A Rewriting of Fife's Theorem about Overlap-free Words *

J. Berstel
LITP, Institut Blaise Pascal
Université Pierre et Marie Curie
4, place Jussieu
F-75252 Paris Cedex 05

Abstract

The purpose of this expository paper is to present a self-contained proof of a famous theorem of Fife that gives a full description of the set of infinite overlap-free words over a binary alphabet. Fife's characterization consists in a parameterization of these infinite words by a set of infinite words over a ternary alphabet. The result is that the latter is a regular set. The proof is by the explicit construction of the minimal automaton, obtained by the method of left quotients.

Introduction

One of the first results about avoidable regularities in words was Axel Thue's proof of the existence of an infinite overlap-free words over two letters. In two important papers [16, 17], Thue derived a great number of results in this and related topics. His papers were overseen for a long time (see [6] for a discussion) and his results have been rediscovered several times (e. g. by Morse [10]), when interest in combinatorics on words, both stimulated by symbolic dynamics and computer science, became more important.

Axel Thue also looked for a complete description of all overlap-free and square-free words. His main tools were morphisms and codes (in contemporary terminology). His aim was to express sets of infinite words as homomorphic images of what is now called a *minimal* set. He achieved this very quickly for overlap-free two-sided infinite words (since they form a minimal set), and in his second paper, obtained such a description for large families of square-free infinite words as a result of a more than thirty pages long investigation.

The description of one-sided infinite words, either square-free or overlap-free, is much more involved. It was E. D. Fife [4] who gave, among other deep results,

*Partially supported by PRC "Mathématiques et Informatique" and by ESPRIT BRA working group 6317 – ASMICS 2.

the first full "description" of the set of infinite overlap-free words. His clever method consists in decomposing each such word in longer and longer blocks, where each block is obtained from preceding ones by exactly one among three rules. Coding each rule by a new symbol, he obtains a "description" by an infinite word over a new, ternary alphabet. The truly remarkable result is that the set of all words obtained in this way is regular, that is recognized by a finite automaton (with five states, as we shall see).

The proof of this result is not quite easy. In the terminology of automata theory, it consists in computing the minimal automaton by the well-known method of derivatives (or left quotients). The purpose of this paper is to present this proof in this context. The paper is aimed to be self-contained, excepted for some basic facts on overlap-free words that can be found in Lothaire [9] and Salomaa [14].. After some preliminaries, we give two general, basic lemmas on overlap-free words. In the next section, we present the result of Fife. The last section is devoted to the proof.

Recently, two results have given new insights in this topic. J. Cassaigne [2] and A. Carpi [1] have presented encodings of finite overlap-free words that are similar to Fife's. Both act simultaneously on both ends of the words to be described. J. Cassaigne succeeded in giving explicit recurrence equations for the number of overlap-free words of a given length, a problem that was open for a while; A. Carpi also constructs automata but which are different from Cassaigne's for the description of overlap-free words.

1 Preliminaries

An *alphabet* is a finite set (of *symbols* or *letters*). A *word* over some alphabet A is a (finite) sequence of elements in A. The length of a word w is denoted by $|w|$. The *empty word* of length 0 is denoted by ε. An *infinite* word is a mapping from the set of nonnegative integers into A.

A *factor* of a word w is any word u that occurs in w, i. e. such that there exist word x, y with $w = xuy$. A *square* is a nonempty word of the form uu. A word is *square-free* if none of its factors is a square. Similarly, an *overlap* is a word of the form $xuxux$, where x is nonempty. The terminology is justified by the fact that xux has two occurrences in $xuxux$, one as a *prefix* (initial factor) one as a *suffix* (final factor) and that these occurrences have a common part (the central x). As before, a word is *overlap-free* if none of its factors is an overlap.

The set of words over A is denoted by A^*. A function $h : A^* \longrightarrow B^*$ is a *morphism* if $h(uv) = h(u)h(v)$ for all words u, v. If there is a letter a such that $h(a)$ starts with the letter a, then $h^n(a)$ starts with the word $h^{n-1}(a)$ for all $n > 0$. If the set words $\{h^n(a)) \mid n \geq 0\}$ is infinite, the morphism is *prolongeable* in a and defines a unique infinite word say $\mathbf{x}$ by the requirement that all $h^n(a)$ are prefixes of $\mathbf{x}$. The word $\mathbf{x}$ is said to be obtained by iterating h on a, and $\mathbf{x}$ is also denoted by $h^\omega(a)$. Clearly, $\mathbf{x}$ is a fixed point of h. For a detailed discussion and results on iterating morphisms, see [3].

2 The Thue-Morse sequence

In this section, we recall some basic properties concerning the Thue-Morse sequence. Other properties and proofs can be found in Lothaire [9] and Salomaa [14].

Let $A = \{a, b\}$ be a two letter alphabet. Consider the morphism μ from the free monoid A^* into itself defined by

$$\mu(a) = ab, \qquad \mu(b) = ba$$

Setting, for $n \geq 0$,

$$u_n = \mu^n(a), \qquad v_n = \mu^n(b)$$

one gets

$$\begin{array}{ll} u_0 = a & v_0 = b \\ u_1 = ab & v_1 = ba \\ u_2 = abba & v_2 = baab \\ u_3 = abbabaab & v_3 = baababba \\ \quad \cdots & \end{array}$$

and more generally

$$u_{n+1} = u_n v_n, \qquad v_{n+1} = v_n u_n$$

and

$$u_n = \overline{v}_n, \qquad v_n = \overline{u}_n$$

where $\overline{w}$ (the *opposite* of w) is obtained from w by exchanging a and b. Words u_n and v_n are *Morse blocks.* It is easily seen that u_{2n} and v_{2n} are palindromes, and that $u_{2n+1} = \tilde{v}_{2n+1}$, where $w^\sim$ is the reversal of w. The morphism μ can be extended to infinite words; it has two fixed points

$$\mathbf{t} = abbabaabbaababbabaab \cdots = \mu(\mathbf{t})$$

$$\overline{\mathbf{t}} = baababbaabbabaababba \cdots = \mu(\overline{\mathbf{t}})$$

and u_n (resp. v_n) is the prefix of length 2^n of $\mathbf{t}$ (resp. of $\overline{\mathbf{t}}$). It is equivalent to say that $\mathbf{t}$ is the limit of the sequence $(u_n)_{n \geq 0}$ (for the usual topology on finite and infinite words), obtained by iterating the morphism μ.

The *Thue-Morse sequence* is the word $\mathbf{t}$. There are several other characterizations of this word. For instance, let t_n be the n-th symbol in $\mathbf{t}$, starting with $n = 0$. Then $t_n = a$ or $t_n = b$ according to the parity of the number of bits equal to 1 in the binary expansion of n. For instance, $bin(19) = 10011$, consequently $d_1(19) = 3$, and indeed $t_{19} = a$.

Theorem 2.1 [17](Satz 6) *The sequence* $\mathbf{t}$ *is overlap-free.*

What Thue actually shows is that a word w is overlap-free iff $\mu(w)$ is overlap-free.

3 Factorization of overlap-free words

The following lemmas have been given by many peoples independently (e. g. Shelton and Soni [15], Kobayashi [8], Restivo and Salemi [11], Kfoury [7].)

Lemme 3.1 *("Progression Lemma") Let $n \geq 0$ and let $x = uvwc$ be an overlap-free word of length $1 + 3 \cdot 2^n$, with $|u| = |v| = |w| = 2^n$ and $c \in A$. If u and v are Morse blocks, then w is a Morse block.*

Proof. By induction on n. The result is clear for $n = 0$. Assume $n \geq 1$. By assumption, x has the form

$$x = UVUVBCc, \quad \text{or} \quad x = UVVUBCc$$

where U and V are the Morse blocks of size 2^{n-1} and $|B| = |C| = 2^{n-1}$. By induction, both B and C are Morse blocks. It remains to show that $BC \neq UU$ and $BC \neq VV$.

If $x = UVUVBCa$, then $BC \neq UU, VV$ since otherwise x has an overlap. If $x = UVVUBCa$, then clearly $BC \neq UU$. Suppose $BC = VV$. Then $x = (UVV)^2 a$, but a can be neither the first letter of U nor the first letter of V without producing an overlap in x. The proof is complete. ■

Lemme 3.2 *("Factorization Lemma") Let x be an overlap-free word. There exist three words u, v, y, with $u, v \in \{\varepsilon, a, b, aa, bb\}$, such that*

$$x = u\mu(y)v.$$

Moreover, the triple (u, y, v) is unique if $|x| \geq 7$.

Proof. The result is straightforward by inspection if $|x| \leq 5$. Suppose $|x| \geq 6$. We show that x contains two consecutive Morse blocks ab or ba. The result then follows from the progression lemma.

By symmetry, we may suppose that x starts with a. The possible prefixes of x, developed up to an encounter of two consecutive Morse blocks ab or ba are:

$$aabaab, aabab, aabba, abaab, abab, abba$$

This shows that the prefixes are of the required form. To prove uniqueness, consider two triples (u, y, v) and (u', y', v') such that $x = u\mu(y)v = u'\mu(y')v'$. Since $|x| \geq 7$, one has $|y|, |y'| \geq 2$. But then the occurrences of $\mu(y)$ and $\mu(y')$ cannot overlap without being equal. This shows uniqueness. ■

As an illustration, we mention the following result, already known to A. Thue (for a related result, see T. Harju [5]):

Theorem 3.3 *The overlap-free squares over A are the words*

$$u_n^2, \quad (u_n v_n u_n)^2$$

for $n \geq 0$, their opposites, and their conjugates.

As a consequence, if xx is an overlap-free square, then $|x| = 3 \cdot 2^n$ or $|x| = 2^n$ for some n.

4 Fife's Theory

Let $X_n = \{u_n, v_n\}$ denote the set of Morse blocks of length 2^n and set $X = \bigcup_{n \geq 0} X_n$.

Let $w \in A^*X_1$. Thus w ends with ab or ba. The *canonical decomposition* of w is the triple

$$(z, y, \bar{y})$$

where y is the longest word in X such that

$$w = zy\bar{y}$$

In other terms, $(z, y, \bar{y})$ is the canonical decomposition of w iff $\bar{y}y$ is not a suffix of z. As an example, the canonical decomposition of $aabaabbabaab$ is

$$(aaba, abba, baab)$$

and that of $aabaabbabaababbaabbabaabbaababbaabbabaab$ is

$$(aabaabbabaababbaabbabaab, baababba, abbabaab)$$

Define now three mappings $\alpha, \beta, \gamma : A^*X_1 \longrightarrow A^*X_1$, written on the right of their arguments like actions, as follows: let $w \in A^*X_1$ have the canonical decomposition $(z, y, \bar{y})$, then

$$\begin{aligned}
w \cdot \alpha &= zy\bar{y} \cdot \alpha = zy\bar{y}yy\bar{y} = wyy\bar{y} \\
w \cdot \beta &= zy\bar{y} \cdot \beta = zy\bar{y}y\bar{y}\bar{y}y = wy\bar{y}\bar{y}y \\
w \cdot \gamma &= zy\bar{y} \cdot \gamma = zy\bar{y}\bar{y}y = w\bar{y}y
\end{aligned}$$

Setting

$$B = \{\alpha, \beta, \gamma\}$$

the word $w \cdot f$ is well defined for all $f \in B^*$. Since w is a prefix of $w \cdot \alpha$, $w \cdot \beta$ and $w \cdot \gamma$, the infinite word $w \cdot \mathbf{f}$ is well defined of any infinite word $\mathbf{f}$ over B. A finite or infinite word f over B is called a *description* of the finite or infinite word x if $x = ab \cdot f$ or $x = aab \cdot f$ (or symmetrically $x = ba \cdot f$ or $x = bba \cdot f$). Here are some examples:

$$\begin{aligned}
&ab \cdot \alpha = abaab \\
&ab \cdot \beta = ababba \\
&ab \cdot \gamma = abba \\
&aba\ ab\ ba \cdot \alpha = aba\ ab\ ba\ ab\ ab\ ba = abaabbaababba \\
&aba\ ab\ ba \cdot \beta = aba\ ab\ ba\ ab\ ba\ ba\ ab = abaabbaabbabaab \\
&ab \cdot \gamma^\omega = \mathbf{t} \\
&aab \cdot \alpha = aabaab = a(ab \cdot \alpha) \\
&aab \cdot \alpha^3\gamma = aabaabbabaababbaabbabaabbaababbaabbabaab
\end{aligned}$$

The word

$$\begin{aligned}
u &= aaba\ abbabaab\ abbabaab\ baababba\ baababba\ abbabaab\ baababba\ baabbaa \\
&= (aab \cdot \alpha\alpha\beta\gamma)baababba\ baabbaa
\end{aligned}$$

of Restivo and Salemi [11] admits no description. As we shall see, this means that it is not the prefix of an infinite overlap-free word.

Proposition 4.1 *Every infinite overlap-free word admits a unique description.*

Proof. This is a simple application of the progression lemma. ∎

Let

$$I = \{\alpha, \beta\}(\gamma^2)^*\{\beta\alpha, \gamma\beta, \alpha\gamma\}$$

and consider the set

$$F = B^\omega - B^* I B^\omega$$

of infinite words over B having no factor in I, and the set

$$G = \{\mathbf{f} \mid \beta\mathbf{f} \in F\}$$

Theorem 4.2 ("Fife's Theorem") *Let* $\mathbf{x}$ *be an infinite word over* A.

(1) *If* $\mathbf{x}$ *starts with* ab, *then* $\mathbf{x}$ *is overlap-free iff its description is in* F;

(2) *If* $\mathbf{x}$ *starts with* aab, *then* $\mathbf{x}$ *is overlap-free iff its description is in* G.

The set F of Fife's words is recognized by an automaton with 5 states, given in the following figure.

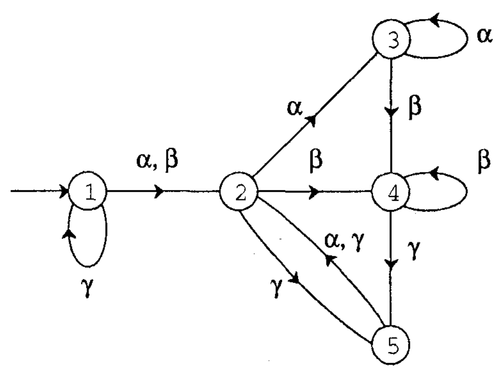

Fig. 1 *Fife's automaton.*

Fife's theorem has a number of consequences. Call a word w *infinitely extensible* if it is a prefix of an infinite overlap-free word. Then one has:

Corollaire 4.3 *A word* w *is infinitely extensible iff it is a prefix of a finite word that admits a description which is a prefix of a word in* F *or* G. *It is decidable whether a word is infinitely extensible.*

Indeed, it is easily seen that if w is a prefix of a word x that admits a (finite) description, then $|x| \leq 2|w|$. Another consequence is:

Corollaire 4.4 *The Thue-Morse* $\mathbf{t}$ *is the greatest infinite overlap-free word, for the lexicographic order, among those starting with the letter* a.

Proof. If one chooses $a < b$ and $\alpha < \beta < \gamma$ then indeed $\mathbf{f} \le \mathbf{f}'$ implies $ab{\cdot}\mathbf{f} \le ab{\cdot}\mathbf{f}'$. Now the greatest word in F is γ^ω and $\mathbf{t} = ab \cdot \gamma^\omega$. ∎

Observe that this result can also be proved directly, by arguing on the form of overlap-free words, and using the progression lemma.

5 Proof

We observe first that the second statement of the theorem is a consequence of the first statement. Indeed, let x be an infinite overlap-free word starting with aab, and let $\mathbf{f}$ be its description (which exists by the proposition). To prove that $\beta\mathbf{f}$ is in F, observe that

$$\mu(aab \cdot \mathbf{f}) = \mu(aab) \cdot \mathbf{f} = ababba \cdot \mathbf{f} = ab \cdot \beta\mathbf{f}$$

and since $aab \cdot \mathbf{f}$ is overlap-free iff $\mu(aab \cdot \mathbf{f})$ is overlap-free, the word $aab \cdot \mathbf{f}$ is overlap-free iff $ab \cdot \beta\mathbf{f}$ is overlap-free, thus iff $\beta\mathbf{f} \in F$.

It is convenient to use, for the proof, the notation n for $u_n = \mu^n(a)$, and symmetrically $\overline{n}$ for $v_n = \mu^n(b)$. (Consider n as a shorthand for μ^n.) For example

$$\begin{array}{ll} 0 = a, & \overline{0} = b \\ 1 = ab, & \overline{1} = ba \\ 2 = abba, & 3 = abbabaab \end{array}$$

It follows that

$$\begin{array}{ll} 1 \cdot \alpha = 0\overline{2}, & n \cdot \alpha = (n-1)\overline{(n+1)} \\ 1 \cdot \beta = 12, & n \cdot \beta = n(n+1) \\ 1 \cdot \gamma = 2, & n \cdot \gamma = n+1 \end{array}$$

We denote by P the set of finite overlap-free words over A and by W those words over B that are description of words in P starting with $1 = ab$:

$$W = \{f \in B^* \mid 1 \cdot f \in P\}$$

Recall that

$$I = \{\alpha, \beta\}(\gamma^2)^*\{\beta\alpha, \gamma\beta, \alpha\gamma\}$$

Fife's theorem is a straightforward extension to infinite words of the following:

Theorem 5.1 *One has* $W = B^* - B^* I B^*$.

We start with a useful observation:

Proposition 5.2 *The set* W *is factorial : if* $1 \cdot fgh$ *is overlap-free, then* $1 \cdot g$ *is overlap-free.*

Proof. We show first that W is suffix-closed, by showing that if $\alpha f \in W$, then $f \in W$, and similarly for β and γ. Now

$$\begin{aligned} 1 \cdot \alpha f &= 0\overline{2} \cdot f = 0\overline{2 \cdot f} = 0\overline{\mu(1 \cdot f)} \\ 1 \cdot \beta f &= 12 \cdot f = 1\mu(1 \cdot f) \\ 1 \cdot \gamma f &= \mu(1 \cdot f) \end{aligned}$$

This shows that in all three cases, the word $1 \cdot f$ is overlap-free.

We now prove that W is prefix-closed. Let $fg \in W$ and set $w = 1 \cdot fg$ and $u = 1 \cdot f$. Then $w = u \cdot g$ and u is a prefix of w. Consequently u is overlap-free and $f \in W$. This completes the proof. ■

For the proof of 5.1, we compute the minimal automaton of the set W. This will be done by the method of quotients. For a word u and a set Y, we definie

$$u^{-1}Y = \{w \mid uw \in Y\}$$

We shall see that the minimal automaton of W is the automaton of the figure which recognizes $B^* - B^*IB^*$. This shows the theorem.

We start by the following easy properties:

Lemme 5.3 $(\alpha^2\gamma)^{-1}W = (\alpha\beta\alpha)^{-1}W = (\alpha\gamma\beta)^{-1}W = \emptyset$.

Proof. It suffice to verify that the words $1 \cdot \alpha^2\gamma$, $1 \cdot \alpha\beta\alpha$ and $1 \cdot \alpha\gamma\beta$ all have an overlap. Indeed:

$$\begin{aligned} 1 \cdot \alpha^2\gamma &= abaabb\ abaabb\ a\ ababba \\ 1 \cdot \alpha\beta\alpha &= abaabbaababb\ abaabbaababb\ a \\ 1 \cdot \alpha\gamma\beta &= abaababb\ abaababb\ a\ abbabaab \end{aligned}$$

■

The following equations are more difficult:

Proposition 5.4 *The following equations hold for W :*

(*i*) $W = \gamma^{-1}W$;
(*ii*) $\alpha^{-1}W = \beta^{-1}W = (\alpha\gamma\alpha)^{-1}W = (\alpha\gamma\gamma)^{-1}W$;
(*iii*) $(\alpha^2)^{-1}W = (\alpha^3)^{-1}W$;
(*iv*) $(\alpha\beta)^{-1}W = (\alpha^2\beta)^{-1}W$;
(*v*) $(\alpha\gamma)^{-1}W = (\alpha\beta\gamma)^{-1}W$.

Let P_a be the set of overlap-free words that have no prefix that is a square ending with the letter a. Thus $w \in P_a$ iff for each prefix $xcxc$ of w with c a letter, on has $c = b$. We show that $aw \in P \iff w \in P_a$, that is

$$P_a = a^{-1}P.$$

Indeed, let $w \in P_a$. If aw has an overlap, this overlap is a prefix of aw, and has the form $axaxa$. But then $xaxa$ is a prefix of w, a contradiction. Thus $w \in P_a$. The converse is straightforward. The set P_b is defined similarly. Set

$$W_a = \{f \in W \mid 1 \cdot f \in P_a\}, \quad W_b = \{f \in W \mid 1 \cdot f \in P_b\}$$

Then:

Proposition 5.5 *The following relations hold:*

(1) $f \in W \iff \gamma f \in W$;
$f \in W_a \iff \gamma f \in W_b$;
$f \in W_b \iff \gamma f \in W_a$;
(2) $\alpha f \in W \iff \alpha f \in W_b \iff f \in W_a$;
(3) $\beta f \in W \iff f \in W_a$;
(4) $\alpha^2 f \in W \Rightarrow \alpha^2 f \in W_a$;
(5) $\alpha\beta f \in W \Rightarrow \alpha\beta f \in W_a$;
(6) $\beta\gamma f \in W \Rightarrow \beta\gamma f \in W_a$.

Proof of proposition 5.4.
(i). From (1).
(ii). From (2) and (3), it follows that $\alpha f \in W \iff \beta f \in W$. Next

$$\begin{aligned}\alpha f \in W &\iff \alpha f \in W_b \iff \gamma\alpha f \in W_a \iff \alpha\gamma\alpha f \in W \\ \alpha f \in W &\iff f \in W_a \iff \gamma\gamma f \in W_a \iff \alpha\gamma\gamma f \in W\end{aligned}$$

(iii). From (4) and (1), one obtains

$$\alpha^2 f \in W \Rightarrow \alpha^2 f \in W_a \Rightarrow \alpha^3 f \in W$$

the converse implication holds because W is prefix-closed.
(iv). From (5),

$$\alpha\beta f \in W \Rightarrow \alpha\beta f \in W_a \Rightarrow \alpha^2\beta f \in W$$

the converse implication holds because W is suffix-closed.
(v). From (ii),(6) and (2), one gets

$$\alpha\gamma f \in W \iff \beta\gamma f \in W \Rightarrow \beta\gamma f \in W_a \Rightarrow \alpha\beta\gamma f \in W \qquad \blacksquare$$

It remains to prove proposition 5.5. For this, we use the following lemma:

Lemme 5.6 *Let w be a word in P. Then*
(a) *if $w \in abaabbaX_1^*$, then $w \in P_a$;*
(b) *if $w \in aabbaX_1^*$, then $w \in P_b$;*
(c) *if $w \in abaabX_1^*$, then $w \in P_b$.*

Proof of proposition 5.5.

(1). First $1 \cdot f \in P \iff \mu(1 \cdot f) = 1 \cdot \gamma f \in P$. Next, let $f \in W_a$ and suppose $\mu(1 \cdot f) = ububv$. Then $|ub| \neq 3$, since otherwise $u = ab$ and $ubub = abbabb \notin X_1^*$, or $u = ba$ and $ubub = babbab \notin X_1^*$. Thus $|ub|$ is even, and $1 \cdot f \notin P_a$. The converse is immediate.

(2). One has $w = 1 \cdot \alpha f = 0\overline{2} \cdot f = abaabv$ for some $v \in X_1^*$, and by (c) of the lemma, one has $w \in P_b$. Thus $\alpha f \in W_b$. Next

$$\begin{aligned}\alpha f \in W_b &\iff 0\overline{2} \cdot f \in P_b \\ f \in W_a &\iff 1 \cdot F \in P_a \iff \overline{1} \cdot f \in P_b \iff \overline{2} \cdot f \in P_a\end{aligned}$$

Thus it remains to show that $0\overline{2} \cdot f \in P_b \iff \overline{2} \cdot f \in P_a$. If $0\overline{2} \cdot f \in P_b$ then $\overline{2} \cdot f \in P_a$ since otherwise $\overline{2} \cdot f$ has an overlap. Conversely, if $\overline{2} \cdot f \in P_a$, then $0\overline{2} \cdot f = a\overline{2} \cdot f$ is overlap-free and, again by (c) of the lemma, it is in P_b.

(3). One has $w = 1 \cdot \beta f = 12 \cdot f = ababba \cdot f = \mu(aab \cdot f) = \mu(a(1 \cdot f))$. If $\beta f \in W$, then $w \in P$, whence $a(1 \cdot f) \in P$, and $f \in W$, and even $f \in W_a$. Conversely, if $f \in W_a$, then $a(1 \cdot f) \in P$, whence $w \in P$ and $\beta f \in W$.

(4). One has $w = 1 \cdot \alpha^2 f = 0\overline{1}3 \cdot f \in abaabbaX_1^* \cap P$, and by (a) of the lemma, $\alpha^2 f \in W_a$.

(5). One has $w = 1 \cdot \alpha\beta f = 0\overline{23} \cdot f \in abaabbaX_1^* \cap P$, and by (a) of the lemma, $\alpha\beta f \in W_a$.

(6). One has $w = 1 \cdot \beta\gamma f = 1\mu(2 \cdot f) = \mu(aabbav)$ for some $v \in X_1^*$. By statement (b) of the lemma, $aabbav \in P_b$, whence $w \in P_a$. ∎

Proof of the lemma.

(a). Suppose the conclusion is false. Then

$$w = abaabbaw' = uuv$$

where u end with an a. The word u has not length 3, hence it has even length, and is of the form $u = au'a$, with u' of even length. But then $u'aa$ is in X_1^*, a contradiction.

(b). Suppose the conclusion is false. Then

$$w = aabbaw' = uuv = (au'b)(au'b)v$$

Again, u is not of length 3, hence it has even length. Since $u'bau'b$ has odd length, the word bv is in X_1^*, and v starts by a letter a and w has an overlap, contradiction.

(c). Suppose the conclusion is false. Then

$$w = abaabw' = uuv = (au'b)(au'b)v$$

Again, u has even length because its length is not 3, and $bv \in X_1^*$, thus v starts with an a and w has an overlap, contradiction. ∎

This ends the proof of Fife's theorem. Let us mention again two finitary versions of this result, which are more complicated, due to J. Cassaigne and A. Carpi.

References

[1] A. Carpi, Overlap-free words and finite automata, *Theoret. Comput. Sci.* **115** (1993), 243–260.

[2] J. Cassaigne, Counting overlap-free binary words, in: *STACS'93*, Enjalbert, Finkel, Wagner (eds), Lect. Notes Comp. Sci.,**665**, Springer-Verlag, 1993, 216–225.

[3] K. CULIK II AND A. SALOMAA, On infinite words obtained by iterating morphisms, *Theoret. Comput. Sci.* **19** (1982), 29–38.

[4] E. D. FIFE, Binary sequences which contain no BBb, *Trans. Amer. Math. Soc.* **261** (1980), 115–136.

[5] T. HARJU, On cyclically overlap-free words in binary alphabets, *The Book of L*, Springer-Verlag, 1986, 123–130.

[6] G.A. HEDLUND, Remarks on the work of Axel Thue, *Nordisk Mat. Tidskr.* **15** (1967), 148–150.

[7] R. KFOURY, A linear time algorithm to decide whether a binary word contains an overlap, *Theoret. Inform. Appl.* **22** (1988), 135–145.

[8] Y. KOBAYASHI, Enumeration of irreductible binary words, *Discrete Appl. Math.***20** (1988), 221–232.

[9] M. LOTHAIRE, *Combinatorics on Words*, Addison-Wesley, 1983.

[10] M. MORSE, Recurrent geodesics on a surface of negative curvature, *Transactions Amer. Math. Soc.* **22** (1921), 84–100.

[11] A. RESTIVO, S. SALEMI, Overlap-free words on two symbols, in: *Automata on infinite words*, Nivat, Perrin (eds), Lect. Notes Comp. Sci. .,**192**, Springer-Verlag, 1985, 198–206.

[12] G. ROZENBERG, A. SALOMAA, *The Mathematical Theory of L-Systems*, Academic Press, 1980.

[13] A. SALOMAA, Morphisms on free monoids and language theory, in *Formal Language Theory : Perspectives and Open Problems*, pp. 141–166, Academic Press, 1980.

[14] A. SALOMAA, *Jewels of Formal Language Theory*, Computer Science Press, 1981.

[15] R. SHELTON, R. SONI, Chains and fixing blocks in irreducible sequences, *Discrete Math.***54** (1985), 93–99.

[16] A. THUE, Über unendliche Zeichenreihen, *Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl.*, Christiana 1906, Nr. 7.

[17] A. THUE, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl.*, Christiana 1912, Nr. 10.

[18] A. THUE, *Selected Mathematical Papers*, edited by T. Nagell, A. Selberg, S. Selberg, K. Thalberg, Universitetsforlaget, Oslo 1977.