Fife's Theorem for $\frac{7}{3}$-Powers

Narad Rampersad
Department of Mathematics, University of Liège, Grande Traverse, 12 (Bat. B37), 4000 Liège, Belgium
narad.rampersad@gmail.com

Jeffrey Shallit
School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada
shallit@cs.uwaterloo.ca

Arseny Shur
Department of Algebra and Discrete Mathematics, Ural State University, Ekaterinburg, Russia
arseny.shur@usu.ru

We prove a Fife-like characterization of the infinite binary $\frac{7}{3}$-power-free words, by giving a finite automaton of 15 states that encodes all such words. As a consequence, we characterize all such words that are 2-automatic.

1 Introduction

An *overlap* is a word of the form *axaxa*, where *a* is a single letter and *x* is a (possibly empty) word. In 1980, Earl Fife [8] proved a theorem characterizing the infinite binary overlap-free words as encodings of paths in a finite automaton. Berstel [4] later simplified the exposition, and both Carpi [6] and Cassaigne [7] gave an analogous analysis for the case of finite words.

In a previous paper [13], the second author gave a new approach to Fife's theorem, based on the factorization theorem of Restivo and Salemi [12] for overlap-free words. In this paper, we extend this analysis by applying it to the case of $\frac{7}{3}$-power-free words.

Given a rational number $\frac{p}{q} > 1$, we define a word w to be a $\frac{p}{q}$-power if w can be written in the form $x^n x'$ where $n = \lfloor p/q \rfloor$, x' is a (possibly empty) prefix of x, and $|w|/|x| = p/q$. The word x is called a *period* of w, and p/q is an *exponent* of w. If p/q is the largest exponent of w, we write $\exp(w) = p/q$. We also say that w is $|x|$-*periodic*. For example, the word `alfalfa` is a $\frac{7}{3}$-power, and the corresponding period is `alf`. Sometimes, as is routine in the literature, we also refer to $|x|$ as the period; the context should make it clear which is meant.

A word, whether finite or infinite, is β-*power-free* if it contains no factor w that is an α-power for $\alpha \geq \beta$. A word is β^+-*power-free* if it contains no factor w that is an α-power for $\alpha > \beta$. Thus, the concepts of "overlap-free" and "2^+-power-free" coincide.

2 Notation and basic results

Let Σ be a finite alphabet. We let Σ^* denote the set of all finite words over Σ and Σ^ω denote the set of all (right-) infinite words over Σ. We say y is a *factor* of a word w if there exist words x, z such that $w = xyz$.

If x is a finite word, then x^ω represents the infinite word $xxx\cdots$.

P. Ambrož, Š. Holub and Z. Masáková (Eds.):
8th International Conference WORDS 2011
EPTCS 63, 2011, pp. 189–198, doi:10.4204/EPTCS.63.25

From now on we fix $\Sigma = \{0,1\}$. The most famous infinite binary overlap-free word is $\mathbf{t}$, the Thue-Morse word, defined as the fixed point, starting with 0, of the Thue-Morse morphism μ, which maps 0 to 01 and 1 to 10. We have

$$\mathbf{t} = t_0t_1t_2\cdots = 0110100110010110\cdots .$$

The morphism μ has a second fixed point, $\bar{\mathbf{t}}$, which is obtained from $\mathbf{t}$ by applying the complementation coding defined by $\overline{0} = 1$ and $\overline{1} = 0$.

We let $\mathscr{F}_{7/3}$ denote the set of (right-) infinite binary $\frac{7}{3}$-power-free words. We point out that these words are of particular interest, because $\frac{7}{3}$ is the largest exponent α such that there are only polynomially-many α-power-free words of length n [9]. The exponent $\frac{7}{3}$ plays a special role in combinatorics on words, as testified to by the many papers mentioning this exponent (e.g., [10, 14, 9, 11, 1, 5]).

We now state a factorization theorem for infinite $\frac{7}{3}$-power-free words:

Theorem 1. *Let $\mathbf{x} \in \mathscr{F}_{7/3}$, and let $P = \{p_0, p_1, p_2, p_3, p_4\}$, where $p_0 = \varepsilon$, $p_1 = 0$, $p_2 = 00$, $p_3 = 1$, and $p_4 = 11$. Then there exists $\mathbf{y} \in \mathscr{F}_{7/3}$ and $p \in P$ such that $\mathbf{x} = p\mu(\mathbf{y})$. Furthermore, this factorization is unique, and p is uniquely determined by inspecting the first 5 letters of $\mathbf{x}$.*

Proof. The first two claims follow immediately from the version for finite words, as given in [9]. The last claim follows from exhaustive enumeration of cases. □

We can now iterate this factorization theorem to get

Corollary 2. *Every infinite $\frac{7}{3}$-power-free word $\mathbf{x}$ can be written uniquely in the form*

$$\mathbf{x} = p_{i_1}\mu(p_{i_2}\mu(p_{i_3}\mu(\cdots))) \tag{1}$$

with $i_j \in \{0,1,2,3,4\}$ for $j \geq 1$, subject to the understanding that if there exists c such that $i_j = 0$ for $j \geq c$, then we also need to specify whether the "tail" of the expansion represents $\mu^\omega(0) = \mathbf{t}$ or $\mu^\omega(1) = \bar{\mathbf{t}}$. Furthermore, every truncated expansion

$$p_{i_1}\mu(p_{i_2}\mu(p_{i_3}\mu(\cdots p_{i_{n-1}}\mu(p_{i_n})\cdots)))$$

is a prefix of $\mathbf{x}$, with the understanding that if $i_n = 0$, then we need to replace p_{i_n} with either 1 (if the "tail" represents $\mathbf{t}$) or 3 (if the "tail" represents $\bar{\mathbf{t}}$).

Proof. The form (1) is unique, since each p_i is uniquely determined by the first 5 characters of the associated word. □

Thus, we can associate each infinite binary $\frac{7}{3}$-power-free word $\mathbf{x}$ with the essentially unique infinite sequence of indices $\mathbf{i} := (i_j)_{j\geq 0}$ coding elements in P, as specified by (1). If $\mathbf{i}$ ends in 0^ω, then we need an additional element (either 1 or 3) to disambiguate between $\mathbf{t}$ and $\bar{\mathbf{t}}$ as the "tail". In our notation, we separate this additional element with a semicolon so that, for example, the string $000\cdots;1$ represents $\mathbf{t}$ and $000\cdots;3$ represents $\bar{\mathbf{t}}$.

Of course, not every possible sequence of $(i_j)_{j\geq 1}$ of indices corresponds to an infinite $\frac{7}{3}$-power-free word. For example, every infinite word coded by an infinite sequence beginning $400\cdots$ has a $\frac{7}{3}$-power.Our goal is to characterize precisely, using a finite automaton, those infinite sequences corresponding to $\frac{7}{3}$-power-free words.

Next, we recall some connections between the morphism μ and the powers over the binary alphabet. Below x is an arbitrary finite or right-infinite word.

Lemma 3. *If the word $\mu(x)$ has a prefix zz, then the word x has the prefix $\mu^{-1}(z)\mu^{-1}(z)$.*

Proof. Follows immediately from [3, Lemma 1.7.2]. ☐

Lemma 4. *(1) For any real $\beta > 1$, we have $\exp(x) = \beta$ iff $\exp(\mu(x)) = \beta$.*
(2) For any real $\beta \geq 2^+$, the word x is β-power-free iff $\mu(x)$ is β-power-free.

Proof. For (1), see [14, Prop. 1.1]. For (2), see [14, Prop. 1.2] or [9, Thm. 5]. ☐

Lemma 5. *Let p be a positive integer. If the longest p-periodic prefix of the word $\mu(x)$ has the exponent $\beta \geq 2$, then the longest $(p/2)$-periodic prefix of x also has the exponent β.*

Proof. Let zzz' (where $|z| = p$ and z' is a possibly empty prefix of z^ω) be the longest p-periodic prefix of $\mu(x)$. Lemma 3 implies that p is even. If $|z'|$ is odd, let a be the last letter of z'. The next letter b in $\mu(x)$ is fixed by the definition of μ: $b \neq a$. By the definition of period, another a occurs p symbols to the left of the last letter of z'. Since p is even, this a also fixes the next letter b. Hence the prefix $zzz'b$ of $\mu(x)$ is p-periodic, contradicting the definition of zzz'. Thus $|z'|$ is even. Therefore x begins with the β-power $\mu^{-1}(z)\mu^{-1}(z)\mu^{-1}(z')$ of period $p/2$.

It remains to note that if x has a $(p/2)$-periodic prefix y of exponent $\alpha > \beta$, then by Lemma 4 (1), the p-periodic prefix $\mu(y)$ of $\mu(x)$ also has the exponent α, contradicting the hypotheses of the lemma. ☐

3 The main result

For each finite word $w \in \{0,1,2,3,4\}^*$, $w = i_1 i_2 \cdots i_r$, and an infinite word $\mathbf{x} \in \{0,1\}^\omega$, we define

$$C_w(\mathbf{x}) = p_{i_1}\mu(p_{i_2}\mu(p_{i_3}\mu(\cdots \mathbf{x} \cdots))) \text{ and}$$
$$F_w = \{\mathbf{x} \in \Sigma^\omega \; : \; C_w(\mathbf{x}) \in \mathscr{F}_{7/3}\}.$$

Note that $F_w \subseteq \mathscr{F}_{7/3}$ for any w in view of Lemma 4 (2).

Lemma 6. *The sets F_w satisfy the equalities listed in Fig. 1. In particular, there are only 15 different nonempty sets F_w; they are*

$$F_\varepsilon, F_1, F_{11}, F_{13}, F_{130}, F_2, F_{20}, F_{203}, F_3, F_{31}, F_{310}, F_{33}, F_4, F_{40}, F_{401}.$$

Proof. Due to symmetry, it is enough to prove only the 30 equalities from the upper half of Fig. 1 and the equality $F_0 = F_\varepsilon$. We first prove the emptiness of 15 sets from the upper half of Fig. 1.

Four sets: F_{21}, F_{22}, F_{201}, and F_{202}, consist of words that start 000.

Eight sets consist of words that contain the factor $0\mu(11) = 01010$ (F_{14}, F_{24}, F_{133}, F_{134}, F_{1303}, and F_{1304}), its μ-image (F_{114}), or the complement of its μ-image (F_{132}).

Two sets: F_{2031} and F_{2032}, consist of words that start $00\mu^2(1)0 = 0010010$. Finally, the words from the set F_{200} have the form $00\mu^3(\mathbf{x})$; each of these words starts either 000 or 0010010.

Each of the 16 remaining equalities has the form $F_{w_1} = F_{w_2}$. We prove them by showing that for an arbitrary $\mathbf{x} \in \mathscr{F}_{7/3}$, the words $u_1 = C_{w_1}(\mathbf{x})$ and $u_2 = C_{w_2}(\mathbf{x})$ are either both $\frac{7}{3}$-power-free or both not. In most cases, some suffix of u_1 coincides with the image of u_2 under some power of μ. Then by Lemma 4 (2) the word u_1 can be $\frac{7}{3}$-power-free only if u_2 is $\frac{7}{3}$-power-free. In these cases, it suffices to study u_1 assuming that u_2 is $\frac{7}{3}$-power-free.

When we refer to a "forbidden" power in what follows, we mean a power of exponent $\geq \frac{7}{3}$.

$F_0 = F_\varepsilon$: By Lemma 4 (2), $u_1 = \mu(\mathbf{x})$ is $\frac{7}{3}$-power-free iff $u_2 = \mathbf{x}$ is $\frac{7}{3}$-power-free.

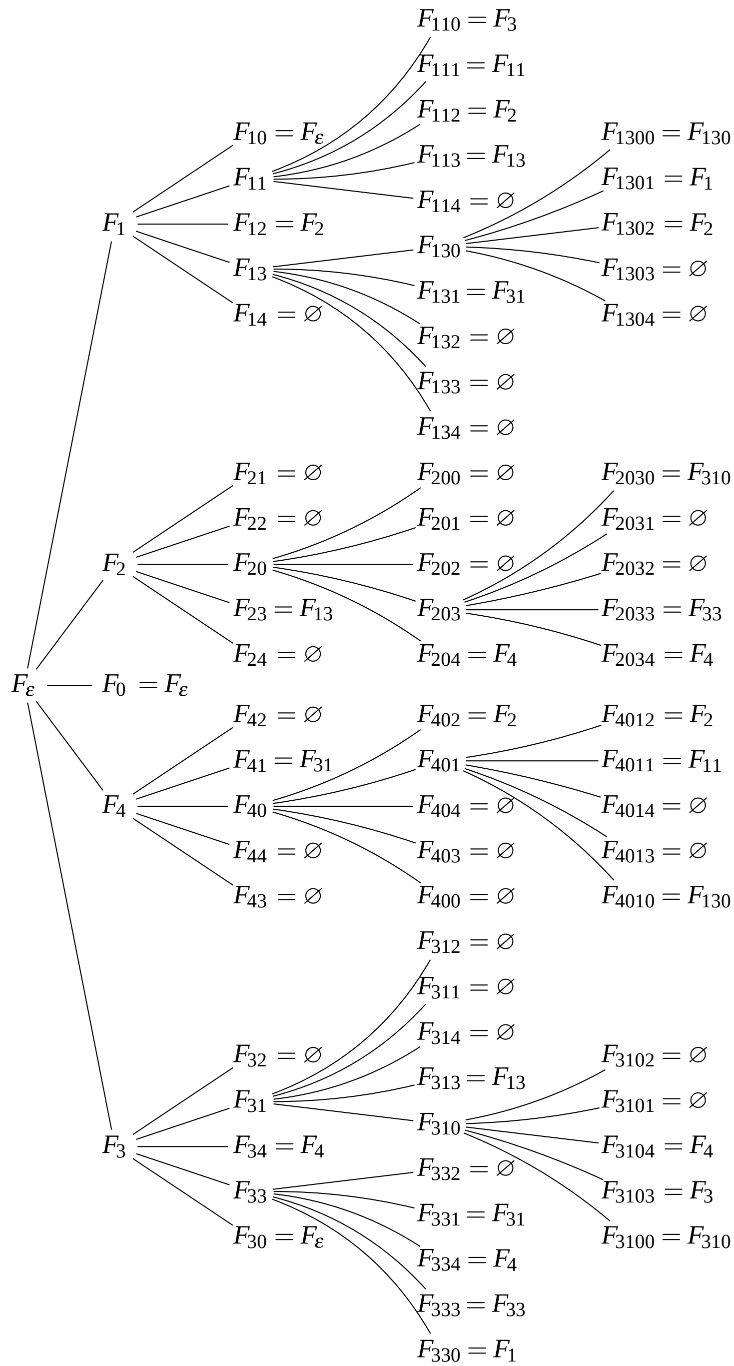

Figure 1: *Equations between languages* F_w.

$F_{10} = F_\varepsilon$: The word $u_1 = 0\mu(\mu(\mathbf{x}))$ contains a μ^2-image of $u_2 = \mathbf{x}$. If $\mathbf{x}$ is $\frac{7}{3}$-power-free, then so is $\mu^2(\mathbf{x})$. Hence, if u_1 has a forbidden power, then this power must be a prefix of u_1.

Now let $\beta < 7/3$ be the largest possible exponent of a prefix of $\mathbf{x}$ and q be the smallest period of a prefix of exponent β in $\mathbf{x}$. Write $\beta = p/q$. Then the word $\mu^2(\mathbf{x})$ has a prefix of exponent β and of period $4q$ by Lemma 4 (1), but no prefixes of a bigger exponent or of the same exponent and a smaller period by Lemma 5. Hence u_1 has no prefixes of exponent greater than $(4p+1)/(4q)$. Since p and q are integers, we obtain the required inequality:

$$\frac{p}{q} < \frac{7}{3} \Longrightarrow 3p < 7q \Longrightarrow 3p + \frac{3}{4} < 7q \Longrightarrow \frac{4p+1}{4q} < \frac{7}{3}.$$

$F_{12} = F_2$: The word $u_1 = 0\mu(00\mu(\mathbf{x}))$ contains a μ-image of $u_2 = 00\mu(\mathbf{x})$. Suppose that u_2 is $\frac{7}{3}$-power-free. Then it starts 0010011. Since the factor 001001 cannot occur in a μ-image, we note that

($\star$) the word $00\mu(\mathbf{x})$ has only two prefixes of exponent 2 (00 and 001001) and no prefixes of bigger exponents.

By Lemma 5, the word $\mu(u_2)$ has only two prefixes of exponent 2 ($\mu(00)$ and $\mu(001001)$) and no prefixes of bigger exponents. Thus, the word $u_1 = 0\mu(u_2)$ is obviously $\frac{7}{3}$-power-free.

$F_{23} = F_{13}$: We have $u_1 = 00\mu(1\mu(\mathbf{x})) = 0u_2$. Suppose the word u_2 is $\frac{7}{3}$-power-free; then it starts 010011. A forbidden power in u_1, if any, occurs at the beginning and hence contains 0010011. But 00100 does not occur later in this word, so no such forbidden power exists.

$F_{110} = F_3$: The word $u_1 = 0\mu(0\mu(\mu(\mathbf{x}))) = 001\mu^3(x)$ is a suffix of the μ^2-image of $u_2 = 1\mu(\mathbf{x})$. Hence, if u_2 is $\frac{7}{3}$-power-free, then by Lemma 4 (2) u_1 is $\frac{7}{3}$-power-free as well.

For the other direction, assume u_1 is $\frac{7}{3}$-power-free and then $\mu(\mathbf{x})$ is $\frac{7}{3}$-power-free. So, if u_2 contains some power yyy' with $|y'| \geq |y|/3$, then this power must be a prefix of u_2. Put $y = 1z$ and $y' = 1z'$. The word $\mu(\mathbf{x})$ starts $z1z1z'$. Hence $\mathbf{x}$ starts $\mu^{-1}(z1)\mu^{-1}(z1)$ by Lemma 3. So we conclude that $|z1| = |y|$ is an even number. Now let $|y| = q$ and $p = |yyy'|$ so that $p/q \geq 7/3$. Thus the word $1u_1 = \mu^2(u_2)$ starts with a (p/q)-power of period $4q$. Since u_1 is $\frac{7}{3}$-power-free, we have $(4p-1)/4q < 7/3$.

This gives us the inequalities $3p \geq 7q$ and $3p - 3/4 < 7q$. Since p and q are integers this means $3p = 7q$ and hence q is divisible by 3. On the other hand from above q is even. So q is divisible by 6. Now $|y'| = |y|/3$ so $|y'|$ is even. But then z' is odd, and begins at an even position in a μ-image, so the character following z' is fixed and must be the same character as in the corresponding position of z, say a. Thus $z1z1z'a$ is a $(7/3)$-power occurring in $\mu(\mathbf{x})$, a contradiction.

$F_{111} = F_{11}$: The word $u_1 = 0\mu(0\mu(0\mu(\mathbf{x}))) = 0010110\mu^3(\mathbf{x})$ contains a μ-image of $u_2 = 0\mu(0\mu(\mathbf{x})) = 001\mu^2(\mathbf{x})$. Suppose u_2 is $\frac{7}{3}$-power-free but to the contrary $u_1 = 0\mu(u_2)$ has a forbidden power. By Lemma 4 (2), this power must be a prefix of u_1. Note that this power can be extended to the left by 1 (not by 0, because a μ-image cannot contain 000). Hence the word $1u_1 = \mu^3(1\mathbf{x})$ starts with a forbidden power. This induces a forbidden power at the beginning of $1\mathbf{x}$; this power has a period q and some exponent $p/q \geq 7/3$. Then u_1 has a prefix of exponent $(8p-1)/8q \geq 7/3$. On the other hand the word $\mu(u_2)$ is $\frac{7}{3}$-power-free, whence $(8p-2)/q < 7/3$. So we get the system of inequalities $3p - 3/8 \geq 7q$, $3p - 3/4 < 7q$. This system has no integer solutions, a contradiction.

$F_{112} = F_2$: We have $u_1 = 0\mu(0\mu(00\mu(\mathbf{x}))) = 001\mu^2(00\mu(\mathbf{x})) = 001\mu^2(u_2)$. In view of ($\star$), one can easily check that if u_2 is $\frac{7}{3}$-power-free, then so is u_1.

$F_{113} = F_{13}$: We have $u_1 = 0\mu(0\mu(1\mu(\mathbf{x}))) = 0\mu(u_2)$. Suppose u_2 is $\frac{7}{3}$-power-free. Then $\mu^2(\mathbf{x})$ starts 01101001. Assume to the contrary that u_1 has a forbidden power. By Lemma 4 (2), this power must be a prefix of u_1. Again, this power can be extended to the left by 1, not by 0. Hence the word $1u_1 = \mu^2(11\mu(\mathbf{x}))$ starts with a forbidden power, thus inducing a forbidden power at the beginning of $u = 11\mu(\mathbf{x}) = 110110\cdots$. The word u has only two squares as prefixes (11 and 110110, cf. $(\star)$). Hence u has the prefix 11011010 and no forbidden factors except for the $(7/3)$-power prefix 1101101. Therefore, the word u_1 has no prefixes of exponent $\geq 7/3$.

$F_{131} = F_{31}$: We have $u_1 = 0\mu(1\mu(0\mu(\mathbf{x}))) = 0\mu(u_2)$. Suppose u_2 is $\frac{7}{3}$-power-free. Then the word $u = 11\mu(0\mu(\mathbf{x}))$ is $\frac{7}{3}$-power-free by the equality $F_{41} = F_{31}$, which is symmetric to $F_{23} = F_{13}$ proved above. But u_1 is a suffix of $\mu(u)$, whence the result.

$F_{204} = F_4$: We have $u_1 = 00\mu(\mu(11\mu(\mathbf{x}))) = 00\mu^2(u_2)$. Suppose u_2 is $\frac{7}{3}$-power-free. Using the observation symmetric to $(\star)$, we check by inspection that u_1 contains no forbidden power.

$F_{1300} = F_{130}$: Neither one of the words $u_1 = 0\mu(1\mu(\mu(\mu(\mathbf{x})))) = 010\mu^4(\mathbf{x})$, $u_2 = 0\mu(1\mu(\mu(\mathbf{x}))) = 010\mu^3(\mathbf{x})$ contains an image of the other. The proofs for both directions are essentially the same, so we give only one of them. Let u_1 be $\frac{7}{3}$-power-free; then the words $\mu^4(\mathbf{x})$, $\mathbf{x}$, and $\mu^3(\mathbf{x})$ are $\frac{7}{3}$-power-free as well, and x starts 0. A simple inspection of short prefixes of u_2 shows that if this word is not $\frac{7}{3}$-power-free, then some β-power with $\beta \geq 2$ is a prefix of $\mu^3(\mathbf{x})$. By Lemma 5, the word $\mathbf{x}$ also starts with a β-power. The argument below will be repeated, with small variations, for several identities.

$(*)$ Consider a prefix yyy' of $\mathbf{x}$ which is the longest prefix of $\mathbf{x}$ with period $|y|$. Then $|y'| < |y|/3$. By Lemma 5, the longest prefix of the word $\mu^3(\mathbf{x})$ having period $8|y|$ is $\mu^3(yyy')$. If some word $z\mu^3(yyy')$ also has period $8|y|$, then z should be a suffix of a μ^3-image of some word. Since the word 010 is not such a suffix, then $10\mu^3(yyy')$ is the longest possible $(8|y|)$-periodic word contained in u_2. Let us estimate its exponent. Since $|z|$ and $|y'|$ are integers, we have

$$8|y'| < 8|y|/3 \Longrightarrow 24|y'| < 8|y| \Longrightarrow 24|y'| + 6 < 8|y| \Longrightarrow 8|y'| + 2 < 8|y|/3,$$

whence $\exp(10\mu^3(yyy')) < 7/3$. Since we have chosen an arbitrary prefix yyy' of $\mathbf{x}$, we conclude that the word u_2 is $\frac{7}{3}$-power-free.

$F_{1301} = F_1$: The word $u_1 = 0\mu(1\mu(\mu(0\mu(\mathbf{x})))) = 010\mu^3(0\mu(\mathbf{x}))$ contains a μ^3-image of $u_2 = 0\mu(\mathbf{x})$. Suppose u_2 is $\frac{7}{3}$-power-free. It suffices to check that the prefix 010 of u_1 does not complete any prefix of $\mu^3(u_2)$ to a forbidden power. For short prefixes, this can be checked directly, while long prefixes that can be completed in this way should have exponents ≥ 2. By Lemma 5, a prefix of period p and exponent $\beta \geq 2$ of the word $\mu^3(u_2)$ corresponds to the prefix of the word u_2 having the exponent β and the period $p/8$. So, we repeat the argument $(*)$ replacing $\mathbf{x}$ with u_2 to obtain that u_1 is $\frac{7}{3}$-power-free.

$F_{1302} = F_2$: We have $u_1 = 0\mu(1\mu(\mu(00\mu(\mathbf{x})))) = 010\mu^3(00\mu(\mathbf{x})) = 010\mu^3(u_2)$. Suppose u_2 is $\frac{7}{3}$-power-free. By $(\star)$ and Lemma 5, among the prefixes of $\mu^3(u_2)$ there are only two squares, $\mu^3(00)$ and $\mu^3(001001)$, and no words of bigger exponent. By direct inspection, u_1 is $(7/3)$-free.

$F_{2030} = F_{310}$: Neither one of the words $u_1 = 00\mu(\mu(1\mu(\mu(\mathbf{x})))) = 001001\mu^4(\mathbf{x})$ and

$$u_2 = 1\mu(0\mu(\mu(\mathbf{x}))) = 101\mu^3(\mathbf{x})$$

contains an image of the other. If the word u_1 is assumed to be $\frac{7}{3}$-power-free, then the proof repeats the proof of the identity $F_{1300} = F_{130}$, up to renaming all 0's to 1's and vice versa. Let u_2 be $\frac{7}{3}$-power-free. The words $\mu^4(\mathbf{x})$ and $\mathbf{x}$ are also $\frac{7}{3}$-power-free, and $\mathbf{x}$ begins with 1, assuring that there are no short forbidden powers in the beginning of u_1. Concerning long forbidden powers, we consider, similar to $(*)$, a prefix yyy' of $\mathbf{x}$ which is the longest prefix of $\mathbf{x}$ with period $|y|$. The longest possible $(16|y|)$-periodic word contained in u_1 is $01001\mu^4(yyy')$, because 001001 is not a suffix of a μ^4-image. As in $(*)$, we obtain $16|y'|+5 < 16|y|/3$, implying $\exp(01001\mu^4(yyy')) < 7/3$. Hence the word u_1 is $\frac{7}{3}$-power-free.

$F_{2033} = F_{33}$: The word $u_1 = 00\mu(\mu(1\mu(1\mu(x)))) = 00\mu^2(110\mu^2(x))$ contains a μ^2-image of $u_2 = 1\mu(1\mu(x)) = 110\mu^2(x)$. Again, if the word u_2 is $\frac{7}{3}$-power-free, then so is $\mu^2(u_2)$, and it suffices to check that the prefix 00 of u_1 does not complete any prefix of $\mu^2(u_2)$ to a forbidden power. Similar to $(*)$, consider a prefix yyy' of u_2 which is the longest prefix of u_2 with period $|y|$. The longest possible $(4|y|)$-periodic word contained in u_1 is $0\mu^2(yyy')$, because 00 is not a suffix of a μ^2-image. As in $(*)$, we see that $4|y'|+1 < 4|y|/3$, implying $\exp(0\mu^2(yyy')) < 7/3$, and conclude that the word u_1 is $\frac{7}{3}$-power-free.

$F_{2034} = F_4$: The word $u_1 = 00\mu(\mu(1\mu(11\mu(x)))) = 001001\mu^3(11\mu(x))$ contains a μ^3-image of $u_2 = 11\mu(x)$. Suppose u_2 is $\frac{7}{3}$-power-free. Using $(\star)$ and Lemma 5, we conclude that among the prefixes of $\mu^3(u_2)$ there are only two squares, $\mu^3(11)$ and $\mu^3(110110)$, and no words of bigger exponent. By direct inspection, u_1 is $(7/3)$-free. $\square$

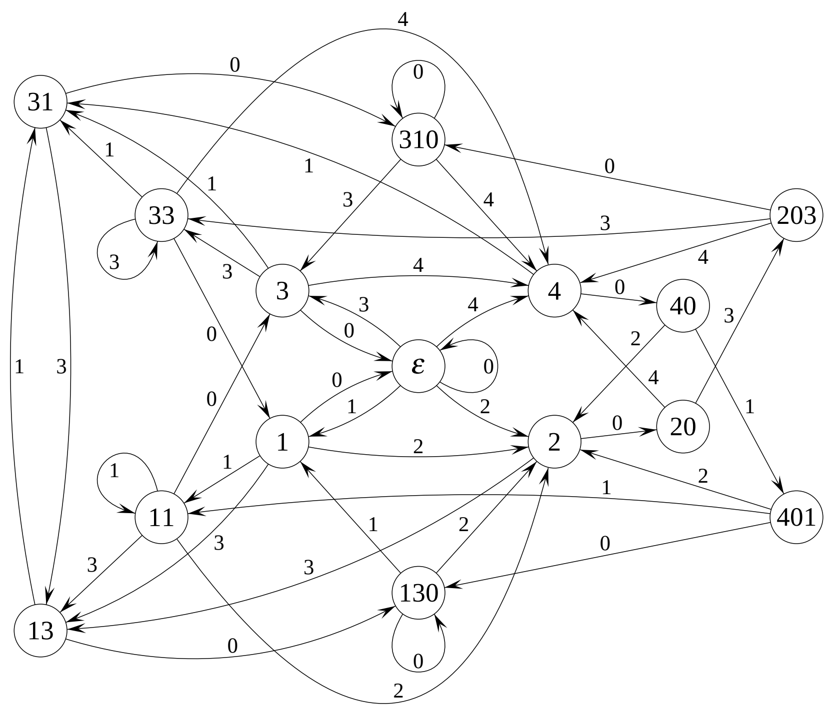

Figure 2: Automaton coding infinite binary $\frac{7}{3}$-power-free words

From Lemma 6 and the results above, we get

Theorem 7. *Every infinite binary $\frac{7}{3}$-power-free word* $\mathbf{x}$ *is encoded by an infinite path, starting in F_ε, through the automaton in Figure 2.*

Every infinite path through the automaton not ending in 0^ω codes a unique infinite binary $\frac{7}{3}$-power-free word $\mathbf{x}$*. If a path* $\mathbf{i}$ *ends in 0^ω and this suffix corresponds to a cycle on state F_ε then* $\mathbf{x}$ *is coded by either* $\mathbf{i};1$ *or* $\mathbf{i};3$*. If a path* $\mathbf{i}$ *ends in 0^ω and this suffix corresponds to a cycle on F_{310}, then* $\mathbf{x}$ *is coded by* $\mathbf{i};3$*. If a path* $\mathbf{i}$ *ends in 0^ω and this suffix corresponds to a cycle on F_{130}, then* $\mathbf{x}$ *is coded by* $\mathbf{i};1$.

Remark 8. Blondel, Cassaigne, and Jungers [5] obtained a similar result, and even more general ones, for finite words. The main advantage to our construction is its simplicity.

Corollary 9. *Each of the 15 sets F_ε, F_1, F_2, F_3, F_4, F_{11}, F_{33}, F_{13}, F_{31}, F_{20}, F_{40}, F_{130}, F_{310}, F_{203}, F_{401} is uncountable.*

Proof. It suffices to provide uncountably many distinct paths from each state to itself. By symmetry, it suffices to prove this for all the states labeled ε or below in Figure 2. These are as follows:

- ε: $(0+10)^\omega$
- 1: $(01+001)^\omega$
- 2: $(0402+030402)^\omega$
- 11: $(0011+00011)^\omega$
- 13: $(013+0013)^\omega$
- 20: $(4020+34020)^\omega$
- 401: $(10401+203401)^\omega$
- 130: $(0+104010)^\omega$.

□

Corollary 10. *For all words $w \in \{0,1,2,3,4\}^*$, either F_w is empty or uncountable.*

4 The lexicographically least $\frac{7}{3}$-power-free word

Theorem 11. *The lexicographically least infinite binary $\frac{7}{3}$-power-free word is* $001001\bar{\mathbf{t}}$.

Proof. By tracing through the possible paths through the automaton we easily find that $2030^\omega;1$ is the code for the lexicographically least sequence. □

Remark 12. This result does not seem to follow directly from [2] as one referee suggested.

5 Automatic infinite binary $\frac{7}{3}$-power-free words

As a consequence of Theorem 7, we can give a complete description of the infinite binary $\frac{7}{3}$-power-free words that are 2-automatic [3]. Recall that an infinite word $(a_n)_{n\geq 0}$ is k-automatic if there exists a deterministic finite automaton with output that, on input n expressed in base k, produces an output associated with the state last visited that is equal to a_n. Alternatively, $(a_n)_{n\geq 0}$ is k-automatic if its k-kernel

$$\{(a_{k^i n+j})_{n\geq 0} \ : \ i \geq 0 \text{ and } 0 \leq j < k^i\}$$

consists of finitely many distinct sequences.

Theorem 13. *An infinite binary $\frac{7}{3}$-power-free word is* 2*-automatic if and only if its code is both specified by the DFA given above in Figure 2, and is ultimately periodic.*

First, we need a lemma:

Lemma 14. *An infinite binary word $\mathbf{x} = a_0a_1a_2\cdots$ is* 2*-automatic if and only if $\mu(\mathbf{x})$ is* 2*-automatic.*

Proof. Proved in [13]. □

Now we can prove Theorem 13.

Proof. Suppose the code of $\mathbf{x}$ is ultimately periodic. Then we can write its code as yz^{ω} for some finite words y and z. Since the class of 2-automatic sequences is closed under appending a finite prefix [3, Corollary 6.8.5], by Lemma 14, it suffices to show that the word coded by z^{ω} is 2-automatic.

The word z^{ω} codes a $\frac{7}{3}$-power-free word $\mathbf{w}$ satisfying $\mathbf{w} = t\varphi(\mathbf{w})$, where t is a finite word and $\varphi = \mu^k$. Hence, by iteration, we get that $\mathbf{w} = t\varphi(t)\varphi^2(t)\cdots$. It is now easy to see that the 2-kernel of $\mathbf{w}$ is contained in

$$S := \{u\mu^i(v)\mu^{i+k}(v)\mu^{i+2k}(v)\cdots \ : \ |u| \le |t| \text{ and } v \in \{t,\bar{t}\} \text{ and } 1 \le i \le k\},$$

which is a finite set.

On the other hand, suppose the code for $\mathbf{x}$ is not ultimately periodic. Then we show that the 2-kernel is infinite. Now it is easy to see that if the code for $\mathbf{x}$ is $a\mathbf{y}$ for some letter $a \in \{0,1,3\}$ then one of the sequences in the 2-kernel (obtained by taking either the odd- or even-indexed terms) is either coded by $\mathbf{y}$ or its complement is coded by $\mathbf{y}$. On the other hand, if the code for $\mathbf{x}$ is $a\mathbf{y}$ with $a \in \{2,4\}$, then $\mathbf{y}$ begins with $0,1$, or 3, say $\mathbf{y} = b\mathbf{z}$. It follows that the subsequences obtained by taking the terms congruent to $0,1,2,$ or 3 (mod 4) is coded by $\mathbf{z}$, or its complement is coded by $\mathbf{z}$. Since the code for $\mathbf{x}$ is not ultimately periodic, there are infinitely many distinct sequences in the orbit of the code for $\mathbf{x}$, under the shift. By the infinite pigeonhole principle, infinitely many correspond to a sequence in the 2-kernel, or its complement. Hence $\mathbf{x}$ is not 2-automatic. ☐

6 Acknowledgments

We are grateful to the referees for their helpful suggestions.

References

[1] A. Aberkane & J. D. Currie (2005): *Attainable lengths for circular binary words avoiding k powers*. *Bull. Belg. Math. Soc.* 12, pp. 525–534.

[2] J.-P. Allouche, J. Currie & J. Shallit (1998): *Extremal infinite overlap-free binary words*. *Electron. J. Combinatorics* 5(1). Available at `http://www.combinatorics.org/Volume_5/Abstracts/v5i1r27.html`.

[3] J.-P. Allouche & J. Shallit (2003): *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press.

[4] J. Berstel (1994): *A rewriting of Fife's theorem about overlap-free words*. In J. Karhumäki, H. Maurer & G. Rozenberg, editors: *Results and Trends in Theoretical Computer Science*, *Lecture Notes in Computer Science* 812, Springer-Verlag, pp. 19–29, doi:`10.1007/3-540-58131-6_34`.

[5] V. D. Blondel, J. Cassaigne & R. M. Jungers (2009): *On the number of α-power-free binary words for $2 < \alpha \le 7/3$*. *Theoret. Comput. Sci.* 410, pp. 2823–2833, doi:`10.1016/j.tcs.2009.01.031`.

[6] A. Carpi (1993): *Overlap-free words and finite automata*. *Theoret. Comput. Sci.* 115, pp. 243–260, doi:`10.1016/0304-3975(93)90118-D`.

[7] J. Cassaigne (1993): *Counting overlap-free binary words*. In P. Enjalbert, A. Finkel & K. W. Wagner, editors: *STACS 93, Proc. 10th Symp. Theoretical Aspects of Comp. Sci.*, *Lecture Notes in Computer Science* 665, Springer-Verlag, pp. 216–225, doi:`10.1007/3-540-56503-5_24`.

[8] E. D. Fife (1980): *Binary sequences which contain no BBb*. *Trans. Amer. Math. Soc.* 261, pp. 115–136, doi:`10.1090/S0002-9947-1980-0576867-5`.

[9] J. Karhumäki & J. Shallit (2004): *Polynomial versus exponential growth in repetition-free binary words*. *J. Combin. Theory. Ser. A* 105, pp. 335–347, doi:10.1016/j.jcta.2003.12.004.

[10] R. Kolpakov & G. Kucherov (1997): *Minimal letter frequency in n-th power-free binary words*. In I. Privara & P. Ružička, editors: *Proc. of the 22nd Symposium, Math. Found. Comput. Sci. (MFCS) 1997*, *Lecture Notes in Computer Science* 1295, Springer-Verlag, pp. 347–357, doi:10.1007/BFb0029978.

[11] N. Rampersad (2005): *Words avoiding* 7/3*-powers and the Thue-Morse morphism*. *Int. J. Found. Comput. Sci.* 16, pp. 755–766, doi:10.1142/S0129054105003273.

[12] A. Restivo & S. Salemi (1985): *Overlap free words on two symbols*. In M. Nivat & D. Perrin, editors: *Automata on Infinite Words*, *Lecture Notes in Computer Science* 192, Springer-Verlag, pp. 198–206, doi:10.1007/3-540-15641-0_35.

[13] J. Shallit: *Fife's theorem revisited*. Available at http://arxiv.org/abs/1102.3932. To appear, DLT 2011.

[14] A. M. Shur (2000): *The structure of the set of cube-free Z-words in a two-letter alphabet*. *Izv. Ross. Akad. Nauk Ser. Mat.* 64(4), pp. 201–224. In Russian. English translation in *Izv. Math.* 64 (2000), pp. 847–871.