A

Narad Rampersad
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1 (Canada)
nrampersad@math.uwaterloo.ca

Abstract

U. Schmidt proved that any binary pattern of length at least 13 is avoided by either the Thue–Morse word or the Fibonacci word. We give a somewhat simpler proof of this result.

1 Introduction

In combinatorics on words, the notion of an avoidable/unavoidable pattern was first introduced (independently) by Bean, Ehrenfeucht, and McNulty [1] and Zimin [14]. Let Σ and Δ be alphabets; the alphabet Δ is the *pattern alphabet*. A *pattern* p is a non-empty word over Δ. A word w over Σ matches p if there exists a non-erasing morphism $h : \Delta^* \to \Sigma^*$ such that $h(p) = w$. A pattern p is *avoidable* if there exists an infinite word $\mathbf{x}$ over a finite alphabet such that no subword of $\mathbf{x}$ matches p. Otherwise, p is *unavoidable*. Schmidt [10] proved the following theorem.

Theorem 1. *Any binary pattern of length at least* 13 *is avoided by either the Thue–Morse word*

$$\mathbf{t} = 0110100110010110\cdots$$

or the Fibonacci word

$$\mathbf{f} = 0100101001001010\cdots .$$

It is easy to check—for instance, by exhaustive search—that any binary word of length at least 13 contains an occurrence of one of the patterns xxx, $xyxyx$, $xyyxyxxy$, or $xyyxxyyx$. It thus suffices to show that each of these 4 patterns is avoided by either the Thue–Morse word or the Fibonacci word. Since it is well-known that the Thue–Morse word avoids the patterns xxx and $xyxyx$ [12]—in fact, Guaiana [4] and Shur [11] completely characterized the set of binary patterns avoided by the Thue–Morse word—it is enough to show that the Fibonacci word avoids the patterns *xyyxyxxy*, or *xyyxxyyx*.

Observe, however, that the patterns *xyyxyxxy* and *xyyxxyyx* are abelian 4-powers. An *abelian k-power* is a word of the form $w_1 w_2 \cdots w_k$, where each w_i is a non-empty word, and for all i, j, w_i is a permutation of the symbols of w_j. Dekking [3] proved that there exists an infinite binary word

$$\mathbf{a} = 0110001000101101101100001 \cdots$$

containing no abelian 4-powers. Dekking's word is obtained by iterating the morphism $0 \rightarrow 011$, $1 \rightarrow 0001$. Since **a** avoids abelian 4-powers, it avoids the patterns *xyyxyxxy* and *xyyxxyyx*. It follows that every binary pattern of length at least 13 is avoided by either the Thue–Morse word or the word **a**. Dekking's paper was published in 1979; Schmidt's was published in 1989. Had she been aware of Dekking's result, Schmidt might have immediately deduced that all binary patterns of length at least 13 are avoidable on the binary alphabet. Nevertheless, proving that the Fibonacci word avoids the patterns *xyyxyxxy* and *xyyxxyyx* allows for an interesting application of several well-known properties of the Fibonacci word. We present a somewhat simpler proof of this result below

Note that the bound of 13 in Theorem 1 is not optimal: if one applies Cassaigne's result [2, Theorem 5.2] that the Thue–Morse word avoids the pattern *xxyxxy*, it is possible to replace 13 with 12. Restivo and Salemi [7, Proposition 1] stated (without proof) that 13 can be replaced by 11.

Subsequent to Schmidt's work, Roth [8] showed that every binary pattern of length at least 6 is avoided by some infinite binary word; however, more than just the Thue–Morse and Fibonacci words are needed to witness the avoidability of some of these patterns. Cassaigne [2] completed the work of Roth by determining exactly the set of binary patterns that are avoidable by some infinite binary word. This classification was also obtained independently by Vaniček [13] (see [5]). The survey chapter in Lothaire [6, Chapter 3] gives a good overview of the main results concerning avoidable patterns.

2 Preliminaries

We are concerned only with patterns over the alphabet $\{x, y\}$ and words over the alphabet $\{0, 1\}$. We begin with some definitions related to the well-studied Fibonacci word. We define the following sequence $(s_n)_{n \geq -1}$ of words:

$$s_{-1} = 1, \quad s_0 = 0, \quad s_n = s_{n-1} s_{n-2}.$$

For $n \geq 0$, we call s_n a *standard word*. The infinite word $\mathbf{f} = lim_{n\to\infty} s_n$ is the *Fibonacci word*. We also define the sequence $(q_n)_{n\geq -1}$, where $q_n = |s_n|$. Clearly, for $n \geq 1$, we have

$$q_n = q_{n-1} + q_{n-2}. \tag{1}$$

Observe that for $n \geq 0$, $q_n = F_{n+2}$, where F_n denotes the n-th Fibonacci number.

Next, we define the morphism ϕ by $\phi(0) = 01$ and $\phi(1) = 0$. For $n \geq 0$, we see that $s_n = \phi^n(0)$, so that $\mathbf{f} = \phi^\omega(0)$. We need the following properties of the Fibonacci word.

- The word $\mathbf{f}$ does not contain the subwords 11, 000, 10101, or 00100100.
- The word $\mathbf{f}$ is *recurrent*; that is, every subword of $\mathbf{f}$ occurs infinitely often in $\mathbf{f}$.
- If u is a subword of $\mathbf{f}$, then so is its reversal, denoted u^R.
- A subword u of $\mathbf{f}$ is *left special* if $0u$ and $1u$ are both subwords of $\mathbf{f}$. The Fibonacci word has exactly one left special subword of each length and this subword is a prefix of $\mathbf{f}$. In particular, the left special subwords are prefixes of the standard words.
- If uu is a non-empty subword of $\mathbf{f}$, then u is a conjugate of a standard word (words w and z are *conjugates* if one can write $w = xy$ and $z = yx$ for some words x and y).

All of the preceding properties of $\mathbf{f}$, except perhaps the last, are well-known and can be found, for example, in Lothaire [6, Chapter 2]. The characterization of the squares of the Fibonacci word seems to have first been proved by Séébold [9].

3 The Fibonacci word avoids *xyyxyxxy* and *xyyxxyyx*

Proposition 2. *The Fibonacci word avoids the pattern xyyxyxxy.*

Proof. Suppose to the contrary that $\mathbf{f}$ contains a subword $w = uvvuvuuv$ of minimal length. Since w contains the squares uu, vv, and $vuvu$, the words u, v, and vu are each conjugates of a standard word. We thus have $|vu| = q_n$ for some $n \geq 2$. Further, either $|u| = q_{n-1}$ and $|v| = q_{n-2}$, or vice-versa. If $n = 2$, then either $(u, v) = (0, 01)$ or $(u, v) = (01, 0)$. In either case, we see that w contains 000, a contradiction. We suppose then that $n > 2$.

Case 1: u and v end with different letters. Then u and v are left special, and hence are both standard words (since the left special subwords are the prefixes of the standard words). Letting $u' = \phi^{-1}(u)$ and $v' = \phi^{-1}(v)$, it follows that $\phi^{-1}(w) = u'v'v'u'v'u'u'v'$ is a subword of $\mathbf{f}$, contradicting the minimality of w.

Case 2: u and v begin with different letters. Then, since the set of subwords of $\mathbf{f}$ is closed under reversal, we may apply the argument of Case 1 to w^R to derive a contradiction.

Case 3: u and v both begin and end with 1. Then w contains 11, a contradiction.

Case 4: u and v both begin with 0 and end with 1. Then letting $u' = \phi^{-1}(u)$ and $v' = \phi^{-1}(v)$, it follows that $\phi^{-1}(w) = u'v'v'u'v'u'u'v'$ is a subword of $\mathbf{f}$, contradicting the minimality of w.

Case 5: u and v both begin with 1 and end with 0. Then we may apply the argument of Case 4 to w^R to derive a contradiction.

Case 6: u and v both begin and end with 0. Note that since 000 is not a subword of w, $u0^{-1}$ and $v0^{-1}$ both end with 1. Suppose that $0w$ is a subword of $\mathbf{f}$. Then, letting $u' = \phi^{-1}(0u0^{-1})$ and $v' = \phi^{-1}(0v0^{-1})$, it follows that $\phi^{-1}(0w0^{-1}) = u'v'v'u'v'u'u'v'$ is a subword of $\mathbf{f}$, contradicting the minimality of w. Similarly, if $w0$ is a subword of $\mathbf{f}$, we may apply the preceding argument to $0w^R$ to derive a contradiction.

It remains to consider the case that $1w1$ is a subword of $\mathbf{f}$. If $u = 010$ or $v = 010$, then $vuuv$ or $uvvu$, respectively, contains the subword 00100100, a contradiction. Otherwise, since $\mathbf{f}$ does not contain any occurrence of 11, 000, or 10101, one easily checks that u must begin with 0100 and v must end with 0010. But then vu contains the subword 00100100, a contradiction.

In all cases we have derived a contradiction. We conclude that $\mathbf{f}$ avoids the pattern $xyyxyxxy$, as required. □

Proposition 3. *The Fibonacci word avoids the pattern xyyxxyyx.*

Proof. Suppose to the contrary that $\mathbf{f}$ contains a subword $w = uvvuuvvu$, where u and v are non-empty. Since w contains the squares uu, vv, and $(uvvu)^2$, the words u, v, and $uvvu$ are each conjugates of a standard word. Thus $|u| = q_i$, $|v| = q_j$, and $|uvvu| = q_k = 2(q_i + q_j)$, where $i, j < k$. We first show that the equation $q_k = 2(q_i + q_j)$ has only the solution given by $q_4 = 2(q_2 + q_0) = 2(3 + 1) = 8$. Observe first that i and j cannot be consecutive integers. Suppose without loss of generality that $i > j$. Recall that by Zeckendorf's theorem, $q_k/2$ can be represented uniquely as the sum of non-consecutive Fibonacci numbers, and this representation is given by the greedy algorithm. Since $q_k/2 = q_i + q_j$ must therefore be the unique Zeckendorf expansion of $q_k/2$, we see from applying the greedy algorithm that we must have $i = k - 2$ and $j \leq k - 4$. However, one easily proves by induction on k that

$$q_{k-2} + q_{k-5} < q_k/2 < q_{k-2} + q_{k-4}$$

for all $k \geq 5$. Thus we must have $k = 4$, $i = 2$, and $j = 0$, as noted earlier. Then either $v = 0$ and u is a conjugate of 010, or vice-versa. In either case we have

000 as a subword of w, a contradiction. We conclude that $\mathbf{f}$ avoids $xyyxxyyx$, as required. ◻

Theorem 1 now follows.

Acknowledgments

Thanks to Jeffrey Shallit for suggesting a more elegant formulation of our original proof of Proposition 3.

References

[1] D. R. Bean, A. Ehrenfeucht, G. F. McNulty, "Avoidable patterns in strings of symbols", *Pacific J. Math.* **85** (1979), 261–294.

[2] J. Cassaigne, "Unavoidable binary patterns", *Acta Inform.* **30** (1993), 385–395.

[3] F. M. Dekking, "Strongly non-repetitive sequences and progression-free sets", *J. Combin. Theory Ser. A* **27** (1979), 181–185.

[4] D. Guaiana, *On the Binary Patterns of the Thue–Morse Infinite Word*. Internal Report, University of Palermo, 1996.

[5] P. Goralčik, T. Vaniček, "Binary patterns in binary words", *Inter. J. Algebra Comput.* **1**, 387–391.

[6] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge, 2002.

[7] A. Restivo, S. Salemi, "Words and patterns". In DLT 2001, LNCS 2295, pp. 117–129, 2002.

[8] P. Roth, "Every binary pattern of length six is avoidable on the two-letter alphabet", *Acta Inform.* **29** (1992), 95–107.

[9] P. Séébold, *Propriétés Combinatoires des Mots Infinis Engendrés par Certains Morphismes*. Thèse de 3ieme cycle, Rapport LITP 85-16, Paris, 1985.

[10] U. Schmidt, "Avoidable patterns on two letters", *Theoret. Comput. Sci.* **63** (1989), 1–17.

[11] A. M. Shur, "Binary words avoided by the Thue–Morse sequence", *Semigroup Forum* **53** (1996), 212–219.

[12] A. Thue, "Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen", *Kra. Vidensk. Selsk. Skrifter. I. Math. Nat. Kl.* **1** (1912), 1–67.

[13] T. Vaniček, *Unavoidable Words*. Diploma thesis, Charles University, Prague, 1989.

[14] A. I. Zimin, "Blocking sets of terms", *Math. USSR Sbornik* **47** (1984), 353–364.