Repetitions in Words

Thesis submitted to the
Indian Institute of Technology Guwahati
for the award of the degree

of

Doctor of Philosophy

Submitted by

Maithilee Patawar

Under the guidance of

Dr. Kalpesh Kapoor

Dr. Benny George K

Department of Computer Science and Engineering

Indian Institute of Technology Guwahati

April 2024

Declaration

I certify that:

- The work contained in this thesis is original and has been done by myself and under the general supervision of my supervisors.
- The work reported herein has not been submitted to any other Institute for any degree or diploma.
- Whenever I have used materials (concepts, ideas, text, expressions, data, graphs, diagrams, theoretical analysis, results, etc.) from other sources, I have given due credit by citing them in the text of the thesis and giving their details in the references. Elaborate sentences used verbatim from published work have been clearly identified and quoted.
- I also affirm that no part of this thesis can be considered plagiarism to the best of my knowledge and understanding and take complete responsibility if any complaint arises.
- I am fully aware that my thesis supervisors are not in a position to check for any possible instance of plagiarism within this submitted work.

Date:
Place: Guwahati **(Maithilee Patawar)**

Acknowledgements

I am thankful to everyone who played a role in my academic journey and the completion of this thesis. I extend my heartfelt appreciation to my supervisor, Prof. Kalpesh Kapoor, for his invaluable guidance and unwavering support. Prof. Kapoor's calm and composed demeanour has always been a source of inspiration, allowing me to navigate through challenging phases with confidence. His encouragement and belief in my abilities have motivated me to push boundaries and explore unconventional perspectives. I consider myself truly fortunate to have had the privilege of working under his supervision.

I would also like to express my gratitude to Prof. Benny George, my co-supervisor, for his valuable insights and suggestions, which greatly contributed to my academic journey. Additionally, I extend my appreciation to the members of the Doctoral Committee: Prof. Sahu, Prof. Wagh, and the DC chair, Prof. Sajith, for their insightful guidance and expertise throughout my research.

I am grateful to the Department of Computer Science and Engineering at IIT Guwahati for providing me with the necessary resources and opportunities to pursue my studies. Special thanks go to my friends and family for their constant support, love, and encouragement throughout the ups and downs of this journey. Without all of you, this work would not have been possible. Thank you.

Certificate

This is to certify that this thesis entitled, **Repetitions in Words**, being submitted by **Ms. Maithilee Patawar**, to the Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, for partial fulfilment of the award of the degree of Doctor of Philosophy, is a bonafide work carried out by her under my supervision and guidance. The thesis, in my opinion, is worthy of consideration for the award of the degree of Doctor of Philosophy in accordance with the regulation of the institute. To the best of my knowledge, it has not been submitted elsewhere for the award of the degree.

Date:
Place: Guwahati

....	....
Dr. Kalpesh Kapoor	**Dr. Benny George K**
Supervisor	Co-supervisor

Abstract

Repetitions are fundamental properties of words, and different types of repetitions have been explored in the area of word combinatorics. This thesis investigates two types of repetitions: squares and antisquares. We investigate the square conjecture that anticipates the number of distinct squares in a word is less than its length. It is known that any location of a word can be mapped to at most two rightmost squares, and a pair of these squares was referred to as an FS-double square. For simplicity, we will refer to the longer square in this pair as an FS-double square throughout this thesis. We examine the properties of words containing FS-double squares and explore the consecutive locations starting with FS-double squares. We observe that FS-double squares introduce no-gain locations where no rightmost squares occur. The count of these no-gain locations in words with a sequence of FS-double squares demonstrates that the square density of such words is less than $\frac{11}{6}$. Furthermore, we investigate words that possess FS-double squares and maintain an equivalent number of such squares when reversed. We prove that the maximum number of FS-double squares in such a word is less than $\frac{1}{11}^{\text{th}}$ of the length of the word. Another aspect of our research involves counting squares in a repetition. A non-primitive word has a form u^k for some non-empty word u and some positive integer k such that $k > 2$. With no-gain locations and FS-double squares in these words, we conclude that the square density of such words approaches $\frac{1}{2}$ as k increases. Also, we work on the lower bound of the square conjecture. The previous lower bound is obtained using a structure that generates words containing a high number of distinct squares. We identify similar structures but produce words with more distinct squares. We also study antisquare, a special repetition of the form $u\bar{u}$ where u is a binary word, and $\bar{u}$ is its complement. We show that a word w can contain at most $\frac{|w|(|w|+2)}{8}$ antisquares, and the lower bound for the number of distinct antisquares in w is $|w| - 1$.

Contents

List of Figures

List of Tables

1
Introduction

Combinatorics on words is a discipline that studies the properties of words, that are sequences of symbols. The topic of repetition or periodicity remains of interest when exploring the characteristics of words. A fundamental concept in the field is the notion of a square, which refers to the smallest repetition in a word with the form "uu". Axel Thue, a pioneer in the field, discovered an infinite sequence of letters without any squares in his work [66]. This was a breakthrough discovery, as it was previously believed that every sequence of symbols contained squares. Thue's work on the properties of words has been influential in many fields of mathematics, including number theory, computer science, and cryptography [2, 13, 23]. For instance, a number theoretic problem discovered by Prouhet has a connection with the Thue-Morse sequence, because of which the sequence is also known as the Prouhet-Thue-Morse sequence [58]. The definition of Prouhet has a connection with finite automata and automatic words identified by Cobham [18]. Figure 1.1 lists some of such problems and areas that have a connection with the word combinatorics and the study of words.

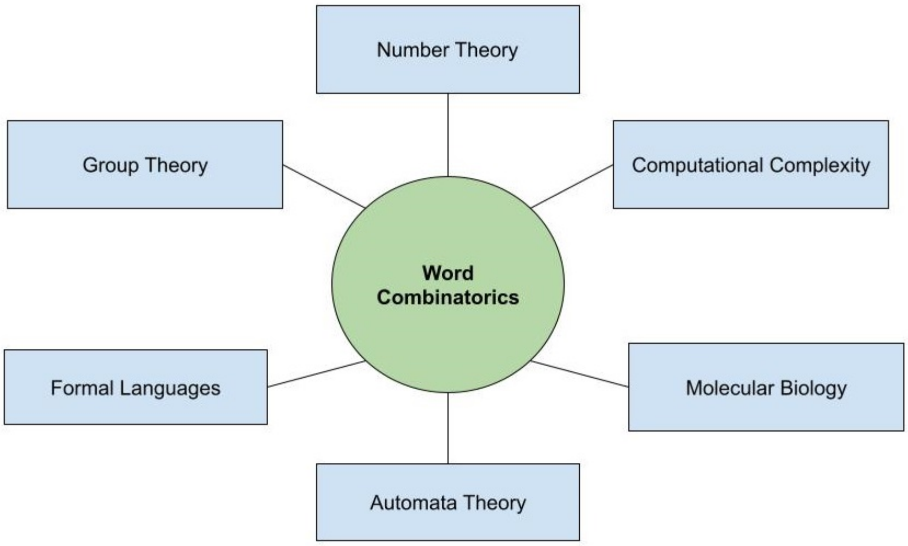

Figure 1.1: *Fields related to word combinatorics*

The properties of words explored in combinatorics have a variety of applications. The study of DNA sequences, gene prediction, and the analysis of genomes are some areas where combinatorics on words plays an important role. For example, identifying repeats in the genome sequence is a key challenge in bioinformatics, and this can be done using combinatorial algorithms based on periodicity analysis of words [6, 67]. The words and their properties are also used in data compression algorithms such as the Lempel-Ziv-Welch (LZW) algorithm, which employs dictionary-based encoding of repeated patterns in the input text [51]. Another fundamental problem in computer science that benefited from the word combinatorics is the identification of patterns and similarities in strings. Word combinatorics provides several algorithms and techniques for solving this problem [21]. The analysis and design of cryptographic systems often involve combinatorial problems on words, such as identifying repeated patterns or constructing sequences with specific properties [41]. In natural language processing, some tasks like identifying word collocations or generating text with specific properties utilize the results of word combinatorics [55].

Apart from square-free words, studying different types of repetitions is a key area of research in word combinatorics. Squares, cubes and runs are some examples of repetitions. They help identify the structural properties of words. A repetition of form u^k concatenates k copies of a word u. Here, u is called the root, and k is called an exponent of the repetition. In such a repetition, the root is duplicated within the resulting word, and the exponent, whether rational or integer, determines the frequency of this duplication. A cube consists of three consecutive copies of the same word, showing an exponent of three. In contrast, the exponent of a square is two. Squares and cubes are specific types of repetitions having integer exponents. Some repetitions can be expressed using multiple exponents. For example, the word $w = abababab$ can be written as $(ab)^4$ or $(abab)^2$. In the context of repetitions, the period refers to the length of the root word. So far, our discussion has focused on repetitions with integer exponents. Another type of repetition, a run, involves a rational exponent. Kolpakov and Kucherov [46] defined a run as a maximal repetition in a word that, extending by a letter at the beginning or at the end, generates a new word with a higher period. An instance of a run in the word $u = aba\underline{ababa}abaab$ is the repetition $ababa$, characterized by a period of 2, a root of ab, and an exponent of 2.5.

In another type of repetition, two instances of the same word are separated by a different word. It is known as gapped repeat. An example is a word $w = abaaab$ where the word 'aa' separates the two occurrences of the word 'ab'. The characterisation of words containing gapped repetitions and their sub-repetitions is discussed in [47]. In addition, the concept of anti-powers has been introduced in [32], which defines them as the concatenation of distinct words of equal length. An antisquare is a special type of anti-power defined over a binary alphabet where the word is concatenated to its complement [12]. In this thesis, we study antisquares that are defined for a binary alphabet. We now discuss some notation and conjectures related to repetitions in the next section.

1.1 Notation and Related Conjectures

Let Σ be an alphabet. A word w is a finite sequence of letters drawn from Σ. Any non-empty subsequence of consecutive letters in w is a factor of w. The length of a word w is the number of symbols in it, and it is denoted by $|w|$. The symbol ϵ represents an empty sequence called an empty word.

A binary operation *concatenation* combines two non-empty words to create a new word. The operation is denoted by the symbol "$\cdot$" or simply by juxtaposing the words together. The concatenation of words u and v is denoted as $u \cdot v = uv$. We use $\mathbb{N}$ to denote the set of non-negative integers. Let Σ^n be the set of words of length $n \in \mathbb{N}$ defined over an alphabet Σ. The set Σ^* is the free monoid generated by Σ under the concatenation operation. Another symbol, Σ^+ represents the set of all non-empty words over Σ. Let $u, v \in \Sigma^*$ and $w = uv$. We say that u is a prefix (resp. v is a suffix). A prefix or a suffix is proper if $u, v \in \Sigma^+$. We use the term $lcp(x, y)$ to denote the longest common prefix shared by two non-empty words x, y. The word vu is a conjugate of the word w.

In this thesis, the term "repetition" refers to a concatenation of multiple equal-length non-empty words. A repetition u^k concatenates k copies of a word u. Here, u is a non-empty word referred to as the root of the repetition. The number k is an exponent or a power of the repetition that satisfies the relation $k > 1$. It can be either an integer or a rational number. The number $|u|$ is a period of the repetition. The repetition of the form u^k is explored in Chapter 3 to Chapter 6, where the exponent of the repetition is an integer. If a word w is repetition u^k where k is an integer such that $k \geq 2$, then w is also known as non-primitive. Any word that is not non-primitive is a primitive word. For instance, '$abab$' and 'aba' are non-primitive and primitive.

A square is a repetition of form u^2. Squares can be classified into two types, primitive squares and non-primitive squares, based on their roots. A *primitive square* is a square whose root is a primitive word. A repetition u^k

where $k = 3$ is labelled as a cube. An i^{th} letter of w is denoted as $w[i]$ and i is the position on w. Denote by $w[i, j]$ the factor $w[i]w[i+1] \dots w[j]$ of w. The factor $w[i, j]$ is called a run if a period of the factor $w[i, j]$ is shorter than the period of $w[i-1, j]$ for $i > 1$, and if $j < n$, then the period of $w[i, j]$ must be shorter than that of $w[i, j+1]$. Chapter 2 describes in detail the results obtained for these types of repetitions. Other types of repetition are anti-powers and antisquares. An anti-power of order k is a concatenation of k different words where the size of each word is equal. Chapter 7 elaborated on these terms in detail.

The results discussed in this thesis are obtained by solving the word equations and comparing the overlaps between words. The next two theorems of Lyndon and Schützenberger [3] are the fundamental results on words that are used to solve the word equations.

Theorem 1.1. *([3]) Let $uv = vw$ where $v \in \Sigma^*$ and $u, w \in \Sigma^+$. Then, there exists $x, y \in \Sigma^*$ and an integer $e \geq 0$ such that $u = xy$, $w = yx$ and $v = (xy)^e x$.*

Theorem 1.2. *([3]) Let the non-empty words u, v satisfy the relation $uv = vu$. Then, there exists a non-empty word z and positive integers i, j such that $u = z^i$ and $v = z^j$.*

The primary emphasis of this thesis lies on the square conjecture. The square conjecture predicts a bound for the count of distinct squares in a word. A word may contain multiple instances of the same square, such as '*ss*' in the English word '*possess*'. So, counting distinct squares needs to consider the unique occurrence of each square. The notation $s_i(w)$ is taken from the work [34], and it indicates the number of distinct squares starting from a location i in word w for the last time. So, we also refer to such squares as the rightmost squares starting at location i. Another term, FS-double square, refers to a longer square starting at location i when $s_i(w) = 2$. The ratio of the number of distinct squares in a word to the number of letters in a word indicates the square density of the word.

Now, we see how the square conjecture is connected to other conjectures related to repetitions. The repetitions with higher exponents like cubes and runs always contain a square. Also, the presence of a square ensures the presence of all conjugates of the root. For instance, the square '*abcabc*' has all conjugates of '*abc*'. This shows that the conjectures focusing on cubes, runs, and conjugates can potentially benefit from the results on squares.

Conjecture	**Year**	**Statement**
Square Conjecture [34]	1998	The number of distinct squares in w is less than $\lvert w\rvert$.
Runs conjecture [46]	1999	The maximum number of runs in w is less than $\lvert w\rvert$
Primitive square conjecture [28]	2011	For $w \in \Sigma^+$, the number of distinct primitive squares is less than $\lvert w\rvert - \lvert\Sigma\rvert$.
Cube conjecture [49]	2013	The number of distinct cubes in w is less than $\lvert w\rvert$
Stronger square conjecture [44]	2014	An improved bound, $\frac{2k-1}{2k+2}$, for the square conjecture. Here k is the count of least appearing letter in a binary word w.
Squares in circular words [5]	2017	The number of distinct squares in a word w and in all of its conjugates is less than $3.14\lvert w\rvert$.

Table 1.1: *Some conjectures on repetitions*

The square conjecture and a list of related conjectures are mentioned in Table 1.1. Some of these conjectures are named after their statement, such as 'cube conjecture', that anticipates the number of distinct cubes in a finite word. This table aims to help the reader understand the context and motivation behind the research questions and how they fit into the larger field of study. These conjectures will be discussed in detail in the related work chapter, further exploring their significance.

1.2 Thesis Outline

The thesis investigates the topic of "repetitions in words," with an emphasis on studying repetitions in finite words. It addresses the problem of determining the number of distinct squares in a finite word to enhance both lower and upper bounds. This problem is studied for specific types of words to gain further insight. Antisquare, a concept dual to the concept of squares, is also discussed. This section provides an overview of the main chapters and their contents.

Chapter 2 provides a critical review of existing literature related to repetitions in a finite word.

Chapter 3 discusses the square conjecture and results related to FS-double squares. The structure of an FS-double square is studied to identify all possible structures that generate words having FS-double squares starting at consecutive locations. Finally, the length of the longest sequence of such locations is determined.

Chapter 4 characterises words with a sequence of locations beginning with FS-double squares. The locations that do not start with any squares are detected in such words. This information is then used to compute their square densities.

Chapter 5 explores two special words, namely bordered FS-double squares and non-primitive words. The count of FS-double squares in a word may differ from that in its reverse. In this regard bordered FS-double square is identified. The bordered FS-double square can be reversed to create another FS-double square of the same size. The chapter also describes the types of squares appearing in non-primitive words and their properties.

Chapter 6 describes various structures of words whose square densities are approaching the value one. These are referred to as patterns. The chapter also provides various methods for generating patterns, along with corresponding notions to compare them.

Chapter 7 presents a study on antisquares. Antisquare is a special type of repetition where a word is concatenated with its complement. The chapter provides basic results on antisquares and a lower bound for the number of distinct antisquares in a word.

Chapter 8 provides a summary of the key discoveries and implications of the research. It also lists some open problems and potential future directions for further research.

2

Related Work

In the field of word combinatorics, the terms repetition, periodicity, and regularity are used to refer to the same concept. Repetitions can be defined in various ways, and this chapter focuses on the following types of repetitions: squares, cubes and runs, and anti-powers. We already discussed that a repetition consists of a root and an exponent. Squares are the smallest repetitions and are always present in any repetition with an exponent greater than two. Squares and cubes are examples of repetitions where the exponent is an integer, while runs are repetitions with rational number exponents. Antisquares, on the other hand, are a type of anti-powers. An anti-power of order k indicates the concatenation of k distinct words. The literature review is organized according to the types of repetitions mentioned in Figure 2.1. The next section presents the known results about squares.

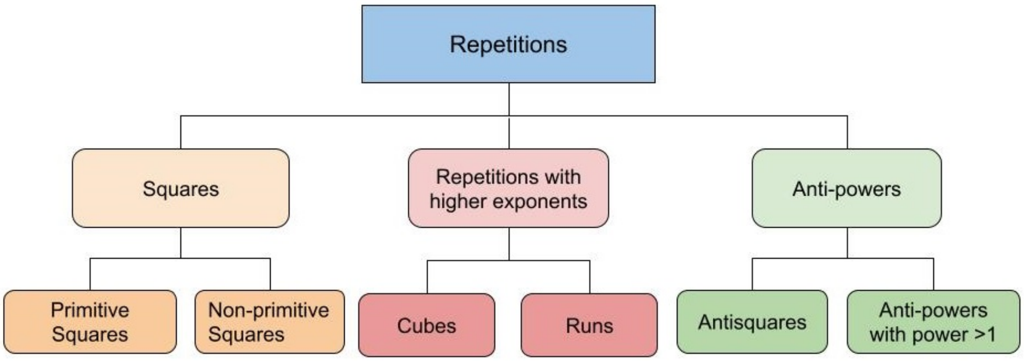

Figure 2.1: *Types of repetitions*

2.1 Squares in Words

The distribution of repetitions in a word reveals interesting properties of words. A square is a basic repetition obtained by concatenating two identical words. The study on squares initially involved detecting all instances of squares within a word, leading to the development of various algorithms for this purpose [19, 64]. Researchers have used the combinatorial aspects of squares to improve these algorithms, leading to further investigation into their properties. Every even-length sequence of letters from a word a^n is a square, so it contains the maximum number of squares. The number of squares in it is $\frac{n^2}{4}$ or $\frac{n^2-1}{4}$. This count takes into account both primitive and non-primitive squares, including their repeated occurrences. Recall a primitive square is a square having a primitive root. Table 2.1 presents a compilation of the various types of squares that have been the subject of research, along with the corresponding research findings.

The study of distinct squares in words has begun by counting them in infinite binary words. The initial results stated that every binary sequence of length greater than or equal to 18 always contain squares[30]. Later, in 1994, Fraenkel and Simpson investigated the maximum length of binary words with at most k distinct squares [36]. Their findings revealed that it is possible to construct a word of any length containing only three distinct squares. The

	Distinct Squares		All Squares	
	Lower bound	Upper bound	Lower bound	Upper bound
Primitive squares	Deza's Words [27] ($< n$)	$n - o(n)$ [34] $n - \lvert\Sigma\rvert$ [28]	Fibonacci words [19]	$O(n \log n)$ [53]
All Squares	Q Words [34] ($< n$)	$\lfloor \frac{11n}{6} \rfloor$ [29]	a^n	$\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor}(n - 2i + 1)$

Table 2.1: *Known bounds for different square conjectures*

same authors investigated another problem of identifying distinct squares in a finite word. They examined the last occurrence of a square, which they referred to as the "rightmost occurrence" and introduced the s_i notation [34]. They demonstrated $s_i(w) \leq 2$ for all $1 \leq I \leq |w|$. In addition, they discovered a structure that generates words with a large number of distinct squares. This structure generates words of varying lengths, with the number of distinct squares increasing as the word length increases. However, the number of distinct squares is always less than the word length. The square conjecture is based on these properties of structure-generated words, which state that the length of the word is always greater than the number of distinct squares it contains.

Ilie in [42] simplified the proof for the inequality $s_i(w) \leq 2$. The proof takes advantage of the fact that all the conjugates of a primitive word, as described in Section 1.1, are distinct. In subsequent work, Ilie attempted to find a better bound for the conjecture [43]. To achieve this, he studied a sequence of locations satisfying $s_i(w) = 2$, and obtained an upper bound of $2n - \log n$ for the square conjecture.

A well-known *Periodicity Lemma* given by Fine and Wilf [33] is often used in the exploration of the square conjecture and the runs conjecture. The lemma establishes the relation between multiple periods and their greatest common divisors. The definition of a period, given in the previous chapter, is discussed here in detail. A word u^r is a repetition for $u \in \Sigma^+$ and $r \geq 2$. Furthermore, the repetition u^r in w is a factor $w[i, i + ur - 1]$ for $1 \leq i \leq |w| + 1 - ur$ and $w[i, i + u - 1]$ is not a repetition. We call $|u|$ is the period of w. The term $gcd(x, y)$ represents the greatest common divisor of two positive integers x, y in the following lemma.

Lemma 2.1 (Periodicity Lemma [33]). *If p, q be two periods of w such that $|w| > p + q - gcd(p, q)$ where $gcd(p, q)$ is a greatest common divisor of p, q. Then, $gcd(p, q)$ is also a period of w.*

The lemma has been extended to yield various results. For example, the work in [17] examined words containing three periods. Another study by Fraenkel and Simpson [35] focused on the structure of longer periods of word w when it consists of two periods. Further, the lemma used in [62] considered a square v^2 starting with another square and obtained a result on a specific factor of v^2. The words starting with two distinct squares were analysed further in [31] that defines a lemma named 'New Periodicity Lemma' (NPL). The lemma investigates the structure of a word starting with two squares and determines the specific locations that do not start with any squares. The proofs discussed in this work considered a special class of words by restricting the types of squares with which these words can begin. The paper also described a general version of the problem of having three overlapping squares to identify specific locations that do not start with any squares. The proof required to validate 14 cases out of which a sub-case 4 is solved in [37]. The work also presented proofs for sub-cases $11 - 14$.

A canonical factorization of two squares starting at the same location is obtained in [9] to broaden the scope of the new periodicity lemma. The structure is then used to generalise Crochemore and Rytter's lemma in [8]. Under certain conditions, the NPL can also be used to map the locations with zero s_i values to a location whose s_i value is two. This is valid for some special words beginning with two squares in which the shorter square is 'regular'. The term regular represents a square whose proper prefix cannot be a square. The result of NPL is further extended in [10] for a square v^2 that starts with another shorter square u^2. This work identified some properties of squares that appear in v^2.

The study of locations whose s_i values are two is conducted in [50]. It showed that the density of any word is at most $\frac{95}{48}$. Later, the term FS-double square was coined in [29] to represent the longer square starting at such a location. The work categorized FS-double squares based on their lengths and starting locations and identified five categories called $\alpha, \beta, \ldots, \epsilon$-mates. Further analysis revealed that at most $\frac{5|w|}{6}$ locations of any word w can begin with FS-double squares, leading to an upper bound of $\frac{11|w|}{6}$ for the square conjecture. Thierry extended the investigation of the five mates and obtained an improved bound of $1.5n$ in an archived paper [65].

A related problem of identifying distinct primitive squares in a word is investigated in [28]. The study proposes a novel approach that takes into account both the word length n and the size of the alphabet d. A (d, n) table is plotted, and several regularities in the values of the table are observed to get new results on primitive squares. Also, a new conjecture is proposed that anticipates the distinct primitive squares in a word are always less than $n-d$. Further, in [26], an extension of the work introduces a computationally efficient framework. This framework focuses on identifying words that contain the maximum number of distinct primitive squares.

In another study [44], a more stringent bound for the square conjecture was proposed. However, it should be noted that this bound was limited to the case of a binary alphabet. According to it, the number of distinct squares in a word depends on the number of the least frequent letters occurring in a binary word. The work anticipated that $|DS(w)| \leq \frac{2k-1}{2k+2}|w|$, where k is the frequency of the least frequent letter in w and $DS(w)$ is the set of distinct squares in w. Meanwhile, the work in [14] conjectured that if a word has a sequence of n indices starting with distinct FS-double squares, then it must have a chain of $2n$ empty indices, indicating that the upper bound on the number of FS-double squares in a word may be further reduced. In [54], the relationship between the density of distinct squares in words over varying alphabets is analyzed. The next theorem presented in this study ensures that solving the square conjecture for binary words alone is sufficient.

Theorem 2.2. *Let $\rho(w)$ be the ratio of distinct squares in a word w to $|w|$. For any word w over a ternary or larger alphabet, there exists a binary word u such that $\rho(u) > \rho(w)$.*

The best words known to have many distinct squares use properties of primitive squares and their conjugates [34]. Another study on conjugates of words [5] considered solving the problem of finding distinct squares in circular words. The circular word of a word w represents a word containing all cyclic rotations of w. For example, *abaab* is a circular word for a word $w = aba$ since the first three locations of the word start with a distinct conjugate of w. A simple way to get a circular word is to extend any word w by its first $|w| - 1$ letters. The number of distinct squares in a circular word of size n is conjectured to be $3.14n$, and it is proved that the lower bound is $1.25n$ [5]. A conjecture on repetitions with higher exponents predicts that the number of distinct repetitions of exponent k in w is $\frac{|w|-|\Sigma|}{k-2}$ [52]. We study repetitions with higher exponents and compute the square density in Chapter 5. Recently, the proof to solve the square conjecture is published in a pre-print in [15]. Unlike the previous works that consider the rightmost occurrences of squares, this work converted a word into Rauzy graphs. The proof of the conjecture established an injection between the number of distinct squares in a word and circuits in the Rauzy graphs.

2.2 Cubes and Runs in Words

Beyond the study of squares, researchers have investigated other repetitions, such as counting the maximum distinct cubes in a word. A relation between the occurrences of cubes and non-primitive words has been observed while determining the combinatorial properties of cubes in words. In this regard, authors in [48] studied the problem of counting distinct cubes or cubic factors in a word. Their results revealed that the maximum number of distinct cubes in a word w falls within the range $\left[\frac{|w|}{2}, \frac{4|w|}{5}\right]$. The study was extended in [49] to demonstrate that the count of cubes in w is any value between the range of $\left[\frac{|w|}{2} - 2\sqrt{n}, \frac{4|w|}{5}\right]$. The authors achieved this by computing the non-primitive squares in the word, which were shown to be at most $\left\lfloor \frac{|w|}{2} \right\rfloor$.

The runs conjecture, which states that the maximum number of runs in a string of length n is less than n, has been the focus of many investigations in word combinatorics. Kolpakov and Kucherov first proved that the maximum number of runs in a word of length n is a linear function of n [46]. However, their proof did not give a specific constant factor. Rytter [61] was the first to give an explicit constant, showing that the number of runs in a word w is less than $5|w|$. Later the constant term 5 was improved to 3.48 in [59]. The authors Crochemore and Ilie succeeded to getting closer to the predicted value of the conjecture by further improving the constant factor to 1.6 [20]. These authors also suggested using computer verification for better bounds. Giraud in [40] was able to reduce the value to $1.52n$ and to $1.29n$ for binary words. All the proofs obtained to get the better constant factor for the conjecture used the periodicity lemma (see Lemma 2.1).

Efforts were made to enhance the lower bound of the runs conjecture, leading to several developments. Initially, it was proven that the number of runs in a word could be greater than $0.927n$, where n represents the length of the word [38]. This number was believed to be optimal at first. However, later a research conducted by Matsubara et al. [56] indicated that the number could be higher than $0.944565n$. Another modification in this regard improved the constant value to $0.944575712n$ [63].

Finally, Bannai et al. solved the runs conjecture using Lyndon words [11].

2.3 Anti-powers and Antisquares

Ramsey theory is a branch of mathematics that studies the emergence of order within large, complex, and apparently disordered structures [60]. The theory has found applications in combinatorics on words through some important results stating the existence of unavoidable regularities, for example, Ramsey, van der Waerden and Shirshov theorems [24]. In [32], authors used the notion of anti-powers to obtain an anti-Ramsey result in the context of combinatorics on words. An anti-power of order k is a concatenation of k equal-length words, each of which is distinct. The study in [32] shows that such repetition is unavoidable in an infinite word. Further, it is shown that if an infinite word contains no anti-power of order 3, then the word is ultimately periodic.

In another research on the lengths of anti-powers appearing in a word w, a function $ap(w, k)$ and Thue-Morse words are explored [25]. Here, $ap(w, k)$ is the minimum m for which the prefix of a word w of length km is a k-anti-power. It is proved that the function $ap(w, k)$ grows linearly for a Thue-Morse word. The conjectures proposed in this study are then explored in [57] to identify the distribution of anti-power prefix lengths in the Thue-Morse word. The work derived several properties of these lengths, including their asymptotic behaviour and their distribution modulo small integers. The paper also introduces a new function that counts the number of anti-power prefix lengths of a given length in the Thue-Morse word. Gaetz [39] has since extended Defant's results to factors of words. Burcroff [16] studied the avoidability of k antipowers in infinite words, generalizing Fici et al.'s results in [32].

Several studies have been conducted to design algorithms for finding anti-powers and antiperiods in words. The algorithmic study of antipowers in words is initiated in [7] where an optimal algorithm is described for locating all factors of a word that are anti-powers of a specified order. A combinatorial lemma is used to show the optimality of the proposed algorithm. It is also shown that a word w contains $\theta(\frac{n^2}{k})$ distinct anti-powers of order k. Recently Kociumaka et al. [45] have shown an output-sensitive algorithm for the same problem with running time $O(nk\log k + c)$, where c is the number of reported antipowers. They also show that they can be counted within time $O(nk\log k)$. The paper [1] described algorithms for computing the smallest antiperiod and all the antiperiods of a word w. The algorithm presented takes $O(n\log n)$ time for computing all antiperiods of a given word. Further, an algorithm is described that computes the smallest antiperiod of the word in $O(n)$ time. These algorithms are offline algorithms, and the same problem is also studied in an online setting in [4]. Here, the algorithms use arrays that compactly and incrementally store anti-powers and antiperiods of words. The space and time requirements of these algorithms is $O(n\log n)$ where the size of the input word is n.

In [12], a more restrictive version of anti-power is defined for squares with the term 'antisquare'. An antisquare is a binary word having a form $u\bar{u}$ where $\bar{u}$ is the complement of u. Here, the authors studied the infinite binary words that do not contain arbitrarily large antisquares. They computed the repetition threshold of the binary sequences containing exactly two distinct antisquares. The concept of antisquares in finite words is relatively new, and it is unexplored in previous works. To address this research gap, Chapter 7 of this thesis investigates antisquares further.

2.4 Summary and Key Findings

The previous sections' studies are summarized through the figures and tables to visually represent the connections between the different works related to the research topic. The following table summarizes the improvements made to the upper and lower bounds of the runs conjecture. The term $R(n)$ represents the number of runs in an n length word. The first entry in the

Authors (Year)	$R(n)$
Kolpakov et al. (1999)	$= O(n)$
Rytter (2006)	$< 5n$
Puglisi et al.	$< 3.48n$
Crochemore et al. (2008)	$< 1.6n$
Giraud (2008)	$< 1.52n$
Franek et al. (2008)	$> 0.92n$
Matsubara et al (2008)	$> 0.944565n$
Simpson (2010)	$> 0.944575712n$
Bannai et al. (2017)	$< n$

Table 2.2: *Improvements in bounds of runs conjecture*

table pertains to the work that first proposed the conjecture. The subsequent four entries detail research on improving the upper bound of the conjecture. The remaining entries focus on efforts to improve the lower bound, with the exception of the final entry, which describes the paper that ultimately resolved the runs conjecture.

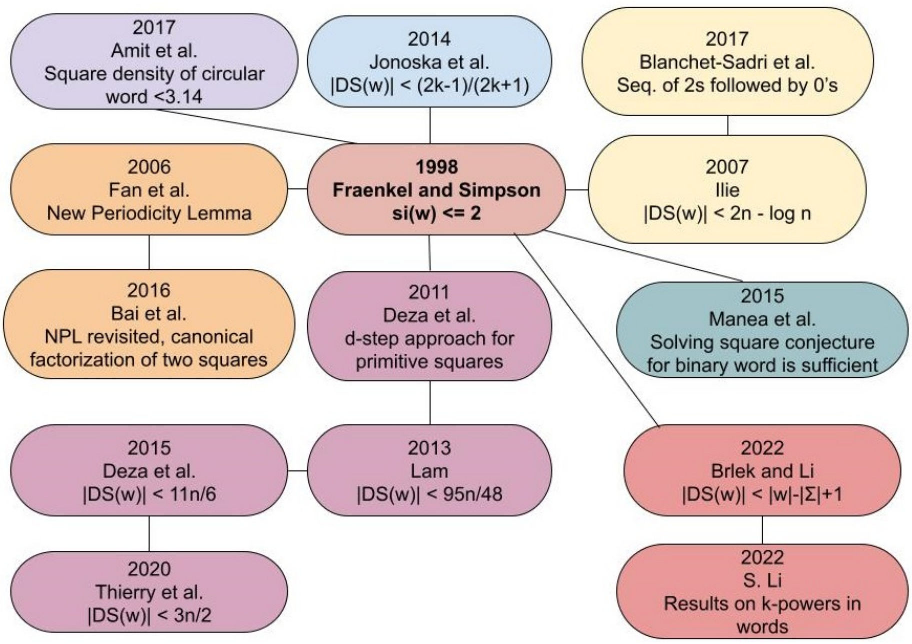

Figure 2.2: *Overview of the research on distinct squares*

Figure 2.2 summarizes the works related to squares. The figure represents each paper as a node with three associated fields: the publication year, author name(s), and a brief work description. The lines connecting the nodes represent the relationships between the papers. The term **square density** used in the figure refers to the ratio $\frac{|DS(w)|}{n}$ where n is the length of word w.

As illustrated in the figure, in a work published in 1998, the value of $\rho(w) = \frac{|DS(w)|}{n}$ was predicted to be one along with a proof for $s_i \leq 2$ [34]. Later results on a number of FS-double squares in w have shown further improvement on $\rho(w)$, where its upper bound is shown to be at most $\frac{11}{6}$ [29]. Moreover, several studies on repetitions have consistently anticipated that $\rho(w)$ is one [10, 31] and proving the square conjecture for binary words is sufficient [54]. It was also predicted that a '2' in any s_i sequence leads to at least two '0' in the same sequence [14]. Further, our observation of the s_i sequences has indicated that the presence of a 2 implies the presence of many $0'$s and the number of $2'$s are definitely less than $\frac{5}{6}^{\text{th}}$ times of the word length. Also, we noticed that the number of FS-double squares may be affected when a word is reversed.

Based on the considerations outlined above, this research addressed the problem of counting distinct squares in a word by focusing on the following research objectives.

1. Investigate the count of $2'$s and $0'$s in the s_i sequence and establish the relationship between them.

2. Identify words that, when reversed, do not reduce the number of FS-double squares.

3. Explore the relationship between square density and repetitions with higher exponents to justify why primitive squares maximize square density.

4. Improve the lower bound for the square conjecture.

3
FS-double Squares

A square is the smallest form of repetition with a structure uu. A trivial word a^n contains at most $\frac{n^2-(n \mod 2)}{4}$ squares. Note that this count includes all the occurrences of each repeating square. The relationship between the length of a word and the total number of squares in it is quadratic. In the last chapter, we discussed the notation s_i and the result $s_i(w) \leq 2$. This shows that the number of distinct squares in a word is a linear function of its length. A square $w = u^2$ with $s_i(w) = 2$ is an FS-double square. The result on $s_i(w)$ was presented by A. Fraenkel and J. Simpson, from whose initials the abbreviation 'FS' is derived. A binary word $w = abaababaab$ is the shortest FS-double square such that the first location of the word begins with two squares $(aba)^2, (abaab)^2$.

This chapter addresses the first research gap identified in the last chapter. We explore the structure of FS-double squares and identify the ways to pack as many FS-double squares as possible into a word. We see some properties of FS-double squares leading to certain structures of words and use these structures in the subsequent chapters to generate words containing many distinct squares. The next section presents the required notation.

3.1 Notation

A primitive word cannot be expressed as a repetition of form u^k for some non-empty u and an integer $k > 1$. A square u^2 with a primitive root u is called a primitive square in [49] and a primitively rooted square in [28]. The former notation is followed throughout this thesis. Similarly, a non-primitive square refers to a square whose root is non-primitive. The next property of primitive words is important to understand the relation between primitive squares and their conjugates. The property is also called the 'synchronization principle' in [10].

Lemma 3.1. *All the conjugates of a primitive word are distinct.*

For a binary word $w = abaa$, its conjugate is any word from the set $\{baaa, aaab, aaba\}$. square u^2 whose final appearance in the word starts at location i is denoted as u_i^2. For a location i in w, the number of rightmost distinct squares that begins at i is given by $s_i(w)$. If $s_i(w) = 2$, the location begins with an FS-double square. The following lemma describes the structure of an FS-double square.

Lemma 3.2 ([29])**.** *The roots, sq_i and SQ_i, of an FS-double square starting at location i have the following structure:*

$$sq_i = (xy)^{p+q}(x) \qquad\qquad SQ_i = (xy)^{p+q}(x)(xy)^p \tag{3.1}$$

where p and q are integers such that $p \geq 1, q \geq 0$, and the words $x, y \neq \epsilon$ and xy is a primitive word.

For clarity in notation, we will use subscripts with the roots of the squares in an FS-double square to indicate the location at which it begins. Further, the notations sq_i^2 and SQ_i^2 are used to distinguish the shorter and longer squares, respectively.

In [29], the concept of a "mate" for an FS-double square was introduced. The classification of these mates based on the lengths and locations of FS-double squares in the word is described below.

Given two FS-double squares SQ_1^2 and SQ_k^2 in a word where $k > 1$, Deza et al. [29] categorized SQ_k^2 into five types based on the value of k and the sizes of roots sq_1, SQ_1, sq_k, SQ_k. Let $SQ_1^2 = (xy)^{p+q}x(xy)^p$ and for $k < (p + q - 1)|xy| + |lcp(xy, yx)|$, SQ_k^2 is considered one of the following mates of SQ_1^2 if it meets the required conditions. Here, the term $lcp(x, y)$ is the longest common prefix of two non-empty words x, y. It is ϵ if x, y starts with different letters. In case of $x = y$, $lcp(x, y) = x$. A conjugate of a word w is $\tilde{w}$, and we use this notation to define various mates in the following definitions.

(a) **$\boldsymbol{\alpha}$-mate**: The root SQ_k is a conjugate of SQ_1 which gives $|SQ_1| = |SQ_k|$. The condition $|sq_1| = |sq_k|$ is further added in [65]. The FS-double squares starting at the first two locations of the below word are α-mates where $SQ_1^2 = ((aab)^2a(aab))^2$ and $SQ_2^2 = ((aba)^2(a)(aba))^2$.

$$w = a\,\underbrace{abaabaaabaabaabaaab \cdot a}_{SQ_2^2}$$

(b) **$\boldsymbol{\beta}$-mate**: These are the FS-double squares starting at any location k where $1 < k \leq |xy|$ which satisfy the relation $|SQ_1| = |SQ_k|$ with the root $sq_k = (\widetilde{xy})^i\widetilde{x}$ for some integer $i \in [2, p + q - 1]$. The words $\widetilde{xy}$ and $\widetilde{x}$ are conjugates of xy and x, respectively. It implies that $|sq_1| > |sq_k|$. The following word has γ-mates where the FS-double squares starts at locations 1 and 4 in w.

$$w = aab\,\underbrace{aabaabaaabaabaabaabaaab \cdot aab}_{SQ_4^2}$$

(c) **$\boldsymbol{\gamma}$-mate**: Here, $k < p + q|xy|$ and $|sq_k| = |SQ_1|$. In the below word, the first FS-double square starts at location 1 and ends just before the symbol '·'. The FS-double SQ_3^2 is its $\gamma-$mate.

$$w = ab\,\underbrace{aabaabaabaaabaabaabaabaabaabaaabaabaa \cdot baaabaabaabaabaabaaba}_{SQ_3^2}$$

(d) **δ-mate**: The lengths of the roots satisfy $|sq_k| > |SQ_1|$. The words with δ-mates are longer in size compare to the mates discussed above. An example of such a word is described before Lemma 3.16.

Three other types of mates are named ϵ, ζ, and η-mates mentioned in [65]. The FS-double squares in these types do not start at consecutive locations. The specifics of these mates are not necessary for the work discussed in this chapter and, thus, are excluded.

3.2 Smallest 2FS Squares

This section analyses words with consecutive locations starting with FS-double squares. We start by examining the structure of a word in which the first two locations start with FS-double squares. Such a word is referred to as a 2FS square.

Definition 3.1 (2FS square). *A word w is called a 2FS square if $s_1(w) = s_2(w) = 2$ and the FS-double square starting at location 2 is the suffix of w.*

The properties of 2FS squares are better understood by studying FS-double squares and primitive squares, as presented in the following results.

Lemma 3.3 (Two Squares Factorization Lemma [10]). *Let an FS-double square, SQ_i^2, begins with a shorter square, sq_i^2, such that $sq_i = (xy)^{p+q}x$ and $SQ_i = sq_i(xy)^p$ where $x, y \in \Sigma^+$ and the integers p, q satisfy $p \geq 1, q \geq 0$. Then, the following two statements hold.*

(a) SQ_i is primitive, and

(b) sq_i is primitive if $p + q > 1$.

Next, a lemma describes a method for generating distinct squares of equal lengths where these squares begin at consecutive locations. It is derived from the well-known Periodicity Lemma [33] and applies to primitive squares.

Lemma 3.4 ([22])**.** *Let u^2 be a primitive square. Then, appending u^2 by a proper prefix v of the root u introduces $|v|$ conjugates of u^2.*

The conjugates that are introduced in the above lemma are distinct. Also, each of these conjugates is a square. Similarly, it is possible to extend a non-primitive square to introduce new squares. However, the newly introduced squares may not always be distinct in such a case. The squares in non-primitive words are explored in detail in chapter 5.

Lemma 3.5. *Let u^2 and v^2 be equal-length squares beginning at consecutive locations in a word. Then, $u = \widetilde{v}$ and $uu = \widetilde{v}\widetilde{v}$.*

In section 3.1, we stated different mates that are given in [29]. Accordingly, an α-mate of an FS-double square SQ_1^2 refers to an FS-double square SQ_k^2 where $k > 1$ and $|sq_1| = |sq_k|, |SQ_1| = |SQ_k|$. The next subsection deals with the case where $k = 2$ and the lengths of the longer roots SQ_1, SQ_k are the same.

Equal 2FS squares

Recall a 2FS square is a word that starts with two consecutive FS-double squares and ends with the FS-double square that starts at the second location. Consider a 2FS square in which $s_1(w) = s_2(w) = 2$ with (sq_1, SQ_1) and (sq_2, SQ_2) being the two respective pairs of the roots of FS-double squares. The 2FS square is an equal 2FS square if it satisfies the relation $|sq_1| = |sq_2|$ and $|SQ_1| = |SQ_2|$. Otherwise, we call it an unequal 2FS square. According to the structure of an FS-double square provided in Lemma 3.2, any 2FS square satisfies the relation $|sq_1| < |SQ_1|$ and $|sq_2| < |SQ_2|$. In this section, we further compare the lengths of the roots sq_1, sq_2, SQ_1 and SQ_2 in detail.

Now, assume that the FS-double squares SQ_1^2 and SQ_2^2 have the following structures for the given 2FS square,

$$\begin{aligned} SQ_1 = &\ (xy)^{p+q}(x)(xy)^p \\ SQ_2 = &\ (uv)^{p'+q'}(u)(uv)^{p'} \end{aligned}$$

In the rest of the chapter, we assume that SQ_1 begins with a letter 'a' such that $x = ax'$.

Lemma 3.6. *Let w be a 2FS square with $|SQ_1| = |SQ_2|$. Then, $|sq_1| = |sq_2|$.*

Proof. Assume SQ_1^2 begins with a letter 'a', so $x = ax'$ for some word x'. Since xy is a primitive word, its conjugate $x'ya$ is also primitive. See the structure of SQ_2^2 shown below.

$$SQ_2^2 = (x'ya)^{p+q}(x'a)(x'ya)^p \cdot (x'ya)^{p+q}(x'a)(x'ya)^q \tag{3.2}$$

If $|sq_2| < |sq_1|$, then $|sq_1| \neq (p+q)|x'ya|$. Consider the next equation, with LHS representing the first appearance of sq_2 and RHS representing the last appearance of the root sq_2 in sq_2^2.

$$sq_2 : (x'ya)^{p+q} = x'a(x'ya)^p(x'ya)^{q-1}u \tag{3.3}$$

In Equation (3.3), u is a word over Σ. If it is an empty word, the prefixes in (3.3) violate Lemma 3.1 unless $x'ya = x'a$, but this means $|y| = 0$, which is not allowed. For the case where u is a proper prefix of $x'ya$, equating the suffixes of length $|x'ya|$ shows that two conjugates of a primitive word are equal. This again violates Lemma 3.1. The next relation can be derived under the constraint $|sq_2| < (p+q)|x'ya|$ assuming $0 < k < (p+q)$ and $u_1, u_2, u_3 \in \Sigma^*$.

$$(x'ya)^{p+q-k}u_1 = u_2(x'ya)^{k-1}(x'a)(x'ya)^r u_3 \tag{3.4}$$

To comply with Lemma 3.1, the relation $u_1 = u_3$ must hold and $u_2 = x'ya$ if $k > 1$. For the latter condition, we get a relation $x'ya = x'a$ which is contradictory as y is a non-empty word. Thus, $k = 1$ must satisfy Equation (3.4), which yields the following relation.

$$(x'ya)^{p_1-1}\cancel{u_1} = s_2x'a(x'ya)^r\cancel{u_3} \tag{3.5}$$

Here, the words u_1, u_2 and u_3 are non-empty words such that $u_1u_2 = x'ya$ (see the structure of SQ_2^2 in Equation (3.2)). The words u_1, u_3 need to be the same to avoid overlapping of conjugates of $x'ya$. While a non-empty u_1 violates Lemma 3.1, the relation $|u_1| = 0$ again implies $|y| = 0$ contradicting the given assumption. Hence, $|sq_2| \geq |sq_1|$.

Figure 3.1, depicts the structure of SQ_2^2 when $|sq_2| > |sq_1|$. A dotted line represents the last occurrence of sq_2 in sq_2^2. The root sq_2, shown as a thick line in the figure, must have a suffix $x'ya$ to meet the requirements of Lemma 3.1.

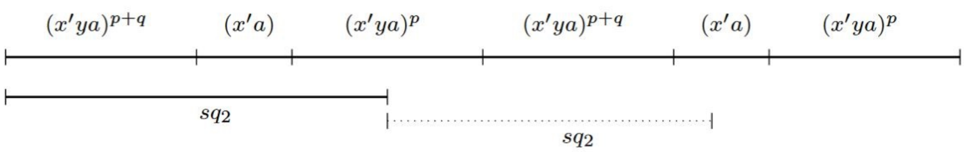

Figure 3.1: *FS-double square SQ_2^2 when $|sq_2| > |sq_1|$*

The first occurrence of the root sq_2 in sq_2^2 ends with $(x'ya)^{p_1}$ for some integer t_1 such that $1 \leq t_1 < p$. The next equations give the possible structures of the root sq_2 where the LHS and the RHS of each equation indicate the first and the last occurrence, respectively, of sq_2. Assume $r < 2p + q - t_1$ and $t_2 < p$.

$$(x'ya)^{p+q}(x'a)(x'ya)^{t_1} = (x'ya)^{p+q+r} \tag{3.6}$$

$$(x'ya)^{p+q}(x'a)(x'ya)^{t_1} = (x'ya)^{2p+q-p_1}(x'a)(x'ya)^{t_2} \tag{3.7}$$

The second occurrence of sq_2 marked with dotted lines in Figure 3.1 cannot end somewhere in the factor $x'a$. This is because the two occurrences of sq_2 end with different conjugates of $x'ya$. To avoid the overlap of $x'ya$ with any of its conjugate, Equations (3.6) and (3.7) must satisfy the relation $x'a = x'ya$ implying $|y| = 0$. This in unacceptable; thus the FS-double square SQ_2^2 never satisfies the relation $|sq_2| > |sq_1|$. Therefore, the only possibility is that $|sq_1| = |sq_2|$. □

Lemma 3.7. *Let w be a 2FS square where $|SQ_1| = |SQ_2|$. Then, SQ_1^2 and SQ_2^2 are α-mates.*

Proof. Since $|SQ_1| = |SQ_2|$, so SQ_1^2 and SQ_2^2 can be α or β-mates. A β-mate is possible in case $|sq_1| > |sq_2|$. From Lemma 3.6, the roots of the shorter squares only satisfy the relation $|sq_1| = |sq_2|$. Further, Lemma 3.5 gives SQ_1^2 and SQ_2^2 are α-mates. □

The following theorem is a consequence of the above lemma, which states that if the lengths of the longer roots of two consecutive FS-double squares are equal, then the lengths of their shorter roots must also be equal.

Theorem 3.8 (Equal 2FS Squares)**.** *In a 2FS square, $|SQ_1| = |SQ_2|$ if and only if $|sq_1| = |sq_2|$.*

Proof. This follows from Lemma 3.6 and 3.7. □

Corollary 3.9. *Let w be a 2FS square. Then, SQ_1^2 and SQ_2^2 are conjugates if and only if sq_1^2 and sq_2^2 are conjugates.*

Proof. The statement follows from Lemma 3.5 and Theorem 3.8. □

We now explore the condition which needed to have an equal 2FS square. Suppose w is an FS-double square with sq_1^2 and SQ_1^2. It is possible to extend w to obtain a 2FS square. According to Lemmas 3.5 and 3.7, the shorter squares among the two consecutive FS-double squares in an equal 2FS square are conjugates. Therefore, the word w must have a square at the second location, which is a conjugate of sq_1^2.

$$\begin{aligned} SQ_1^2 &= (ax'y)^{p+q} \cdot (ax') \cdot (ax'y)^p \cdot (ax'y)^{p+q} \cdot (ax') \cdot (ax'y)^p \\ &= \underline{(ax'y)^{p+q} \cdot (ax') \cdot (ax'y)^{p+q} \cdot (ax'y})^p \cdot (ax') \cdot (ax'y)^p \\ &= a \cdot \underbrace{(x'ya)^{p+q} \cdot (x'a) \cdot (x'ya)^{p+q} \cdot x'(y}_{sq_2^2 \implies y \text{ begins with } a} ax')^p \cdot (ax'y)^p \end{aligned} \tag{3.8}$$

The result of Theorem 3.8 can also be deduced from Lemma 2 given in [43] (see Lemma 3.10 below). An obvious question would be why go into the details of the structures of consecutive FS-double squares while the said lemma rules out the possibility of starting a β-mate adjacent to a location starting with the FS-double square. The aim is to highlight the relationship between consecutive FS-double squares and the words x, y shown in Equation (3.8).

Lemma 3.10 ([43])**.** *Assume a word w begins with two rightmost squares SQ_1^2 and sq_1^2 such that $|SQ_1| > |sq_1|$. If u^2 is the rightmost square beginning from the second location of w, then $|u| \in \{|sq_1|, |SQ_2|\}$ or $|u| \geq 2|SQ_1|$.*

From Equation (3.8), it can be observed that $lcp(x, y)$ must be non-empty to have equal 2FS squares. The next lemma determines the highest number of conjugate FS-double squares that start at successive locations.

Lemma 3.11. *Let a word w that begins with i consecutive FS-double squares such that $|SQ_1| = |SQ_2| = \cdots = |SQ_i|$ where $SQ_1 = (xy)^{p+q}x(xy)^p$ and some integers i, p, q satisfy $p \geq 1, i \geq 1, q \geq 0$. Here, $x, y \in \Sigma^+$. Then, the next relation holds.*

$$i \leq \begin{cases} |lcp(xy, yx)| + 1 & \text{if } q > 0 \\ min(|lcp(xy, yx)| + 1, |x|) & \text{otherwise} \end{cases}$$

Proof. Since consecutive FS-double squares are of equal lengths, they are conjugates (see Lemma 3.5). Similarly, the shorter squares in these FS-double squares are also conjugates. From Lemma 3.4, conjugates of sq_1^2 at consecutive locations are possible if sq_1^2 is extended with one of its proper prefixes.

$$\begin{aligned} SQ_1^2 =& (xy)^{p+q}(x)(xy)^p \cdot (xy)^{p+q}(x)(xy)^p \\ =& \underbrace{(\underline{xy})^{p+q}(x)(xy)^{p+q}x}_{sq_1^2} \cdot (\underline{yx})^p(xy)^p \end{aligned}$$

The value of i depends on the longest common prefix of the underlined words in the above structure. Since $xy \neq yx$, the number of conjugates, in this case, is $|lcp(xy, yx)|$. Given any nonempty word u, we can get at most $|u| - 1$ conjugates of a square uu by appending the square by its prefix such that these conjugates start at consecutive locations. So, there can be $|sq_1| - 1$ conjugates that are possible for sq_1^2 and $|lcp(xy, yx)| < |sq_1| - 1$. It shows that the total number of FS-double squares is $|lcp(xy, yx)| + 1$. We know that $q \geq 0$ (see Lemma 3.2) and this value of i holds for $q > 0$.

It is also necessary to preserve earlier FS-double squares while extending the larger square SQ_1^2. For $q = 0$, the square sq_1^2 repeats if SQ_1^2 is extended with one of its prefixes of size $|x|$ or more. For example, sq_1^2 is a suffix of the word $SQ_1^2.(x)$ as shown below.

$$\begin{aligned} SQ_1^2 =& (xy)^p(x)(xy)^p \cdot (xy)^p(x)(xy)^p \\ SQ_1^2.(x) =& (xy)^p(x)(xy)^p \cdot \boxed{(xy)^p(x)(xy)^p(x)} \implies sq_1^2 \end{aligned}$$

In a word $SQ_1^2.x$, the first location starts with only one rightmost square, contradicting the assumption that SQ_1^2 is an FS-double square. So, $i \leq min(|lcp(xy, yx)| + 1, |x|)$. $\square$

3.3 Unequal 2FS Squares

Given a word starting with an FS-double square, the FS-double square that begins at the adjacent location can be one of the four mates described in section 3.1. After verifying the possibility of α and β-mates in the earlier section, we now explore the possibility of the second FS-double square in a 2FS square being a γ or a δ-mate.

Lemma 3.12. *Given a 2FS square where $|SQ_1| \neq |SQ_2|$. Then, SQ_1^2 and SQ_2^2 cannot be γ-mates. Moreover, these squares are δ-mates.*

Proof. Let w be a word that begins with two consecutive FS-double squares SQ_1^2 and SQ_2^2 such that SQ_2^2 is a suffix of w. Assume sq_1^2 is the shortest rightmost square starting at location 1 and sq_2^2 is the shortest rightmost square starting at location 2. According to the definitions in section 3.1, SQ_2^2 is either a γ-mate or a δ-mate of SQ_1^2. According to Lemma 3.10, the size of a square starting at the second location is equal to either $|sq_1^2|$, $|SQ_1^2|$ or $2|SQ_1^2|$. If SQ_2^2 is a γ-mate of SQ_1^2, then $|sq_2| = |SQ_1|$. Further, the relation $|SQ_2| \geq 2|SQ_1|$ holds. According to the definition described in Lemma 3.2, we have $|SQ_2| < 2 * |sq_1|$. Here, $|SQ_2| = 2|SQ_1| = 2|sq_2|$ contradicts the condition $|SQ_2| < 2 * |sq_1|$. Thus, SQ_1^2 and SQ_2^2 cannot be γ-mates.

Consider the word $w = a((abaaabaabaaabb)(ab)(abaaabaabaaabb))^2$ to see that δ-mates can begin at adjacent positions. □

Based on the results obtained in Theorem 3.8, Lemma 3.10 and 3.12, the following theorem summarizes the types of 2FS squares.

Theorem 3.13 (2FS Square)**.** *A 2FS square belongs to one of the following types:*

(a) Equal 2FS square with $|sq_1| = |sq_2|$ and $|SQ_1| = |SQ_2|$, or

(b) Unequal 2FS square with $|sq_1| < |SQ_1|$ and $2|SQ_1| \leq |sq_2| < |SQ_2|$.

Proof. The proof of part (a) is obtained from Lemma 3.10. The lemma deals with the possibilities where either of the pairs $(|sq_1|, |sq_2|)$ or $(|SQ_1|, |SQ_2|)$ are equal. The rest of the possible cases are verified in Lemma 3.12, which proves part (b) of the theorem. □

Corollary 3.14. *The two FS-double squares starting at adjacent locations are α or δ-mates.*

While there is a unique equal 2FS square for a given FS-double square, it turns out that the words with δ-mates have different structures. In other words, the FS-double SQ_2^2 is not unique for SQ_1^2 when SQ_1^2 and SQ_2^2 are δ-mates. The results of Lemma 3.10 and 3.12 lead to the subsequent lemma.

Lemma 3.15. *The following statements hold for an unequal 2FS square.*

(a) $|sq_2| \geq 2|SQ_1|$*, and*

(b) $|SQ_2| > 2|SQ_1|$.

A proof has been presented for a word that begins with consecutive FS-double squares, demonstrating that any two consecutive squares have the structure of either an equal or an unequal 2FS square. The s_i sequence of such a word has a chain of $2'$s in the beginning. A word w has a sequence of $2'$s if $s_i(w) = s_{i+1}(w) = \cdots = s_j(w) = 2$ where the integers i and j satisfy $1 \leq i < j < |w|$. It is possible to extend an FS-double square to get an arbitrarily long sequence of $2'$s. One way to achieve this is described in Lemma 3.11, where an FS-double square is appended by its prefix. In this case, all the consecutive FS-double squares at the beginning of a word are conjugates. The number of such FS-double squares is finite, and the length of a sequence of $2'$s is limited. However, it is always possible to introduce an unequal 2FS square to increase the length of the sequence of $2'$s. Thus, we can extend an FS-double square to get a sequence of $2'$s of any desired length by introducing a new equal or an unequal 2FS square.

Under specific conditions, a single letter is added to the FS-double square to introduce a new equal 2FS square. In contrast, the FS-double square is appended by many letters to get a new unequal 2FS square. Let us see some equal and unequal 2FS squares. Next is an example of an equal 2FS square and its s_i sequence. Here, $SQ_1^2 = ((aba)^1(ab)(aba)^1)^2$ and $SQ_2^2 = ((baa)^1(ba)(baa)^1)^2$.

$$\begin{aligned} w \ &= a\,b\,a\,a\,b\,a\,b\,a\,a\,b\,a\,a\,b\,a\,b\,a\,a \\ s_i(w) &= 2\,2\,0\,0\,0\,0\,1\,1\,1\,0\,0\,1\,1\,0\,0\,1\,0 \end{aligned}$$

The word, w, if continued to be extended further with the prefix of SQ_1^2, then sq_1^2 repeats after the first location. It reduces the value of s_1 to one. In such words, it is necessary to introduce an unequal 2FS square to extend the sequence of 2′s further. Unlike equal 2FS squares, the length of an unequal 2FS square varies. There are different ways to extend an FS-double square to get an unequal 2FS square. To elaborate on this further, an FS-double square is extended in two different ways to get two different unequal 2FS squares. Let $SQ_1^2 = aabaaabaabaaab$ be the FS-double square which is extended to get two unequal 2FS squares w_1 and w_2, where

$$w_1 = a((abaaabaabaaabb)(ab)(abaaabaabaaabb))^2 \tag{3.9}$$

$$w_2 = a((abaaabaabaaabb)(abaaabaabaaabbab)(abaaabaabaaabb))^2 \tag{3.10}$$

and the respective s_i sequences are,

$$\begin{aligned} w_1 \ &= \mathrm{a\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b} \\ &\quad\ \mathrm{a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b} \\ s_i(w_1) &= 2\,2\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ &\quad\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,0\,0\,1\,1\,1\,0\,0\,0\,0\,1\,0\,1\,0 \end{aligned}$$

$$\begin{aligned} w_2 \ &= \mathrm{a\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b} \\ &\quad\ \mathrm{a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,a\,b\,a\,a\,a\,b\,b} \\ s_i(w_2) &= 2\,2\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1 \\ &\quad\ 1\,1\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,0\,0\,1\,1\,1\,0\,0\,0\,0\,1\,0\,1\,0 \end{aligned}$$

A word can be extended to get an unequal 2FS square at any location. Moreover, it is possible to yield an unequal 2FS square at a particular location l such that it does not affect the s_i value of another location m where $0 < m < l$. This new 2FS square almost doubles the overall word length, though. So, the relationship between the length of the longest sequence of 2′s and the word length is investigated. It is evident that the ratio of the longest sequence of $2'$s in a word to its length is higher for equal 2FS squares. The following lemma gives the ratio for the sequence of $2'$s such that any two consecutive FS-double squares in the sequence have the structure of an equal 2FS square.

Lemma 3.16 (Longest sequence of $2'$s with Equal 2FS Squares). *Let T be the longest sequence of consecutive FS-double squares in w such that any two consecutive FS-double squares in T are conjugates. Then, $\frac{|T|}{|w|} \leq \frac{1}{7}$.*

Proof. Assume that the first FS-double square SQ_1^2 in T is $((xy)^{p+q}(x)(xy)^p)^2$ where $x, y \in \Sigma^+$ and integers p, q satisfy the relation $p+q \geq 1, q \geq 0$. From Lemma 3.11, we know that the length of T depends on the values of p and q. The highest value of the ratio $|T|/|w|$ when $q = 0$ is computed below.

$$\begin{aligned} \frac{|T|}{|w|} &= \frac{\min(|lcp(xy, yx)| + 1, (|x| - 1))}{2((p+p+1)|x| + (p+p)|y|) + |lcp(xy, yx)|} \\ &= \frac{|x| - 1}{(4p+3)|x| + 4p|y| - 1} \leq \frac{1}{7} \end{aligned}$$

The following equation shows that the ratio $|T|/|w|$ approaches $\frac{1}{7}$ when $q > 0$ and x reaches ∞.

$$\begin{aligned} \frac{|T|}{|w|} &= \frac{|lcp(xy, yx)| + 1}{2((2+1+1)|x| + (2+1)|y|) + |lcp(xy, y, x)|} \\ &= \frac{|x| + |y| - 2}{8|x| + 6|y|) + |x| + |y| - 2} = \frac{|x| + |y| - 2}{9|x| + 7|y|) - 2} \leq \frac{1}{7} \end{aligned}$$

□

The bound for $|T|$ is also computed in Lemma 4 of the paper [43] where it is shown that $|T| < \frac{|w|}{2}$. However, the value $|lcp(xy, yx)|$ can be used to show that $\frac{|T|}{|w|}$ is at most $\frac{1}{7}$. The sequence of consecutive FS-double squares obtained in Lemma 3.16 can be further extended by adding a new unequal 2FS square. So, another way to generate a long sequence of 2's is to start with an FS-double square and extend it to add all possible conjugates of the square. At this point, we can append the word to generate an unequal

2FS square so that the sequence of 2's continues to grow. Thus, a sequence increases either with an equal or an unequal 2FS square. The length of such a sequence in a word with respect to the word length is computed in the following lemma.

Lemma 3.17 (Longest Sequence of $2'$s with Equal and Unequal 2FS squares)**.** *Let T be the longest sequence of FS-double squares in a word w that contains at least one equal and at least one unequal 2FS square. Then,* $\frac{|T|}{|w|} \leq \frac{6}{55}$.

Proof. The length of T can be increased by adding a new 2FS square. Every new equal 2FS square increments the value of both $|T|$ and $|w|$ by 1. This improves the value of $\frac{|T|}{|w|}$. However, it is not always possible to introduce an equal 2FS square (see Lemma 3.11) and, therefore, an unequal 2FS square is required to get a longer T. Unlike an equal 2FS square, a new unequal 2FS square decreases the value of $\frac{|T|}{|w|}$. To understand this, suppose an unequal 2FS square begins at location one where SQ_1^2 and SQ_2^2 are two consecutive FS-double squares with shorter squares sq_1^2 and sq_2^2, respectively. Lemma 3.15 gives the relation $|SQ_2| > 2|SQ_1|$. Accordingly, SQ_1^2 is appended by a word containing at least $2|SQ_1|$ letters to make $s_2 = 2$. Thus, the value of $\frac{|T|}{|w|}$ decreases significantly after introducing an unequal 2FS square. So, we can obtain the best ratio from the word that has a maximum number of equal 2FS squares and some unequal 2FS squares.

Given a word with $s_1 = s_2 = \cdots = s_i = 2$ such that the location i starts with an FS-double square SQ_i^2. From Lemma 3.16, the sequence of $2'$s can be extended to get at most $\frac{|SQ_i|}{7}$ new equal 2FS squares. An unequal 2FS square must be introduced to continue the sequence of $2'$s further. The ratio $\frac{|T|}{|w|}$ for the smallest FS-double square $w = (abaab)^2$ is $\frac{1}{10}$. We can use the above method to extend w. The respective $\frac{|T|}{|w|}$ obtained after introducing a new unequal 2FS square results into the following sequence.

$$\frac{1+\frac{10}{7}+1}{10+\frac{10}{7}+(10+10)}, \frac{1+\frac{10}{7}+1+\frac{20}{7}+1}{10+20+\frac{10}{7}+\frac{20}{7}+(20+20)},$$

$$\frac{1+\frac{10}{7}+1+\frac{20}{7}+1+\frac{40}{7}+1}{10+20+40+\frac{10}{7}+\frac{20}{7}+\frac{40}{7}+(40+40)}, \cdots$$

The n^{th} term of the above sequence is

$$\frac{(n+1)+\frac{10}{7}\sum_{i=0}^{n-1} 2^i}{\frac{10}{7}\sum_{i=0}^{n-1} 2^i + 10\sum_{i=0}^{n-1} 2^i + 10 * 2^n}$$

This is a decreasing sequence where the first term is $\frac{6}{55}$ and it converges to $\frac{1}{15}$. □

Now, we will see the highest value of the ratio $\frac{|T|}{|w|}$ in the next theorem.

Theorem 3.18. *Let T be the longest sequence of $s_i = 2's$ in a word w. Then, $|T| \leq \frac{|w|}{7}$.*

Proof. The computation in Lemmas 3.16 and 3.17 shows that the best value of $\frac{|T|}{|w|}$ where T contains either equal length FS-double squares or a combination of equal and unequal 2FS squares. The best value is obtained in the former case, that is, $\frac{1}{7}$. We now compare this value with the sequence of T where every two consecutive FS-double squares follow the structure of an unequal 2FS square.

Suppose SQ_1^2 and SQ_2^2 result in an unequal 2FS square at the beginning of a word. Then, $|SQ_2| > 2|SQ_1|$ (see Lemma 3.15). Thus, to introduce a new unequal 2FS square at location i, it is required to append at least $2|SQ_i|$ letters to the FS-double square SQ_i^2. Allowing only unequal 2FS squares in T, we compute the ratio of the length of the longest sequence of $2'$s in a word to its word length as follows.

$$\frac{1}{10}, \frac{1+1}{10+20}, \frac{1+1+1}{10+20+40}, \cdots, \frac{n}{10*(2^n-1)}, \cdots$$

The value of the ratio decreases as n increases and the ratio has the maximum value of $\frac{1}{15}$ for $n = 2$. We ignore the value with $n = 1$ as the sequence will have only one FS-double square. Therefore, $|T| \leq \frac{|w|}{7}$. □

3.4 Conclusions

We investigated the ways to get a sequence of FS-double squares by extending a given FS-double square. In this regard, the term 2FS square is introduced. It is a word starting with two consecutive FS-double squares. A 2FS square is characterized by two types, viz. equal and unequal 2FS square. The former has a single letter added to an existing FS-double square to obtain a new FS-double square. In contrast, the FS-double square is appended by a word of its length (or longer) to yield an unequal 2FS square. Though getting a new equal-length FS-double square is easy, but only a finite number of such squares can be added. The results demonstrated that an FS-double square could be extended to generate up to $\frac{|w|}{7}$ equal 2FS squares. In fact, the same ratio is also obtained in [14] for the shortest possible words that start with equal-length consecutive FS-double squares.

On the contrary, it is shown that a square w could be extended to have any number of new FS-double squares at consecutive locations by introducing unequal 2FS squares. The overall length of the resulting word increases significantly with the inclusion of such squares. So, the maximum number of successive FS-double squares in a word is compared with its length. During the chapter, it is demonstrated that the ratio of the number of consecutive FS-double squares in an n length word is less than $\frac{6n}{55}$ in the presence of unequal 2FS squares. The best ratio, $\frac{n}{7}$, is possible only with equal 2FS squares.

4
2FS Squares

In this chapter, we examine the characteristics of words with a long sequence of positions starting with FS-double squares. The results in the subsequent sections show that the words with these positions also contain some positions that do not start with any rightmost square. These findings are then combined and applied to determine the square density of these words.

4.1 Motivation

The Q words described in [34] are the lower bound for the square conjecture. These words do not have any FS-double squares. We observed that the s_i sequences of words with a large number of distinct squares contain either no $2'$s or an equal number of $0'$s and $2'$s. Existing works also suggest that a sequence of $2'$s leads to a much longer sequence of $0'$s [14]. These observations motivate further exploration of FS-double squares. The number of distinct squares in a word is a sum of s_i value of each of its locations. So, if a 2 in an s_i sequence leads to $0'$s, it does not help maximize the number of distinct squares in a word beyond its length. The $0'$s in the s_i sequence are referred to as no gain locations while a 2 is associated with an FS-double square. We see the relationship between the number of FS-double squares and no-gain locations in the following sections.

The structures of words given in chapter 6 are the extension of Q words. These words reveal that words with a large number of distinct squares have conjugates of squares starting at consecutive locations. So, the words having consecutive FS-double squares are explored here, and the square density of such words is computed by counting no-gain locations within them.

4.2 Notation

For a word w, recall the definition of $s_i(w)$ given in section 1.1. The notation gives the number of rightmost squares starting at i in w. Denote by **2FS square** is a word w where $s_1(w) = s_2(w) = 2$ and w ends with the FS-double square that starts at its second location. For instance, see the following example of 2FS square.

$$w = aaaabaaaaabaaaabaaaaaaba$$

$$s_i(w) = 221000011111000001010100 0$$

If $s_i(w) = 0$, then location i in w is refer to no-gain location. Let $DS(w)$ be the set containing distinct squares in w. We can obtain the size of the set as $|DS(w)| = \sum_{i=0}^{|w|} s_i(w)$. The square density of w is the ratio $\frac{|DS(w)|}{|w|}$. It is denoted by $\rho(w)$.

It can be observed that in the example shown above the word has many no-gain locations. These are discussed in detail in the next section.

4.3 No-gain Locations in FS-double Squares

The square density of a word is determined by the number of locations with non-zero s_i values in the word. As mentioned earlier, a location that starts with no rightmost square is considered a no-gain location. If the number of distinct FS-double squares in a word is equal to the number of no-gain locations, then the square density of the word is at most one. This section explores words in which successive locations start with FS-double squares and identifies the no-gain locations within them. To achieve this, the structures of rightmost squares that start in the prefix of an FS-double square are studied. Recall the structure of an FS-double square, $((xy)^{p+q}(x)(xy)^p)^2$, where $p \geq 1, q \geq 0$ (see Lemma 3.2).

The last location of a word is always a no-gain location, and it is possible to find many such locations in a word by analyzing the structure of the rightmost squares. A trail of no-gain locations is often observed in some FS-double squares, particularly when $q = 0$, and the size of the trail typically increases with the value of p. This chapter explores all possible structures of the rightmost squares beginning in the highlighted part of an FS-double square $\boldsymbol{(xy)^p}x(xy)^{2p}x(xy)^p$, which we refer to as u^2 for convenience. Throughout the chapter, we use the notation $u(1)$ and $u(2)$ to indicate the first and last occurrence of the root u in the given u^2.

An FS-double square begins with two squares where sq_1 and SQ_1 are the roots of these squares. We are interested in a rightmost occurrence of a square, u^2, in an FS-double square. In this regard, the following claim is taken from Lemma 17 in [29].

Lemma 4.1 ([29])**.** *Let u^2 be the rightmost square that starts in the marked prefix S of an FS-double square shown in Fig. 4.1 and $sq_1 = (xy)^p(x)$. Then,* $|u| \leq |sq_1|$.

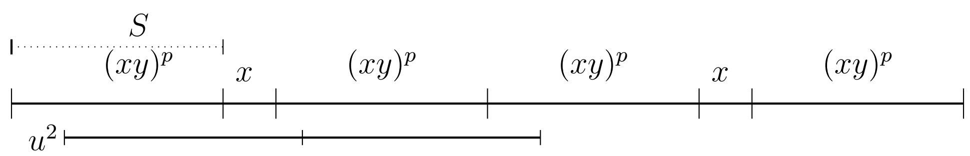

Figure 4.1: *u^2 in the prefix of an FS-double square*

Note that in the above lemma $|u| \not\geq |SQ_1|$, else the size of u^2 would exceed $|SQ_1^2|$. Consider $(xy)^p = (xy)^2(xy)^{p-2}$ where $p-2>0$, we divide the rightmost squares starting in the prefix $(xy)^p$ of $w = (xy)^p(x)(xy)^{2p}(x)(xy)^p$ into two categories:

(a) square starting in $2|(xy)|$ length prefix of $(xy)^p$, and

(b) squares starting in the suffix $(xy)^{p-2}$ of $(xy)^p$.

The aforesaid squares are discussed in Lemma 4.4. The squares of type (b) are explored in Lemma 4.2. Later theorem 4.3 shows that the s_i value of every location k in the given word w is zero where $k \in [2|xy|, (p-2)|xy|]$.

Lemma 4.2. *Let w be a word as shown in Equation 4.1. Assume the rightmost occurrence of u^2 starts somewhere in the marked location S and ends in T (see Equation (4.1)). Then, xy is a non-primitive word.*

$$w = (xy)^2 \underbrace{(xy)^{p-2}}_{S}(x)(xy)^p \underbrace{(xy)^p}_{T}(x)(xy)^p \tag{4.1}$$

Proof. *(By contradiction)* Let w has u^2 satisfy the given conditions and xy be a primitive word. With this assumption, w represents an FS-double square. We write $x = x_1x_2$ and $y = y_1y_2$ for some non-empty words x_1, x_2, y_1, y_2. According to Lemma 4.1, u^2 satisfies $|u| \leq (p|xy| + |x|)$, and, therefore, every u^2 that begins at some location in S as given in Equation (4.1) ends somewhere in T. The roots $u(1)$ and $u(2)$ of u^2 indicates the first, respectively, the last occurrence of u in u^2. The structures of these roots are given below.

$$u(1) = beg_1(xy)^j(x)(xy)^k end_1 \qquad u(2) = beg_2(xy)^l end_2 \tag{4.2}$$

In Equation (4.2), the exponents of (xy) satisfy $j, k, l > 0, l = k + j$ and $|u| < (p|xy| + |x|)$. The words beg_1, beg_2 are (possibly empty) prefixes of xy or are empty words. The words end_1 and end_2 are the suffixes of xy or can be empty words. The roots of u^2 show that xy is non-primitive if beg_1 and

beg_2 are different. For example, if $beg_1 = x_2y$ and $beg_2 = y_2$ then we get a relation contradicting Lemma 3.1 as follows.

$$\underline{x_2y}(xy)^j(x)(xy)^k end_1 = \underline{y_2(xy)}^l end_2 \implies x_2yx_1 = y_2xy_1$$

This results in the relation $beg_1 = beg_2$. By substituting this in Equation (4.2), we get $xy = yx$ as,

$$beg_1\underline{(xy)^j(x)(xy)}^k end_1 = beg_1\underline{(xy)}^l end_2 \tag{4.3}$$

$$\implies (xy)^j(x)xy\cdots = (xy)^j(xy)^{l-j}\cdots \tag{4.4}$$

Thus, u^2 starting somewhere in location S of Equation (4.1) implies that xy is a non-primitive word. For a word w to be an FS-double square, the factor xy must be a primitive word. Under this condition, it is not possible for the square u^2 to have its last occurrence as specified. □

The results obtained in Lemma 4.2 show that certain locations of an FS-double square do not start with any rightmost square. The number of such locations and the constraint on an FS-double square in such cases is described in the next theorem.

Theorem 4.3. *Let FS-double square* $w = (xy)^p(x)(xy)^{2p}(x)(xy)^p$ *such that* $p > 2$. *Then,* $s_t(w) = 0$ *where* $t \in [2|xy|, (p-2)|xy|]$. *That is,*

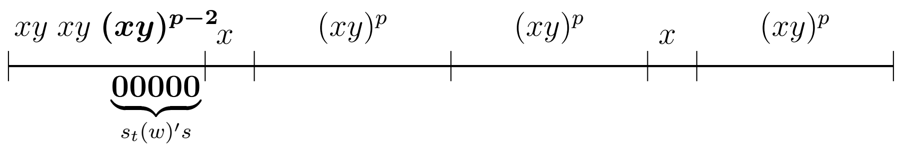

Figure 4.2: *Sequence of* $0'$*s in FS-double square*

Proof. Lemma 4.1 gives the precise lengths of the rightmost squares that begin at any location in the prefix $(xy)^p$ of the given words. This information is used in Lemma 4.2 to show that if any rightmost square starts at a location t, then it implies xy is non-primitive. According to the definition of FS-double square, xy is primitive. Thus, $s_t(w) = 0$. □

Let us now examine an example of a word that illustrates the theorem presented above. Consider an FS-double square $w = ((xy)^p x(xy)^p)^2$ where $x = abaa$, $y = aba$, and $p = 3$. As proved, the $s_i(w)$ has $(p-2)|xy|$ locations with zero s_i values. The trail of zeroes and the corresponding locations in w are marked below.

$$
\begin{aligned}
w =& abaaabaabaaaba\underline{abaaaba}abaaabaaabaabaaabaabaaabaabaaaba\ldots\\
s_i(w) =& 2111100000010\underline{0000000}00000011111110000000111111100001\ldots
\end{aligned}
$$

We can observe that the rightmost squares in the prefix of FS-double square have specific characteristics. The following lemma enumerates them in detail.

Lemma 4.4. *Let w be an FS-double square where $q = 0$. If u^2 is a square that starts at location i such that $s_i(w) = 1$ and $2 \le i \le 2|xy|$, then the following statements hold.*

(a) The number of such i's where $|u| = |sq_1|$ is $|lcp(xy, yx)|$.

(b) If $|u| < |sq_1|$, then the number such squares is less than $|xy|$.

Proof. (a) According to Lemma 4.1, if $s_i(w) = 1$ and u^2 starts at location i, then $|u| \le |sq_1|$. For the given w, the following two squares start at the beginning,

$$
\begin{aligned}
SQ_1^2 &= (xy)^p(x)(xy)^p(x).(yx)^p(xy)^p\\
sq_1^2 &= (xy)^p(x)(xy)^p(x)
\end{aligned}
$$

Let $x = x_1x_2$ and $y = y_1y_2$ for some non-empty words x_1, x_2, y_1, y_2. If $|u| = |sq_1|$ and $i \le |lcp(xy, yx)|$, then u^2 has one of the next structure.

$$
\begin{aligned}
u^2 =& x_2y_1y_2(xy)^{p-1}(x)(xy)^p(x)(x_1) && \text{or}\\
u^2 =& y_1y_2(xy)^{p-1}(x)(xy)^p(x)(x) && \text{or}\\
u^2 =& y_2(xy)^{p-1}(x)(xy)^p(x)(xy_1)
\end{aligned}
$$

From the above equation set, u^2 is a conjugate of sq_1^2, and the number of such squares is at most $|lcp(xy, yx)|$. When u^2 starts after $|lcp(xy, yx)|$ locations, it has either of the following structures.

$$(xy)^{p-k}(x)(xy)^k = (xy)^p(x)$$
$$(x_2y)(xy)^{p-k}(x)(xy)^{k-1}(x_1) = (x_2y)(xy)^{p-1}(x_1)end_1$$
$$(y)(xy)^{p-k}(x)(xy)^{k-1}(x) = (y)(xy)^{p-1}(x)end_2$$
$$(y_2)(xy)^{p-k}(x)(xy)^{k-1}(xy_1) = (y_2)(xy)^{p-1}(xy_1)end_3$$

In these words, k is an integer such that $1 < k < p$ and end_1 is $|x|$ length prefix of x_2yx_1. Similarly, words end_2 and end_3 are prefixes of yx and y_2xy_1, respectively where $|end_2| = |end_3| = |x|$. The suffixes of roots of u^2 in all of the above words give the relation $xy = yx$, and it contradicts Lemma 3.1. So, the valid u^2 where $|u| = |sq_1|$ is conjugate of sq_1^2.

(*b*) We now verify all possible u^2 where $|u| < |sq_1|$. Since u^2 is the rightmost square, it must end after the first occurrence of sq_1. So, the first occurrence of u starts in the prefix xy and ends somewhere in the highlighted part of Equation (4.5).

$$\begin{aligned} SQ_1^2 &= (xy)^p(x)(xy)^p(xy)^p(x)(xy)^p \\ &= (xy)(xy)^{p-1}(x)\boldsymbol{(xy)}(xy)^{p-1}(xy)^p(x)(xy)^p \end{aligned} \tag{4.5}$$

Assuming $x = x_1x_2 = x_3x_4$ and $y = y_1y_2 = y_3y_4$ where $x_i, y_i \in \Sigma^+, i \in [1, 4]$ and $x_1 \neq x_3, y_1 \neq y_3$. The two occurrences of u in u^2 always start with two different prefixes. A relation found in equating these prefixes implies that two different conjugates of xy are equal. For instance, if the first occurrence of u starts with x_2yx_1, then the second occurrence starts with x_3yx_4. It shows that $x_2yx_1 = x_3yx_4$, which is a contradiction. Therefore, u^2 cannot start anywhere in the $|xy|$ length prefix of the given FS-double square. The roots of valid u^2 starting after $|xy|$ locations are shown below, where S is a prefix of yx.

$$
\begin{aligned}
x_2y(xy)^{p-2}(x)x_1 &= x_2y(xy)^{p-2}(x)y_1 && |y_1| = |x_1| \text{ or}\\
y(xy)^{p-2}(x)x &= y(xy)^{p-2}(x)S && |S| = |x| \text{ or}\\
y_2(xy)^{p-2}(x)xy_1 &= y_2(xy)^{p-2}(x)S && |S| = |xy_1|
\end{aligned}
$$

The length of u is the same in all the above words and is $(p-1)|xy|+|x|$. The above squares start after $|xy|$ locations. Since the first such square starts with $x_2y(xy)\cdots$, the maximum number of such squares is less than $|xy|$. □

It is not the case that every FS-double square $w = ((xy)^p(x)(xy)^p)^2$ have squares starting at locations $[2, 2|xy|]$. From Lemma 4.4, the square in w whose last occurrence begins at any location in $[2, |lcp(xy, yx)|]$ is a conjugate of sq_1^2. In case of $s_i(w) = 1$ where $|xy| < i < 2|xy|$, the square starting at i is shorter than sq_1^2. Further, the structures of the roots of these squares given in the lemma show that the $lcp(x, y)$ is a non-empty word. An example of an FS-double square having these two types of squares has $x = aaba, y = aab, p = 3$ and the s_i sequence of $w = ((xy)^3x(xy)^3)^2$ is given below.

$$
\begin{aligned}
w =&((aaba \cdot aab)(aaba \cdot aab)(aaba \cdot aab)(aaba)(aaba \cdot aab)^3)^2\\
s_i(w) =&((2111 \cdot 110)(0000 \cdot \underline{11}0)(0000 \cdot 000)(0000)0001111 \cdots)
\end{aligned}
$$

In the next section, the no-gain locations in words having a sequence of FS-double squares are computed. The result is then used to compute the square density of such words.

4.4 Square Density of Words with 2FS Squares

We have seen that two consecutive occurrences of $2'$s in the s_i sequence of a word are only possible in the following two scenarios.

(I) Equal 2FS square: The lengths of two consecutive FS-double squares are equal.

(II) Unequal 2FS square: The shorter FS-double square always starts before the longer FS-double square.

Given an FS-double square w, an equal 2FS square is obtained by appending a single letter to w where the new FS-double square is a conjugate w. However, w needs to be extended by a long word to introduce another FS-double square and to construct an unequal 2FS square. Some properties of unequal 2FS squares are given in Lemma 3.15.

Suppose the s_i sequence of a word is "$222222 \cdots 1011 \cdots$", then the trail of $2's$ is a result of either all equal length FS-double squares or a combination of equal and unequal 2FS squares. Here, the aim is to detect no-gain locations or a sequence of $0's$ in the given s_i sequence. The locations in all possible words adhering to the given criteria are found and their square densities are computed in Theorem 4.7. Before that, the square density of words containing only equal 2FS square is computed in Lemma 4.6. The lemma is supported by the following existing property of an equal 2FS square obtained from Lemma 3.7.

Lemma 4.5. *Let the last letter of an equal 2FS square $w = SQ_1^2.a$ adds two new squares t, v. Then, $t = SQ_2^2$ and $|v| = |sq_1|$.*

Lemma 4.6. *Let w begins with a sequence of equal-length consecutive FS-double squares where the last FS-double square of the sequence is a suffix of w. If the exponents of the first FS-double square are equal and greater than two, then $|DS| < 1.64|w|$.*

Proof. Let $SQ_1 = (xy)^p(x)(xy)^p$ be the root of first FS-double square where $x, y \in \Sigma^+$ and integer $p > 2$. as noted in Lemma 3.11 that there are at most $|lcp(xy, yx)|$ consecutive equal-length FS-double squares. By Lemma 4.4, there can be at most $|xy|$ rightmost squares that start at locations in $[|xy| + 1, 2|xy|]$. However, these squares reappear after the first $|SQ_1|$ locations in w when w is extended by a non-empty word k to obtain the sequence of $2'$s in the beginning. Thus, $|k| \leq |lcp(xy, yx)|$. For SQ_1^2, the s_i values of each locations in $[2|xy|, p|xy|]$ is zero (see Theorem 4.3). As per Lemma 4.5, if SQ_1^2 is extended to construct w, then the s_i values of the locations marked in the next structure can possibly change:

$$SQ_1^2 = \boldsymbol{(xy)}(xy)^{p-1}(x)\boldsymbol{(xy)^p}(xy)^p(x)(xy)^p$$

So, no new square with $s_i(w) = 1$ is getting added where $2|xy| \leq i \leq p|xy|$ when SQ_1^2 is extended to get w. Thus, the only structure of words possible is shown below assuming $w = SQ_1^2 \cdot k$ where $k \in \Sigma^+$.

$$w = ((xy)^p(x)(xy)^p)^2).k$$
$$s_i(w) = \underbrace{22222\cdots}_{|lcp(xy,yx)|}\underbrace{00000\cdots}_{(p-1)|xy|}10111010\cdots$$

Assuming $|lcp(xy, yx)| \approx |xy|$, we compute the square density of w as follows.

$$\rho(w) = \frac{2|\boldsymbol{lcp(xy, yx)}| + 0(p-1)|xy| + \frac{11}{6}(3p|xy| + |\boldsymbol{k}| + 2|x|)}{4p|xy| + |\boldsymbol{k}| + 2|x|}$$
$$= \frac{2|xy| + \frac{11}{6}((3p+1)|xy| + 2|x|)}{(4p+1)|xy| + 2|x|}$$
$$\text{Case } |x| = |y| : \rho(w) = \frac{4|x| + \frac{11}{6}((6p+2)|x| + 2|x|)}{(8p+2)|x| + 2|x|} = \frac{(11p + \frac{34}{3})|x|}{(8p+3)|x|} < 1.641$$
$$\text{Case } |x| \gg |y| : \rho(w) = \frac{2|x| + \frac{11}{6}((3p+1)|x| + 2|x|)}{(4p+1)|x| + 2|x|} = \frac{5.5p + 7.5}{4p+3} < 1.6$$
$$\text{Case } |x| \ll |y| : \rho(w) = \frac{2|y| + \frac{11}{6}((3p+1)|y|)}{(4p+1)|y|} = \frac{\frac{11}{2}p + \frac{23}{6}}{4p+1} < 1.563$$

□

It is possible to extend the example discussed after Lemma 4.4 to get a sequence of $2's$ as mentioned in the above theorem. The word $w = ((xy)^3x(xy)^3)^2$, where $x = aaba, y = aab$ can be appended by 'aab' to get four consecutive FS-double squares in the beginning. We get $|w| = 95$, and the total number of distinct squares in w is 51. The $s_i(w)$ has a sequence of zeroes as specified in the above lemma.

$$
\begin{aligned}
w &= aabaaab\underline{aabaaabaabaaab}aabaaabaaabaabaaabaabaaabaabaaabaa\ldots \\
s_i(w) &= 2222110\underline{00000000000000}000000001111111000000011111100001\ldots
\end{aligned}
$$

The next theorem computes the square density of words that starts with a sequence of FS-double squares.

Theorem 4.7. *Let w be the word containing k consecutive FS-double squares such that $s_1(w) = s_2(w) = \cdots = s_k(w) = 2$, where w ends with SQ_k^2. If the exponents of individual FS-double square are equal and are more than 2, then $\rho(w) < 1.64$.*

Proof. The square density of an FS-double square $w = ((xy)^p(x)(xy)^p)^2$ where $p > 2$ and $x, y \in \Sigma^+$ is given in Equation (4.6). The first location of w starts with two rightmost squares, and the number of rightmost squares starting in the next $2|xy|$ locations is $< 2|xy|$. Thus, at most $2|xy|$ squares start in the prefix $(xy)^2$. According to Theorem 4.3, the locations $[2|xy|, p|xy|]$ start with no rightmost squares.

$$
\begin{aligned}
\rho(w) &= \frac{2|xy| + (p-1)0 + \frac{11}{6}(3p|xy| + 2|x|)}{4p|xy| + 2|x|} \\
(p = 3) : &< \frac{2|xy| + \frac{11}{6}(9|xy| + 2|x|)}{12|xy| + 2|x|} = \frac{133|x| + 111|y|}{84|x| + 72|y|} \\
&< max(\frac{61}{39}, \frac{133}{81}, \frac{111}{72}) = 1.641 \qquad (4.6)
\end{aligned}
$$

The three fractions in Equation (4.6) are obtained by assuming $|x| \approx |y|$, $|x| \gg |y|$ and $|x| \ll |y|$, respectively. The highest square density is obtained for the case where $|x| \gg |y|$. We divide the words containing a sequence of consecutive FS-double squares into the following three types.

(I) Sequence of Equal 2FS squares: The lengths of FS-double squares in the sequence of $2'$s are equal,

(II) Sequence of Unequal 2FS squares: For $i < j$, the length of FS-double square starting at j is more than that of the FS-double square at i,

(III) Sequence of Equal and Unequal 2FS squares: The lengths of FS-double square are the combination of case (I) and case (II).

The square density of words given in (I) is less than $\frac{133}{81}$ as given in Lemma 4.6. Let w be a word as described in either (II) or (III). The word w can be rewritten as $w'SQ_k^2$ where w' is a prefix of length $k-1$, and two rightmost squares start from any location of w'. The length of w' determines $\rho(w)$ as shown below.

$$\rho(w) = \frac{2|w'|}{|w|} + \frac{|DS(SQ_k^2)|}{|w|} \tag{4.7}$$

The value of the fraction $\frac{|DS(SQ_k^2)|}{|w|}$ is the highest when $\frac{|x|}{|y|} \approx 1$ (see Equation (4.6)). So, we compute $\frac{|w'|}{|w|}$ for the same case. Assume $SQ_1 = (xy)^p(x)(xy)^p$ and $SQ_2 = (uv)^q(u)(uv)^q$ are the roots of two FS-double squares in w where the exponents p and q are greater than 2. The following relation is obtained using Lemma 4.4.

$$\begin{aligned} |(uv)^q(u)| &> |(xy)^p(x)(xy)^p| + |(xy)^p(x)| + |(xy)^{p-1}| \\ |(u)^{2q+1}| &> |(x)^{8p}| \implies |u| > \frac{8p}{2q+1}|x| \end{aligned} \tag{4.8}$$

The highest value of $\rho(w)$ is obtained when the ratio $\frac{8p}{2q+1}$ in Equation (4.8) has the smallest value. So, $p = q = 3$ and it gives $|u| > \frac{24}{7}|x|$.

Length of w'	$s_i(w) = 0'$**s**	$\lvert w \rvert$
1	$\lvert xy\rvert = 2\lvert x\rvert$	$12\lvert xy\rvert + 2\lvert x\rvert = 26\lvert x\rvert$
$1+1$	$\lvert uv\rvert = 2(\frac{24}{7})\lvert x\rvert$	$> 26(\frac{24}{7})\lvert x\rvert$
$1+1+1$	$2(\frac{24}{7})^2\lvert x\rvert$	$> 26(\frac{24}{7})^2\lvert x\rvert$
$1+1+1+1$	$2(\frac{24}{7})^3\lvert x\rvert$	$> 26(\frac{24}{7})^3\lvert x\rvert$
$1+1+1+1+1$	$2(\frac{24}{7})^4\lvert x\rvert$	$> 26(\frac{24}{7})^4\lvert x\rvert$

Table 4.1: *FS-double squares and no-gain locations in words of type (II)*

Length of w'	$s_i(w) = 0'$s	$\lvert w \rvert$
$\lvert xy \rvert$	$\lvert xy \rvert$	$14\lvert xy \rvert + 2\lvert x \rvert$
$\lvert xy \rvert + 1 + 3\lvert xy \rvert$	$2 * 3\lvert xy \rvert$	$> 13 * 3^2\lvert xy \rvert$
$\lvert xy \rvert + 1 + 3\lvert xy \rvert + 1 + 3^2\lvert xy \rvert$	$2 * 3^2\lvert xy \rvert$	$> 13 * 3^4\lvert xy \rvert$
$\lvert xy \rvert + 1 + 3\lvert xy \rvert + 1 + 3^2\lvert xy \rvert + 1 + 3^3\lvert xy \rvert$	$2 * 3^3\lvert xy \rvert$	$> 13 * 3^6\lvert xy \rvert$

Table 4.2: *FS-double squares and no-gain locations in words of type (III)*

Let us consider a word with a sequence of FS-double squares that follows the type (II) where $SQ_1^2 = ((xy)^3(x)(xy)^3)^2$ and $|SQ_1| = 12|xy| + 2|x|$. The next FS-double square starting at location 2 must be $SQ_2^2 = ((uv)^3u(uv)^3)^2$ satisfying the relation $|u| > \frac{24}{7}|x|$ which gives $|w| = 26 * \frac{24}{7}|x| + 1$. The length of w and the number of no gain lengths increase with $|w'|$ as described in Table 4.1. The highest value of the ratio $\frac{|w'|}{|w|}$ obtained in this case is $\frac{2}{26*\frac{24}{7}} \approx 0.224$.

We can construct a word with a combination of equal and unequal 2FS squares as given in (III). However, the number of no-gain lengths in such cases is always twice the number of consecutive FS-double squares introduced (see the proof of Lemma 4.6 and Table 4.2). The ratio obtained in this case is $\frac{|w'|}{|w|} \leq \frac{2}{26*3} \approx 0.025$. If $SQ_k^2 = ((mn)^3m(mn)^3)^2$ is the last FS-double square in the sequence, then in the $3|mn|$ length prefix of SQ_k^2 at least $2|mn|$ locations are no-gain locations. So, with this information we use Equation (4.7) to get $\rho(w) = (0.025 + 0.833) < 1.61$. □

The square density of the above words is less than the best upper bound obtained for the square conjecture, which is $\frac{11}{6}$ [29]. Consider the smallest FS-double square, $w = abaababaab$ where $s_i(w) = 2001100100$ and, its reverse $w^r = baababaaba$ where $s_i(w^r) = 1011101000$. A word and its reversal always contain the same number of distinct squares, but the number of distinct FS-double square(s) in each may vary. It is not always the case that the distinct square density of a word with many FS-double squares is more than the words with no or fewer FS-double squares. For example, the structure given by Sadri and Osborn in [14] produces words with a distinct square density less than $\frac{5}{6}$.

In contrast, some patterns with a distinct square density equal to one do not have any FS-double square [34]. In the case of a structure introducing FS-double squares, it is observed that such a structure produces words containing a sequence of FS-double squares. However, when the word is reversed, these FS-double squares are distributed to other locations. To put it in simple terms, these words contain some locations with two s_i values and many locations with zero s_i values. After reversing the words, $2'$s get distributed with $0'$s resulting in a word with many locations having one s_i value. Therefore, the structure of words that begins with k consecutive FS-double squares such that the first k locations in the reverse of such words also start with two distinct squares. This structure can be used to build words with high square densities, which is discussed in the next chapter.

4.5 Conclusions

The square conjecture predicted that the square density of a word is less than one. An upper bound for the square density is obtained by counting the maximum number of locations that start with FS-double squares, assuming that every location that does not initiate an FS-double square begins with exactly one rightmost square. We introduced no-gain locations, which are the locations that do not begin with any rightmost squares. A proof is presented that demonstrates the existence of no-gain locations in the prefix of an FS-double square $((xy)^p(x)(xy)^p)^2$, with the first $2|xy|$ locations having non-zero s_i values and the rest being no-gain locations.

Furthermore, it is shown that extending the given FS-double square to add equal-length FS-double squares at neighbouring locations does not change the number of no-gain locations. This result implies that the square density of words containing only equal-length FS-double squares at the beginning of a word is less than one. In addition, the chapter discussed the results on words with FS-double squares at consecutive locations, where the exponents of each FS-double square are equal. By identifying the no-gain locations in these words, it was shown that the number of distinct squares in a word is at most $\frac{133}{81}$ times its length.

5

Squares in Some Special Words

The chapter delves into the topic of square density for two types of words: words that contain special FS-double squares and non-primitive words. The previous chapters introduced the concept of FS-double squares and discussed their structures and properties while also focusing on finding properties of primitive words and squares.

For any word, the count of distinct squares in a word is maintained after reversing the complete word itself. However, the number of FS-double squares may be affected. The first part of this chapter elaborates on the connection between a word, its reversal, and the number of FS-double squares in each of them.

The second part of the chapter focuses on exploring the limitations of using non-primitive squares to increase the square density. As discussed in the first part and in the previous chapters, there exists a relationship between the number of distinct squares and primitive squares in a word. The structure of a primitive square allows to closely pack conjugates of squares, maximizing the square density of words with a large number of distinct squares. Interestingly, FS-double squares also qualify as primitive squares. However, the literature does not adequately explain why non-primitive words cannot be used to increase square density. This background information sets the stage for the investigation carried out in the second part of the chapter.

5.1 Motivation

The value of s_i represents the number of rightmost squares starting at location i, but it does not necessarily mean that the location corresponds to two different squares when the word is reversed. To illustrate this, further consider the smallest FS-double square, $w = abaababaab$ where $s_i(w) = 2001100100$ and its reverse $w^r = baababaaba$ where $s_i(w^r) = 1011101000$. Although a word and its reversal always have the same number of distinct squares, the number of distinct FS-double squares in each can vary. Also, the square density of a word is not always higher when it contains many FS-double squares. For example, the pattern given by Blanchet-Sadri et al. [14] produces words with a square density less than $\frac{5}{6}$ whereas some patterns with a square density one do not generate words containing an FS-double square [34]. One reason for the lower value of the distinct square density of words of an earlier pattern is that the words of this pattern have a series of FS-double squares, and these squares are then distributed to other locations when the word is reversed. Consequently, these words have numerous locations with $s_i = 0$ values. The characteristics of such words are described, and their square density is computed in this chapter.

The main focus of this thesis is to examine the properties and structures of words that maximize the number of distinct squares in a word. The early chapters discuss primitive squares and their properties, which are key in this regard. It is important to note that an FS-double square is a type of primitive square. The non-primitive squares are the repetitions with an exponent greater than 2. The analysis in section 5.4 demonstrates that non-primitive squares do not have high square density. The maximum square density of these words is computed, emphasising the importance of primitive squares in maximizing the number of distinct squares in a word.

5.2 Bordered FS-double Squares

The square density of w is $\rho(w) = \frac{|DS(w)|}{|w|}$. The result $s_i \leq 2$ implies that any location can begin with at most two distinct squares. Using the same result, a location can end with the last letter of two distinct squares in the word. We use the notation e_i to show the number of leftmost distinct squares ending at location i. The result on s_i also applies on e_i and $e_i(w) \leq 2$. Thus, extending a word by a letter can add at most two new distinct squares. However, for any FS-double square w, $s_1 = 2$ need not always imply $e_{|w|} = 2$. It is also possible that extending letters to an FS-double square may not add any new square. We will see a way to extend an FS-double square where each added letter adds two new distinct squares. By finding this structure, a special set of words is obtained where the maximum number of locations of these words can be mapped to two distinct squares. The next section describes such words and finds the limit on such types of extensions.

Recall the structure of FS-double square $(xy)^{p+q}x(xy)^p$ where $x, y \in \Sigma^+$ and integers p, q satisfies $p \geq 1, q \geq 0$. The square always has a non-empty border $(xy)^p$. The symbol $\mathbb{B}(w)$ represents the set of borders of the word w. For $w = aabaa$, the set $\mathbb{B}(w) = \{a, aa\}$. As given in Lemma 3.2, the roots of an FS-double square are obtained from the non-empty words x, y. An FS-double square, SQ^2, begins with two distinct squares, and these squares end after the first instance of SQ. This section identifies the structure of FS-double squares, where each terminal letter contributes to two distinct squares. The additional condition is that the length of each of these squares is greater than $|SQ|$. For example, consider an FS-double square $SQ^2 = awb$ where $a, b \in \Sigma, w \in \Sigma^+$. The difference between the number of distinct squares in SQ^2 and the number of distinct squares in aw (or wb) must be two. Some of the existing results of FS-double squares and properties of primitive words are used to obtain such words. lemma 5.1describes the types of squares that start at the beginning of an FS-double square.

Lemma 5.1 ([10]). *The following statements hold for an FS-double square* $SQ^2 = ((xy)^{p+q}x(xy)^p)^2$.

(a) SQ^2 *is a primitive square, and*

(b) sq^2 *is a primitive square for some positive integer* p *that holds* $p > 1$.

The following lemma describe some basic results on words.

Lemma 5.2 (Squares in squares). *Let* $w^2 = uv'uv'u'$ *for some* $u, v, u', v', x \in \Sigma^+$ *such that* $|u'| = |u|$ *and* $|v'| = |v|$. *Then,* $u = x^2$.

Proof. As $|w| = |v| + 1.5|u|$, we get $w = uv'x_1 = x_2v'u'$ for $u = x_1x_2$ and $|x_1| = |x_2|$. From the prefixes of the two expressions, $x_1 = x_2 = x$. □

Fan et al. [31] described "The new periodicity lemma" that classifies the squares in an FS-double square based on their structures and locations. This lemma is revisited in [10] and presented as follows.

Lemma 5.3 ([10]). *Let* u^2 *be square in an FS-double square* SQ^2. *Then, one of the statements holds: (a)* $|u| = |SQ|$, *(b)* $|u| < |sq|$, *(c) If* $|SQ| > |u| \geq |sq|$, *then the primitive root of* u *is a conjugate of* xy. *Here,* $SQ = (xy)^{p+q}x(xy)^p$, $sq = (xy)^{p+q}x$ *where* $x, y \in \Sigma^+$ *and an integer* $p \geq q \geq 1$.

Theorem 5.7 identifies the structure of FS-double squares that ends with two distinct squares. The results obtained in Lemmas 5.4 to 5.6 are used to prove the theorem.

Lemma 5.4. *For some non-empty words* x, y, *let* $SQ^2 = (xyxxy)^2$ *be an FS-double square that ends with a rightmost square,* v^2, *such that* $|SQ| < 2|v| \neq 0$. *Then,* $|v| = |sq|$ *and* $x \in \mathbb{B}(xy)$ *where* $sq = xyx$.

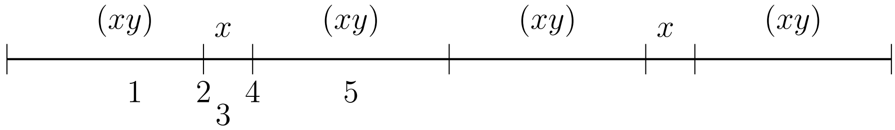

Figure 5.1: *Starting location of v^2 in SQ^2*

Proof. The square v^2 can begin at one of the five locations. Given $x = x_1x_2$ and $y = y_1y_2$ where x_1, x_2, y_1, y_2 are some non-empty words. There are nine possible structures for v^2 where $|xyx| < |v|$. We use Lemma 5.2 along with Lyndon and Schützenberger theorem (Theorem 1.5.2 from [3]) to discard the cases where the root SQ non-primitive. Table 5.1 divides the valid and invalid cases.

Sr. no.	v^2	Is valid?
1	$yx.yx.xy$	×
2	$xyx.y.xxy$	×
3	$xxy.xy.xxy$	×
4	$yxxy.x.yxxy$	×
5	y_2xyxxy	✓
6	$x_1yxyxxy$	×
7	$x_2xyxyxxy$	×
8	$y_2xxyxyxxy$	×

Table 5.1: *Possible structures of v^2*

The ninth structure of $v^2 = x_2yxxy.x.yxxy$ is possible if $xy = yx$. Here, the only possible square obtained from the structure mentioned in Table 5.1 has $v = y_2xy = xxy$. Thus, v^2 is conjugate of sq. □

Lemma 5.5. *Let v^2 be a suffix of an FS-double square, $SQ^2 = ((xy)^px(xy)^p)^2$ for some $x, y \in \Sigma^+$ and a positive integer p. If $2|v| > |SQ|$, then $x \in \mathbb{B}(xy)$ and $|v| = |sq|$. Here, $sq = (xy)^px$.*

Proof. The given statement holds for $p = 1$ (refer Lemma 5.4). There are five

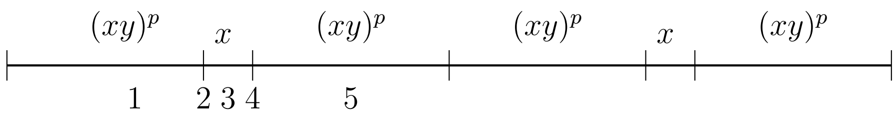

Figure 5.2: *Starting location of v^2 in SQ^2 for $q = 0$*

possibilities for v^2 to start in SQ^2 as shown in Figure 5.2 for $p > 1, q = 0$. The number in the figure indicates the beginning of v^2 in SQ^2 and the respective case number. As v^2 is a suffix of SQ^2 and $|v| > p|xy|$, the second occurrence of v in all of the five cases ends with $(xy)^p$. The first occurrence of v ends with either of the suffixes: (a) xy, (b) y_2xy_1 where $y = y_1y_2$ or, (c) x_2yx_1 where $x = x_1x_2$. According to Lemma 5.3, the primitive root of v is a conjugate of xy. Thus, the relations obtained in cases (b) and (c) imply that two conjugates of xy are equal. This contradicts Lemma 3.1, so we discard the words with these two cases. The possible structures of v^2 for case (a) are verified in Table 5.2 where $|v| = |sq|$ is the only valid case. The structure of v^2, in this case, implies $x \in \mathbb{B}(xy)$.

Sr. no.	**Possible structure of base v**	**Condition**	**Remark**
1	$v = (xy)^q(x)(xy)^{p+s} = (xy)^s(xy)^q(x)(xy)^p$	$p > q > s > 0$	$xy = yx$
2	$v = x(xy)^p(xy)^q = (xy)^qx(xy)^p$	$p > q > 0$	$xy = yx$
3	$v = x_2(xy)^p(xy)^q = (xy)^{p-q}(x_1x_2)(xy)^p$	$p, q, p - q > 0$	$xy = yx$
4	$v = (xy)^p(xy)^s = (xy)^{p-s}x(xy)^p$	$p, s, p - s > 0$	$xy = yx$
5	$v = y_2(xy)^p = x(xy)^p$	$p > 0$	$x \in \mathbb{B}(xy)$

Table 5.2: *First occurrences of v ending with xy in SQ^2*

□

Lemma 5.6. *Given an FS-double square SQ^2 with $q > 1$ that ends with v^2 where $|SQ| < 2|v|$. Then, $|v| = |SQ|$.*

Proof. We have $SQ = (xy)^{p+q}x(xy)^p$. In Figure 5.3, we marked all the possible starting locations of v^2 in SQ^2. Here, we show that every case leads to the relation $xy = yx$ and the relation contradicts Lemma 3.1. Similar to Lemma 5.5, the first occurrence of v ends with either x_2yx_1, y_2xy_1 or xy assuming $x = x_1x_2, y = y_1y_2$. We discard the first two types of squares since occurrences of v's violate Lemma 3.1. The first occurrence of v that starts at one of the marked locations $1, 3, 5, 7$ or 8 never ends with xy. If v^2 begins

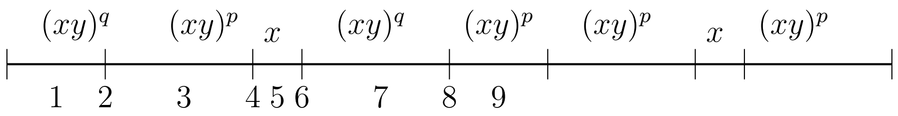

Figure 5.3: *Beginning of v^2 in SQ^2 where $p_1 > p_2$*

at location 2, 4 or 6, then equating the structures of two v's always gives $xy = yx$. □

Theorem 5.7 (Bordered FS Square)**.** *Let $SQ^2 = s.a$ be an FS-double square, where $s \in \Sigma^+$ and $a \in \Sigma$. Then, $|DS(SQ^2)| - |DS(s)| = 2$ if and only if SQ^2 ends with a conjugate of sq^2.*

Proof. **(If)** The statement follows from Lemma 5.5 and 5.6.
(Only if) The last letter of every FS-double square is a part of the FS-square itself. We assume SQ^2 ends with a conjugate of sq^2, say v^2. So, xy in Equation (5.1) ends with x. Thus, we can write $xy = y'x$ for $y' \in \Sigma^+$ and $|y'| = |y|$. Now, the reverse of SQ^2 is a word that starts with two distinct squares, and these two squares satisfy the premise of Lemma 3.2. Thus, v^2 is a unique square and removing the last letter of SQ^2 removes two distinct squares.

$$SQ^2 = (xy)^p(x)(xy)^{p-1}\underbrace{xy.(xy)^p(x)(xy)^p}_{y'v^2} \tag{5.1}$$

□

The bordered FS square is a word in which removing either the first or the last letter removes two distinct squares. In the following section, we show the words in which many letters in the suffix and the prefix are mapped to two distinct squares.

5.3 Squares in Bordered FS Squares

The following lemma computes the maximum number of equal-length bordered FS squares in a word. The result is based on the work described in the section 4.4.

Lemma 5.8. *Let w begin with k equal length consecutive bordered FS squares. If the first bordered FS square is $SQ^2 = (xyxxy)^2$, then $k = |LCP(x, y)| + 1$ and $|y| > |x|$.*

Proof. Given $SQ^2 = (xyxxy)^2$ and Theorem 5.7 shows that $x \in \mathbb{B}(xy)$. So, either x is a suffix of y or $|y| < |x|$. Let SQ^2 and $\overline{SQ}^2$ be two consecutive bordered FS squares. Assume x begins with a letter 'a' such that $x = ax'$. So, $\overline{SQ} = (\boldsymbol{x'y}a)(x'a)(x'ya)$ and $x'a \in \mathbb{B}(x'ya)$. The latter condition holds provided y begins with 'a' (refer to the prefix in bold in the structure of $\overline{SQ}$) and ya ends with $x'a$. Thus, the value of $|LCP(x, y)|$ must be at least one to get two consecutive bordered FS squares. Similarly, the conjugate of $\overline{SQ}^2$ adjacent to it is bordered FS square if $|LCP(x, y)| = 2$. Thus, k consecutive bordered FS squares are possible when $|LCP(x, y)| = k - 1$. In case of $|y| < |x|$, the two bases SQ and $\overline{SQ}$ are non-primitive. This contradicts Lemma 3.3. So, $|y| > |x|$. □

We now see the computation of the maximum number of equal-length consecutive bordered FS squares in a word with respect to the length of the word. The proof of the theorem is based on the results obtained in Lemma 5.8.

Theorem 5.9. *Let w contains k equal length consecutive bordered FS squares. Then, $11k < |w|$.*

Proof. The value of $\frac{k}{|w|}$ is maximum if k consecutive bordered FS squares are at the beginning of w and k^{th} bordered FS square is a suffix of w. So, assume w begins with an FS-double square SQ^2 followed by $k-1$ consecutive bordered FS squares of size $2|SQ|$. From Equation (5.2), the value of $\frac{k}{|w|}$ is maximum for $p = 1$. Lemma 3.11 shows that SQ^2 can be extended with at-most $|x| - 1$ letters to obtain consecutive FS-double squares. Here, $k = |x|$ and $SQ = (xy'x)(x)(xy'x)$ where y' is some non-empty word (refer Lemma 5.8). The ratio is computed below.

$$\frac{k}{|w|} = \frac{|x|}{2((p+1)|x| + p|y|) + |x| - 1} < \frac{|x|}{2(5|x| + 2|y'|) + |x| - 1} < \frac{1}{11} \quad (5.2)$$

□

The exponents of bordered FS-double squares are identical, meaning that the structure of an FS-double square $((xy)^{p_1}x(xy)^{p_2})^2$ satisfies $p_1 = p_2 = p$. Chapter 4 detected no-gain locations for $p > 2$. Thus, bordered FS-double squares also contain no-gain locations as specified in the chapter. Moreover, the ratio $\frac{|w|}{11}$ obtained for a word w in the theorem mentioned above indicates that the sequence of 2s obtained with bordered FS-double squares is shorter than the longest possible sequence of 2s, which is $\frac{|w|}{7}$. This demonstrates that the square density of the words mentioned in Theorem 5.9 is less than $\frac{133}{81}$. To compute the density precisely, this work can be extended further.

The works that are discussed to this point identified various properties of primitive squares, which also include FS-doubles as a special case. Interestingly, both the upper bounds predicted for the square conjecture [34] and the conjecture on primitive squares [28] are similar. Chapter 6 discusses the words that contain the maximum number of distinct squares, and it is observed that these words have a significantly higher number of primitive squares compared to non-primitive squares. To understand the relationship between maximizing primitive squares in a word to increase the distinct squares, it is essential to examine the properties of non-primitive squares. The subsequent section of this chapter identifies various types of squares in non-primitive words and calculates the number of no-gain locations.

5.4 Squares in Non-primitive Words

If we take a root u, then the smallest non-primitive word that we can form is the square u^2. If we increase the exponent while keeping the same root, then the resulting non-primitive word would be u^3. Before discussing the square density of a non-primitive word u^k, it is important to understand the different types of squares that appear in u^3, as we will see in the next lemma.

Lemma 5.10. *Let u be a primitive word. Then, every location in $[2, |u|-1]$ starts with a conjugate of u^2 that appears once in u^3. Further, $s_i(u^3) \leq 1$ for all $i \in \left[1, \frac{|u|}{2}\right]$.*

Proof. Let v^2 be a rightmost square starting at a location in $[2, |u|-1]$. Then, there are at most two possible values of $|v|$ per location. The word u has $|u|-1$ distinct conjugates apart from itself (see Lemma 3.1 (a)). Since u^2 is a primitive square, each location in $[2, |u|-1]$ starts with a conjugate of u^2 that appears once in u^3. Thus, one of the values of $|v|$ is $|u|$.

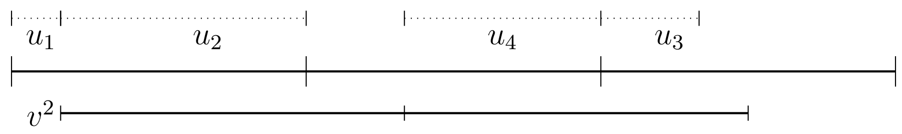

Figure 5.4: *Square v^2 in the prefix of u^3*

Now, consider the other two possibilities for a rightmost square that begins at a location i satisfying $2 \leq i \leq \frac{|u|}{2}$.

Case $|v| > |u|$: Comparing the two copies of v, the first and the last v in v^2 as shown in the Figure 5.4, we get $u_2u_1 = u_4u_3$, where $u = u_1u_2 = u_3u_4$ and $|u_1| \neq |u_3|$. This implies that the two conjugates of u are identical and u is non-primitive, contradicting the assumption for u. Therefore, $|v| \ngtr |u|$.

Case $|v| < |u|$: In this case, the square v^2 reappears in the suffix of u^3. Thus, the occurrence of v^2 in any of the first half locations of u cannot be the rightmost occurrence. Therefore, $|v| \not< |u|$.

So, the locations $[2, |u|-1]$ start with conjugates of u^2 and $s_i(u^3) = 1$ for all $i \in \left[2, \frac{|u|}{2}\right]$.

If the first location of u^3 starts with a square of size $\leq 2|u|$, then the square has another instance in u^3. A square resulting $s_1(u^3) > 0$ must be of a size greater than $2|u|$. However, it leads to the arguments in the above case where $|v| > |u|$ and u need to be a non-primitive word to have such a square. Hence, $s_1(u^3) = 0$. □

In a square, u^2, the s_i values of the first $\frac{|u|}{4}$ locations is at most one [5]. The number of such locations increases in u^3 and according to Lemma 5.10, there are a total of $(\frac{|u|}{2} + \frac{|u|}{4})$ such locations in u^3. We observed that in some cases, the number of new distinct squares introduced by appending u^3 with u's is not always $\geq |u|$. For example, consider a word $w = u^k$ where $u = abaa$ and $k > 2$. Here, the number of new distinct squares that are getting added is either 3 or 1 as the value of k increases. Whereas the number of letters that are newly getting added is always four. In general, if k is even, the k^{th} u in u^k adds the square with root $u^{k/2}$. Otherwise, the k^{th} u adds all conjugates of $u^{\frac{k-1}{2}}$ which is equal to $|u|-1$. Thus, new $|u|$ squares are introduced per $2|u|$ new letters. So, a group of locations having zero s_i values can be found. The next lemma verifies and counts the number of such locations.

Lemma 5.11. *Let u be a primitive word. Then, the word u^k where $k > 3$ has at least $\lceil\frac{k-3}{2}\rceil|u|$ locations that do not start with any rightmost square.*

Proof. Assume m is an even integer such that $2 < m \leq k$ and $u^k = u^{k-m}u^m$. Let v^2 be a rightmost square that starts in $(k-m+1)^{th}$ occurrence of u in u^k on counting from the left to right. Such a v^2 must end in the last u of u^k. Let $u = u_1u_2 = u_3u_4$ and u_5 be one of the prefixes of u, we get v^2 as follows.

$$\begin{aligned} v^2 &= u_2(u_1u_2)^{m-2}u_5 \\ &= (u_2u_1)^{\frac{m-2}{2}}u_2(u_1u_2)^{\frac{m-2}{2}}u_5 \end{aligned} \tag{5.3}$$

From Equation (5.3), the length of a root v is $|(u_2u_1)^{\frac{m-2}{2}}| + \frac{|u_2|+|u_5|}{2}$. Here, the term $\frac{m-2}{2}$ is always a positive integer. If $|u_2|$ is zero, then $u_5 = u_5'u_5''$ such that $|u_5'| = |u_5''|$ and we get $v = u^{\frac{m-2}{5}}u_5' = u_5''u^{\frac{m-2}{5}}$. This contradicts the assumption that u is primitive since one of the prefixes of v's shows that the two conjugates of u are equal. We get a similar case when u_5 is an empty word or when u_2, u_5 are empty words. On the other hand, the condition $|u_2| = |u_5|$ leads to a relation $u_2u_1 = u_1u_2$, which is again unacceptable. Similarly, we can discard the possibility of v^2 when $|u_2| \lessgtr |u_5|$.

Hence, the first letter of u^m starts with the rightmost square z^2 where $z = u^{\frac{m}{2}}$. The remaining $|u| - 1$ locations in the first u of u^m start with no rightmost squares. The total number of even powers of u in u^k is $\lfloor \frac{m}{2} \rfloor$ and the powers that are greater than 3 are $\lceil \frac{k-3}{2} \rceil$. So, there are at least $\lceil \frac{k-3}{2} \rceil |u|$ locations whose s_i values are zero. □

The above lemma marks the locations in u^k where s_i values are zero. The squares in u^k that are not present in u^3 are of a specific structure. The following lemma explains it in detail.

Lemma 5.12. *Let m be a positive odd integer such that $u^k = u^{k-m}u^m$ and $m > 3$. If u is a primitive word and v^2 is a rightmost square that begins in $(k-m+1)^{th}$ u of u^k. Then, v^2 is a conjugate of u^{m-1}.*

Proof. Consider the suffix u^m of u^k. Let v^2 be a rightmost square that satisfies the criteria mentioned in the statement, so it begins at any of the first $|u|$ locations of u^m and ends in the last u of u^m. Assuming $u = u_1u_2 = u_3u_4$, we can write $v^2 = u_2(u)^{m-2}u_3$. The word u_3 is non-empty; else v^2 would no longer be the rightmost. Using this relation, we obtain the first and the last occurrence of the root v in v^2 as follows.

$$v = u_2(u)^{\frac{m-3}{2}}u_5 = u_6(u)^{\frac{m-3}{2}}u_3, \text{ where}$$

$$u = u_5u_6$$

Here, v^2 must be a conjugate of u^{m-1}. Otherwise, two conjugates of u would be the same. □

Theorem 5.13. *For $k > 3$, $\rho(u^k) \leq \frac{1}{2} + \frac{3}{2k}\left(2\rho(u^3) - 1\right)$.*

Proof. Let us divide the u^k into two parts as $u^k = u^{k-3}.u^3$. We consider the rightmost occurrences of squares to count the number of distinct squares in the word. Let m be an integer that satisfies $2m \leq k$ and $2m+1 \leq k$. From Lemma 5.11 and 5.12, we obtain the s_i values of the locations of the first u in factors u^{2m+1} and u^{2m} as shown in Figure 5.5.

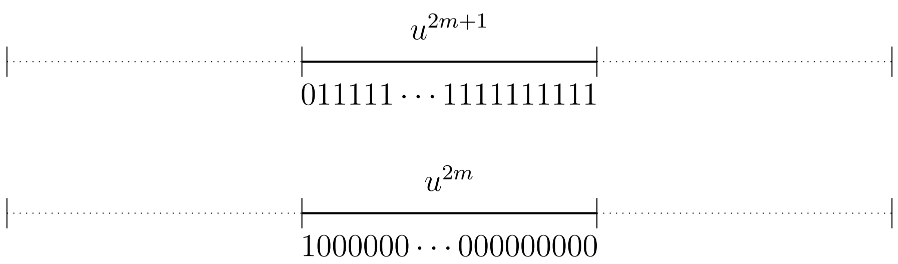

Figure 5.5: *s_i values of factors of u^k*

Thus, for every odd power of u^k greater than 3, there are exactly $|u| - 1$ rightmost squares, whereas only one rightmost square is obtained per even power of u^k. There are total $\lfloor \frac{k-3}{2} \rfloor$ odd powers (excluding the value 3) in u^k to give $\lfloor \frac{k-3}{2} \rfloor |u|$ distinct squares. This gives the following ratio,

$$\begin{aligned} \rho(u^k) = \frac{|DS(u^k)|}{|u^k|} &= \frac{|DS(u^3)| + \lfloor \frac{k-3}{2} \rfloor |u|}{k|u|} \\ &\leq \frac{|DS(u^3)|}{k|u|} + \frac{1}{2} - \frac{3}{2k} = \frac{1}{2} + \frac{3}{2k}\left(2\rho(u^3) - 1\right) \end{aligned} \tag{5.4}$$

□

Theorem 5.14. *Let u be a primitive word and $k > 1$ be a positive integer. Then, the value of $\rho(u^k)$ approaches $\frac{1}{2}$ as k increases.*

Proof. The relation obtained in Theorem 5.13 indicates that $\rho(u^k)$ depends on $\rho(u^3)$ and k. Since the value of $\rho(u^3)$ cannot be more than two [34], it is sufficient to show that $\rho(u^k)$ approaches $\frac{1}{2}$ as k approaches infinity. □

We now obtain a relation between densities of u^k for varying k's.

Lemma 5.15. *Let u be a non-empty word and k be an integer greater than 3. Then, $\rho(u^k) > \rho(u^{k+2})$ provided $\rho(u^k) > \frac{1}{2}$. Further, $\rho(u^k) > \rho(u^{k+1})$ whenever k is an odd integer and $|u| \neq 1$.*

Proof. The relation (5.4) gives us:

$$\rho(u^{k+2}) = \frac{|DS(u^3)| + \lfloor \frac{k-1}{2} \rfloor |u|}{(k+2)|u|} \tag{5.5}$$

We substitute the numerator and the denominator in Equation (5.4) by x and y, respectively, and rewrite Equation (5.4) and (5.5) as follows.

$$\rho(u^k) = \frac{x}{y} \qquad\qquad \rho(u^{k+2}) = \frac{x + |u|}{y + 2|u|}$$

So, the relation $\rho(u^k) > \rho(u^{k+2})$ holds if $\frac{x}{y} > \frac{1}{2}$.

From the proof of Theorem 5.13, we also have $|DS(u^{k+1})| - |DS(u^k)| = 1$ whenever k is an odd number. With $\rho(u^k) = \frac{x}{y}$, we get $\rho(u^{k+1}) = \frac{x+1}{y+|u|}$. Thus, $\rho(u^k) > \rho(u^{k+1})$ for $|u| \geq 2$. □

Note that the square densities of u and u^3 are generally not comparable. For example, $u = abaab$ gives $\rho(u) < \rho(u^3)$ while the square density of u for $u = aababbabba$ is more than $\rho(u^3)$.

5.5 Conclusions

The study of words containing FS-double squares is extended in this chapter. It is shown in the previous chapter that the square density of words having a series of locations starting with FS-double square is less than $\frac{133}{81}$. These words are then categorized into two types by considering the number of FS-double squares in the reversed words. In this chapter, it is found that the words and reversal of the words may have a different number of FS-double squares. So, a notation $e_i(w)$ is introduced to verify the count of FS-double in the word obtained by reversing w. If a word and its reversal have the same number of consecutive FS-double squares, then the number of locations in w whose s_i values are two are exactly the number of locations in w whose e_i values are two. The chapter elaborated upon the structure of such words and labelled them as bordered FS-double squares. It is also shown that the number of bordered FS-double squares is less than $\frac{1}{11}^{\text{th}}$ of the length of the word.

A study of no-gain locations is also conducted for non-primitive squares and non-primitive words. For a non-primitive word u^k where $k > 3$, it is shown that the word contains at least $\lceil \frac{k-3}{2} \rceil |u|$ no-gain locations. Additionally, an absence of a location starting with an FS-double square is shown in the prefix u^{k-3} of u^k. These results imply that the square density of non-primitive words converges to $\frac{1}{2}$ as the exponent of the repetition increases.

6

Dense Patterns

In this chapter, the focus is on the investigation of the structures that generate words with the maximum number of distinct squares, which are considered the best-known words for the square conjecture [34]. In addition, the chapter also explores the structure of words containing the maximum number of distinct primitive squares, as described in [28]. The structures are called patterns and are analyzed to create even better ones. The chapter describes different types of patterns and introduces a method to calculate the square density of each pattern. Additionally, the chapter presents a criterion for comparing various patterns, with the goal of gaining a better understanding of how maximizing primitive squares in a word can result in an increase in the number of distinct squares. By analyzing the patterns, researchers hope to identify the most efficient methods to construct words containing the maximum number of distinct or primitive squares.

6.1 Motivation

The structure of a square is such that it can also give rise to partially formed squares, which can then be extended to form new squares. For instance, in the square $u^2 = abaaba$, the suffix *baaba* can be seen as a partially formed square. By appending a single letter 'a' to the square u^2, a new square *baabaa* is formed in the resulting word. Notably, primitive squares have the potential to be extended to introduce more distinct squares than non-primitive squares of the same length. The previous chapter has already provided an explanation for why non-primitive words do not aid in increasing square density. A specific structure given in [34] generates words in which the

Types of squares	**Lower bound**
Repeated squares	a^n
Repeated primitive squares	Fibonacci words [34]
Distinct primitive squares	Deza's words [28]
Distinct squares	Q words [34]
Distinct FS-double squares	Sadri's words [14]
Distinct squares in circular words	Amit's words [5]

Table 6.1: *Lower bounds for different square conjectures*

number of distinct squares increases as the length of the words increases. The words generated by this structure pack a large number of primitive squares. Similar structures are used to obtain the lower bounds for counting various types of squares. Table 6.1 lists some of them.

By studying and extending the words generated by these structures, it is possible to find structures that favour the distinct square density of words. These properties can be combined with the various types of FS-double squares discussed earlier to generate even better words.

6.2 Notation

As explained, a word w is a finite sequence of letters drawn from Σ. Any non-empty subsequence of consecutive letters in w is a factor of w. The word, w, with $|w| = 0$ is called an empty word and is denoted by ϵ. A symbol $\mathbb{N}$ denotes the set of non-negative integers. As mentioned before, the concatenation of two words x and y is the word $x.y$ or simply xy. The concatenation of a collection of words $w_1, \ldots, w_k$ is denoted by $\bigodot_{i=1}^{k} w_i$. A factor u is a border of w if it is both prefix and suffix of w. Denote by $MaxNS(n)$ the maximum number of distinct square contained within an n length word. A number $n \in \mathbb{N}$ is referred to as a no-gain length if $MaxNS(n)$ is equal to $MaxNS(n-1)$, else it is a gain length. Let $MaxDS(n)$ be the set of words with length n having the maximum number of distinct squares. Formally,

$$MaxDS(n) = \{w \mid w \in \Sigma^n \text{ and } \forall u \in \Sigma^n \cdot |DS(w)| \geq |DS(u)|\}.$$

The set $DS(w)$ contains the distinct squares in a word w. The number of locations in w having zero s_i values is represented by $NG(w)$. In other words, $NG(w)$ counts the number of no-gain locations in w. Suppose $w = w_1 w_2$ for some w_1, $w_2 \in \Sigma^+$. We refer to the squares whose last occurrence starts and ends in w_1 and w_2, respectively, by $cross(w_1, w_2)$.

The following section presents an overview of the words with the highest square densities described in the literature and their corresponding structures.

6.3 Existing Structures

This section discusses several existing structures, called patterns, that generate words with high square density. The first two patterns discussed below share similarities in the types of squares contained by the pattern-generated words. The remaining patterns are intended to generate words with specific properties.

Pattern Q

The pattern Q and the square conjecture are proposed in [34]. It is the first word generator and has the highest square density. The words generated by the structure below are referred to as Q words.

$$Q(x) = (a^2ba^1ba^2b)(a^3ba^2ba^3b)(a^4ba^3ba^4b)...(a^xba^{x-1}ba^xb)$$

The length of a word generated by the pattern can be calculated as follows.

$$|Q(x)| = \sum_{i=1}^{x}(3i+2) = \frac{3x^2+7x}{2} \tag{6.1}$$

A square in a word $Q(x)$ contains either 0, 2 or 4 b*'s* [34]. The total number of distinct squares in the given word can be computed using the following equation.

$$|DS(Q(x))| = \frac{1}{2}(3x^2+2x-10-(x \mod 2)) \tag{6.2}$$

Every position i in a Q word satisfies the condition $s_i(Q(x)) \leq 1$. This implies that the square count in $Q(x)$ is always less than $|Q(x)|$. Therefore,

$$|Q(x)| = DS(Q(x)) + NG(Q(x))$$

This leads to another expression, $NG(Q(x)) = 3x + 5 - \lfloor \frac{x}{2} \rfloor$. The number of additional no-gain locations introduced upon generating $Q(x+1)$ is first calculated. Depending on the x, this number can be 3 or 2. Specifically, if x is even, then three no-gain locations are newly introduced, whereas if x is odd, then the number is 2. For $x \mod 2 = 0$, we get $\lfloor \frac{x}{2} \rfloor = \lfloor \frac{x+1}{2} \rfloor = \frac{x}{2}$. It leads to the following computation.

$$NG(Q(x+1)) - NG(Q(x)) = 3$$

For odd x, $\lfloor \frac{x+1}{2} \rfloor = \lfloor \frac{x}{2} \rfloor + 1$. The next value is computed using this relation.

$$NG(Q(x+1)) - NG(Q(x)) = 2$$

The following section presents a pattern that is utilized to suggest a more stringent upper limit for the square conjecture.

Pattern JMS

A stricter bound to the distinct square conjecture is proposed in [44] along with the JMS pattern. It is the simplest pattern with the structure shown below.

$$w_{jms}(x) = a^1ba^2b..a^xb$$

The following relations calculate the length and the distinct squares count of a JMS word, respectively.

$$|w_{jms}(x)| = \frac{x^2 + 3x}{2} \tag{6.3}$$

$$DS(w_{jms}(x)) = \begin{cases} \frac{x^2}{2}, & \text{if } x \mod 2 = 0 \\ \frac{x^2-1}{2}, & \text{otherwise} \end{cases}$$

Similar to pattern Q, words generated by the pattern JMS have no FS-double squares. The words of the pattern, therefore, follow the next relation.

$$|w_{jms}(x)| = DS(w_{jms}(x)) + NG(w_{jms}(x))$$

The number of no-gain locations in $w_{jms}(x)$ can be computed using the following relation.

$$NG(w_{jms}(x)) = \lfloor \frac{3x}{2} \rfloor$$

Every word generated by a pattern have no-gain locations. The difference between no-gain locations of consecutive words gives the number of newly added no-gain locations, which can be computed using the following equation.

$$NG(w_{jms}(x+1)) - NG(w_{jms}(x)) = \begin{cases} 2, & \text{if } (x+1) \mod 2 = 0 \\ 1, & \text{otherwise} \end{cases}$$

Given a pattern $T(x)$ accepting an integer value x to generate a corresponding pattern-generated word. Define $newZeroes(T(x+1))$ by the difference between $NG(T(x+1))$ and $NG(T(x))$. It is used to analyze the rate of no-gain locations in the pattern-generated words as x increases. It is further used to compare dense patterns defined later in this chapter.

Below are some existing patterns that possess specific characteristics of words. Although they do not necessarily produce words with the maximum number of distinct squares, they are included here for the sake of completeness on the topic.

Pattern generating words with FS-double squares

The words that have a sequence of FS-double squares in their prefixes are studied in [14]. The following pattern generated words as mentioned.

$$w_{FSD}(x) = (a^{x-1}baa^{x-1}ba^{x-1}ba)^2a^{x-1}$$

The next expressions can be used to count the number of distinct squares for the above words. We can use the following equation if x is an even number.

$$DS(w_{FSD}(x)) = 4.5x + 1$$

In the case of an odd value of x, the number of distinct squares in $w_{FSD}(x)$ can be computed as shown below.

$$DS(w_{FSD}(x)) = 4.5x + 0.5$$

It has been computed in the same work mentioned above that the square density of words generated by the pattern is always less than $\frac{5}{6}$.

Pattern maximizing distinct squares in circular words

Another study on the number of distinct squares in a circular word conducted in [5] predicts the number of distinct squares in a circular word of length n

is at most $3.14n$. It considers the number of distinct squares in all cyclic rotations of an n length word. The lower bound for this problem if found by proposing the following structure of the words.

$$w_c(x) = a(ba)^{x+1}a(ba)^{x+2}a(ba)^{x+1}a(ba)^{x+2}$$

The number of distinct squares in the word obtained from the above pattern and in all of its conjugates is counted. It is shown that the number of distinct squares, in this case, is $10x + 16 - (x \mod 2)$.

The following section explores the words containing the maximum number of distinct squares for their respective lengths. This exploration aims to use the characteristics of such words to incorporate a pattern to generate the best possible words.

6.4 Square Maximal Words

It is conjectured that the maximum number of distinct squares is achieved for a binary alphabet [54]. In the rest of the chapter, we assume that the underlying alphabet is binary, containing letters 'a' and 'b'. The exact characterization of the set $MaxDS(n)$ is unknown. In other words, to check if a word belongs to the set $MaxDS(n)$, it is required to exhaustively search in the set of all 2^n possible words. A word is called a square-maximal word if it belongs to a set $MaxDS(n)$.

The function $MaxNS$ is non-decreasing, and the difference between two successive values of $MaxNS$ is at most two. Let $n \in \mathbb{N}$ be a no-gain length. Then, for any word $w \in MaxDS(n-1)$, both lw and wl are in $MaxDS(n)$, where $l \in \{a, b\}$. Thus, for a no-gain length n, the cardinality of the set $MaxDS(n)$ is always more than that of the set $MaxDS(n-1)$. Thus, it is important to characterize words for gain lengths because it enables generating words with no-gain lengths. For this reason, some of the following results are obtained for the gain lengths.

Consider a word w in $MaxDS(n)$, where n is a gain length. Every letter in the word w must be part of some square in the set $DS(w)$. Otherwise, removing a letter that is not part of any square will give a smaller word with $MaxNS(n)$ squares, which is not possible as n is a gain length. Suppose $w = w_1w_2$ for some $w_1,\ w_2 \in \Sigma^+$. We refer to the squares whose last occurrence starts and ends in w_1 and w_2, respectively, by $cross(w_1, w_2)$.

Lemma 6.1. *Let $n \in \mathbb{N}$ be a gain length and w be a word in $MaxDS(n)$. Further, let $w = w_1w_2$ where $w_1,\ w_2 \in \Sigma^*$ such that $|DS(w)| = |DS(w_1)| + |DS(w_2)|$. Then, $|DS(w_1) \cap DS(w_2)| = k \geq 0$ if and only if $|cross(w_1, w_2)| = k$.*

Proof. We have, $|DS(w)| = |DS(w_1| + |DS(w_2)| - |DS(w_1) \cap DS(w_2)|$

(if) Suppose the sets $DS(w_1)$ and $DS(w_2)$ have k squares in common. Then, it must be the case that at least k rightmost squares starts in w_1 and ends in w_2 to satisfy the premise $|DS(w)| = |DS(w_1)| + |DS(w_2)|$.

(only if) Suppose there are k rightmost squares that begin and end in words w_1 and w_2, respectively. Define

$$sq_{w_1} = DS(w_1) - DS(w_2), \quad sq_{w_2} = DS(w_2) - DS(w_1),$$
$$sq_{comm} = DS(w_1) \cap DS(w_2), \quad sq_{cross} = DS(w) - (DS(w_1) \cup DS(w_2))$$

Note that sq_{cross} is the set of distinct squares that begin in w_1 and end in w_2 which are not present in w_1 or w_2. So, the number of distinct squares in w is, $|DS(w)| = sq_{w_1} + sq_{w_2} + sq_{comm} + sq_{cross}$. Since $|DS(w_1)| + |DS(w_2)| = sq_{w_1} + sq_{w_2} + 2 * sq_{comm}$, we get $|sq_{cross}| = |sq_{comm}|$

□

Lemma 6.2. *Let a word $w = w_1w_2 \in MaxDS(n)$ for some gain length $n \in \mathbb{N}$ such that $|w_1| \leq |w_2|$. If $|cross(w_1, w_2)| = 0$, then the length of the smallest border of the word w is greater than $|w_1|$.*

Proof. The condition $|cross(w_1, w_2)| = 0$ implies $|DS(w_1w_2)| = |DS(w_2w_1)|$. No proper suffix of the word w_1 can be a prefix of the word w_2. Otherwise, $w_1 = w_1'u$ and $w_2 = uw_2'$ will imply $DS(w_1'u.uw_2') = DS(w_1'uw_2')$ for some non-empty words w_1', w_2', u. Similarly, no proper suffix of the word w_2 can be a prefix of the word w_1. Therefore, the length of the smallest border of the words w_1w_2 and w_2w_1 must be greater than $|w_1|$. □

Lemma 6.3. *Let $n \in \mathbb{N}$ be a gain length and w be a word in $MaxDS(n)$ with $w = w_1w_2$ such that $|DS(w)| = |DS(w_1)|+|DS(w_2)|$ and $DS(w_1) \cap DS(w_2) = \emptyset$. Then, $\{a^2, b^2\} \subseteq DS(w)$.*

Proof. Assume $DS(w)$ contains at most one square from $\{aa, bb\}$. If both a^2 and b^2 are not in w then the word w must be of the form $(ab)^k$ for some positive integer $k > 0$. However, such a word cannot be in $MaxDS(n)$.

Consider the case in which only one among a^2 or b^2 is a factor of w. Without loss of generality, assume $aa \notin DS(w_1)$, $aa \notin DS(w_2)$ and $bb \in w$ implying a square bb is either in set $DS(w_1)$ or $DS(w_2)$. If b^2 is a factor of w_1, then the structure of w_2 depends on the initial and final letters of w_1. As a result, w_2 is either $b(ab)^j$ or $(ab)^j$ for some $j \in \mathbb{N}$. If the factor w_1 starts and ends with the same letter, say 'a' then $w_1 = aua$ where $u \in \Sigma^+$ and $w_2 = b(ab)^j$. However, as the word $w_3 = (ab)^j.a$ contains $|DS(w_2)|$ distinct squares, the word $w' = w_1w_3$ will also be in the set $MaxDS(|w|)$. Here w_3 does not satisfy the constraint on border given in Lemma 6.2. Therefore, the length $|w|$ must be a no-gain length, which is a contradiction.

Now, suppose w_1 starts and ends with different letters, say $w_1 = aub$. In this case, the only possible structure for w_2 is $(ab)^j$. Similar to the previous case, we have a contradiction as the word $aub.b(ab)^{j-1}$ will also be in the set $MaxDS(|w|)$. □

Lemma 6.4. *Let $n \in \mathbb{N}$ be a gain length and w be a word in $MaxDS(n)$ with $w = w_1w_2$ such that $|DS(w)| = |DS(w_1)|+|DS(w_2)|$ and $DS(w_1) \cap DS(w_2) =$*

$\emptyset$. Then, for some integer $k > 2$ and $u_1, u_2 \in \Sigma^+$, we have $w_1 = a^k u_1 a^k$ and $w_2 = b^k u_2 b^k$. Further, the factorization of a word w as $w_1.w_2$ is unique.

Proof. We know from Lemma 6.3 that, for a given gain length, any word containing the maximum number of distinct squares must have the trivial squares, viz. a^2 and b^2. These squares can be in either w_1 or in w_2. Accordingly, we consider two cases for w_1 and w_2 depending on whether they start with the same or the different letters. We now see that among all possible structures, only one structure mentioned in Case II satisfies all the given conditions.

Case I Assume w_1 and w_2 start with the same letter. Suppose the words a^2 and b^2 are in the set $DS(w_1)$. Then, these trivial squares cannot be in the set $DS(w_2)$. Thus, the word w_2 must be of the form $(ab)^k$ for some integer $k > 2$. Also, the square $(ba)^2$ cannot be in the set $DS(w_1)$, otherwise, the square $(ab)^2$ will also be in the set $DS(w_1)$. So the squares $(ab)^2$ and $(ba)^2$ cannot be in the word w_1. But then we have another word $w' = w_1.(ba)^k$, which has $|DS(w)|$ number of distinct squares implying that w is a no-gain length since the word w' has a border whose length is less than $|w_1|$ which contradicts the assumption.

Consider another alternative wherein the trivial squares aa and bb are factors of the words w_1 and w_2, respectively. Then the factors w_1 and w_2 must end and begin with 'ab', which again does not satisfy the assumption that n is a gain length.

Case II Suppose w_1 and w_2 start with two different letters. To satisfy the constraint on border mentioned in Lemma 6.2, suppose w_1 begins and ends with the letter 'a'. Let $a^2 \in DS(w_2)$, then the word w_1 begins with ab and the word w_2 ends with ab, thereby the word $w_2 w_1$ has a border of length smaller than $|w_1|$. Therefore, it must be the case that the structures of the word w_1 and w_2 are $a^k u_1 a^k$ and $b^k u_2 b^k$, respectively, such that $k > 1$ and $DS(u_1) \cap DS(u_2) = \emptyset$. Note that if there is a factorization of the factor w_1 as $w_{11}w_{12}$ such that $cross(w_{11}, w_{12}) = \emptyset$, then a^2 will be a factor of w_{11} and not of w_{12}. However, every factor w_{12} of length more than one will always end with an a^2. So, the factor $w_1 = w_{11}w_{12}$ cannot have $cross(w_{11}, w_{12}) = \emptyset$. Now, another factorization $w = w_3w_4$ that satisfies the condition $|DS(w)| = |DS(w_3)| + |DS(w_4)|$ is possible if $w_3 = a^k u 1 a^{k-1}$ and $w_4 = a.b^k u_2 b^k$. However, in such a case, by Lemma 6.2, n will be a no-gain length. Therefore, the factorization of w as $w_1.w_2$ is unique.

□

Let Σ be an alphabet and $w = l_1 \cdot l_2 \cdot \ldots \cdot l_n$ be a word, where $l_i \in \Sigma$ for $1 \leq i \leq n$. The letters l_1 and l_n as terminal letters of w. Any letter that is not a terminal letter is a non-terminal letter of w. In Lemma 6.4, it is shown that a square-maximal word, say $w = w_1w_2$ of length n can have at most one w_1 that satisfy $cross(w_1, w_2) = \emptyset$. We now explore a gain length n for which $MaxNS(n) > MaxNS(n_1) + MaxNS(n - n_1)$ and identify the structure of a square-maximal word.

Lemma 6.5. *Let $n \in \mathbb{N}$ such that $MaxNS(n) > MaxNS(n_1)+MaxNS(n-n_1)$ for some integer $n_1 \in \{1, \ldots, n-1\}$ and $w \in MaxDS(n)$. The following statements hold:*

(a) $|DS(w)| > |DS(w_1)| + |DS(w_2)|$ *for all w_1, w_2 such that $w = w_1w_2$.*

(b) $cross(w_1, w_2) \neq \emptyset$ *for all w_1, w_2 such that $w = w_1w_2$.*

Proof. (a) Suppose $|w_1| = n_1$. Then, the maximum value of the expression $|DS(w_1)| + |DS(w_2)|$ is $|MaxDS(n_1)| + |MaxDS(n - n_1)|$. As $MaxNS(n) > MaxNS(n_1) + MaxNS(n-n_1)$, the relation $|DS(w)| > |DS(w_1)| + |DS(w_2)|$ follows.

(b) We conclude from (a) that the rightmost square starts in w_1 and ends in w_2 for all factors w_1. So, the set $cross(w_1, w_2)$ is non-empty for every w_1.

□

The following lemma inspects the characteristics of terminal letters in a square-maximal word.

Lemma 6.6. *Let $n \in \mathbb{N}$ be a gain length and $w \in MaxDS(n)$.*

(a) If $MaxNS(n) = MaxNS(n-1) + 1$, then every terminal letter of w is the terminal letter in exactly one rightmost square of w.

(b) If $MaxNS(n) = MaxNS(n-1) + 2$, then w begins with an FS double square and the last letter of w is a terminal letter of two squares in $DS(w)$.

Proof. As n is a gain length, $MaxNS(n) - MaxNS(n-1) = i \in \{1, 2\}$. The first letter of a word $w \in MaxDS(n)$ must be a part of exactly i distinct squares. Otherwise, removing the first letter will result in a word of $n-1$ length containing more than $MaxNS(n-1)$ distinct squares, which is not feasible. A similar argument applies to the last letter of a word w. □

We can observe, from Lemma 6.6, that the square-maximal words for successive gain lengths always begin and end with a square. It has been observed in the manual inspection of square-maximal words for lengths up to 40 that if such a word ends with the longest primitive square, then extending it further with the prefix of its square base results in a longer square-maximal word. The following lemma explains one such way to introduce new square(s) using a prefix of a square base.

Lemma 6.7. *Let $w = uu$ be a primitive base square such that $|u| > 1$ and v be a proper prefix of u. Then, $|DS(w.v)| \geq |DS(w)| + |v|$.*

Proof. Assume, $u = u_1 u_2 ... u_n$. A square uu has all conjugates of u, and every letter of the first u in w begins with a distinct conjugate. Similarly, for a word $uu.v$, every letter of the word v adds a new square, that is, a conjugate of uu.

The number of new squares added by a prefix v is more than $|v|$ if the word begins with an FS double square. An FS double square is a primitive square that begins with two rightmost squares. As explained in the previous chapters, the structure of the longer root of an FS double square is known to be $(xy)^{p_1}(x)(xy)^{p_2}$. Here, x and $y \in \Sigma^+$ and the integers p_1, p_2 satisfy $p_1 \geq p_2 \geq 0$. A word beginning with an FS double square introduces $|v|+|v'|$ new distinct squares for some non-empty longest common prefix v' of the words x and y, where $|v'| \leq |v|$. □

We have seen the definition of a pattern and some existing patterns in the previous section. A pattern is a way to represent a family of words sharing similar characteristics. The following section employs the properties of square-maximal words identified above to define a dense pattern.

Dense patterns

We use the notation $T(x)$ to denote a function from $\mathbb{N} \to \Sigma^+$. For example, $T(x) = a^x b$ generate words $\{b, ab, a^2b, a^3b, \ldots\}$. We refer to a function $T(x)$ as a pattern. For a word, w, the square density, $\alpha(w)$ is defined as the ratio $\frac{|DS(w)|}{|w|}$ and it is known that no upper bound on $\alpha(w)$ is sharp [54]. We extend the definition of the square density to a pattern, $T(x)$, and define it as

$$\alpha_T = \lim_{x \to \infty} \frac{|DS(T(x))|}{|T(x)|}$$

The square density of a pattern depends on the number of no-gain lengths between two successive words generated by the pattern. A high square density indicates more gain lengths or equivalently fewer no-gain lengths. The difference between lengths of successive words generated by a pattern need not be a constant. For every positive integer, x, a pattern $T(x)$ introduces $|DS(T(x))| - |DS(T(x-1))|$ new squares. The number of no-gain lengths introduced in $T(x-1)$ to obtain $T(x)$ is defined as $\mathcal{N}_T(x) = (|T(x)| - |T(x-1)|) - (|DS(T(x))| - |DS(T(x-1))|)$. A good pattern should minimize the value of $\mathcal{N}_T(x)$. In Section 6.5, we use $|DS(T(x))|$ and $\mathcal{N}_T(x)$ to compare different patterns.

As mentioned before, the aim of this chapter is to characterise the words in the set $MaxDS(n)$, where n is a gain length. To do so, the properties of square-maximal words are used. A word $w = w_1w_2 \in MaxDS(n)$ satisfies either $|DS(w)| > |DS(w_1)| + |DS(w_2)|$ or $|DS(w)| = |DS(w_1)| + |DS(w_2)|$. For the latter case, note that $cross(w_1, w_2) = \phi$, else $|DS(w)| > |DS(w_1)| + |DS(w_2)|$. For this case, Lemma 6.4 shows that such a word has a unique factorization where the factor $w_1 = w_{11}w_{12}$ always satisfies the relation $|DS(w_1)| > |DS(w_{11})| + |DS(w_{12})|$. The only possible structures for this case have $w_1 = a^k u_1 b^k$ and $w_2 = b^k u_2 b^k$. These factors cannot have any squares in common, so given a factor w_1, it is easy to find w_2. Therefore the relation $|DS(w)| > |DS(w_1)| + |DS(w_2)|$ is used to obtain a dense pattern. The interpretation of the relation given in Lemma 6.5 is included in the next definition.

Definition 6.1 (Dense Pattern). *A pattern, T, is considered dense if and only if it satisfies the following conditions.*

(a) $\alpha_T \geq 1$, and

(b) For all $x \in \mathbb{N}$, if $T(x) = w_1w_2$ then $cross(w_1, w_2) \neq \emptyset$, where w_1, $w_2 \in \Sigma^+$.

A word produced by a dense pattern is known as a dense word. The next lemma provides an aid to verify the second condition in Definition 6.1.

Lemma 6.8. *Let $w = u_1u_2 \ldots u_k$ be a word such that for all $i \in \{2, \ldots, k-1\}$, the factor $sf_{i-1}.u_i.pr_{i+1}$ is a rightmost square in $DS(w)$ and u_1pr_2 ($sf_{k-1}u_k$)*

is the first (the last) rightmost square of w *for some non-empty prefix and suffix,* pr_i *and* sf_i*, respectively, of* u_i*. Then,* $cross(w_1, w_2) \neq \emptyset$ *for all* w_1*,* w_2 *such that* $w = w_1 w_2$*.*

Proof. The word w begins and ends with a rightmost square. For $1 < i < k$, the structure of a rightmost square $sf_{i-1}.u_i.pr_{i+1}$ ensures that every non-terminal letter in w is also a non-terminal letter in any rightmost square of w. Thus, for all w_1 and w_2 such that $w = w_1 w_2$ implies $cross(w_1, w_2) \neq \emptyset$. □

Now, a pattern P is defined as follows:

$$P(x) = a.(a^1ba^2b...a^y).\left\{\bigodot_{i=y-1}^{x-2}(ba^iba^{i+1}ba^{i+2})\right\}.(ba^{x-1}ba^xba^{x-1}bab)a \quad (6.4)$$

where x and y are positive integers and $y = \lceil \frac{x}{2} \rceil \geq 4$. Similar to the squares in a word obtained from the pattern Q described in [34], the words generated by the pattern P have three types of distinct squares. These are (i) trivial squares having only letter a, (ii) squares with exactly two b's, and (iii) squares with exactly four b's. All the squares in the last two types are primitive squares.

x mod 2	$\|P(x)\|$	$\|DS(P(x))\|$
0	$\frac{1}{8}(10x^2 + 36x + 40)$	$\frac{1}{8}(10x^2 + 20x + 24)$
1	$\frac{1}{8}(10x^2 + 32x + 38)$	$\frac{1}{8}(10x^2 + 16x + 22)$

Table 6.2: *Properties of the words generated by the pattern P*

A set of primitive squares and all their conjugates in the word has the structure described in Lemma 6.7. Refer to Table 6.2 for the length and the number of distinct squares in a word that can be obtained by the pattern P. We check the pattern against the definition of dense patterns. For this, we first verify that the factors of the words generated by this pattern satisfy the criterion (b) of the Definition 6.1.

Lemma 6.9. *The following factor, w, of a word generated by the pattern P satisfies cross*$(w_1, w_2) \neq \emptyset$ *for any w_1, w_2, where $w = w_1w_2$, $y \geq 4$ and $x > 6$.*

(a) $w = a^1ba^2b \dots a^y$

(b) $w = \left\{ \bigodot_{i=y-1}^{x-2} (ba^iba^{i+1}ba^{i+2}) \right\} (ba^{x-1}ba^xba^{x-1}bab)a$

Proof. (a) Consider a set of squares $S = \{(aba)^2, (abaa)^2, (abaaa)^2 \dots, (aba^{y-2})^2\}$. We can recreate w using the squares from S as the word described in Lemma 6.8. Note that, the squares here are the rightmost. Thus, the relation holds true for given w.

(b) Consider the rightmost instances of the squares in a subset R of $DS(w)$, where $R = \{(a^iba^{i+1}ba^2)^2, (a^{i+1}ba^{i+2}ba^2)^2, ..., (a^{x-1}ba^xba^2)^2, (ba^i)^2, (ba)^2\}$. We can use the squares in R to rewrite the word in the structure mentioned in Lemma 6.8 in which w begins and ends with squares $(ba^i)^2$ and $(ba)^2$, respectively.

□

Lemma 6.10. *The pattern P is a dense pattern.*

Proof. Refer to Table 6.2 to count the square density of the pattern P as follows.

$$\alpha_P = \lim_{x \to \infty} \frac{|DS(P(x))|}{|P(x)|} = \lim_{x \to \infty} \frac{10x^2 + 20x + 24}{10x^2 + 16x + 22} = 1$$

The pattern satisfies the first condition given in Definition 6.1. Equation (6.4) shows that the word generated by the pattern P is the concatenation of two factors given in Lemma 6.9. According to this lemma, these factors individually qualify the last condition of the Definition 6.1. We use the same subset of rightmost squares to write $P(x)$ according to Lemma 6.8. Thus, P is a dense pattern. □

It is possible to modify a pattern to convert it into a dense pattern. Accordingly, the structure of some existing patterns is changed to make them dense patterns. It is discussed later in Section 6.5. Before that, it is first shown that infinitely many dense patterns exist using a pattern generator.

Theorem 6.11. *There are infinitely many dense patterns.*

Proof. The following pattern generator generates infinitely many dense patterns.

$$Gen(x, y) = a.(a^1b.a^2...b.a^y). \left\{ \bigodot_{k=(y-1)}^{x-2} (ba^kba^{k+1}ba^{k+2}) \right\} .(ba^{x-1}b.a^x)(ba^{x-1}b.ab).a \tag{6.5}$$

where $x, y \in \mathbb{N}$ such that $3 \leq y \leq (x - 3)$. Every value of y gives a different pattern, and we use $G_3, G_4, ..., G_y$ to denote these patterns.

$$|G_y(x)| = \frac{1}{2}(3x^2 + 9x - 2y^2 + 10) \tag{6.6}$$

$$|DS(G_y(x))| = \frac{1}{2}(3x^2 + 4x - 2y^2 + 2y + 6 - (x \mod 2)) \tag{6.7}$$

Equation (6.6) and (6.7) show that the square density of each G_y is one. Further, the factors of any of these patterns are as given in Lemma 6.9. Thus, G_y is a dense pattern. □

Note that every G_y supports the 'stronger' square conjecture [44]. Also, it is possible to get more dense patterns by replacing the letters (a, b) in the generator explained in Theorem 6.11 with certain words. The discussion on these patterns continued where the existing best-known patterns are compared with the pattern P.

6.5 Pattern P vs. Existing Patterns

The patterns described in Section 6.3 have varying square densities. The Definition 6.1 is used to verify existing patterns for a dense pattern. Accordingly, the square density of a pattern must approach one. This condition leads to omitting the patterns of lower densities. Patterns given in [5, 14] have a square density of less than one, while the square density of patterns in [34, 44] approaches one. So, only patterns Q and JMS are verified against the definition of a dense pattern. Pattern Q is defined in [34] as follows.

$$Q(x) = \bigodot_{i=2}^{x} a^i b a^{i-1} b a^i b$$

The square density of pattern Q is computed below using the Equations (6.8) and (6.9).

$$|Q(x)| = \frac{1}{2}(3x^2 + 7x - 10) \tag{6.8}$$

$$|DS(Q(x))| = \frac{1}{2}(3x^2 + 2x - 10 - (x \mod 2)) \tag{6.9}$$

$$\alpha_Q = \lim_{x \to \infty} \frac{|DS(Q(x))|}{|Q(x)|} = 1 \tag{6.10}$$

For $|w_1| = 1$, the word $Q(x) = w_1 w_2$ satisfy $cross(w_1, w_2) = \emptyset$. So, Q is not a dense pattern. However, the pattern, Q', obtained by removing the first letter from $Q(x)$ makes it a dense pattern.

$$Q'(x) = (ababa^2b). \bigodot_{i=3}^{x} a^i b a^{i-1} b a^i b$$

We have $|DS(Q'(x))| = |DS(Q(x))|$ and $|Q'(x)| = |Q(x)| - 1$, therefore, the square density of Q' is one. A set of rightmost squares as mentioned in Lemma 6.8 exists for Q', that is, $R \subset DS(Q')$ where

$$R = \{(ab)^2, (aba^2baa)^2, (aba^3baa)^2, \ldots, (aba^{i-1}baa)^2, (a^{i-1}ba^ib)^2\}$$

Hence, the pattern Q' is a dense pattern.

A stricter bound for the number of distinct squares is conjectured in [44] and is supported by a pattern called JMS. It is a simple pattern with the structure:

$$JMS(x) = \bigodot_{i=1}^{x} a^i b$$

The following equations give the length and the number of distinct squares in $JMS(x)$.

$$|JMS(x)| = \frac{1}{2}(x^2 + 3x) \tag{6.11}$$

$$|DS(JMS(x))| = \frac{1}{2}(x^2 - 2 - (x \mod 2)) \tag{6.12}$$

The square density of the pattern obtained with the above equations is $\alpha_{JMS} = 1$. Similar to $Q(x)$, a word $JMS(x) = w_1w_2$ satisfies $cross(w_1, w_2) = \phi$ for $|w_2| = 1$. We remove the last letter of $JMS(x)$ to get a word that satisfies the condition (b) of Definition 6.1:

$$JMS'(x) = \left\{ \bigodot_{i=3}^{x-1} a^i b \right\} .a^x$$

The square density of pattern JMS' is one, and we can use Lemma 6.9 to show that it is a dense pattern. Both the patterns Q' and JMS' construct words using the same principle to increase the number of distinct squares. They maximize the distinct primitive squares to achieve a higher square density, as mentioned in Lemma 6.7. Let us see a criterion to compare the dense patterns.

Comparing dense patterns

Two dense patterns Q' and JMS' are obtained from the existing patterns. Also, the newly proposed pattern P met all the conditions defined for a dense pattern. A pattern that reaches its square density quickly is the best pattern. It is evident that if a pattern introduces a lot of no-gain lengths between its successive words, then it will move slowly towards its density. Therefore, a notation β_T is used to determine the rate of a pattern T to arrive at its square density. The notation is valid for a pattern that has at least one no-gain length between its successive words.

Definition 6.2 (Gain lengths per no-gain length)**.** *Let $x \in \mathbb{N}$. The term $\beta_T(x)$ is the ratio of number of distinct squares in $T(x)$ that are not in $T(x-1)$*

to the number of no-gain lengths between $T(x)$ and $T(x-1)$, that is,

$$\beta_T(x) = \frac{|DS(T(x))| - |DS(T(x-1))|}{\mathcal{N}_T(x)}$$

We get $\beta_{Q'}(x)$ and $\beta_{JMS'}(x)$ from Equations (6.8), (6.9), (6.11) and (6.12) as follows:

$$\beta_{Q'}(x) = \frac{3x}{2} \text{ or } \frac{3x-1}{3} \qquad \text{and} \qquad \beta_{JMS'}(x) = \frac{x}{1} \text{ or } \frac{x-1}{2} \tag{6.13}$$

Lemma 6.12. *For all positive integers $x > 4$, there exists $y \in \mathbb{N}$ with $|JMS'(x)| > |Q'(y)|$ and $|DS(JSM'(x))| < |DS(Q'(y)|$.*

Proof. The statement holds since $\beta_{JMS'}(x) < \beta_{Q'}(x)$ (see Equation (6.13)). □

Theorem 6.13. *Pattern P is the lower bound for $MaxNS(n)$.*

$\boldsymbol{\beta_T(x)}$	$\boldsymbol{x \mod 2 = 0}$	$\boldsymbol{x \mod 2 = 1}$
$\beta_{Q'}(x)$	$1.5x$	$x - 0.33$
$\beta_P(x)$	$1.5x + 0.5$	$x + 0.5$

Table 6.3: *New distinct squares per new no-gain length*

Proof. The Lemma 6.12 shows that the pattern Q' is better than the pattern JMS'. The β values of patterns Q' and P are listed in Table 6.3. It shows that the rate of approaching the square density of pattern P is faster than that of Q'. □

Corollary 6.14. *For every word, $Q'(x)$, there exists a word, $P(y)$, such that $|Q'(x)| > |P(y)|$ and $|DS(Q'(x))| < |DS(P(y))|$ where x and $y \in \mathbb{N}$ and $x > 5$.*

6.6 Patterns to Generate FS-double Squares

The square density of a word depends on the arrangements of the squares in it. So, finding the favourable distribution of squares is necessary to maximize the square density. In this regard, the words generated by patterns Q [34] and P are the best words obtained so far. Note that the lower bound for the square conjecture is a family of words obtained from a pattern Q that is then generalized to get a better pattern, pattern P.

A pattern represents a group or a family of words having some similar structure. It can be seen as a function that accepts an integer and produces a unique word. For example, consider the following pattern N.

$$N(i) = a^i b \bigodot_{t=1}^{i} (ab^t)$$

The symbol $\bigodot$ is used to concatenate the factors that are grouped by the parenthesis. The pattern N generates the words *abab* and *aabababb* for $i = 1$ and $i = 2$, respectively.

Highest square density of words with 2FS squares

We see a structure to obtain words containing a sequence of FS-double squares. The pattern D generates a word $D(i)$ containing 'i' consecutive FS-double square. The structure of pattern D is given below.

$$D(i) = aa \left\{ \bigodot_{t=2}^{\lceil \frac{i}{2} \rceil} (ba^t) \bigodot_{k=(\lceil \frac{i}{2} \rceil - 1)}^{i-2} (ba^k ba^{k+1} ba^{k+2}) \right\} (ba^i ba^{i-1} ba^i)^2 \qquad (6.14)$$

In a word $D(i)$, the highest exponent of a letter 'a' is i. The suffix $(a^{i-1}b)a^iba^{i-1}ba^iba^iba^{i-1}ba^i$ starts with $|i|$ consecutive FS-double squares. The first FS-double square is $(a^{i-1}ba^iba^{i-1}ba)^2$ and is appended by the prefix a^{i-1} to get $|i| - 1$ conjugates. Comparing the structure with the definition of FS-double square given in Lemma 3.2, we get $x = a^{i-1}b, j = a, p_1 = p_2 = 1$.

The value of $|lcp(xy, yx)|$ is $(i-1)$ which is required to introduce new $(i-1)$ conjugates. Some examples of pattern D words with their s_i sequences are listed below.

$$
\begin{aligned}
D(5) =& aabaabaaabaabaaabaaaabaaabaaaabaaaaabaaaaabaaaabaaaaa \\
& baaaaabaaaabaaaaa \\
s_i(D(5)) =& 1111110111111111110111111111111112222200000000000111111 \\
& 0011111100000001010 \\
D(6) =& aabaabaaabaabaaabaaaabaaabaaaabaaaaabaaaabaaaaabaaaaaab \\
& aaaaaabaaaaabaaaaaabaaaaaabaaaaabaaaaaa \\
s_i(D(6)) =& 1111110111111111110111111111111101111111111111111111222222 \\
& 0000000000001111111001111111000000101010
\end{aligned}
$$

The size of a word generated by a pattern D and the number of distinct squares in it can be computed using Equation (6.15) and (6.16), respectively.

$$
|D(i)| = \frac{1}{2}(3i^2 + 15i - 2y^2 + 8) \tag{6.15}
$$

$$
\begin{aligned}
|DS(D(i))| &= \frac{3i^2 + 11i}{2} - (y^2 + 2y + 2 + i \mod 2) \\
&= \frac{2i^2 + 10i - 2}{2}, \frac{2i^2 + 8i - 10}{2}
\end{aligned} \tag{6.16}
$$

Lemma 6.15. *The distinct square density of pattern D is one.*

Proof. From Equations (6.15) and (6.16), the distinct square density of pattern D is one.

$$
\alpha_D = \lim_{i \to \infty} \frac{|D(i)|}{|DS(D(i))|} = 1 \tag{6.17}
$$

□

According to Theorem 4.3, an FS-double square $SQ_1^2 = ((xy)^p(x)(xy)^p)^2$ where $p > 2$ contains more than $(p-2)|xy|$ no-gain locations and $\rho(SQ_1^2) < \lfloor \frac{11}{6} \rfloor$. Similarly, a word w with a sequence of FS-double squares where these squares have structures similar to that of SQ_1^2 has $\rho(w) < 1.64$ (refer Theorem 4.7). The structure of FS-double squares is embedded in which the values of exponents are set to one to increase the square density of words by minimizing no-gain locations. However, a sequence of no-gain locations is observed at some locations that are not explored in Theorem 4.3. Further investigation of an FS-double square will assist in concluding that every word with k FS-double squares contains k no-gain locations.

6.7 Conclusions

In this chapter, the objective was to identify a pattern that could generate words with the maximum number of distinct squares. The focus was on studying square maximal words to derive properties that could increase square density. In such words, it was observed that extending a square with its root generates a conjugate of the square for each newly added letter. Based on various properties of square maximal words, a new term, "dense pattern", was defined, and a new dense pattern, pattern P, was proposed. The chapter also presented a pattern generator to produce infinitely many dense patterns. Different patterns in the literature are studied, and some existing patterns were modified to meet the requirements of a dense pattern. They were then compared with a pattern P. The chapter concluded that the pattern P is the new lower bound for the square conjecture. Since primitive squares have the maximum number of conjugates, it is conjectured that such a structure would have the maximum number of distinct primitive squares. The proposed structure for building dense patterns introduces at most one distinct square per letter. The chapter also examined other patterns that generate words containing FS-double squares occurring at consecutive positions. The square density of all these patterns is one.

7

Antisquares

This chapter explores a different form of repetition compared to the previous works. It focuses on a specific type of repetition known as an "antisquare", which is defined as a binary word of the form $u\bar{u}$, where $\bar{u}$ is obtained by complementing the letters of a non-empty word u. While antisquares have been explored for infinite words in previous literature [12], this chapter investigates their properties in finite words. The term $u^{\bar{2}}$ is used to refer to the antisquare where $\bar{2}$ indicates that the given word is concatenated with its complement. As mentioned in the previous chapters, a notation $s_i(w)$ determines the number of distinct rightmost squares of w beginning at location i. So, the notation $\bar{s}_i(w)$ is used to represent the number of distinct rightmost antisquares starting at location i. The set $DA(w)$ is the set of all distinct antisquares in w. The next section investigates the properties of antisquares similar to those of distinct squares.

7.1 Properties of Antisquares

In the context of squares, we discussed various counting problems, such as determining the number of repeated squares, repeated primitive squares, distinct squares and distinct primitive squares in words. Similarly, we now focus on examining analogous problems concerning antisquares in words. In the following lemma, we obtain an upper bound on the number of repeated antisquares in words.

Lemma 7.1 counts the upper limit for the number of repeated antisquares in a word.

Lemma 7.1. *The number of antisquares in any binary word w is at most* $\frac{|w|^2+2|w|}{4}$.

Proof. The mid location of an antisquare $u\bar{u}$ is the location $|u|$. Let $w = a_1a_2\ldots a_n$. A word location i can be the mid location of at most i different length antisquares where $i \in \left[1, \left\lfloor\frac{n}{2}\right\rfloor\right]$. To understand this, consider the word $aabb$. The second location of the word is a mid location for two antisquares $\{ab, aabb\}$. Likewise, every location j where $j \in \left[\left\lfloor\frac{n}{2}\right\rfloor + 1, \ldots, n\right]$ can be the mid location of at most $n - j$ different length antisquares. So, the total number of antisquares in a word is less than or equal to $2(1+2+3+\ldots+\frac{n}{2}) = \frac{n^2+2n}{4}$. □

In Chapter 6, we discussed words with the highest square density and the occurrences of distinct primitive squares in such words. It was demonstrated in Chapter 5 that increasing the number of specific primitive squares could help to pack distinct squares compactly. The following lemma identifies the structure of primitive and non-primitive antisquares. Note that a primitive antisquare is a word that is both primitive and an antisquare.

Lemma 7.2. *Let $x^{\bar{2}}$ be an antisquare. The antisquare is non-primitive if and only if $x = (w\bar{w})^i w$, where $w \in \Sigma^+$ and an integer $i > 0$.*

Proof. Consider an antisquare $x\bar{x} = u^k$ where $x, u \in \Sigma^+$ and integer $k > 1$. If k is even, then both x and $\bar{x}$ begins with u. This is a contradiction. If k is odd, we get $x = u^{\frac{k-1}{2}}u_1$ where $u = u_1u_2$. This gives $\bar{x} = u_2u^{\frac{k-1}{2}}$ and following set of relations.

$$\begin{aligned} u^{\frac{k-1}{2}}u_1 &= \bar{u_2}\bar{u}^{\frac{k-1}{2}} \\ u^{\frac{k-1}{2}}u_1 &= (\bar{u_2}\bar{u_1})^{\frac{k-1}{2}}\bar{u_2} \\ \implies |u_1| &= |u_2|, u_1 = \bar{u_2} \end{aligned}$$

We get the structure of the antisquare as $(w\bar{w})^iw$ on replacing u_1 by w where i is any positive integer. □

Let us begin the discussion with some properties of words. The next lemma describes a basic result on words.

Lemma 7.3. *The relation $ua \neq bu$ holds for any word over a binary alphabet Σ where $\Sigma = \{a, b\}$.*

Proof. The statement follows from Theorem 1.1. □

As shown in the above lemma, $ua \neq bu$ for any word u. However, the relation holds if we replace one of the u's with its complement. It is shown in the following lemma. Further, the result is extended in Lemma 7.15.

Lemma 7.4. *Let u be a non-empty word over $\Sigma = \{a, b\}$ such that $w = bu = \bar{u}a$. Then, $w = (ba)^k$ for some integer $k > 0$.*

Proof. Clearly, u starts and ends with the letter 'a'. The lemma statement holds for $|w| = 2$. For $|w| > 2$, we can write $u = au_1a$ which gives $b(au_1a) = b\bar{u_1}ba$. Now, u_1 starts and ends with a letter 'b' giving $u_1 = bu_2b$. Similary, u_2 can be further divided like u_1 where u_2 starts and ends with the letter 'a' Thus, the words $u, u_2, u_4, \ldots$ start and end with 'a' while $u_1, u_3, u_5, \ldots$ start and end with 'b'. So, $w = (ba)^k$. □

Counting antisquares in a word is observed to be more challenging and intricate than counting squares. While counting distinct squares in simple words like a^k and $(ab)^k$ is relatively easy, the next lemma presents a method for counting the distinct antisquares in one such word. It demonstrates that counting antisquares in a trivial word $(ab)^k$ can be complex.

Lemma 7.5. *The number of antisquares in $w = (ab)^m$ is $(p+1)(2m - 2p - 1)$ where $m > 2$ and $p = \lfloor \frac{m-1}{2} \rfloor$. Further, w contains at most $\frac{|w|^2+2|w|}{8}$ antisquares.*

Proof. Define $A(w)$ as the count of all antisquares within the word w. It is important to note that $A(w)$ accounts for repeated occurrences of an antisquare in w. For example, if $w = abaab$, then $A(w)$ equals three because w contains two instances of the antisquare ab and one instance of ba. In a word $(ab)^m$, every length $2(2k+1)$ factor of w beginning from any location is an antisquare where k is an integer such that $k \geq 0$. Here, a factor of a word is a sequence of characters within that word that appears consecutively. Since the length of a factor could be at most the size of the word, we get $2(2k+1) \leq 2m$. Thus, $0 \leq k \leq \frac{m-1}{2}$. The total number of factors of length $2(2k+1)$ in w are $2m - (2(2k+1)) + 1$. We use this information to count the number of antisquares in w as follows.

$$
\begin{aligned}
A(w) &= \sum_{k=0}^{\lfloor \frac{m-1}{2} \rfloor} 2m - (2(2k+1)) + 1 \\
&= \sum_{k=0}^{\lfloor \frac{m-1}{2} \rfloor} 2m - 4k - 1 \\
&= (2m-1) + (2m-1)\left\lfloor \frac{m-1}{2} \right\rfloor - 2\left[\left\lfloor \frac{m-1}{2} \right\rfloor \left(\left\lfloor \frac{m-1}{2} \right\rfloor + 1\right)\right]
\end{aligned}
$$

We get $A(w) = (p+1)(2m-2p-1)$ by substituting $p = \lfloor \frac{m-1}{2} \rfloor$. The length of w is $2m$ and $\lfloor \frac{m-1}{2} \rfloor \leq \frac{m-1}{2}$. With this information, the highest value of antisquares obtained using the above equation is $\frac{|w|^2+2|w|}{8}$. □

In the case of squares, it is possible for a location in a word w to begin with squares of every possible even length. For example, a location i in a word a^k can start with squares of lengths $2, 4, \ldots, 2j$, where $2j \leq n - i$. The same word is also an example if a location starts with a square of a particular length, then its consecutive location can start with squares of any length. However, the structure of antisquares imposes constraints on the lengths of antisquares starting at consecutive locations. The next lemma explains these constraints in detail.

Lemma 7.6. *Let a word start with two antisquares $u^{\bar{2}}$ and $v^{\bar{2}}$ such that $|v| - |u| = 1$. If the word starts with another antisquare $w^{\bar{2}}$ where $|w| > |v|$, then $|w| > |v| + 1$.*

Proof. The statement needs to be verified only for the case where $|w| = |v|+1$. Let $w^{\bar{2}}$ begins with two shorter antisquares $u^{\bar{2}}, v^{\bar{2}}$ such that $|w| = |v| + 1$ and $|v| = |u| + 1$ where $|u| \geq 1$. We consider the case where $|u| > 2$. Without loss of generality, the word u in antisquare $u^{\bar{2}}$ begins with either aa or ab such that $u = aau_1$ or $u = abu_1$ for some word u_1. When $u = aau_1$, we get the following set of equations.

$$u\bar{u} = aau_1.bb\bar{u}_1 \tag{7.1}$$

$$v\bar{v} = aau_1b.bb\underline{\bar{u}_1a} \tag{7.2}$$

$$w\bar{w} = aau_1bb.b\underline{b\bar{u}_1}aa \tag{7.3}$$

Since $v\bar{v}$ is one of the prefixes of $w\bar{w}$, we get the relation $\bar{u}_1a = b\bar{u}_1$, which is not feasible (see Lemma 7.3).

Now, for $u\bar{u} = abu_1.ba\bar{u}_1$, we get $v\bar{v} = \underline{abu_1b.ba}\bar{u}_1a$ and $w = \underline{abu_1ba}.ba\bar{u}_1ab$. Here, the underlined prefixes of $v\bar{v}$ and $w\bar{w}$ overlap as $v\bar{v}$ is one of the prefixes of $w\bar{w}$. This shows $abu_1bb = abu_1ba$ implying $a = b$ and is not acceptable.

The lemma statement also holds when $|u| = 2$. The following set of equations explains the case.

$$u\bar{u} = aabb \qquad\qquad u\bar{u} = abb\boxed{a}$$

$$v\bar{v} = aabb\underline{a}a \qquad\qquad v\bar{v} = abb\boxed{b}aa$$

$$w\bar{w} = aabb\underline{b}baa$$

The underlined and boxed letters in the above equations overlapped, violating the condition $a \neq b$. □

The trivial word a^n is an example of a word containing the maximum number of squares. The next lemma shows the word containing the maximum number of antisquares.

Lemma 7.7. *The number of antisquares in any binary word of length* $2k$ *is at most the number of antisquares in a word* $(ab)^k$ *where the integer* k *satisfies* $k > 0$.

Proof. Every location of $(ab)^k$ begins with antisquares in which the lengths of the roots are consecutive odd numbers. According to Lemma 7.6, a word location cannot begin with three antisquares whose root lengths are n, $n+1$, and $n+2$. So, each location of the given word begins with the maximum number of antisquares. □

The remaining part of the chapter is devoted to discussing lemmas aimed at finding the bounds for the problem of determining the number of distinct antisquares in a binary word. The first two lemmas presented below describe structures of words that maximize the number of distinct antisquares in the resulting words. Based on our observation, we conjecture that the number of distinct antisquares in a word w is at most $4|w|$. The next lemma is a lower bound for the conjecture as it shows the existence of a binary word w that contains $|w| - 1$ distinct antisquares.

Lemma 7.8. *For the word* $w = a^{k+3}b^{k+3}a^{k+2}b^{k+1}a^k$, $|DA(w)| = |w| - 1$.

Proof. The length of the word is $2(k+3)+3k+3 = 5k+9$. Let us find $|DA(w)|$. The prefix $a^{k+3}b^{k+3}$ of w contains $k+3$ antisquares and they are: $\{ab, aabb, \ldots, a^{k+3}b^{k+3}\}$. In a similar way, the factor $b^{k+2}a^{k+2}$ gives another set of antisquares $\{ba, bbaa, \ldots, b^{k+2}a^{k+2}\}$ where the size of the set is $k+2$. Together these are $2k+5$ distinct antisquares.

Now, consider a prefix $a^{k+3}b^{k+3}\underline{a^{k+2}}$ of w. The underlined part a^{k+2} results into $k+2$ new antisquares that are conjugtes of $a^{k+3}b^{k+3}$. These includes $\{a^{k+2}b.b^{k+2}a, a^{k+1}b^2.b^{k+1}a^2, \ldots, ab^{k+2}.ba^{k+2}\}$. Similarly, the factor $b^{k+2}a^{k+2}b^{k+1}$ gives a total $k+1$ distinct conjugates of $b^{k+2}a^{k+2}$. Further, k distinct conjugates of $a^{k+1}b^{k+1}$ are obtained from the factor $a^{k+1}b^{k+1}a^k$. Note that none of the conjugates obtained from these three factors are the same. The total number of distinct antisquares in w is $2k+5+3k+3 = 5k+8$. □

Lemma 7.9. *Let $w = a^k b^k a^k b^j$ where the positive integers j, k satisfy $j < k$. Then, $|DA(w)| = |w| - 1$.*

Proof. The length of the given word is $|w| = 3k + j$. Antisquares appearing in the different factors of w are explained below.

$$\begin{aligned} a^k b^k \text{ contains } \{ab, aabb, \ldots, a^k b^k\}, \text{ total antisquares } &= k \\ b^k a^k \text{ contains } \{ba, bbaa, \ldots, b^k a^k\}, \text{ total antisquares } &= k \\ a^k b^k a^{k-1} \text{ contains } \{(a^{k-1}b)^{\bar{2}}, \ldots, (ab^{k-1})^{\bar{2}}\}, \text{ total antisquares } &= k-1 \\ b^k a^k b^j \text{ contains } \{(b^{k-1}a)^{\bar{2}}, \ldots, (b^{k-j}a^j)^{\bar{2}}\}, \text{ total antisquares } &= j \end{aligned}$$

Thus, the total number of distinct antisquares in w is $|w| - 1$. □

Lemma 7.10. *There exists at least one binary word of length n containing exactly $n-1$ distinct antisquares.*

Proof. The structure of words described in Lemma 7.9, $a^k b^k a^k b^j$, for $k > 1$ and $j = 1, 2, 3$ produces every n length word such that $n > 6$ containing $n-1$ distinct antisquares. The words $\{a, ab, aba, abba, aabba, aabbaa\}$ complete the proof. □

7.2 Rightmost Antisquares in Words

In this section, we see some results on structures of the rightmost antisquares. An upper bound on the number of distinct squares in a word is obtained by counting the rightmost squares [29]. Similarly, we discuss some results below that can be further extended to get an upper bound on the number of distinct antisquares in a word.

The square conjecture was recently proven by Brlek and Li in [15] by proving the next lemma.

Lemma 7.11. *[15] The number of distinct squares in a word w over an alphabet Σ follows the below relation where $|\Sigma|$ is the number of distinct letters in w.*

$$|DS(w)| \leq |w| - |\Sigma| + 1$$

According to Lemma 7.11, the maximum number of distinct squares in a binary word w is at most $|w|-1$. However, the lower bound for the number of distinct antisquares in a word w is $|w|-1$ (see Lemma 7.10 and the conjecture proposed before the lemma). Further, the relation $s_i(w) \leq 2$ is used to get the upper bound for the square conjecture. In case of antisquares, it is observed that a word could begin with three distinct rightmost antisquares meaning a word w exists satisfying the relation $\bar{s_i}(w) = 3$. Some examples of such words are $\{(ababbababaabaa)^{\bar{2}}, (aabbba)^{\bar{2}}\}$. The following lemma is an attempt to explore a word w with $\bar{s_i}(w) = 4$. The lemma tries to identify the relation between the lengths of four distinct antisquares starting at the same location.

Lemma 7.12. *Let an antisquare $x\bar{x}$ begins with three shorter rightmost antisquares $u^{\bar{2}}, v^{\bar{2}}, w^{\bar{2}}$ such that $|u| \leq |v| \leq |w|$. Then, either $2|w| > |x|$ or $2|u| > |v|$.*

Proof. Consider a case where neither of the given constraints holds. So, in the following arrangement of antisquares, we have $2|w| \leq |x|$ and $2|u| \leq |v|$.

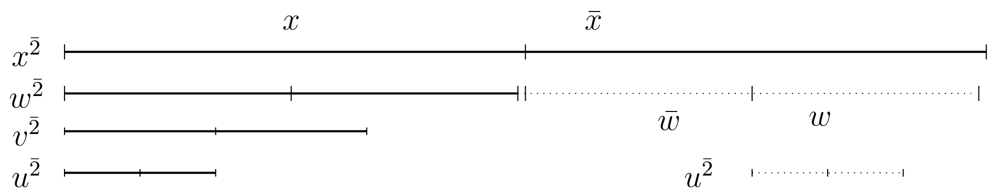

Figure 7.1: *Arrangement of shorter antisquares in* $x^{\bar{2}}$

As shown in Figure 7.1, the antisquare $u^{\bar{2}}$ repeats in the suffix $\bar{x}$ of $x^{\bar{2}}$ which gives $\bar{s}_1(x\bar{x}) = 3$. This leads to a contradiction. Thus, an antisquare $x\bar{x}$ do not satisfy $2|w| \leq |x|$ and $2|u| \leq |v|$ at the same time. □

The above lemma asserts that when a word location begins with four rightmost distinct antisquares, then there always exists an antisquare whose root is longer than the root of at least one antisquare. If we consider only these two antisquares, then the shorter antisquare ends in the second half of the longer antisquare. The subsequent part of the chapter deals with the study of such rightmost occurrences of distinct antisquares. We are interested in expanding these structures to incorporate more antisquares starting at the same location.

We see the structure of words in which a location starts with two such occurrences of antisquares in the following lemma.

Lemma 7.13. *If $v\bar{v}$ is an antisquare that begins with another shorter rightmost antisquare such that $\bar{s}_1(v\bar{v}) = 2$, and the shorter antisquare ends in $\bar{v}$. Then, v has one of the following structures:*

(a) $v = xx\bar{x}$, $u = xx$

(b) $v = xyx\bar{x}\bar{y}$, $u = xyx$ or

(c) $v = xxy\bar{x}$, $u = xxy$ where $y <_p xy$

Here, u is a non-empty prefix of v, and the expression $y <_p xy$ indicates that y is a prefix of xy.

Proof. Let $u\bar{u}$ be the shorter rightmost square in $v^{\bar{2}}$, so $2|u| > |v|$. Here, $u\bar{u}$is a prefix of $v\bar{v}$, and $v = u\bar{u}_1$ such that $u = u_1u_2$ for $u_1, u_2 \in \Sigma^+$ as shown in Figure 7.2.

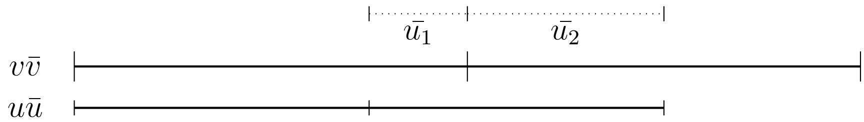

Figure 7.2: *$u^{\bar{2}}$ in the prefix of $v^{\bar{2}}$*

Case I: Let $|u_1| = |u_2|$. We have $v = u_1u_2\bar{u}_1$ and $\bar{u}_2 <_p \bar{v}$. As u_1, u_2 are of same length, $v = u_1u_1\bar{u}_1$ and $u = u_1u_1$.
Case II: Let $|u_1| > |u_2|$. From the above structure, $v = u\bar{u}_1$ and $\bar{v} = \bar{u}_2v_2$ for

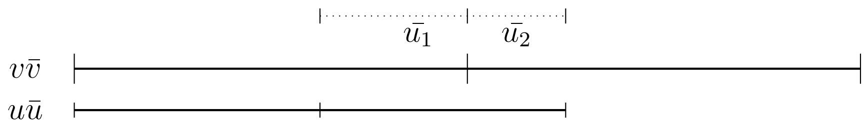

Figure 7.3: *$u^{\bar{2}}$ with $|u_1| > |u_2|$*

some non-empty word v_2. We get $u_1 = u_2u_3$ by complementing one of these equations and equating it with the other. Here, u_3 is a non-empty word. So, it gives $v = u\bar{u}_1 = u_1u_2\bar{u}_1 = u_2u_3u_2\bar{u}_2\bar{u}_3$ and $u = u_1u_2 = u_2u_3u_2$.
Case III: Let $|u_1| < |u_2|$. From Figure 7.2, v begins with u_1u_2 and $\bar{v}$ begins with $\bar{u}_1$. So, v also begins with u_2, and $u_2 = u_1u_3$ where $u_3 \in \Sigma^+$.

$$v = u_1u_2\bar{u}_1 \qquad \bar{v} = \bar{u}_2\bar{v}_2$$
$$= u_1u_1u_3\bar{u}_1 \qquad \bar{v} = \bar{u}_1\bar{u}_3\bar{v}_2$$

Thus, $v = u_1u_1u_3\bar{u}_1$ and $u = u_1u_1u_3$. Since $u^{\bar{2}} <_p v^{\bar{2}}$, $\bar{u}_3$ is one of the prefixes of $\bar{u}_1\bar{u}_3$. □

As mentioned in the above lemma, there are three possible structures of words with $\bar{s}_i(w) = 2$. We now verify the feasibility of the first structure to add more rightmost antisquare at the beginning.

Lemma 7.14. *Let $w^{\bar{2}} = (xx\bar{x})^{\bar{2}}$ where $x \in \Sigma^+$. Then, $\bar{s}_1(w) < 3$.*

Proof. The value of $\bar{s}_1(w)$ is at least two (see Lemma 7.13). Assume $w^{\bar{2}}$ starts with another two shorter rightmost antisquares $u^{\bar{2}}, v^{\bar{2}}$ such that $|v| < |u|$. The rightmost appearance of these antisquares implies $2|u| > |w| < 2|v|$ and $v = xx$.The word u is a proper prefix of $xx\bar{x}$ where $u \neq xx$. There are two conditions: $|x| < |u| < 2|x|$ and $2|x| < |u| < 3|x|$.

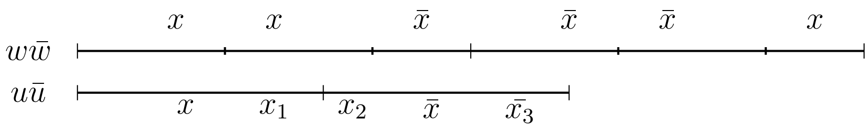

Figure 7.4: *Antisquare $u^{\bar{2}}$ in $(xx\bar{x})^{\bar{2}}$ with $|x| < |u| < 2|x|$*

Refer Figure 7.4 for the first case, $u = xx_1$ where $x = x_1x_2$ for some non-empty word x_1, x_2. So, $\bar{u} = x_2\bar{x}\bar{x_3}$ for some non-empty prefix x_3 of x. We then equate two structures of u which give the following equations.

$$xx_1 = \bar{x_2}xx_3 \qquad \Longrightarrow |x_1| = |\bar{x_2}| + |x_3| \tag{7.4}$$

$$x_1x_2\underline{x_1} = \bar{x_2}x_1\underline{x_2x_3} \qquad \Longrightarrow x_1 = x_2x_3 \tag{7.5}$$

Eq. 7.4 and Eq. 7.5 show that x_1 begins with $\bar{x_2}$ and x_2 respectively. This is a contradiction since $x_2 \neq \bar{x_2}$.

Similarly for case $2|x| < |u| < 3|x|$, we get the relation $xx\bar{x_1} = x_2xx_3$ from the structure of antisquare $u^{\bar{2}}$ as shown in Figure 7.5. Here, $x_1, x_3 \in \Sigma^+$ such

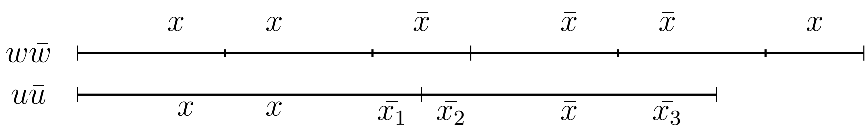

Figure 7.5: *$u^{\bar{2}}$ in $(xx\bar{x})^{\bar{2}}$ with $2|x| < |u| < 3|x|$*

that $x = x_1x_2 = x_3x_4$. We further analyse the relation below.

$$\begin{aligned} x_1x_2x_1x_2\bar{x}_1 &= x_2x_1x_2x_3 & &\implies x_1x_2 = x_2x_1 \\ z^iz^jz^iz^j\bar{z}^i &= z^jz^iz^jx_3 & &\text{as } x_1 = z^i, x_2 = z^j \text{ for } i, j \in \mathbb{Z} \end{aligned}$$

We can conclude that $x = x_1x_2 = z^{i+j}$ and $x_3 = z^i\bar{z}^i$ from the above equation set. In this case, x_3 is one of the prefixes of x, and this shows $z = \bar{z}$. So, this case is invalid, and it shows that $\bar{s}_1(w) < 3$ where the structure of w follows certain constraints. □

The next two lemmas discuss some basic results that hold under certain constraints. We encounter these types of relationships while exploring the structures described in Lemma 7.13.

Lemma 7.15. *Let u, v be two non-empty unequal words such that $uv = \bar{v}\bar{u}$. Then, $uv = (\bar{m}m)^{2(k+1)}$ and u is $(\bar{m}m)^{k+1}\bar{m}$ for $|u| > |v|$ or $u = (\bar{m}m)^k\bar{m}$ for $|u| < |v|$.*

Proof. Without loss of generality, assume $|u| > |v|$. Consider the following structure describing the given condition.

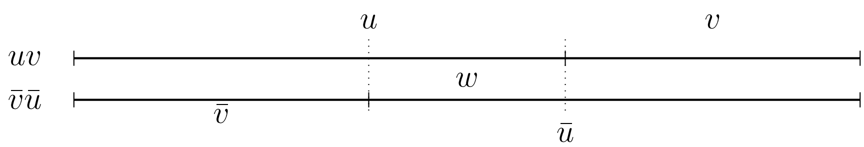

Figure 7.6: *Structure of words holding the relation $uv = \bar{v}\bar{u}$*

We get $u = \bar{v}w$ and $\bar{u} = wv$. Thus, $v\bar{w} = wv$. Using Lyndon and Schützenberger theorem (Theorem 1.5.2 from [3]), we get $w = mn, \bar{w} = nm, v = (mn)^km$. We then substitute these words to get the following equations.

$$\begin{aligned} u &= \bar{v}w & \bar{u} &= wv \\ &= (\bar{m}\bar{n})^k\bar{m}mn & u &= \bar{w}\bar{v} \\ & & u &= nm(\bar{m}\bar{n})^k\bar{m} \\ \implies mn &= \bar{n}\bar{m} & & \end{aligned} \tag{7.6}$$

We get the solution for Eq. 7.6 when $|m| = |n|$; otherwise, exploring the relation further leads to a similar relation. If $|m| = |n|$, then $n = \bar{m}$ and we get $u = (\bar{m}m)^{k+1}\bar{m}, v = (m\bar{m})^k m$.

If $|u| < |v|$, we get $u = (\bar{m}m)^k\bar{m}$ and $v = (m\bar{m})^{k+1}m$. □

Lemma 7.16. *Let $xy = \bar{y}x$ for some non-empty binary words x, y. Then, the longer and the shorter words between x and y are $(m\bar{m})^{2(k+1)}$ and $(m\bar{m})^k m$ respectively, where $m \in \Sigma^+$, k is a positive integer.*

Proof. The relation does not hold for $|x| = |y|$ as it gives two relations $x = \bar{y}$ and $x = y$. Consider the case $|x| < |y|$ as shown in Figure 7.7. Let c be a word such that $y = cx$ and $\bar{y} = xc$. This gives $cx = \bar{x}\bar{c}$.

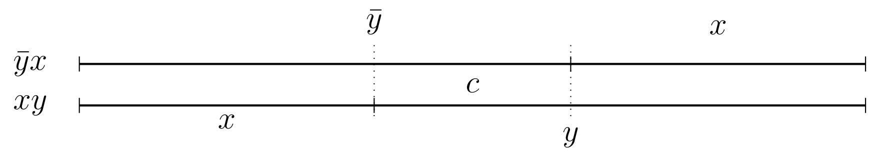

Figure 7.7: *Word arrangement for $xy = \bar{y}x$*

We use Lemma 7.15 to get the following sets of relations.

$$x = (m\bar{m})^k m \qquad c = (\bar{m}m)^{k+1}\bar{m}$$

$$y = xc = (m\bar{m})^{2(k+1)} \qquad xy = [(m\bar{m})^k m]^2\bar{m}(m\bar{m})^{k+1}$$

The arrangement of words in the case of $|x| > |y|$ gives a similar relation $ty = \bar{y}t$ assuming $x = ty$. Here, we get $x = (m\bar{m})^{2(k+1)}$ and $y = (m\bar{m})^k m$. □

The $\bar{s}_i$ value of a location can be three, and we discussed examples of such words (see the paragraph before Lemma 7.12). However, the shortest antisquare in such examples is observed to be shorter than half of the longest antisquare. It will be interesting to explore and identify the properties of a non-empty word $x\bar{x}$ where two shorter antisquares starting at the first location end in $\bar{x}$.

7.3 Conclusions

We proved that the word $(ab)^k$ contains the maximum number of antisquares, and the highest value is $\frac{k^2+k}{2}$. We also identified the structure of antisquares that are non-primitive. The problem of counting distinct antisquares in a word is then explored with the notation $\bar{s}_i(w)$. The lower bound obtained in this regard shows that the number of distinct antisquares in a word is $|w| - 1$ and is more than the number of distinct squares. It was also shown that the value of $\bar{s}_i(w)$ could be three, but no upper bound for $\bar{s}_i(w)$ has been identified yet. We explored the structural aspects of a word w when $\bar{s}_i(w) = 4$. Some basic results are obtained in this direction, but further investigation is needed to confirm the value. However, based on our analysis, we suspect that the value of $\bar{s}_i(w)$ is at most three.

8

Conclusions and Future Work

The contributions presented in this dissertation focus on exploring the square conjecture, particularly through precisely counting the number of distinct FS-squares in a word. We explored the structure of FS-double squares and found that the length of the longest sequence of consecutive FS-double squares is at most $\frac{1}{7}^{th}$ of the length of the word. However, we observed that such FS-double squares do not maximize the square density. To prove this, we obtained a result stating an FS-double square inevitably contains a certain number of no-gain locations. This result is then extended to show that the square density of words with consecutive FS-double squares is at most $\frac{133}{81}$.

Later, we worked on the characteristics of words whose square densities exceed the value of one. In this regard, we found the structure of words where reversing the word does not change the number of FS-double squares. It is shown that the number of FS-double squares appearing at consecutive locations in such words is always less than $\frac{1}{11}^{\text{th}}$ of the overall word length. We also studied the distribution and types of squares in non-primitive words, where we proved that the square density of a non-primitive word is inversely proportional to its exponent.

We additionally discussed several patterns that generate words of increasing square densities. The square density of words generated by these patterns approaches one. A new lower bound for the square conjecture is presented using a pattern.

Finally, we obtained some properties of antisquares where we presented the best-known words containing a maximum number of distinct antisquares. It is also shown that for any word containing k distinct squares, a binary word with at least k distinct antisquares always exists.

8.1 Summary of Results

Table 8.1 elaborates on all the results discussed in the preceding section. It serves as a comprehensive reference for readers to understand better the outcomes of the research conducted in this thesis.

Problem statement	**Result**
The length of the longest sequence of consecutive locations that begin with FS-double squares in w is at most	$\frac{\lvert w\rvert}{7}$
The number of feasible structures for a 2FS square	2
The number of no-gain lengths in an FS-double square $((xy)^p x (xy)^p)^2$ where $p > 2$ is alt least	$(p-2)\lvert xy\rvert$
Square density of words with consecutive FS-double squares is at most	$\frac{133}{81}$
The highest square density of words with consecutive bordered FS-double squares is at most	$\frac{1}{11}$
The square density of a non-primitive word u^k where $k \to \infty$ approaches	$\frac{1}{2}$
The highest square density of a pattern that generates words with an increasing number of distinct squares	1
The lower bound for the known words with the maximum possible number of distinct antisquares in w	$\lvert w\rvert - 1$

Table 8.1: *Summary of research outcomes*

8.2 Future Work

The initial chapters aim to solve the square conjecture by investigating FS-double squares and no-gain locations. The proof method exhaustively examines all possible word structures and relies on case analysis. Since the analysis considers all cases, it also reveals the different distribution of squares that increases and decreases the square density. So, the properties identified to solve the square conjecture can be utilized to solve related conjectures, such as the conjecture proposed for circular words [5]. However, it became apparent that the chosen case analysis method may lead to lengthy and convoluted proof for the square conjecture. Therefore, future investigations should aim to develop more concise and elegant proof methods to solve word equations. Exploring alternative proof techniques, such as the graph-based approach employed by Brlek and Li [15], may offer a promising direction.

We also studied the problem of counting distinct antisquares in finite words. While antisquares have been studied in the context of infinite words in the existing literature [12], our focus on finite words presents an intriguing opportunity for further exploration. Based on the preliminary findings from our case analysis approach, we conjecture that the maximum number of distinct antisquares in a word w is at most $3|w|$. However, continuing to solve the problem with an exhaustive case study would be challenging. Therefore, future research should aim to refine the problem-solving approach or explore alternative techniques.

Exploration of the concept of antisquare density in both primitive and non-primitive words presents another avenue for investigation. Antisquare density refers to the ratio of the number of distinct antisquares in a word to its length. Our research has established the correlation between the square density and the number of primitive squares by computing the square density of non-primitive words. It is interesting to explore the antisquare density of both primitive and non-primitive words.

Publications

Journals

1. **Maithilee Patawar** and Kalpesh Kapoor. “ The Length of the Longest Sequence of Consecutive FS-double Squares in a Word”, *Communications in Combinatorics and Optimization* (CCO), December 2022.

2. **Maithilee Patawar** and Kalpesh Kapoor. “Density of distinct squares in non-primitive words”, *Information Processing Letters* (IPL), January 2023

3. **Maithilee Patawar** and Kalpesh Kapoor. “ The Square Density of Words Having a Sequence of FS-double Squares.”, *Discrete Applied Mathematics SI CALDAM 2021* (DAM), December 2023

Conferences

1. **Maithilee Patawar** and Kalpesh Kapoor, “ Characterization of Dense Patterns Having Distinct Squares”, *7th Annual International Conference on Algorithms and Discrete Applied Mathematics* (CALDAM). February 2021.

2. **Maithilee Patawar** and Kalpesh Kapoor, “Characterization of FS-double Squares Ending with Two Distinct Squares.”, *European Conference on Combinatorics, Graph Theory and Applications* (EUROCOMB). September 2021.

References

[1] H. M. A. Alamro, G. Badkobeh, D. Belazzougui, C. Iliopoulos, and S. J. Puglisi. Computing the antiperiod(s) of a string. In 30^{th} *Annual Symposium on Combinatorial Pattern Matching, CPM*, pages 32–43. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [Pg. 17.]

[2] J.-P. Allouche and J. Shallit. The ubiquitous prouhet-thue-morse sequence. In *Sequences and their Applications: Proceedings of SETA*, pages 1–16. Springer, 1999. [Pg. 1.]

[3] J.-P. Allouche and J. Shallit. *Automatic sequences: theory, applications, generalizations*, pages 10–14. Cambridge University Press, 2003. [Pg. 5, 56, and 101.]

[4] M. Alzamel, A. Conte, D. Greco, V. Guerrini, C. Iliopoulos, N. Pisanti, N. Prezza, G. Punzi, and G. Rosone. Online algorithms on antipowers and antiperiods. In *String Processing and Information Retrieval:* 26^{th} *International Symposium, SPIRE*, pages 175–188. Springer, 2019. [Pg. 17.]

[5] M. Amit and P. Gawrychowski. Distinct squares in circular words. In *International Symposium on String Processing and Information Retrieval*, pages 27–37. Springer, 2017. [Pg. 6, 14, 62, 68, 72, 84, and 106.]

[6] A. Apostolico and N. Pisanti. Combinatorial algorithms for DNA sequence analysis. *Briefings in Bioinformatics*, 1:39–49, 2000. [Pg. 2.]

[7] G. Badkobeh, G. Fici, and S. J. Puglisi. Algorithms for anti-powers in strings. *Information Processing Letters*, 137:57–60, 2018. [Pg. 16.]

[8] H. Bai, A. Deza, and F. Franek. On a lemma of Crochemore and Rytter. *Journal of Discrete Algorithms*, 34:18–22, 2015. [Pg. 12.]

[9] H. Bai, F. Franek, and W. F. Smyth. Two squares canonical factorization. In *Stringology*, pages 52–58, 2014. [Pg. 12.]

[10] H. Bai, F. Franek, and W. F. Smyth. The new periodicity lemma revisited. *Discrete Applied Mathematics*, 212:30–36, 2016. [Pg. 13, 19, 21, 23, and 55.]

[11] H. Bannai, T. I, S. Inenaga, Y. Nakashima, M. Takeda, and K. Tsuruta. The "runs" theorem. *SIAM Journal on Computing*, 46(5):1501–1514, 2017. [Pg. 15.]

[12] A. Baranwal, J. Currie, L. Mol, P. Ochem, N. Rampersad, and J. Shallit. Antisquares and critical exponents. *arXiv preprint arXiv:2209.09223*, 2022. [Pg. 3, 17, 90, and 106.]

[13] J. Berstel and D. Perrin. The origins of combinatorics on words. *European Journal of Combinatorics*, 28(3):996–1022, 2007. [Pg. 1.]

[14] F. Blanchet-Sadri and S. Osborne. Constructing words with high distinct square densities. *arXiv preprint arXiv:1708.06462*, pages 1–15, 2017. [Pg. 13, 19, 36, 37, 50, 53, 68, 72, and 84.]

[15] S. Brlek and S. Li. On the number of squares in a finite word. *arXiv preprint arXiv:2204.10204*, 2022. [Pg. 14, 97, and 106.]

[16] A. Burcroff. (k, λ)-anti-powers and other patterns in words. *arXiv preprint arXiv:1807.07945*, 2018. [Pg. 16.]

[17] M. G. Castelli, F. Mignosi, and A. Restivo. Fine and wilf's theorem for three periods and a generalization of sturmian words. *Theoretical Computer Science*, 218(1):83–94, 1999. [Pg. 12.]

[18] A. Cobham. Uniform tag sequences. *Mathematical systems theory*, 6(1-2):164–192, 1972. [Pg. 1.]

[19] M. Crochemore et al. An optimal algorithm for computing the repetitions in a word. *Information Processing Letters*, 12(5):244–250, 1981. [Pg. 10 and 11.]

[20] M. Crochemore and L. Ilie. Maximal repetitions in strings. *Journal of Computer and System Sciences*, 74(5):796–807, 2008. [Pg. 15.]

[21] M. Crochemore and W. Rytter. *Text Algorithms*. Oxford University Press, 1994. [Pg. 2.]

[22] M. Crochemore and W. Rytter. *Jewels of stringology: text algorithms*, pages 16–17. World Scientific, 2002. [Pg. 24.]

[23] A. de Luca and S. Varricchio. Some combinatorial properties of the thue–morse sequence and a problem in semigroups. *Theoretical Computer Science*, 63(3):333–348, 1989. [Pg. 1.]

[24] A. De Luca and S. Varricchio. *Finiteness and Regularity in Semigroups and Formal Languages*, pages 31–50. Springer Science & Business Media, 2012. [Pg. 16.]

[25] C. Defant. Anti-power prefixes of the thue-morse word. *arXiv preprint arXiv:1607.05825*, 2016. [Pg. 16.]

[26] A. Deza and F. Franek. A d-step approach to the maximum number of distinct squares and runs in strings. *Discrete Applied Mathematics*, 163:268–274, 2014. [Pg. 13.]

[27] A. Deza, F. Franek, and M. Jiang. Primitive square maximal words. `http://optlab.mcmaster.ca/~jiangm5/research/square.html`. Accessed: 2020-05-18. [Pg. 11.]

[28] A. Deza, F. Franek, and M. Jiang. A d-step approach for distinct squares in strings. In *Combinatorial Pattern Matching: 22nd Annual Symposium, CPM*, pages 77–89. Springer, 2011. [Pg. 6, 11, 13, 21, 60, 67, and 68.]

[29] A. Deza, F. Franek, and A. Thierry. How many double squares can a string contain? *Discrete Applied Mathematics*, 180:52–69, 2015. [Pg. 11, 13, 19, 21, 22, 24, 39, 49, and 97.]

[30] R. C. Entringer, D. E. Jackson, and J. Schatz. On nonrepetitive sequences. *Journal of Combinatorial Theory, Series A*, 16(2):159–164, 1974. [Pg. 10.]

[31] K. Fan, S. J. Puglisi, W. F. Smyth, and A. Turpin. A new periodicity lemma. *SIAM Journal on Discrete Mathematics*, 20(3):656–68, 2006. [Pg. 12, 19, and 55.]

[32] G. Fici, A. Restivo, M. Silva, and L. Q. Zamboni. Anti-powers in infinite words. *Journal of Combinatorial Theory, Series A*, 157:109–119, 2018. [Pg. 3 and 16.]

[33] N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, 1965. [Pg. 11, 12, and 24.]

[34] A. S. Fraenkel and J. Simpson. How many squares can a string contain? *Journal of Combinatorial Theory, Series A*, 82(1):112–120, 1998. [Pg. 5, 6, 11, 14, 19, 37, 50, 53, 60, 64, 67, 68, 70, 81, 84, and 87.]

[35] A. S. Fraenkel and J. Simpson. An extension of the periodicity lemma to longer periods. *Lecture Notes in Computer Science*, pages 98–105, 2001. [Pg. 12.]

[36] A. S. Fraenkel and R. J. Simpson. How many squares must a binary sequence contain? *The Electronic Journal of Combinatorics*, pages 1–9, 1995. [Pg. 10.]

[37] F. Franek, R. C. Fuller, J. Simpson, and W. F. Smyth. More results on overlapping squares. *Journal of Discrete Algorithms*, 17:2–8, 2012. [Pg. 12.]

[38] F. Franek and Q. Yang. An asymptotic lower bound for the maximal number of runs in a string. *International Journal of Foundations of Computer Science*, 19(01):195–203, 2008. [Pg. 15.]

[39] M. Gaetz. Anti-power j-fixes of the thue-morse word. *Discrete Mathematics & Theoretical Computer Science*, 23, 2021. [Pg. 16.]

[40] M. Giraud. Not so many runs in strings. In *Language and Automata Theory and Applications: Second International Conference, LATA*, pages 232–239. Springer, 2008. [Pg. 15.]

[41] O. Goldreich. *Modern Cryptography: Theory and Practice*. Prentice Hall, 2008. [Pg. 2.]

[42] L. Ilie. A simple proof that a word of length n has at most $2n$ distinct squares. *Journal of Combinatorial Theory, Series A*, 112(1):163–164, 2005. [Pg. 11.]

[43] L. Ilie. A note on the number of squares in a word. *Theoretical Computer Science*, 380(3):373–376, 2007. [Pg. 11, 28, and 33.]

[44] N. Jonoska, F. Manea, and S. Seki. A stronger square conjecture on binary words. In *International Conference on Current Trends in Theory*

and Practice of Informatics, pages 339–350. Springer, 2014. [Pg. 6, 13, 71, 83, and 84.]

[45] T. Kociumaka, J. Radoszewski, W. Rytter, J. Straszyński, T. Waleń, and W. Zuba. Efficient representation and counting of antipower factors in words. *Information and Computation*, 286:104779, 2022. [Pg. 17.]

[46] R. Kolpakov and G. Kucherov. Finding maximal repetitions in a word in linear time. In 40^{th} *Annual Symposium on Foundations of Computer Science*, pages 596–604. IEEE, 1999. [Pg. 3, 6, and 15.]

[47] R. Kolpakov, M. Podolskiy, M. Posypkin, and N. Khrapov. Searching of gapped repeats and subrepetitions in a word. In *Combinatorial Pattern Matching:* 25^{th} *Annual Symposium, CPM*, pages 212–221. Springer, 2014. [Pg. 3.]

[48] M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń. On the maximal number of cubic subwords in a string. In *Combinatorial Algorithms:* 20^{th} *International Workshop, IWOCA*, pages 345–355. Springer, 2009. [Pg. 15.]

[49] M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń. On the maximum number of cubic subwords in a word. *European Journal of Combinatorics*, 34(1):27–37, 2013. [Pg. 6, 15, and 21.]

[50] N. H. Lam. *On the number of squares in a string*. PhD dissertation, McMaster University, 2013. [Pg. 13.]

[51] A. Lempel and J. Ziv. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1977. [Pg. 2.]

[52] S. Li. On the number of k-powers in a finite word. *Advances in Applied Mathematics*, 139:1–7, 2022. [Pg. 14.]

[53] M. Lothaire. *Applied Combinatorics on Words*, volume 105, pages 1–40. Cambridge University Press, 2005. [Pg. 11.]

[54] F. Manea and S. Seki. Square-density increasing mappings. In *International Conference on Combinatorics on Words*, pages 160–169. Springer, 2015. [Pg. 14, 19, 73, and 79.]

[55] C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT press, 1999. [Pg. 2.]

[56] W. Matsubara, K. Kusano, A. Ishino, H. Bannai, and A. Shinohara. New lower bounds for the maximum number of runs in a string. In *Stringology*, pages 140–145, 2008. [Pg. 15.]

[57] S. Narayanan. Functions on antipower prefix lengths of the thue–morse word. *Discrete Mathematics*, 343(2):111675, 2020. [Pg. 16.]

[58] E. Prouhet. Mémoire sur quelques relations entre les puissances des nombres. *CR Acad. Sci. Paris*, 33(225), 1851. [Pg. 1.]

[59] S. J. Puglisi, J. Simpson, and W. F. Smyth. How many runs can a string contain? *Theoretical Computer Science*, 401(1-3):165–171, 2008. [Pg. 15.]

[60] F. P. Ramsey. On a problem of formal logic. *Classic Papers in Combinatorics*, pages 1–24, 1987. [Pg. 16.]

[61] W. Rytter. The number of runs in a string: Improved analysis of the linear upper bound. In *23rd Annual Symposium on Theoretical Aspects of Computer Science, STACS*, pages 184–195. Springer, 2006. [Pg. 15.]

[62] J. Simpson. Intersecting periodic words. *Theoretical computer science*, 374(1-3):58–65, 2007. [Pg. 12.]

[63] J. Simpson. Modified padovan words and the maximum number of runs in a word. *Australasian Journal of Combinatorics*, 46:129–145, 2010. [Pg. 15.]

[64] A. O. Slisenko. Detection of periodicities and string-matching in real time. *Journal of Soviet Mathematics*, 22:1316–1387, 1983. [Pg. 10.]

[65] A. Thierry. A proof that a word of length n has less than $1.5n$ distinct squares. *arXiv preprint arXiv:*2001.02996, pages 1–30, 2020. [Pg. 13, 22, and 23.]

[66] A. Thue. Über annäherungswerte algebraischer zahlen. 1909(135):284–305, 1909. [Pg. 1.]

[67] Y. Wexler, Z. Yakhini, Y. Kashi, and D. Geiger. Finding approximate tandem repeats in genomic sequences. In *Proceedings of the* 8^{th} *Annual International Conference on Research in Computational Molecular Biology*, pages 223–232, 2004. [Pg. 2.]

Department of Computer Science and Engineering

Indian Institute of Technology Guwahati

Guwahati 781039, India