}^\omega(n+k+1)|
\end{aligned}$$

since the maps δ, δ_L, δ_R, and δ_M are all surjective. Thus we have the following inequalities:

$$\begin{aligned}\tau_{d(\omega)}(2n-1) &= \left|\text{Perm}^{d(\omega)}(2n-1)\right| = \left|\text{Perm}^{d(\omega)}_{ev}(2n-1)\right| + \left|\text{Perm}^{d(\omega)}_{odd}(2n-1)\right| \\ &\leq |\text{Perm}^{\omega}(n+k)| + |\text{Perm}^{\omega}(n+k)| = 2(\tau_\omega(n+k))\end{aligned}$$

$$\begin{aligned}\tau_{d(\omega)}(2n) &= \left|\text{Perm}^{d(\omega)}(2n)\right| = \left|\text{Perm}^{d(\omega)}_{ev}(2n)\right| + \left|\text{Perm}^{d(\omega)}_{odd}(2n)\right| \\ &\leq |\text{Perm}^{\omega}(n+k)| + |\text{Perm}^{\omega}(n+k+1)| = \tau_\omega(n+k) + \tau_\omega(n+k+1)\end{aligned}$$

∎

The maps δ, δ_L, δ_R, and δ_M can be, but are not necessarily, injective maps. To see this, consider the following example. For this example we will use the Thue-Morse word T, defined in Section 5.2, and subpermutations of π_T, the infinite permutation associated with T. There will be 4 classes of C_j words for the Thue-Morse word (namely $C_0 = 00$, $C_1 = 01$, $C_2 = 10$, and $C_3 = 11$), and any factor of length $n \geq 9$ will contain each of these 4 classes. The following example will use subpermutations of length 9, with $n = 7$ and $k = 2$, to keep the example subpermutations short. Examples like this (as in Corollary 5.3.3) can be found for subpermutations of π_T of length $2^r + 1$ for any $r \geq 3$.

$$\begin{aligned}p &= \pi_T[0,8] = (4\,9\,7\,2\,6\,1\,3\,8\,5) \\ q &= \pi_T[12,20] = (5\,9\,7\,2\,6\,1\,3\,8\,4).\end{aligned}$$

So $p \neq q$ and both of these subpermutations have form $T[0,7] = T[12,19] = 01101001$. Let $p' = \delta(p) = \pi_{d(T)}[0,13]$ and $q' = \delta(q) = \pi_{d(T)}[24,37]$. Then by Proposition 5.3.2 we see:

$$p' = (5\,8\,14\,13\,12\,10\,3\,6\,11\,9\,1\,2\,4\,7) = q'.$$

So $\delta(p) = \delta(q)$ which leads to $\delta_L(p) = \delta_L(q)$, $\delta_R(p) = \delta_R(q)$, and $\delta_M(p) = \delta_M(q)$. Thus these 4 maps are not injective in general and the values in Lemma 5.3.7 are only an upper bound. If δ is an injective map it implies δ_R and δ_L are both injective, as shown by the following lemma.

Lemma 5.3.8 *For any aperiodic uniformly recurrent word ω, let p, q, p', and q' be as above. Then*

(a) $p' = q'$ if and only if $R(p') = R(q')$.

(b) $p' = q'$ if and only if $L(p') = L(q')$.

Proof Let $p = \pi_\omega[a, a+n+k-1]$, $q = \pi_\omega[b, b+n+k-1]$, $p' = \delta(p)$, and $q' = \delta(q)$. For both of these cases it should be clear that if $p' = q'$ then each of $R(p') = R(q')$ and $L(p') = L(q')$. Also recall that $n > N_k$.

Also note since $u = \omega[a, a+n-1]$ and $v = \omega[b, b+n-1]$, $d(u) = d(\omega)[2a, 2a+2n-1]$ and $d(v) = d(\omega)[2b, 2b+2n-1]$ so for each $0 \leq i \leq n-1$ we have

$$d(u)_{2i} = d(u)_{2i+1} = u_i \quad \text{and} \quad d(v)_{2i} = d(v)_{2i+1} = v_i$$

We will use the following notation

$$U_j = \{\, i \mid 0 \leq i \leq n-1 \text{ and } \omega[a+i] \text{ has } C_j \text{ as a prefix.} \,\}$$

$$V_j = \{\, i \mid 0 \leq i \leq n-1 \text{ and } \omega[b+i] \text{ has } C_j \text{ as a prefix.} \,\}$$

and due to the length of u and v we know $|U_j| \geq 1$ and $|V_j| \geq 1$ for each j.

(a) Suppose $p' \neq q'$, and assume $R(p') = R(q')$. We will need the following claim about $R(p')$ and $R(q')$ before we proceed.

Claim 5.3.9 *If $R(p') = R(q')$ then $d(u) = d(v)$.*

Proof Suppose $R(p') = R(q')$. For each $0 \leq i \leq 2n-3$, $R(p')_i < R(p')_{i+1}$ if and only if $R(q')_i < R(q')_{i+1}$ and thus $d(u)_{i+1} = d(v)_{i+1}$. Since $d(u)_0 = d(u)_1$ and $d(v)_0 = d(v)_1$, we see $d(u)_0 = d(u)_1 = d(v)_1 = d(v)_0$. In a similar fashion we see $d(u)_{2n-1} = d(u)_{2n-2} = d(v)_{2n-2} = d(v)_{2n-1}$. Thus $d(u) = d(v)$ and $u = v$. ∎

For each pair of real numbers $i \neq j$ where $0 \leq i, j \leq 2n-2$,

$$R(p')_i < R(p')_j \iff R(q')_i < R(q')_j$$

and thus

$$p'_{i+1} < p'_{j+1} \iff q'_{i+1} < q'_{j+1}.$$

Since $p' \neq q'$ there must be some $1 \leq i \leq 2n-1$ so, without loss of generality,

$$p'_0 < p'_i \text{ and } q'_0 > q'_i.$$

There is an $\alpha \in \{0, 1\}$ so $d(u)_1 = d(v)_1 = \alpha$, and so $d(u)_0 = d(v)_0 = \alpha$. If $d(u)_i = d(v)_i \neq \alpha$ we have $p'_0 < p'_i$ if and only if $q'_0 < q'_i$, which would be a contradiction. So $d(u)_i = d(v)_i = \alpha$.

Case a.1: Suppose $1 \leq i \leq 2n-2$ and recall that $p'_0 < p'_i$ and $q'_0 > q'_i$. If $d(u)_{i+1} = d(v)_{i+1} \neq \alpha$ we have $d(u)[0,1] = \alpha\alpha$ and $d(u)[i, i+1] = \alpha\beta$, so $p'_0 < p'_i$ if and only if $q'_0 < q'_i$, which is a contradiction, so $d(u)_{i+1} = d(v)_{i+1} = \alpha$. Thus $d(u)[i, i+1] = d(v)[i, i+1] = \alpha\alpha$ and

$$p'_0 < p'_i \implies p'_1 < p'_{i+1} \implies R(p')_0 < R(p')_i$$
$$q'_0 > q'_i \implies q'_1 > q'_{i+1} \implies R(q')_0 > R(q')_i$$

by Claim 5.1.1 which contradicts the assumption. Therefore $R(p') \neq R(q')$.

Case a.2: Suppose $i = 2n-1$ is the only i so that $p'_0 < p'_i$ and $q'_0 > q'_i$. So as above we have $d(u)[0,1] = d(v)[0,1] = \alpha\alpha$, so $d(u)_{2n-1} = d(v)_{2n-1} = \alpha$ and $d(u)[2n-2, 2n-1] = d(v)[2n-2, 2n-1] = \alpha\alpha$. Thus $p'_0 < p'_{2n-1}$ and $q'_0 > q'_{2n-1}$ imply

$$p'_{2n-1} - 1 = R(p')_{2n-2} = R(q')_{2n-2} = q'_{2n-1}.$$

For each $1 \leq j \leq 2n-2$ we know the following

$$p'_0 < p'_j \iff q'_0 < q'_j$$
$$R(p')_{j-1} = p'_j \iff R(q')_{j-1} = q'_j,$$

and thus $p'_j = q'_j$ for each $1 \leq j \leq 2n-2$, since $R(p')_j = R(q')_j$ and $p'_0 > p'_j$ implies $R(p')_{j-1} = p'_j$. So only $p'_0 \neq q'_0$ and $p'_{2n-1} \neq q'_{2n-1}$. Since $p'_{2n-1} = q'_{2n-1} + 1$, it must be

$$p'_0 = p'_{2n-1} - 1 \text{ and } q'_0 = q'_{2n-1} + 1.$$

Let $1 \leq x \leq 2n$ so that $p'_0 = q'_{2n-1} = x$ and $q'_0 = p'_{2n-1} = x + 1$. Then either $p'_0 < p'_{2n-2}$ or $p'_0 > p'_{2n-2}$.

Case a.2.i: Suppose $p'_0 < p'_{2n-2}$ then $p'_1 < p'_{2n-1}$ and $p_0 < p_{n-1}$. Likewise $q'_0 < q'_{2n-2}$, $q'_1 < q'_{2n-1}$, and $q_0 < q_{n-1}$. Thus

$$q'_1 < q'_{2n-1} < q'_0 < q'_{2n-2}$$

and by Lemma 5.3.1 we know $\alpha = 1$ and there is a j so that both $\omega[b]$ and $\omega[b+n-1]$ have C_j as a prefix. Since both $\omega[b]$ and $\omega[b+n-1]$ have C_j as a prefix and $q_0 < q_{n-1}$, there is some $m \geq 1$ so that

$$q'_0 = x + 1 \qquad q'_1 = x - m \qquad q'_{2n-2} = x + 1 + m \qquad q'_{2n-1} = x.$$

For this to occur q_0 must be less than all other shifts of v which have C_j as a prefix, and q_{n-1} must be greater than all other shifts of v which have C_j as a prefix.

Thus C_j is a prefix of $\omega[a]$, since $u = v$. Since $\omega[a+n-1]$ also begins with a 1 and $p_0 < p_{n-1}$, the order of p'_0, p'_1, p'_{2n-2}, and p'_{2n-2} must be as follows by Lemma 5.3.1

$$p'_1 < p'_0 < p'_{2n-1} < p'_{2n-2}.$$

So we have $\omega[a]$ has C_j as a prefix and for $i > j$ $\omega[a+n-1]$ has C_i as a prefix. If $i > j+1$, there will be some l so that $\omega[a+l]$ and $\omega[b+l]$ each have C_{j+1} as a prefix which is totally contained in u and v. Thus we would have $p_l < p_{n-1}$ and $q_l > q_{n-1}$, so $R(p')_{2l-1} < R(p')_{2n-3}$ and $R(q')_{2l-1} > R(q')_{2n-3}$ which would be a contradiction, therefore $i = j+1$.

Since $p'_i = q'_i$ for each $1 \leq i \leq 2n-2$, we know

$$p'_0 = x \qquad p'_1 = x - m \qquad p'_{2n-2} = x + 1 + m \qquad p'_{2n-1} = x + 1.$$

For this to occur p_0 must be greater than all other shifts of u which have C_j as a prefix, and p_{n-1} must be less than all other shifts of u which have C_{j+1} as a prefix. Based on the construction of p' and q' from Proposition 5.3.2, there will be m shifts of u that have C_j as a prefix, and m shifts of u which have C_{j+1} as a prefix. Since the word C_{j+1} occurs in the word u, there is at least one l so that $\omega[a+l]$ has C_{j+1} as a prefix which is totally contained in u. Thus both $\omega[a+l]$ and $\omega[a+n-1]$ have C_{j+1} as a prefix, $\omega[a+n-1] < \omega[a+l]$, so $p'_{2(a+l)+1} \geq x+2$ and m must be greater than 1.

Since $m > 1$, there is some i so that $q_{n-1} = q_i + 1$ and thus $\omega[b+i]$ has C_j as a prefix and $q'_{2i+1} = x - 1$. In a similar fashion, there is some $\hat{i}$ so that $p_0 = p_{\hat{i}} + 1$ and thus $\omega[a+\hat{i}]$ has C_j as a prefix and $p'_{2\hat{i}} = x - 1$. Thus $q'_{2i+1} = x - 1 = p'_{2\hat{i}}$, and so $p'_{2i+1} \neq q'_{2i+1}$ as well as $p'_{2\hat{i}} \neq q'_{2\hat{i}}$ which gives a contradiction.

Case a.2.ii: Suppose $p'_0 > p'_{2n-2}$ then $p'_1 > p'_{2n-1}$ and $p_0 > p_{n-1}$. Likewise $q'_0 > q'_{2n-2}$, $q'_1 > q'_{2n-1}$, and $q_0 > q_{n-1}$. Thus

$$p'_{2n-2} < p'_0 < p'_{2n-1} < p'_1.$$

and by Lemma 5.3.1 we know $\alpha = 0$ and there is a j so that both $\omega[a]$ and $\omega[a+n-1]$ have C_j as a prefix. Since both $\omega[a]$ and $\omega[a+n-1]$ have C_j as a prefix and $p_0 > p_{n-1}$, there is some $m \geq 1$ so that

$$p'_0 = x \qquad p'_1 = x + 1 + m \qquad p'_{2n-2} = x - m \qquad p'_{2n-1} = x + 1.$$

For this to occur p_0 must be greater than all other shifts of u which have C_j as a prefix, and p_{n-1} must be less than all other shifts of u which have C_j as a prefix.

Thus C_j is a prefix of $\omega[b]$, since $u = v$. Since $\omega[b+n-1]$ also begins with a 0 and $q_0 > q_{n-1}$, the order of q'_0, q'_1, q'_{2n-2}, and q'_{2n-2} must be as follows by Lemma 5.3.1

$$q'_{2n-2} < q'_{2n-1} < q'_0 < q'_1.$$

So we have $\omega[b]$ has C_j as a prefix and for $i < j$ $\omega[b+n-1]$ has C_i as a prefix. If $i < j-1$, there will be some l so that $\omega[a+l]$ and $\omega[b+l]$ each have C_{j-1} as a prefix which is totally contained in u and v. Thus we would have $p_l < p_{n-1}$ and $q_l > q_{n-1}$, so $R(p')_{2l-1} < R(p')_{2n-3}$ and $R(q')_{2l-1} > R(q')_{2n-3}$ which would be a contradiction, therefore $i = j - 1$.

Since $p'_i = q'_i$ for each $1 \le i \le 2n-2$, we know

$$q'_0 = x + 1 \qquad q'_1 = x + 1 + m \qquad q'_{2n-2} = x - m \qquad q'_{2n-1} = x.$$

For this to occur q_0 must be less than all other shifts of v which have C_j as a prefix, and q_{n-1} must be greater than all other shifts of v which have C_{j-1} as a prefix. Based on the construction of p' and q' from Proposition 5.3.2, there will be m shifts of v that have C_j as a prefix, and m shifts of v which have C_{j-1} as a prefix. Since the word C_{j-1} occurs in the word v, there is at least one l so that $\omega[b+l]$ has C_{j-1} as a prefix which is totally contained in v. Thus both $\omega[b+l]$ and $\omega[b+n-1]$ have C_{j-1} as a prefix, $\omega[b+n-1] > \omega[b+l]$, so $q'_{2(a+l)} \le x - 1$ and m must be greater than 1.

Since $m > 1$, there is some i so that $p_i = p_{n-1} + 1$ and thus $\omega[a+i]$ has C_j as a prefix and $p'_{2i+1} = x + 2$. In a similar fashion, there is some $\hat{i}$ so that $q_{\hat{i}} = q_0 + 1$ and thus $\omega[b+\hat{i}]$ has C_j as a prefix and $q'_{2\hat{i}} = x + 2$. Thus $p'_{2i+1} = x + 2 = q'_{2\hat{i}}$, and so $p'_{2i} \ne q'_{2i}$ as well as $p'_{2\hat{i}+1} \ne q'_{2\hat{i}+1}$ which gives a contradiction.

In either case $(a.2.i)$ or $(a.2.ii)$ we find a contradiction, so $R(p') \ne R(q')$. Therefore $p' = q'$ if and only if $R(p') = R(q')$.

(b) Suppose $p' \ne q'$, and assume $L(p') = L(q')$. For each pair of real numbers $i \ne j$ where $0 \le i, j \le 2n-2$,

$$L(p')_i < L(p')_j \iff L(q')_i < L(q')_j$$

and thus

$$p'_i < p'_j \iff q'_i < q'_j.$$

In particular for each pair of real numbers $i \ne j$ where $0 \le i, j \le n-1$,

$$p'_{2i} < p'_{2j} \iff q'_{2i} < q'_{2j}.$$

So we see

$$p'_{2i} < p'_{2j} \implies p_i < p_j \implies L^k(p)_i < L^k(p)_j$$
$$q'_{2i} < q'_{2j} \implies q_i < q_j \implies L^k(q)_i < L^k(q)_j$$

and thus $L^k(p) = L^k(q)$.

We will need the following claim about $L(p')$ and $L(q')$ before we proceed.

Claim 5.3.10 *If $L(p') = L(q')$ then $d(u) = d(v)$.*

Proof Suppose $L(p') = L(q')$ and assume $d(u) \neq d(v)$. Since $L^k(p) = L^k(q)$ we know $\omega[a, a+n-2] = \omega[b, b+n-2]$ and so $d(\omega)[2a, 2a+2n-3] = d(\omega)[2b, 2b+2n-3]$. Thus $d(\omega)[2a+2n-2, 2a+2n-1] \neq d(\omega)[2b+2n-2, 2b+2n-1]$, and so $u_{n-1} \neq v_{n-1}$. Recall that $|u|_0$ is the number of occurrences of the letter 0 in the word u. Without loss of generality, say $u_{n-1} = 0$ and $v_{n-1} = 1$, and so $|u|_0 = |v|_0 + 1$. We also have

$$p'_{2n-2} < p'_{2n-1} \quad \text{and} \quad q'_{2n-2} > q'_{2n-1}$$

and so

$$p'_{2n-2} = L(p')_{2n-2} = L(q')_{2n-2} = q'_{2n-2} - 1.$$

For each $0 \leq i \leq n-2$, if $u_i = v_i = 0$ then $q_{n-1} > q_i$, and so $p_{n-1} > p_i$ because $L^k(p) = L^k(q)$. Likewise if $u_i = v_i = 1$ then $p_{n-1} < p_i$, and so $q_{n-1} < q_i$ because $L^k(p) = L^k(q)$. Moreover, $\omega[a+n-1]$ has 01 as a prefix and $\omega[b+n-1]$ has 10 as a prefix. To see this let $0 \leq i \leq n-3$ (we know $i < n-2$ because $u_{n-1} = 0$) so that $u[i, i+1] = v[i, i+1] = 01$. Since $u_i = v_i = 0$, and $p_{n-1} > p_i$, $\omega[a+n-1]$ can not have 00 as a prefix. A similar argument will show $\omega[b+n-1]$ can not have 11 as a prefix.

Thus there are exactly $|v|_0$ many j so that $q_{n-1} > q_j$, thus $L^k(p)_{n-1} = L^k(q)_{n-1} = |v|_0 + 1 = |u|_0$. By Proposition 5.3.2 we see

$$\begin{aligned} p'_{2n-1} &= L^k(p)_{n-1} + |u|_0 = 2\,|u|_0 = 2\,|v|_0 + 2 \\ q'_{2n-1} &= L^k(q)_{n-1} + |v|_0 = 2\,|v|_0 + 1 \end{aligned}$$

Fix an $0 \leq i \leq n-3$ so that $u[i, i+1] = v[i, i+1] = 01$ and an $0 \leq \hat{i} \leq n-3$ so that $u[\hat{i}, \hat{i}+1] = v[\hat{i}, \hat{i}+1] = 10$. Since $\omega[a+i]$ and $\omega[a+n-1]$ both have 01 as a prefix and $p_{n-1} > p_i$, we have

$$p'_{2n-2} < p'_{2i-1} \leq 2\,|v|_0 + 1$$

by Lemma 5.3.1. Since $\omega[b+\hat{i}]$ and $\omega[b+n-1]$ both have 10 as a prefix and $q_{n-1} < q_i$, we have

$$q'_{2n-2} > q'_{2\hat{i}-1} \geq 2\,|v|_0 + 2$$

by Lemma 5.3.1. Therefore

$$p'_{2n-2} < 2\,|v|_0 + 1 = 2\,|v|_0 + 2 - 1 < q'_{2n-2} - 1$$

and we have a contradiction to $p'_{2n-2} = q'_{2n-2} - 1$. Therefore $d(u) = d(v)$ and $u = v$. ∎

Since $p' \neq q'$ there must be some $0 \leq i \leq 2n-2$ so, without loss of generality,

$$p'_{2n-1} < p'_i \text{ and } q'_{2n-1} > q'_i.$$

There is an $\alpha \in \{0, 1\}$ so $d(u)_{2n-2} = d(v)_{2n-2} = \alpha$, so $d(u)[2n-2, 2n-1] = d(v)[2n-2, 2n-1] = \alpha\alpha$. If $d(u)_i = d(v)_i \neq \alpha$ we have $p'_{2n-1} < p'_i$ if and only if $q'_{2n-1} < q'_i$, which would be a contradiction, so $d(u)_i = d(v)_i = \alpha$. It should be noted that $i \neq 2n-2$, because $d(u)_{2n-2} = d(v)_{2n-2} = \alpha$ so $p'_{2n-2} < p'_{2n-1}$ if and only if $q'_{2n-2} < q'_{2n-1}$.

Case b.1: Suppose for $1 \leq i \leq 2n-2$ we have $p'_{2n-1} < p'_i$ and $q'_{2n-1} > q'_i$.

If $d(u)_{i-1} = d(v)_{i-1} = \alpha$ we have $d(u)[2n-2, 2n-1] = \alpha\alpha$ and $d(u)[i-1, i] = \alpha\alpha$, so

$$p'_{2n-1} < p'_i \implies p'_{2n-2} < p'_{i-1}$$

$$q'_{2n-1} > q'_i \implies q'_{2n-1} > q'_{i-1}$$

which contradicts the assumption $L(p') = L(q')$. So $d(u)_{i-1} = d(v)_{i-1} \neq \alpha$, say $d(u)_{i-1} = d(v)_{i-1} = \beta$. Thus $d(u)[i-1, i+1] = \beta\alpha\alpha$ and i is an even number, so rather than using i we will use $2c$. Now we will consider when $\alpha = 0$ and when $\alpha = 1$.

Case b.1.i: Suppose $\alpha = 0$. Then we know $d(u)[2n-2, 2n-1] = d(v)[2n-2, 2n-1] = 00$, and $d(u)[2c-1, 2c+1] = d(u)[2c-1, 2c+1] = 100$. So $p'_{2n-2} < p'_{2n-1}$ and $q'_{2n-2} < q'_{2n-1}$, and since $L(p') = L(q')$ we see

$$p'_{2n-2} = L(p')_{2n-2} = L(q')_{2n-2} = q'_{2n-2},$$

so $p'_{2n-2} = q'_{2n-2}$.

Since $p'_{2n-1} < p'_{2c}$ and $d(u)[2n-2, 2n-1] = 00$ we have

$$p'_{2n-2} < p'_{2n-1} < p'_{2c} < p'_{2c+1}.$$

Thus by Lemma 5.3.1, we know there are $i < j$ so $\omega[a+n-1]$ has C_i as a prefix and $\omega[a+c]$ has C_j as a prefix. Since $\omega[b+n-1]$ also begins with 0 and $q'_{2n-1} > q'_{2c}$, the order of q'_{2c}, q'_{2c+1}, q'_{2n-2}, and q'_{2n-2} must be as follows by Lemma 5.3.1

$$q'_{2n-2} < q'_{2c} < q'_{2n-1} < q'_{2c+1}.$$

So we have both $\omega[a+n-1]$ and $\omega[a+c]$ have C_j as a prefix. If $j > i+1$, there will be some l so that $\omega[a+l]$ and $\omega[b+l]$ each have C_{i+1} as a prefix which is totally contained in u and v. Thus we would have $p_l < p_{n-1}$ and $q_l > q_{n-1}$ and $L^k(p) \neq L^k(q)$, therefore $j = i+1$.

Since $\omega[a+n-1]$ has C_i as a prefix for each $l \in U_{i+1}$ we know $p_{n-1} < p_l$, and so $q_{n-1} < q_l$ since $L^k(p) = L^k(q)$. Likewise, since $\omega[b+n-1]$ has C_{i+1} as a prefix, for each $l \in V_i$ we know $q_{n-1} > q_l$ and so $p_{n-1} > p_l$. Thus p_{n-1} is the greatest of all occurrences of C_i and q_{n-1} is the least of all occurrences of C_{i+1}. So $L^k(p)_{n-1} = L^k(q)_{n-1}$ and we see

$$L^k(p)_{n-1} = |U_0| + \cdots + |U_i|$$
$$L^k(q)_{n-1} = |V_0| + \cdots + |V_i| + 1.$$

Now we investigate how the size of U_0, ..., U_i are related to the size of V_0, ..., V_i. Since $u = v$, and these words have 0 as a suffix, there is some $m \geq 1$ so that 10^m is a suffix of both u and v. Thus for each $0 \leq h \leq n - m - 1$, there is some j so $\omega[a + h]$ and $\omega[b + h]$ have C_j as a prefix, so $h \in U_j$ and $h \in V_j$. Moreover, this prefix is totally contained within u and v, respectively, because $\omega[a + n - m - 1]$ and $\omega[b + n - m - 1]$ begin with the class 10. For $n - m \leq h \leq n - 1$ we have

$$n - m \in U_{i-m+1}, \qquad n - m + 1 \in U_{i-m+2}, \qquad \cdots \qquad n - 1 \in U_i$$

$$n - m \in V_{i-m+2}, \qquad n - m + 1 \in V_{i-m+3}, \qquad \cdots \qquad n - 1 \in V_{i+1}.$$

For example, if $m = 1$ we have $n - 1 \in U_i$ and $n - 1 \in V_{i+1}$, so

$$|U_i| = |V_i| + 1 \qquad |U_{i+1}| = |V_{i+1}| - 1$$

and $|U_j| = |V_j|$ for all other j. Since $|V_i| \geq 1$ we see $|U_i| \geq 2$.

If $m = 2$ we have $n - 2 \in U_{i-1}$ and $n - 2 \in V_i$; $n - 1 \in U_i$ and $n - 1 \in V_{i+1}$, so

$$|U_{i-1}| = |V_{i-1}| + 1 \qquad |U_i| = |V_i| \qquad |U_{i+1}| = |V_{i+1}| - 1$$

and $|U_j| = |V_j|$ for all other j.

Thus for $m \geq 1$,

$$|U_{i-m+1}| = |V_{i-m+1}| + 1 \qquad |U_{i+1}| = |V_{i+1}| - 1$$

and $|U_j| = |V_j|$ for all other j. Since $|V_{i-m+1}| \geq 1$ we see $|U_{i-m+1}| \geq 2$. Each occurrence of C_{i-m+1} which is contained in u will have C_i as a suffix, and since $n - 1 \in U_i$ we have $|U_i| \geq |U_{i-m+1}| \geq 2$. Thus by Proposition 5.3.2

$$\begin{aligned}
p'_{2n-2} &= L^k(p)_{n-1} + |U_0| + \cdots + |U_{i-1}| = 2(|U_0| + \cdots + |U_{i-1}|) + |U_i| \\
q'_{2n-2} &= L^k(q)_{n-1} + |V_0| + \cdots + |V_{i-1}| + |V_i| = 2(|V_0| + \cdots + |V_{i-1}| + |V_i|) + 1 \\
&= 2(|U_0| + \cdots + |U_{i-1}| + |U_i| - 1) + 1 = 2(|U_0| + \cdots + |U_{i-1}|) + |U_i| + |U_i| - 1 \\
&\geq p'_{2n-2} + 2 - 1 = p'_{2n-2} + 1
\end{aligned}$$

and we see $p'_{2n-2} < q'_{2n-2}$. Therefore we have a contradiction, and $L(p') = L(q')$.

Case b.1.ii: Suppose $\alpha = 1$. This argument is similar to the argument used in Case (*b*.1.*i*). Then we know $d(u)[2n - 2, 2n - 1] = d(v)[2n - 2, 2n - 1] = 11$, and $d(u)[2c - 1, 2c + 1] = d(u)[2c - 1, 2c + 1] = 011$. So $p'_{2n-2} > p'_{2n-1}$ and $q'_{2n-2} > q'_{2n-1}$, and since $L(p') = L(q')$ we see

$$p'_{2n-2} - 1 = L(p')_{2n-2} = L(q')_{2n-2} = q'_{2n-2} - 1,$$

so $p'_{2n-2} = q'_{2n-2}$.

Since $q'_{2n-1} > q'_{2c}$ and $d(v)[2n-2, 2n-1] = 11$ we have

$$q'_{2c+1} < q'_{2c} < q'_{2n-1} < q'_{2n-2}$$

Thus by Lemma 5.3.1, we know there are $i < j$ so $\omega[b+n-1]$ has C_j as a prefix and $\omega[b+c]$ has C_i as a prefix. Since $\omega[b+n-1]$ also begins with 1 and $p'_{2n-1} < p'_{2c}$, the order of p'_{2c}, p'_{2c+1}, p'_{2n-2}, and p'_{2n-2} must be as follows by Lemma 5.3.1

$$p'_{2c+1} < p'_{2n-1} < p'_{2c} < p'_{2n-2}$$

So we have both $\omega[a+n-1]$ and $\omega[a+c]$ have C_i as a prefix. If $j > i+1$, there will be some l so that $\omega[a+l]$ and $\omega[b+l]$ each have C_{i+1} as a prefix which is totally contained in u and v. Thus we would have $p_l > p_{n-1}$ and $q_l < q_{n-1}$ and $L^k(p) \neq L^k(q)$, therefore $j = i+1$.

Since $\omega[a+n-1]$ has C_i as a prefix for each $l \in U_{i+1}$ we know $p_{n-1} < p_l$, and so $q_{n-1} < q_l$ since $L^k(p) = L^k(q)$. Likewise, since $\omega[b+n-1]$ has C_{i+1} as a prefix, for each $l \in V_i$ we know $q_{n-1} > q_l$ and so $p_{n-1} > p_l$. Thus p_{n-1} is the greatest of all occurrences of C_i and q_{n-1} is the least of all occurrences of C_{i+1}. So $L^k(p)_{n-1} = L^k(q)_{n-1}$ and we see

$$L^k(p)_{n-1} = |U_0| + \cdots + |U_i|$$
$$L^k(q)_{n-1} = |V_0| + \cdots + |V_i| + 1.$$

Now we investigate how the size of $U_0, \ldots, U_i$ are related to the size of $V_0, \ldots, V_i$. Since $u = v$, and these words have 1 as a suffix, there is some $m \geq 1$ so that 01^m is a suffix of both u and v. Thus for each $0 \leq h \leq n-m-1$, there is some j so $\omega[a+h]$ and $\omega[b+h]$ have C_j as a prefix, so $h \in U_j$ and $h \in V_j$. Moreover, this prefix is totally contained within u and v, respectively, because $\omega[a+n-m-1]$ and $\omega[b+n-m-1]$ begin with the class 01.For $n-m \leq h \leq n-1$ we have

$$n-m \in U_{i+m-1}, \qquad n-m+1 \in U_{i+m-2}, \qquad \cdots \qquad n-1 \in U_i$$

$$n-m \in V_{i+m}, \qquad n-m+1 \in V_{i+m-1}, \qquad \cdots \qquad n-1 \in V_{i+1}.$$

For example, if $m = 1$ we have $n-1 \in U_i$ and $n-1 \in V_{i+1}$, so

$$|V_{i+1}| = |U_{i+1}| + 1 \qquad |V_i| = |U_i| - 1$$

and $|U_j| = |V_j|$ for all other j. Since $|U_{i+1}| \geq 1$ we see $|V_{i+1}| \geq 2$.

If $m = 2$ we have $n-2 \in U_{i+1}$ and $n-2 \in V_{i+2}$; $n-1 \in U_i$ and $n-1 \in V_{i+1}$, so

$$|V_{i+2}| = |U_{i+2}| + 1 \qquad |V_{i+1}| = |U_{i+1}| \qquad |V_i| = |U_i| - 1$$

and $|U_j| = |V_j|$ for all other j.

Thus for $m \geq 1$,

$$|V_{i+m}| = |U_{i+m}| + 1 \qquad |V_i| = |U_i| - 1$$

and $|U_j| = |V_j|$ for all other j. Since $|U_{i+m}| \geq 1$ we see $|V_{i+m}| \geq 2$. Each occurrence of C_{i+m} which is contained in v will have C_{i+1} as a suffix, and since $n-1 \in V_{i+1}$ we have $|V_{i+1}| \geq |V_{i+m}| \geq 2$. Thus by Proposition 5.3.2

$$\begin{aligned}
p'_{2n-2} &= L^k(p)_{n-1} + |U_0| + \cdots + |U_{i-1}| + |U_i| = 2(|U_0| + \cdots + |U_{i-1}| + |U_i|) \\
q'_{2n-2} &= L^k(p)_{n-1} + |V_0| + \cdots + |V_i| + |V_{i+1}| = 2(|V_0| + \cdots + |V_{i-1}| + |V_i|) + |V_{i+1}| + 1 \\
&= 2(|U_0| + \cdots + |U_{i-1}| + |U_i| - 1) + |V_{i+1}| + 1 = 2(|U_0| + \cdots + |U_{i-1}|) + |V_{i+1}| - 1 \\
&\geq p'_{2n-2} + 2 - 1 = p'_{2n-2} + 1
\end{aligned}$$

and we see $p'_{2n-2} < q'_{2n-2}$. Therefore we have a contradiction, and $L(p') = L(q')$.

Case b.2: Suppose $i = 0$ is the only i so that $p'_{2n-1} < p'_i$ and $q'_{2n-1} > q'_i$. So as above we have $d(u)[0,1] = d(v)[0,1] = \alpha\alpha$, and $d(u)[2n-2, 2n-1] = d(v)[2n-2, 2n-1] = \alpha\alpha$. Thus $p'_{2n-1} < p'_0$ and $q'_{2n-1} > q'_0$ imply

$$p'_0 = L(p')_0 = L(q')_0 = q'_0 - 1.$$

For each $1 \leq j \leq 2n-2$ we know the following

$$p'_{2n-1} < p'_j \iff q'_{2n-1} < q'_j$$

$$L(p')_j = p'_j \iff L(q')_j = q'_j,$$

and thus $p'_j = q'_j$ for each $1 \leq j \leq 2n-2$, because $L(p')_j = L(q')_j$. So only $p'_0 \neq q'_0$ and $p'_{2n-1} \neq q'_{2n-1}$. Since $q'_0 = p'_0 + 1$, it must be

$$p'_{2n-1} = p'_0 - 1 \text{ and } q'_{2n-1} = q'_0 + 1.$$

Let $1 \leq x \leq 2n$ so that $p'_{2n-1} = q'_0 = x$ and $q'_{2n-1} = p'_0 = x + 1$. Then, as in Case $(a.2)$, we find a contradiction to the construction of p' and q'.

Therefore $p' = q'$ if and only if $L(p') = L(q')$. ∎

From the previous lemma, if the map δ is injective,

$$p = q \iff p' = q' \iff L(p') = L(q')$$

$$p = q \iff p' = q' \iff R(p') = R(q').$$

Therefore when δ is injective, δ_R and δ_L are both injective as well. A troubling fact is the map δ being injective does not imply δ_M is injective. As will be shown for the Thue-Morse word T, there are cases of distinct subpermutations p and q where $\delta(p) \neq \delta(q)$ but $\delta(p)_M = \delta_M(q)$. The following sections deal with some different words and we will show when δ and δ_M are injective, but these proofs will use special properties of the words considered.

5.3.4 Permutation Complexity of $d(s)$

In this section we will investigate the permutation complexity of Sturmian words under the doubling map. We now recall from Section 2.4.2 some of the equivalent definitions of Sturmian words. The class of Sturmian words are the aperiodic binary words with minimal factor complexity. So an infinite word s is a Sturmian word if for each $n \geq 0$, s has exactly $n+1$ distinct factors of length n, or $\rho_s(n) = n+1$ (the only factor of length $n = 0$ being the empty-word). An equivalent definition for Sturmian words is that they are the class of aperiodic balanced binary words.

First we will show when the map δ is applied to permutations from a Sturmian word, δ is injective and thus a bijection. Then we show the maps δ_R, δ_L, and δ_M are injective as well and thus also bijections. First we look at the permutation complexity of Sturmian words. Recall from Theorem 5.1.7, that if s is a Sturmian word then $\tau_s(n) = n$ for each $n \geq 1$. If s is a Sturmian word, then $d(s)$ is not Sturmian. The word $d(s)$ will contain both 00 and 11 as factors and is not balanced. Thus we know $\tau_{d(s)}(n) > n$ for some $n \geq 3$.

Fix a Sturmian word s over $\{0, 1\}$. Since s is balanced, there is some $k > 0$ so that for $a, b \in \{0, 1\}$, with $a \neq b$, every a is followed by either k or $k-1$ b's. So consecutive a's will look like either $ab^k a$ or $ab^{k-1}a$. For example consider the Fibonacci word, t, where

$$t = 010010100100101001 01 \ldots$$

In t, consecutive 1's look like either 1001 or 101, and if 000 or 11 were factors then t would not be balanced.

Let $d(s)$ be the image of s under the doubling map. Let π_s be the infinite permutation associated to s, and $\pi_{d(s)}$ be the infinite permutation associated to $d(s)$. We will now calculate the permutation complexity of $d(s)$. By Lemma 5.1.4 we may assume there is a natural number $k > 1$ so that each 1 is followed by either $0^k 1$ or $0^{k-1}1$, because $d(s)$ and $d(\overline{s}) = \overline{d(s)}$ have the same permutation complexity. There will be $k+1$ classes of factors of s, which are

$$\begin{aligned}C_0 &= 0^k\\ C_1 &= 0^{k-1}1\\ &\vdots\\ C_{k-1} &= 01\\ C_k &= 10\end{aligned}$$

For each $i \in \mathbb{N}$, $s[i] = s_i s_{i+1} \cdots$ will have exactly one the above classes of words as a prefix. Since Sturmian words are uniformly recurrent ([13]), there is an $N \in \mathbb{N}$ so that each factor of s of length $n \geq N$ will contain each of C_0, C_1, …, C_k.

Let $u = s[a, a+n-1]$ and $v = s[b, b+n-1]$, $a \neq b$, be factors of s of length $n \geq N$, so C_j is a factor of both u and v for each $0 \leq j \leq k$. For $0 \leq j \leq k$ define

$$U_j = \{\, i \mid 0 \leq i \leq n-1 \text{ and } s[a+i] \text{ has } C_j \text{ as a prefix.} \,\}$$

$$V_j = \{\, i \mid 0 \leq i \leq n-1 \text{ and } s[b+i] \text{ has } C_j \text{ as a prefix.} \,\}$$

and $|U_0| + |U_1| + \cdots + |U_k| = |V_0| + |V_1| + \cdots + |V_k| = n$. Since $|u| = |v| \geq N$ we know for each j there is an occurrence of C_j in both u and v so $|U_j| \geq 1$ and $|V_j| \geq 1$. Let $p = \pi_s[a, a+n+k-1]$ and $q = \pi_s[b, b+n+k-1]$ be subpermutations of π_s. Then define subpermutations $\delta(p) = p' = \pi_{d(s)}[2a, 2a+2n-1]$ and $\delta(q) = q' = \pi_{d(s)}[2b, 2b+2n-1]$ as in Proposition 5.3.2. The following lemma concerns the relationship of p and q to p' and q'.

Lemma 5.3.11 *For the Sturmian word s, let p, q, p', and q' be as above. Then $p = q$ if and only if $p' = q'$.*

Proof If $p = q$, then it follows from Lemma 5.3.5 that $p' = q'$.

Then suppose that $p \neq q$. Thus p and q have a different form by Lemma 5.1.6. Thus there is an $0 \leq i \leq n+k-2$ so that, without loss of generality, $p_i < p_{i+1}$ and $q_i > q_{i+1}$. We will look at the least i where this happens and it will be handled in two cases. First when $0 \leq i \leq n-1$, and then when $n \leq i \leq n+k-2$.

Case a: Suppose $0 \leq i \leq n-1$ is the least i where $p_i < p_{i+1}$ and $q_i > q_{i+1}$. Then $p' \neq q'$ follows from Corollary 5.3.4.

Case b: Suppose $n \leq i \leq n+k-2$ is the least i where $p_i < p_{i+1}$ and $q_i > q_{i+1}$. Thus we know $u = s[a, a+n-1] = s[b, b+n-1] = v$, and so $L^k(p) = L^k(q)$ by Lemma 5.1.6.

If $u_{n-1} = v_{n-1} = 1$, then both u and v are followed by 0^{k-1} so $s[a, a+n+k-2] = s[b, b+n+k-2] = u0^{k-1}$ and $p = q$ contradicting the assumption. Thus $u_{n-1} = v_{n-1} = 0$, and letting $m = i - n + 1$

$$\begin{aligned} s[a] &= u0^m1\cdots \\ s[b] &= u0^{m-1}1\cdots. \end{aligned}$$

For $l = i - n - k + 1$, $s[a+l]$ has C_0 as a prefix and $s[b+l]$ has C_1 as a prefix. Since $L^k(p) = L^k(q)$ and $|V_0| \geq 1$, by Proposition 5.3.2 we have

$$\begin{aligned} p'_{2l} &= L^k(p)_l \\ q'_{2l} &= L^k(q)_l + |V_0| = L^k(p)_l + |V_0| \geq L^k(p)_l + 1, \end{aligned}$$

so $p'_{2l} < q'_{2l}$, and $p' \neq q'$.

Therefore $p = q$ if and only if $p = q$. ∎

Thus the map

$$\delta : \mathrm{Perm}^s(n+k) \mapsto \mathrm{Perm}^{d(s)}_{ev}(2n)$$

is injective when applied to permutations associated with a Sturmian word, and is therefore bijective. When Lemma 5.3.11 is used with Lemma 5.3.8 we see the maps

$$\begin{aligned} \delta_L &: \mathrm{Perm}^\omega(n+k) \mapsto \mathrm{Perm}^{d(\omega)}_{ev}(2n-1) \\ \delta_R &: \mathrm{Perm}^\omega(n+k) \mapsto \mathrm{Perm}^{d(\omega)}_{odd}(2n-1) \end{aligned}$$

are also injective, and thus are bijections. So when $n \geq N$

$$\left|\mathrm{Perm}^{d(s)}_{ev}(2n)\right| = \left|\mathrm{Perm}^s(n+k)\right|.$$

$$\left|\mathrm{Perm}^{d(s)}_{ev}(2n-1)\right| = \left|\mathrm{Perm}^s(n+k)\right|$$

$$\left|\mathrm{Perm}^{d(s)}_{odd}(2n-1)\right| = \left|\mathrm{Perm}^s(n+k)\right|.$$

We will now show the map δ_M is also injective when applied to permutations associated with a Sturmian word.

Lemma 5.3.12 *For the Sturmian word s, let p, q, p', and q' be as above. Then $p' = q'$ if and only if $M(p') = M(q')$.*

Proof It should be clear that if $p' = q'$ then $M(p') = M(q')$.

Note that since $d(u) = d(s)[2a, 2a+2n-1]$ and $d(v) = d(s)[2b, 2b+2n-1]$, for each $0 \leq i \leq n-1$

$$d(u)_{2i} = d(u)_{2i+1} = u_i \quad \text{and} \quad d(v)_{2i} = d(v)_{2i+1} = v_i.$$

We will need the following claim about $M(p')$ and $M(q')$ before we proceed.

Claim 5.3.13 *If $M(p') = M(q')$ then $d(u) = d(v)$.*

Proof Suppose $M(p') = M(q')$. For $0 \leq i \leq 2n-3$, $d(u)_i = d(v)_i$ by Claim 5.3.9. Then assuming $d(u) \neq d(v)$ we find a contradiction by Claim 5.3.10, so $d(u) = d(v)$. Therefore if $M(p') = M(q')$ then $d(u) = d(v)$, and $u = v$. ∎

Suppose $p' \neq q'$, and assume $M(p') = M(q')$. For each pair of real numbers $i \neq j$ where $0 \leq i, j \leq 2n-3$,

$$M(p')_i < M(p')_j \iff M(q')_i < M(q')_j$$

and thus

$$p'_{i+1} < p'_{j+1} \iff q'_{i+1} < q'_{j+1}.$$

Thus we know $d(u) = d(v)$ and $u = v$, so $L^k(p) = L^k(q)$.

From Lemma 5.3.8 we know $R(p') \neq R(q')$ and $L(p') \neq L(q')$ because $p' \neq q'$, but

$$R(L(p')) = M(p') = M(q') = R(L(q')).$$

Thus there is an $1 \leq i \leq 2n-2$ so that $L(p')_0 < L(p')_i$ and $L(q')_0 > L(q')_i$. As in Lemma 5.3.8 Case $(a.1)$, if $1 \leq i \leq 2n-3$ we have a contradiction. Thus we can assume that $i = 2n-2$ is the only i so that $L(p')_0 < L(p')_i$ and $L(q')_0 > L(q')_i$. Thus

$$L(p')_0 < L(p')_{2n-2} \implies p'_0 < p'_{2n-2} \implies p_0 < p_{n-1} \implies L^k(p)_0 < L^k(p)_{n-1}$$

$$L(q')_0 > L(q')_{2n-2} \implies q'_0 > q'_{2n-2} \implies q_0 > q_{n-1} \implies L^k(q)_0 > L^k(q)_{n-1},$$

and $L^k(p) \neq L^k(q)$, and so by Lemma 5.1.6 we see $u \neq v$ and $d(u) \neq d(v)$ which is a contradiction to the assumption. Therefore $M(p') \neq M(q')$.

Therefore $p' = q'$ if and only if $M(p') = M(q')$. ∎

Thus we see, for a Sturmian word s,

$$p = q \iff \delta(p) = \delta(q) \iff \delta_M(p) = \delta_M(q)$$

and thus the map

$$\delta_M : \text{Perm}^s(n+k) \mapsto \text{Perm}^{d(s)}_{odd}(2n-2)$$

is also injective, and thus is a bijection. So when $n \geq N$

$$\left|\text{Perm}^{d(s)}_{odd}(2n-2)\right| = \left|\text{Perm}^s(n+k)\right|.$$

The following theorem will give the permutation complexity of a Sturmian word.

Theorem 5.3.14 *Let s be a Sturmian word over $\mathcal{A}$, where for $a, b \in \mathcal{A}$, $a \neq b$, there are strings of either k or $k-1$ a's between each b, with $k > 1$. There is an N so that each factor of s of length at least N will contain each of a^k, $a^{k-1}b$, $\ldots$, ab, b. For each $n \geq 2N$ the permutation complexity of $d(s)$ is*

$$\tau_{d(s)}(n) = n + 2k + 1$$

Proof Let s be a Sturmian word as in the hypothesis, and let $n \geq 2N$. Then there is $m \geq N$ so that either $n = 2m$ or $n = 2m-1$, and recall $\tau_s(n) = n$ for each $n \geq 1$. Since s is Sturmian, each of

$$\begin{aligned} \delta &: \mathrm{Perm}^s(m+k) \mapsto \mathrm{Perm}^{d(s)}_{ev}(2m) \\ \delta_L &: \mathrm{Perm}^s(m+k) \mapsto \mathrm{Perm}^{d(s)}_{ev}(2m-1) \\ \delta_R &: \mathrm{Perm}^s(m+k) \mapsto \mathrm{Perm}^{d(s)}_{odd}(2m-1) \\ \delta_M &: \mathrm{Perm}^s(m+k+1) \mapsto \mathrm{Perm}^{d(s)}_{odd}(2m) \end{aligned}$$

are bijections, and so

$$\begin{aligned} \left|\mathrm{Perm}^{d(s)}_{ev}(2m-1)\right| &= \left|\mathrm{Perm}^s(m+k)\right| \\ \left|\mathrm{Perm}^{d(s)}_{odd}(2m-1)\right| &= \left|\mathrm{Perm}^s(m+k)\right| \\ \left|\mathrm{Perm}^{d(s)}_{ev}(2m)\right| &= \left|\mathrm{Perm}^s(m+k)\right| \\ \left|\mathrm{Perm}^{d(s)}_{odd}(2m)\right| &= \left|\mathrm{Perm}^s(m+k+1)\right|. \end{aligned}$$

Thus

$$\begin{aligned} \tau_{d(s)}(2m-1) = \left|\mathrm{Perm}^{d(s)}(2m-1)\right| &= \left|\mathrm{Perm}^{d(s)}_{ev}(2m-1)\right| + \left|\mathrm{Perm}^{d(s)}_{odd}(2m-1)\right| \\ &= (m+k) + (m+k) = (2m-1) + 2k + 1 \end{aligned}$$

$$\begin{aligned} \tau_{d(s)}(2m) = \left|\mathrm{Perm}^{d(s)}(2m)\right| &= \left|\mathrm{Perm}^{d(s)}_{ev}(2m)\right| + \left|\mathrm{Perm}^{d(s)}_{odd}(2m)\right| \\ &= (m+k) + (m+k+1) = 2m + 2k + 1 \end{aligned}$$

Therefore for either $n = 2m-1$ or $n = 2m$,

$$\tau_{d(s)}(n) = n + 2k + 1.$$

∎

5.3.5 Permutation Complexity of $d(T)$

In this section we will investigate the permutation complexity of $d(T)$, the image of the Thue-Morse word, T, under the doubling map, d. Recall from Section 5.2, the Thue-Morse word $T = 011010011001\cdots$ is the fixed point of the Thue-Morse morphism $\mu_T : 0 \mapsto 01$, $1 \mapsto 10$.

The calculation of the permutation complexity of $d(T)$ will use the formula for the factor complexity of T. Again, the factor complexity is known ([11, 31]) and we will use the formula calculated by S. Brlek.

Proposition 5.3.15 ([11]) *For $n \geq 3$, the function $\rho_T(n)$ is given by*

$$\rho_T(n) = \begin{cases} 6 \cdot 2^{r-1} + 4p & 0 < p \leq 2^{r-1} \\ 8 \cdot 2^{r-1} + 2p & 2^{r-1} < p \leq 2^r \end{cases}$$

where r and p are uniquely determined by the equation

$$n = 2^r + p + 1, \qquad 0 < p \leq 2^r$$

We also need many properties from Section 5.2, which we now recall. Theorem 5.2.8 stated that subpermutations of π_T have the same form if and only if they are a complementary pair. Proposition 5.2.12 calculated, based on subpermutation length, the type for complementary pairs, and Lemma 5.2.14 which stated how many complementary pairs of type 1 or 2 can arise for each length n. Finally Theorem 5.2.16 gives the permutation complexity function.

The following proposition is a variation of Proposition 5.2.7.

Proposition 5.3.16 *Suppose $p = \pi_T[a, a+n]$ and $q = \pi_T[b, b+n]$ are a complementary pair of type $k \geq 1$.*

(a) $L(p)$ and $L(q)$ are a complementary pair of type $k-1$.

(b) $R(p)$ and $R(q)$ are a complementary pair of type $k-1$.

(c) $M(p)$ and $M(q)$ are a complementary pair of type $k-2$.

Proof Let $p = \pi_T[a, a+n]$ and $q = \pi_T[b, b+n]$ be a complementary pair of type $k \geq 1$.

The argument from Proposition 5.2.7, part (c), will show if p and q are a complementary pair of type k, then $L(p)$ and $L(q)$ are a complementary pair of type $k-1$. In a similar fashion Proposition 5.2.7, part (d), implies $R(p)$ and $R(q)$ are a complementary pair of type $k-1$, and Proposition 5.2.7, part (e), implies $M(p)$ and $M(q)$ are a complementary pair of type $k-2$. ∎

Now to calculate the permutation complexity of $d(T)$ we need to identify the classes of factors of T with blocks of the same letter. Since T is overlap-free, and thus cube-free, we can identify the 4 classes of factors of T, which are

$$C_0 = 00, \quad C_1 = 01, \quad C_2 = 10, \quad C_3 = 11$$

For each $i \in \mathbb{N}$, $T[i] = T_i T_{i+1} \cdots$ will have exactly one of the above classes of words as a prefix. Since the Thue-Morse word is uniformly recurrent ([2]), there is an $N \in \mathbb{N}$ so that each factor of T of length $n \geq N$ will contain each of C_0, C_1, C_2, and C_3. It is readily verified that any factor of length $n \geq 9$ will contain each of these 4 classes of words as a factor.

Let $u = T[a, a+n-1]$ and $v = T[b, b+n-1]$, $a \neq b$, be factors of T of length $n \geq 9$, so C_j is a factor of both u and v for each $0 \leq j \leq 3$. Let $p = \pi_T[a, a+n+1]$ and $q = \pi_s[b, b+n+1]$ be subpermutations of π_T. Then define subpermutations $\delta(p) = p' = \pi_{d(s)}[2a, 2a+2n-1]$ and $\delta(q) = q' = \pi_{d(s)}[2b, 2b+2n-1]$ as in Proposition 5.3.2, with $k = 2$. The following lemma concerns the relationship of p and q to p' and q'.

Lemma 5.3.17 *Let p and q be subpermutations of length $n+2$ of π_T, with $n \geq 9$, and let $p' = \delta(p)$ and $q' = \delta(p)$.*

(a) If $n \notin \{2^r - 1, 2^r \mid r \geq 3\}$, $p = q$ if and only if $p' = q'$.

(b) If $n \in \{2^r - 1, 2^r \mid r \geq 3\}$, p and q have the same form if and only if $p' = q'$.

Proof Let $p = \pi_T[a, a+n+1]$ and $q = \pi_s[b, b+n+1]$, $a \neq b$, be subpermutations of π_T of length $n+2$, with $n \geq 9$, and let $u = T[a, a+n-1]$ and $v = T[b, b+n-1]$. Since the length of u and v is at least 9, each of C_0, C_1, C_2, and C_3 occurs in both of u and v. Then let $p' = \delta(p) = \pi_{d(s)}[2a, 2a+2n-1]$ and $q' = \delta(q) = \pi_{d(s)}[2b, 2b+2n-1]$ as in Proposition 5.3.2.

(a) Suppose $n \notin \{2^r - 1, 2^r \mid r \geq 3\}$. If $p = q$ then $p' = q'$ by Lemma 5.3.5.

Suppose $p \neq q$. Then either p and q have the same form, or they do not have the same form. These cases will be handled independently.

Case (a.1) Suppose p and q have the same form. Since $n \notin \{2^r - 1, 2^r \mid r \geq 3\}$, p and q are a complementary pair of type $k \geq 3$, by Theorem 5.2.8. By Proposition 5.3.16, $L^2(p)$ and $L^2(q)$ are a complementary pair of type $k-2$, where $k-2 \geq 1$, and so $L^2(p) \neq L^2(q)$. Therefore $p' \neq q'$, by Corollary 5.3.4.

Case (a.2) Suppose p and q do not have the same form. Thus there is an $0 \leq i \leq n$ so that, without loss of generality, $p_i < p_{i+1}$ and $q_i > q_{i+1}$. We may say $i = n$ is the only i so that $p_i < p_{i+1}$ and $q_i > q_{i+1}$, because if there is an $0 \leq i \leq n-1$ so $p_i < p_{i+1}$ and $q_i > q_{i+1}$ then $u \neq v$, and $p' \neq q'$ be Corollary 5.3.4. We may also say $L^2(p) = L^2(q)$, because if $L^2(p) \neq L^2(q)$ then $p' \neq q'$ be Corollary 5.3.4. Thus $u = v$ and $p_n < p_{n+1}$ and $q_n > q_{n+1}$.

For $0 \leq j \leq 3$ let

$$U_j = \{\, i \mid 0 \leq i \leq n-1 \text{ and } T[a+i] \text{ has } C_j \text{ as a prefix.} \,\}$$

$$V_j = \{\, i \mid 0 \leq i \leq n-1 \text{ and } T[b+i] \text{ has } C_j \text{ as a prefix.} \,\}$$

and $|U_0| + |U_1| + |U_2| + |U_3| = |V_0| + |V_1| + |V_2| + |V_3| = n$. Since $u = v$ and $|u| = |v| \geq 9$ we know for each j there is an $0 \leq i \leq n-2$ so that $i \in U_j$ and $i \in V_j$, thus $|U_j| \geq 1$ and $|V_j| \geq 1$. Since $u = v$ and $p_n < p_{n+1}$ and $q_n > q_{n+1}$ there is an $\alpha \in \{0,1\}$ so that

$$T[a+n-1, a+n] = \alpha 0$$

$$T[b+n-1, b+n] = \alpha 1$$

so either

$$|U_0| \neq |V_0| \qquad |U_1| \neq |V_1| \qquad |U_2| = |V_2| \qquad |U_3| = |V_3|$$

or

$$|U_0| = |V_0| \qquad |U_1| = |V_1| \qquad |U_2| \neq |V_2| \qquad |U_3| \neq |V_3|\,.$$

If $\alpha = 0$ then $|U_0| \neq |V_0|$ and $|U_1| \neq |V_1|$, and by Proposition 5.3.2

$$\begin{aligned} p'_{2n-2} &= L^2(p)_{n-1} \\ q'_{2n-2} &= L^2(q)_{n-1} + |V_0| \geq L^2(p)_{n-1} + 1 > p'_{2n-2}. \end{aligned}$$

If $\alpha = 1$ then $|U_2| \neq |V_2|$ and $|U_3| \neq |V_3|$, and by Proposition 5.3.2

$$\begin{aligned} p'_{2n-1} &= L^2(p)_{n-1} + |U_0| + |U_1| \\ q'_{2n-1} &= L^2(q)_{n-1} + |V_0| + |V_1| + |V_2| \geq L^2(p)_{n-1} + |U_0| + |U_1| + 1 > p'_{2n-1}. \end{aligned}$$

Therefore in either case $(a.1)$ or $(a.2)$, $p' \neq q'$.

Therefore if p and q are subpermutations of π_T of length $n+2$, with $n \neq 2^r - 1$ or 2^r for any $r \geq 3$, $p = q$ if and only if $p' = q'$.

(b) Suppose $n \in \{2^r - 1, 2^r \mid r \geq 3\,\}$. If p and q do not have the same form, there is an $0 \leq i \leq n$ so that, without loss of generality, $p_i < p_{i+1}$ and $q_i > q_{i+1}$ and $p \neq q$. Thus p and q are as in Case $(a.2)$, and $p' \neq q'$.

Suppose p and q have the same form, so for each $0 \leq i \leq n-1$, there is some j so that both $T[a+i]$ and $T[b+i]$ have C_j as a prefix. We can say $p \neq q$, because if $p = q$ then $p' = q'$ by Lemma 5.3.5. By Theorem 5.2.8 and Proposition 5.2.12, p and q are a complementary pair of type 1 or 2 and $L^2(p) = L^2(q)$ by Proposition 5.3.16. So by Corollary 5.3.3, $p' = q'$.

Therefore if p and q are subpermutations of π_T of length $n+2$, with $n = 2^r - 1$ or 2^r for some $r \geq 3$, p and q have the same form if and only if $p' = q'$. ∎

Thus, for $n \geq 9$, the maps

$$\delta : \text{Perm}^T(n+2) \mapsto \text{Perm}_{ev}^{d(T)}(2n)$$

$$\delta_L : \text{Perm}^T(n+2) \mapsto \text{Perm}_{ev}^{d(T)}(2n-1)$$

$$\delta_R : \text{Perm}^T(n+2) \mapsto \text{Perm}_{odd}^{d(T)}(2n-1)$$

when applied to permutations associated with the Thue-Morse word are injective when $n \notin \{2^r - 1, 2^r \mid r \geq 3\}$ (or when there are no complementary pairs of type 1 or 2), so

$$\left|\text{Perm}_{ev}^{d(T)}(2n)\right| = \left|\text{Perm}^T(n+2)\right|$$

$$\left|\text{Perm}_{ev}^{d(T)}(2n-1)\right| = \left|\text{Perm}^T(n+2)\right|$$

$$\left|\text{Perm}_{odd}^{d(T)}(2n-1)\right| = \left|\text{Perm}^T(n+2)\right|.$$

When $n \in \{2^r - 1, 2^r \mid r \geq 3\}$ the maps δ, δ_R, and δ_L are surjective, but not injective because complementary pairs of type 1 or 2 will give the same subpermutation under δ. In this case, if p and q are subpermutations of π_T of length $n+2$, where p has form u' and q has form v', $|u'| = |v'| = n+1$, $\delta(p) = \delta(q)$ if and only if $u' = v'$. Likewise we see $\delta_L(p) = \delta_L(q)$ and $\delta_R(p) = \delta_R(q)$ if and only if $u' = v'$. Thus the number of subpermutations of $\pi_{d(T)}$ for these lengths are determined by the number of factors of T, or

$$\left|\text{Perm}_{ev}^{d(T)}(2n)\right| = \left|\mathcal{F}_T(n+1)\right|.$$

$$\left|\text{Perm}_{ev}^{d(T)}(2n-1)\right| = \left|\mathcal{F}_T(n+1)\right|$$

$$\left|\text{Perm}_{odd}^{d(T)}(2n-1)\right| = \left|\mathcal{F}_T(n+1)\right|.$$

The following lemma shows when the map δ_M is injective when applied to permutations associated with the Thue-Morse word.

Lemma 5.3.18 *For the Thue-Morse word T, let p, q, p', and q' be as above. Then*

(a) If $n \neq 2^r - 1$, 2^r, or $2^r + 1$ for any $r \geq 3$, $p' = q'$ if and only if $M(p') = M(q')$.

(b) If $n = 2^r - 1$, 2^r, or $2^r + 1$ for some $r \geq 3$, p and q have the same form if and only if $M(p') = M(q')$.

Proof It should be clear for either case that if $p' = q'$ then $M(p') = M(q')$.

Note that since $d(u) = d(s)[2a, 2a + 2n - 1]$ and $d(v) = d(s)[2b, 2b + 2n - 1]$, for each $0 \leq i \leq n - 1$

$$d(u)_{2i} = d(u)_{2i+1} = u_i \quad \text{and} \quad d(v)_{2i} = d(v)_{2i+1} = v_i.$$

If $M(p') = M(q')$ then $d(u) = d(v)$, by Claim 5.3.13, and so $u = v$.

We will again use the notation

$$U_j = \{\, i \mid 0 \leq i \leq n - 1 \text{ and } T[a + i] \text{ has } C_j \text{ as a prefix.} \,\}$$

$$V_j = \{\, i \mid 0 \leq i \leq n - 1 \text{ and } T[b + i] \text{ has } C_j \text{ as a prefix.} \,\}$$

and due to the length of u and v we know $|U_j| \geq 1$ and $|V_j| \geq 1$ for each j.

(a) Let $n \neq 2^r - 1$, 2^r, or $2^r + 1$ for any $r \geq 3$, and $p = \pi_T[a, a + n + 1]$ and $q = \pi_T[b, b + n + 1]$ be subpermutations of π_T of length $n + 2 \geq 11$. Then $p' = \delta(p)$ and $q' = \delta(q)$ by Proposition 5.3.2.

Suppose $p' \neq q'$, and assume $M(p') = M(q')$. For each pair of real numbers $i \neq j$ where $0 \leq i, j \leq 2n - 3$,

$$M(p')_i < M(p')_j \iff M(q')_i < M(q')_j$$

and thus

$$p'_{i+1} < p'_{j+1} \iff q'_{i+1} < q'_{j+1}.$$

Thus we know $d(u) = d(v)$ and $u = v$. There is an $\alpha \in \{0, 1\}$ so that $d(u)_1 = d(v)_1 = \alpha$, and so $d(u)[0, 1] = d(v)[0, 1] = \alpha\alpha$. As in Lemma 5.3.8 we have $d(u)[2n - 2, 2n - 1] = d(v)[2n - 2, 2n - 1] = \alpha\alpha$.

(a.1) Suppose p and q have the same form. By Theorem 5.2.8 and Proposition 5.2.12, p and q are a complementary pair of type $k \geq 4$. By Proposition 5.3.16, $L^2(p)$ and $L^2(q)$ are a complementary pair of type $k - 2 \geq 2$. Thus, without loss of generality, $L^2(p)_{k-2-1} + 1 = L^2(p)_{n-1}$ and $L^2(q)_{n-1} + 1 = L^2(q)_{k-2-1}$. Thus $L^2(p)_{k-3} < L^2(p)_{n-1}$ and $L^2(q)_{k-3} > L^2(q)_{n-1}$, so $p'_{2k-6} < p'_{2n-2}$ and $q'_{2k-6} > q'_{2n-2}$. Thus $M(p')_{2k-5} < M(p')_{2n-3}$ and $M(q')_{2k-5} > M(q')_{2n-3}$ so $M(p') \neq M(q')$ which is a contradiction.

(a.2) Suppose p and q do not have the same form. From Lemma 5.3.8 we know $R(p') \neq R(q')$ and $L(p') \neq L(q')$ because $p' \neq q'$, but

$$R(L(p')) = M(p') = M(q') = R(L(q')).$$

Thus there is an $1 \leq i \leq 2n-2$ so that $L(p')_0 < L(p')_i$ and $L(q')_0 > L(q')_i$. As in Lemma 5.3.8 Case $(a.1)$, if $1 \leq i \leq 2n-3$ we have a contradiction. Thus we can assume that $i = 2n-2$ is the only i so that $L(p')_0 < L(p')_i$ and $L(q')_0 > L(q')_i$. Thus

$$L(p')_0 < L(p')_{2n-2} \implies p'_0 < p'_{2n-2} \implies p_0 < p_{2n-1} \implies L^2(p)_0 < L^2(p)_{n-1}$$

$$L(q')_0 > L(q')_{2n-2} \implies q'_0 > q'_{2n-2} \implies q_0 > q_{2n-1} \implies L^2(q)_0 > L^2(q)_{n-1}$$

so $L^2(p) \neq L^2(q)$, and $u = v$. Thus, by Theorem 5.2.8 and Proposition 5.2.12, $L^2(p)$ and $L^2(q)$ are a complementary pair of type $k \geq 2$. Thus, without loss of generality, $L^2(p)_{k-1} < L^2(p)_{n-1}$ and $L^2(q)_{k-1} > L^2(q)_{n-1}$, so $p'_{2k-2} < p'_{2n-2}$ and $q'_{2k-2} > q'_{2n-2}$. Thus $M(p')_{2k-1} < M(p')_{2n-3}$ and $M(q')_{2k-1} > M(q')_{2n-3}$ so $M(p') \neq M(q')$ which is a contradiction.

Therefore if $n \neq 2^r-1$, 2^r, or 2^r+1 for any $r \geq 3$, $p' = q'$ if and only if $M(p') = M(q')$.

(b) Let $n = 2^r - 1$, 2^r, or $2^r + 1$ for some $r \geq 3$, and $p = \pi_T[a, a+n+1]$ and $q = \pi_T[b, b+n+1]$ by subpermutations of π_T of length $n+2 \geq 11$. Then $p' = \delta(p)$ and $q' = \delta(q)$ as in Proposition 5.3.2.

(b.1) Suppose p and q have the same form. So for each $0 \leq i \leq n$,

$$p_i < p_{i+1} \iff q_i < q_{i+1}.$$

So we know for each i, $T[a+i]$ and $T[b+i]$ both have the same C_j as a prefix, so

$$i \in U_j \iff i \in V_j$$

and so $|U_j| = |V_j|$ for each j.

If $p = q$, then $p' = q'$ and $M(p') = M(q')$, so we can say $p \neq q$. If $n = 2^r-1$ or 2^r then $p' = q'$ by Lemma 5.3.17 and $M(p') = M(q')$, so we can say $n = 2^r+1$ for some $r \geq 3$. Thus p and q are a complementary pair of type 3 by Theorem 5.2.8 and Proposition 5.2.12, and $L^2(p)$ and $L^2(q)$ are a complementary pair of type 1 by Proposition 5.3.16. So, without loss of generality, there is some $1 \leq x \leq n-1$ so that $L^2(p)_0 = L^2(q)_{n-1} = x$ and $L^2(p)_{n-1} = L^2(q)_0 = x+1$, and for each $1 \leq i \leq n-2$ $L^2(p)_i = L^2(q)_i$.

Since p and q are a complementary pair of type 3 we know $T[a, a+1] = T[a+n-1, a+n]$, thus we know $T[b, b+1] = T[b+n-1, b+n] = T[a, a+1]$ because $u = v$. So there is a j so that each of $T[a]$, $T[a+n-1]$, $T[b]$, and $T[b+n-1]$ each have C_j as a prefix. So by Proposition 5.3.2, there are some y and z so that

$$\begin{aligned} p'_0 &= y & q'_0 &= y+1 \\ p'_1 &= z & q'_1 &= z+1 \\ p'_{2n-2} &= y+1 & q'_{2n-2} &= y \\ p'_{2n-1} &= z+1 & q'_{2n-1} &= z \end{aligned}$$

and for each $2 \le i \le 2n-3$, $p'_i = q'_i$. The order of y and z will be either $y < y+1 < z < z+1$ (if $T_a = T_b = 0$) or $z < z+1 < y < y+1$ (if $T_a = T_b = 1$). If $y < y+1 < z < z+1$, then $M(p')_0 = z-1 = M(q')_0$ and $M(p')_{2n-2} = y = M(q')_{2n-2}$. If $z < z+1 < y < y+1$, then $M(p')_0 = z = M(q')_0$ and $M(p')_{2n-2} = y-1 = M(q')_{2n-2}$. In either case we have, for $2 \le i \le 2n-3$,

$$p'_i < y \iff q'_i < y+1 \quad \text{and} \quad p'_i < z+1 \iff q'_i < z$$

so $M(p')_{i-1} = M(q')_{i-1}$. Therefore $M(p') = M(q')$.

Therefore if p and q have the same form then $M(p') = M(q')$.

(b.2) Suppose p and q do not have the same form, and assume $M(p') = M(q')$. If p and q do not have the same form, there is an $0 \le i \le n$ so that, without loss of generality, $p_i < p_{i+1}$ and $q_i > q_{i+1}$ and $p \ne q$. By Lemma 5.3.17, $p' \ne q'$. Then as in Case $(a.2)$ we find a contradiction to the assumption, so $M(p') \ne M(q')$.

Therefore p and q have the same form if and only if $M(p') = M(q')$. ■

Thus, for $n \ge 9$, the map

$$\delta_M : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^{d(T)}_{odd}(2n-2)$$

when applied to permutations associated with the Thue-Morse word are injective when $n \ne 2^r-1$, 2^r, or 2^r+1 for any $r \ge 3$ (or when there are no complementary pairs of type 1, 2, or 3), so

$$\left|\mathrm{Perm}^{d(T)}_{odd}(2n-2)\right| = \left|\mathrm{Perm}^T(n+2)\right|.$$

When $n = 2^r-1$, 2^r, or 2^r+1 for some $r \ge 3$ the map δ_M is surjective, but not injective. In this case, if p and q are subpermutations of π_T of length $n+2$, where p has form u and q has form v, $|u| = |v| = n+1$, $\delta_M(p) = \delta_M(q)$ if and only if $u = v$. Thus the number of subpermutations of $\pi_{d(T)}$ of length $2n-2$ which start in an odd position are determined by the number of factors of T of length $n+1$, or

$$\left|\mathrm{Perm}^{d(T)}_{odd}(2n-2)\right| = \left|\mathcal{F}_T(n+1)\right|.$$

We are now ready to calculate the permutation complexity of $d(T)$.

Theorem 5.3.19 *For the Thue-Morse word T, let $n \ge 9$.*

(a) If $n = 2^r$, then

$$\tau_{d(T)}(2n-1) = 2^{r+2} + 2^{r+1}$$
$$\tau_{d(T)}(2n) = 2^{r+2} + 2^{r+1} + 4$$

(b) If $n = 2^r + p$ for some $0 < p \leq 2^r - 1$, then

$$\tau_{d(T)}(2n-1) = 2^{r+3} + 4p$$

$$\tau_{d(T)}(2n) = 2^{r+3} + 4p + 2$$

Proof Let $n \geq 9$.

(a) Suppose $n = 2^r$. So $2n = 2(2^r) = 2(2^r + 1) - 2$, and from Lemma 5.3.17 and Lemma 5.3.18 each of the maps

$$\delta : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^{d(T)}_{ev}(2n)$$

$$\delta_L : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^{d(T)}_{ev}(2n-1)$$
$$\delta_R : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^{d(T)}_{odd}(2n-1)$$
$$\delta_M : \mathrm{Perm}^T(n+3) \mapsto \mathrm{Perm}^{d(T)}_{odd}(2n)$$

are not injective.

So $n+1 = 2^r+1 = 2^{r-1}+2^{r-1}+1$, and $n+2 = 2^r+2 = 2^r+1+1$. So by Proposition 5.3.15

$$\begin{aligned}\tau_{d(T)}(2n-1) &= \left|\mathrm{Perm}^{d(T)}_{ev}(2n-1)\right| + \left|\mathrm{Perm}^{d(T)}_{odd}(2n-1)\right| = |\mathcal{F}_T(n+1)| + |\mathcal{F}_T(n+1)| \\ &= 8(2^{r-2}) + 2(2^{r-1}) + 8(2^{r-2}) + 2(2^{r-1}) = 2^{r+2} + 2^{r+1}\end{aligned}$$

$$\begin{aligned}\tau_{d(T)}(2n) &= \left|\mathrm{Perm}^{d(T)}_{ev}(2n)\right| + \left|\mathrm{Perm}^{d(T)}_{odd}(2n)\right| = |\mathcal{F}_T(n+1)| + |\mathcal{F}_T(n+2)| \\ &= 8(2^{r-2}) + 2(2^{r-1}) + 6(2^{r-1}) + 4(1) = 2^{r+2} + 2^{r+1} + 4\end{aligned}$$

(b) Suppose $n = 2^r + p$. There will be 3 cases to consider. First when $0 < p \leq 2^r - 3$, next when $p = 2^r - 2$, and finally when $p = 2^r - 1$.

(b.1) Suppose $0 < p \leq 2^r - 3$. So $2n = 2(2^r + p) = 2(2^r + p + 1) - 2$, and from Lemma 5.3.17 and Lemma 5.3.18 each of the maps

$$\delta : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^{d(T)}_{ev}(2n)$$

$$\delta_L : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^{d(T)}_{ev}(2n-1)$$
$$\delta_R : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}^{d(T)}_{odd}(2n-1)$$

$$\delta_M : \mathrm{Perm}^T(n+3) \mapsto \mathrm{Perm}_{odd}^{d(T)}(2n)$$

are injective, and thus are bijective.

So $n+2 = 2^r + p + 2$, and $n+3 = 2^r + p + 3$. So by Theorem 5.2.16

$$\begin{aligned}
\tau_{d(T)}(2n-1) &= \left|\mathrm{Perm}_{ev}^{d(T)}(2n-1)\right| + \left|\mathrm{Perm}_{odd}^{d(T)}(2n-1)\right| \\
&= \left|\mathrm{Perm}^T(n+2)\right| + \left|\mathrm{Perm}^T(n+2)\right| = 2 \cdot 2(2^r + p + 2 - 2) \\
&= 2^{r+2} + 4p
\end{aligned}$$

$$\begin{aligned}
\tau_{d(T)}(2n) &= \left|\mathrm{Perm}_{ev}^{d(T)}(2n)\right| + \left|\mathrm{Perm}_{odd}^{d(T)}(2n)\right| = \left|\mathrm{Perm}^T(n+2)\right| + \left|\mathrm{Perm}^T(n+3)\right| \\
&= 2(2^r + p + 2 - 2) + 2(2^r + p + 3 - 2) = 2^{r+2} + 4p + 2
\end{aligned}$$

(b.2) Suppose $p = 2^r - 2$, so $n = 2^r + 2^r - 2 = 2^{r+1} - 2$. From Lemma 5.3.17 each of the maps

$$\delta : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}_{ev}^{d(T)}(2n)$$

$$\delta_L : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}_{ev}^{d(T)}(2n-1)$$

$$\delta_R : \mathrm{Perm}^T(n+2) \mapsto \mathrm{Perm}_{odd}^{d(T)}(2n-1)$$

are injective, and thus are bijective. Then we have $2n = 2(2^{r+1} - 2) = 2(2^{r+1} - 1) - 2$ and by Lemma 5.3.18 the map

$$\delta_M : \mathrm{Perm}^T(n+3) \mapsto \mathrm{Perm}_{odd}^{d(T)}(2n)$$

is not injective.

So $n+2 = 2^{r+1} = 2^r + 2^r = 2^r + (2^r - 1) + 1$. So by Proposition 5.3.15 and Theorem 5.2.16

$$\begin{aligned}
\tau_{d(T)}(2n-1) &= \left|\mathrm{Perm}_{ev}^{d(T)}(2n-1)\right| + \left|\mathrm{Perm}_{odd}^{d(T)}(2n-1)\right| \\
&= \left|\mathrm{Perm}^T(n+2)\right| + \left|\mathrm{Perm}^T(n+2)\right| = 2(2^{r+1} + 2^r - 2) + 2(2^{r+1} + 2^r - 2) \\
&= 2^{r+3} + 2^{r+2} - 8 = 2^{r+3} + 4(2^r - 2)
\end{aligned}$$

$$\begin{aligned}
\tau_{d(T)}(2n) &= \left|\mathrm{Perm}_{ev}^{d(T)}(2n)\right| + \left|\mathrm{Perm}_{odd}^{d(T)}(2n)\right| = \left|\mathrm{Perm}^T(n+2)\right| + \left|\mathcal{F}_T(n+2)\right| \\
&= 2(2^{r+1} + 2^r - 2) + 8(2^{r-1}) + 2(2^r - 1) = 2^{r+3} + 2^{r+2} - 6 \\
&= 2^{r+3} + 4(2^r - 2) + 2
\end{aligned}$$

(b.3) Suppose $p = 2^r - 1$, so $n = 2^r + 2^r - 1 = 2^{r+1} - 1$. So $2n = 2(2^{r+1} - 1) = 2(2^{r+1}) - 2$, and from Lemma 5.3.17 and Lemma 5.3.18 each of the maps

$$\delta : \text{Perm}^T(n+2) \mapsto \text{Perm}^{d(T)}_{ev}(2n)$$

$$\delta_L : \text{Perm}^T(n+2) \mapsto \text{Perm}^{d(T)}_{ev}(2n-1)$$
$$\delta_R : \text{Perm}^T(n+2) \mapsto \text{Perm}^{d(T)}_{odd}(2n-1)$$
$$\delta_M : \text{Perm}^T(n+3) \mapsto \text{Perm}^{d(T)}_{odd}(2n)$$

are not injective.

So $n + 1 = 2^{r+1} = 2^r + (2^r - 1) + 1$, and $n + 2 = 2^{r+1} + 1 = 2^r + 2^r + 1$. So by Proposition 5.3.15

$$\begin{aligned}\tau_{d(T)}(2n-1) &= \left|\text{Perm}^{d(T)}_{ev}(2n-1)\right| + \left|\text{Perm}^{d(T)}_{odd}(2n-1)\right| = |\mathcal{F}_T(n+1)| + |\mathcal{F}_T(n+1)| \\ &= 8(2^{r-1}) + 2(2^r - 1) + 8(2^{r-1}) + 2(2^r - 1) = 2^{r+3} + 2^{r+2} - 4 \\ &= 2^{r+3} + 4(2^r - 1)\end{aligned}$$

$$\begin{aligned}\tau_{d(T)}(2n) &= \left|\text{Perm}^{d(T)}_{ev}(2n)\right| + \left|\text{Perm}^{d(T)}_{odd}(2n)\right| = |\mathcal{F}_T(n+1)| + |\mathcal{F}_T(n+2)| \\ &= 8(2^{r-1}) + 2(2^r - 1) + 8(2^{r-1}) + 2(2^r) = 2^{r+3} + 2^{r+2} - 2 \\ &= 2^{r+3} + 4(2^r - 1) + 2\end{aligned}$$

■

5.4 Further Research Ideas

In Section 1.2, some possible directions for further research were given.

One direction of further research will be to investigate a possible relationship between morphisms and permutation complexity. For words which are the fixed point of a morphism an initial conjecture could be that permutation complexity can be calculated recursively, but if the morphism is not N-uniform (recall, $|\varphi(a)| = N$ for each $a \in \mathcal{A}$) there may be complications with such a calculation. It may be possible determine an upper bound for the permutation complexity of a fixed point of a morphism based on only the morphism, but it is unclear at the time of this writing if even this will be possible.

Another research direction is to start with a morphism φ on $\mathcal{A}$ and determine the permutation complexity of $\varphi(\omega)$, for some aperiodic binary word ω. After the investigation of permutation complexity and the doubling map, for an arbitrary morphism φ it may not be possible to determine the permutation complexity of $\varphi(\omega)$ based only on φ.

There is a class of morphisms, call it $\mathcal{M}$, so that if $\varphi \in \mathcal{M}$ the permutation complexity of $\varphi(\omega)$ can be calculated recursively for any word ω. The doubling map is not in the class $\mathcal{M}$, but the class $\mathcal{M}$ is not empty either. Trivially the identity map $(a \mapsto a)$ and the complement map $(0 \mapsto 1; 1 \mapsto 0)$ are both in $\mathcal{M}$. A natural question is, are there any other morphisms in $\mathcal{M}$?

In this study of permutation complexity, the main focus has been on counting the number of subpermutations which arise from an infinite permutation induced by a word. In the case of the permutation complexity of the Thue-Morse word we considered a pattern in subpermutations, but this was the depth of considering the subpermutations which arise. At the time of this writing, this author is not aware of an investigation of the subpermutations which arise from an aperiodic word. There are many possible questions in this area, but one question stands out. Given a permutation group G, where the maximum length of any permutation in G is n, is there a word ω so that

$$G = \{ \text{Perm}^{\omega}(i) \mid i \leq n \}?$$

Since there are permutations which will not occur in a binary word, there would be some conditions to impose on G, but what conditions could lead to a possible answer?

Appendix A

Subpermutations of π_T

The subpermutations and their form for factors of length 1 through 8 are shown below.

0 : (12)	1 : (21)

01 : (132)	(231)	00 : (123)
10 : (312)	(213)	11 : (321)

010 : (2413)	(1324)	001 : (1243)	100 : (3124)
101 : (4231)	(3142)	011 : (2431)	110 : (4312)

0011 : (23541)	(13542)	0010 : (12435)	1010 : (52413)
0110 : (25413)	(35412)	0100 : (24135)	1011 : (42531)
1001 : (41253)	(31254)	0101 : (14253)	1101 : (54231)
1100 : (53124)	(43125)		

00110 : (246513)	(136524)	00101 : (125364)	10010 : (412536)
01100 : (364125)	(254136)	01001 : (251364)	10100 : (524136)
10011 : (523641)	(413652)	01011 : (253641)	10110 : (526413)
11001 : (641253)	(531264)	01101 : (365241)	11010 : (652413)

011001 : (3751264) (2641375)

100110 : (6247513) (5137624)

001011 : (1364752)

001100 : (2475136)

001101 : (2476351)

010010 : (2513647)

010011 : (3624751)

010110 : (2637514)

011010 : (4762513)

100101 : (4126375)

101001 : (6251374)

101100 : (5264137)

101101 : (6375241)

110010 : (6412537)

110011 : (6413752)

110100 : (7524136)

0010110 : (13748625)

0011001 : (24861375)

0011010 : (25873614)

0100101 : (25137486)

0100110 : (37258614)

0101100 : (26375148)

0101101 : (37486251)

0110010 : (37512648)

0110011 : (37514862)

0110100 : (48625137)

1001011 : (51374862)

1001100 : (62485137)

1001101 : (62487351)

1010010 : (62513748)

1010011 : (73624851)

1011001 : (62741385)

1011010 : (74862513)

1100101 : (74126385)

1100110 : (75138624)

1101001 : (86251374)

00101101 : (248597361) (148597362)

01001011 : (361485972) (261485973)

01011010 : (485972613) (385972614)

01101001 : (497261385) (597261384)

10010110 : (613849725) (513849726)

10100101 : (725138496) (625138497)

10110100 : (849625137) (749625138)

11010010 : (962513748) (862513749)

00101100 : (137486259)

00110010 : (248613759)

00110100 : (259736148)

01001100 : (372596148)

01011001 : (273851496)

01100101 : (385127496)

01100110 : (386149725)

10011001 : (724961385)

10011010 : (725983614)

10100110 : (837259614)

10110011 : (738514962)

11001011 : (851374962)

11001101 : (862497351)

11010011 : (973624851)

Bibliography

[1] J.-P. Allouche and A. Glen. Extremal properties of (epi)Sturmian sequences and distribution modulo 1, 2009. arXiv:0907.2430v1.

[2] J.-P. Allouche and J. Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In *Sequences and Their Applications, Proc. SETA'98*, pages 1–16. Springer-Verlag, 1999.

[3] J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, 2003.

[4] E. Altman, B. Gaujal, and A. Hordijk. Balanced sequences and optimal routing. *J. ACM*, 47(4):752–775, 2000.

[5] P. Arnoux and G. Rauzy. Représentation géométrique de suites de complexité $2n+1$. *Bull. Soc. Math. France*, 119(2):199–215, 1991.

[6] S.V. Avgustinovich, A.E. Frid, T. Kamae, and P.V. Salimov. Infinite permutations of lowest maximal pattern complexity, 2009. http://arxiv.org/abs/0910.5696v2.

[7] J. Barát and P. P. Varjú. Partitioning the positive integers to seven beatty sequences. *Indag. Math. (N.S.)*, 14(2):149–161, 2003.

[8] J. Berstel. Recent results in sturmian words. In J. Dassaw, editor, *Developments in Language Theory*. World Scientific, Singapore, 1995.

[9] J. Berstel and D. Perrin. The origins of combinatorics on words. *European J. Combin.*, 28(3):996–1022, 2007.

[10] J. Berstel and P. Séébold. A characterization of Sturmian morphisms. In *Mathematical Foundations of Computer Science 1993*, volume 711 of *Lecture Notes in Computer Science*, pages 281–290. Springer-Verlag, 1993.

[11] S. Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Appl. Math*, 24:83–96, 1989.

[12] J. Cassaigne, S. Ferenczi, and L.Q. Zamboni. Imbalances in Arnoux-Rauzy sequences. *Ann. Inst. Fourier*, 50(4):1265–1276, 2000.

[13] E.M. Coven and G.A. Hedlund. Sequences with minimal block growth. *Math. Systems Theory*, 7(2):138–153, 1973.

[14] J. Currie and N. Rampersad. Recurrent words with constant abelian complexity, 2009. arXiv:0911.5151v1.

[15] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2):539–553, 2001.

[16] X. Droubay and G. Pirillo. Palindromes and Sturmian words. *Theoret. Comput. Sci.*, 223(1-2):73–85, 1999.

[17] L. Euler. On the expansion of the power of any polynomial $(1+x+x^2+x^3+x^4+etc)^n$, 2005.

[18] D.G. Fon-Der-Flaass and A.E. Frid. On periodicity and low complexity of infinite permutations. *European J. Combin.*, 28(8):2106–2114, 2007.

[19] A. Fraenkel. Complementing and exactly covering sequences. *J. Combin. Theory Ser. A*, 14(1):8–20, 1973.

[20] A. Glen and J. Justin. Episturmian words: a survey. *RAIRO - Theor. Inform. Appl.*, 43:402–433, 2009.

[21] A. Glen, J. Justin, S. Widmer, and L.Q. Zamboni. Palindromic richness. *European J. Combin.*, 30(2):510–531, 2009.

[22] R. Graham. Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta] : n = 1, 2, ...\}$. *J. Combin. Theory Ser. A*, 15(3):354–358, 1973.

[23] P. Hubert. Propriétés combinatoires des suites définies par le billard dans les triangles pavants. *Theoret. Comput. Sci.*, 164(1-2):165–183, 1996.

[24] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoret. Comput. Sci.*, 276:281–313, 2002.

[25] T. Kamae and L.Q. Zamboni. Maximal pattern complexity for discrete systems. *Ergodic Theory Dynam. Systems*, 22:1201–1214, 2002.

[26] T. Kamae and L.Q. Zamboni. Sequence entropy and the maximal pattern complexity of infinite words. *Ergodic Theory Dynam. Systems*, 22:1191–1199, 2002.

[27] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1983.

[28] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2002.

[29] M. Lothaire. *Applied Combinatorics on Words*, volume 105 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2005.

[30] A.de Luca. Sturmian words: structure, combinatorics, and their arithmetics. *Theoret. Comput. Sci.*, 183(1):45–82, 1997.

[31] A.de Luca and S. Varricchio. Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups. *Theoret. Comput. Sci.*, 63:333–348, 1989.

[32] M.A. Makarov. On permutations generated by infinite binary words. *Sib. Èlektron. Mat. Izv.*, 3:304–311, 2006. (in Russian).

[33] M.A. Makarov. On the permutations generated by the Sturmian words. *Sib. Math. J.*, 50(3):674–680, 2009.

[34] M.A. Makarov. On the infinite permutation generated by the period doubling word. *European J. Combin.*, 31(1):368–378, 2010.

[35] M. Morse. Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.*, 22(1):84–100, 1921.

[36] M. Morse and G.A. Hedlund. Symbolic dynamics. *Amer. J. Math.*, 60(4):815–866, 1938.

[37] M. Morse and G.A. Hedlund. Symbolic dynamics II: Sturmian trajectories. *Amer. J. Math.*, 62(1):1–42, 1940.

[38] G. Paquin and L. Vuillon. A characterization of balanced episturmian sequences. *Electron. J. Combin.*, 14(1), 2007. R33.

[39] G. Rauzy. Suites à termes dans un alphabet fini. *J. Théor. Nombres Bordeaux*, exposé 25:1–16, 1983.

[40] G. Richomme, K. Saari, and L.Q. Zamboni. Abelian complexity in minimal subshifts. *J. London Math. Soc.*, 2010. arXiv:0911.2914v1.

[41] G. Richomme, K. Saari, and L.Q. Zamboni. Balance and abelian complexity of the tribonacci word. *Adv. in Appl. Math.*, 45:212–231, 2010.

[42] C.F. Sainte-Marie. Solution to question nr. 48. *L'intermédiaire des Mathématiciens*, 1:107–110, 1894.

[43] A. Thue. Über unendliche Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.*, 7:1–22, 1906.

[44] A. Thue. Über die gegenseitige lage gleicher Teile gewisser Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.*, 1:1–67, 1912.

[45] R. Tijdeman. Exact covers of balanced sequences and Fraenkel's conjecture. In *Algebraic Number Theory and Diophantine Analysis*, pages 467–483, Graz, Austria, 1998. de Gruyter(2000).

[46] J. Tromp and J. Shallit. Subword complexity of the generalized Thue-Morse word. *Inform. Process. Lett.*, 53:313–316, 1995.

[47] T. van Aardenne-Ehrenfest and N.G. de Bruijn. Circuits and trees in oriented linear graphs. *Bull. Belg. Math. Soc. Simon Stevin*, 28:203–217, 1951.

[48] L. Vuillon. Balanced words. *Bull. Belg. Math. Soc. Simon Stevin*, 10(5):787–805, 2003.

[49] E.W. Weisstein. Composition. *From MathWorld–A Wolfram Web Resource*. http://mathworld.wolfram.com/Composition.html.

[50] S. Widmer. Permutation complexity of the Thue-Morse word. *Adv. in Appl. Math.* 2010. doi:10.1016/j.aam.2010.08.002.