AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **35** (2006), Pages 41–49

On the powers in the Thue-Morse word

A. Aberkane V. Linek S. J. Mor*

Department of Mathematics and Statistics
University of Winnipeg
Winnipeg
Manitoba R3B 2E9
CANADA

Abstract

A word of the form $XX\cdots XY = X^nY$ where X is not empty and Y is a prefix of X is called a $(\frac{n|X|+|Y|}{|X|})$-power, where $|X|$ is the length of the word X. Here we give an explicit description of the powers that occur in the Thue-Morse word.

1 Introduction

If A is a set of symbols then a *word over* A is any finite or infinite string of symbols of A. The set of all finite words over A is denoted A^*, and A^ω is the set of all infinite words over A. Under the operation of concatenation of words A^* is a monoid. If the word w is the concatenation of the words x, u and y, i.e., $w = xuy$, then u is a *subword* of w and we write $u \lhd w$ or $w \rhd u$.

A *morphism* from A_1^* to A_2^* is a mapping $\varphi\colon A_1^* \to A_2^*$ such that $\varphi(wv) = \varphi(w)\varphi(v)$ for all $w, v \in A_1^*$, where juxtaposition of words denotes concatenation. Let $\varphi\colon \{0,1\}^* \to \{0,1\}^*$ be the morphism that is uniquely determined by stipulating that $\varphi(0) = 01$ and $\varphi(1) = 10$. The Thue-Morse word is then defined as $\mathbf{t} = \lim_{n\to\infty} \varphi^n(0) = 0110100110010110\ldots$, that is, $\mathbf{t}$ is the infinite binary word with successively longer prefixes $\varphi(0) = 01$, $\varphi^2(0) = 0110$, $\varphi^3(0) = 01101001$, and so on. Allouche and Shallit [1] give more information on the Thue-Morse word.

A word of the form $XX\cdots XY = X^nY$ where X is not empty and Y is a prefix of X is called a $(\frac{n|X|+|Y|}{|X|})$*-power*, where $|X|$ is the length of the word X. For example, *abbab* is a $(\frac{5}{3})$-power, and *abbcabbcabb* is an $(\frac{11}{4})$-power. A 2-power is usually called a square, a 3-power is a cube and so on. No r-power with $r > 2$ occurs as a subword of $\mathbf{t}$, and this well known fact is usually expressed by saying that $\mathbf{t}$ contains no *overlap*, that is, $\mathbf{t}$ contains no subword of the form $aXaXa$, where a is a single symbol and X is a (possibly empty) word. However $\mathbf{t}$ does contain squares. In fact if XX is a

* Research of all authors supported by NSERC.

subword of $\mathbf{t}$ then $|X| = 2^n$ or $3{\cdot}2^n$ for some $n \geqslant 0$, and these lengths are all achieved (this follows easily from Lemma 3, parts (Y), (N2) and (E1)).

Now if $w = X^nY$ is a $(\frac{p}{q})$-power then we shall say that w is a $(\frac{p}{q})$-power *with multiplier* $\frac{n|X|+|Y|}{p}$. For example, aa is a $(\frac{2}{1})$-power with multiplier 1, but $abcabc$ is a $(\frac{2}{1})$-power with multiplier 3. We have just noted that $\mathbf{t}$ has $(\frac{2}{1})$-powers as subwords but only with the multipliers 2^n or $3\cdot 2^n$ for some $n \geqslant 0$. In particular $\mathbf{t}$ has no $(\frac{14}{7})$-power occurring with a positive integer multiplier, but it does have a $(\frac{14}{7})$-power occurring with multiplier $\frac{1}{7}$. Clearly a fraction $\frac{p}{q}$ with $\gcd(p, q) = 1$ can only occur with a positive integer multiplier. Given a word $\mathbf{w}$ we define $\mathrm{Mult}_{\mathbf{w}}(\frac{p}{q})$ to be the set of all positive multipliers of all $(\frac{p}{q})$-powers that occur in the word $\mathbf{w}$. Note that two different fractions may represent the same rational number and yet have different sets of multipliers for a given word $\mathbf{w}$, e.g., $\mathrm{Mult}_{\mathbf{t}}(\frac{2}{1}) = \{2^n, 3{\cdot}2^n : n \geqslant 0\}$ whereas $\mathrm{Mult}_{\mathbf{t}}(\frac{14}{7}) = \{\frac{2^n}{7}, \frac{3\cdot 2^n}{7} : n \geqslant 0\}$. Clearly if $r \in \mathrm{Mult}_{\mathbf{t}}(\frac{p}{q})$ then (apply the morphism) $2r \in \mathrm{Mult}_{\mathbf{t}}(\frac{p}{q})$, so if $\mathrm{Mult}_{\mathbf{t}}(\frac{p}{q})$ is nonempty then it is infinite, and if $\mathrm{Mult}_{\mathbf{t}}(\frac{p}{q})$ contains a positive integer then it contains infinitely many integers. We are primarily interested in fractions that occur with integer multipliers, so we define the *exponent set* of $\mathbf{w}$, denoted $E(\mathbf{w})$, as

$$E(\mathbf{w}) = \{\frac{p}{q} : \mathrm{Mult}_{\mathbf{w}}(\frac{p}{q}) \cap \mathbb{N} \neq \emptyset\}.$$

Considerable attention has been focused on the critical exponent $e(\mathbf{w}) = \sup E(\mathbf{w})$. For instance Mignosi and Pirillo [8] determine $e(\mathbf{f})$ for the Fibonacci word, $\mathbf{f}$, and Dejean [5] conjectures the value of $\inf e(\mathbf{w})$ over all words $\mathbf{w}$ on an n-letter alphabet. Carpi [4] determines $\mathrm{Mult}_{\mathbf{k}}(\frac{2}{1})$ for the Kolakoski word, $\mathbf{k}$, and as a consequence determines $e(\mathbf{k})$. However, to the best of our knowledge, no one has yet given a complete determination of $\mathbf{w}$ for some (interesting) word $\mathbf{w}$. Our main result is the explicit determination of the powers in $E(\mathbf{t})$.

Theorem 1 *Let $1 < \frac{m}{n} < 2$ be rational. Let $m = n + k$ and $n = 2^{\nu(n)}n_1$ with n_1 odd. A $(\frac{m}{n})$-power occurs in the Thue-Morse word with a positive integer multiplier if and only if $k < 3{\cdot}2^{\nu(n)} + 1$ and*

$$\frac{m}{n} \notin \{\frac{13{\cdot}2^j + a}{11{\cdot}2^j} : 1 \leqslant a \leqslant 2^j, j \geqslant 0\} \cup \{\frac{15{\cdot}2^j + a}{13{\cdot}2^j} : 1 \leqslant a \leqslant 2^j, j \geqslant 0\}$$

Moreover, if a $(\frac{m}{n})$-power occurs in the Thue-Morse word with a positive integer multiplier, then it occurs in the Thue-Morse word with multiplier 1.

2 Basic structural features

We review some basic structural features of the Thue-Morse word.

Let $\varphi(0) = 01$, $\varphi(1) = 10$. Then $\varphi^{2n}(0)$ is a palindrome for $n \geqslant 1$, hence if s is a finite word in $\mathbf{t}$ then the reverse of s is also a word in $\mathbf{t} = \lim_{n\to\infty} \varphi^n(0)$. Alternatively let $\overline{s}$ be the complement of the binary word s (e.g., $\overline{011010} = 100101$), if $s_1 = 0$ and $s_n = s_{n-1}\overline{s_{n-1}}$ for $n > 1$ then we also have $\mathbf{t} = \lim_{n\to\infty} s_n$. Since $\overline{s_n} \lhd \mathbf{t}$ for each n it follows that if $s \lhd \mathbf{t}$ then $\overline{s} \lhd \mathbf{t}$.

For a finite binary word s in t the *orbit* of s, $O(s)$, consists of all other words that can be obtained from s by complementing and reversing, for example $O(10010) = \{10010, 01001, 01101, 10110\}$. For a set, S, of binary words we let $O(S) = \cup_{s\in S} O(s)$. The orbit of the empty word has one element, all other orbits have two or four elements. Then all of the words in an orbit occur in $\mathbf{t}$, or they all do not occur in $\mathbf{t}$.

Note that $\varphi(\mathbf{t}) = \mathbf{t}$. Given $s \lhd \mathbf{t} = t_0t_1t_2t_3\ldots = \varphi(t_0)\varphi(t_1)\varphi(t_2)\ldots$ we say s is *in phase* if $|s|$ is even and s consists of images of φ. More precisely, s is in phase if $s = t_it_{i+1}\ldots t_{i+2n-1}$ and the index i is even, so that $s = \varphi(t_j\ldots t_{j+n-1})$ where $j = i/2$. The important thing about subwords in $\mathbf{t}$ that are in phase is that they are images of shorter subwords of $\mathbf{t}$.

Any subword in $\mathbf{t}$ of odd length is always out of phase, and subwords 00, 11 are never in phase in $\mathbf{t}$. We say $s = t_i\ldots t_k \lhd \mathbf{t}$ *begins in phase* if i is even, and *ends in phase* if k is odd. Following Lothaire [7, p. 114] we call a binary word s *synchronizing* if $s \lhd \mathbf{t}$ and if all of its occurrences in $\mathbf{t}$ begin in phase, or if all of its occurrences in $\mathbf{t}$ begin out of phase. Clearly 00 and 11 are synchronizing words, hence any subword of $\mathbf{t}$ that has 00 or 11 as a subword is also synchronizing. The following lemma is now easy to see.

Lemma 2 *If $s \in \{0, 1, 01, 10, 010, 101\}$ then s may begin in or out of phase in $\mathbf{t}$. Any other binary word s either is not a subword of $\mathbf{t}$ or is a synchronizing word.*

We now proceed to tabulate some common structures of $\mathbf{t}$. The notation $Z = X(n)Y$ indicates that Z has prefix X and suffix Y, and that $|Z| = |X| + n + |Y|$. For example $1001011 = 10(3)11$. In Table 1 take $n \geqslant 0$ in addition to any other restrictions on n listed there. Letters x, y, z stand for single symbols, and X stands for a word. We note that Fitzpatrick [6] also gives some of the entries (f)-(m).

Entries (a)-(e) are checked by observing that any subword of $\mathbf{t}$ of length 5 or less appears in one of $\varphi^2(00)$, $\varphi^2(10)$, $\varphi^2(01)$ or $\varphi^2(11)$.

If for some $\Delta > 0$ we have $t_i = t_{i+\Delta}$ for all i then $\mathbf{t}$ would be periodic, a contradiction. It follows from this and the symmetries of $\mathbf{t}$ that $\mathbf{t}$ contains both $0(n)1$ and $1(n)0$ for each $n \geqslant 0$. Similarly if $t_i \neq t_{i+\Delta}$ for all i then 2Δ would be a period of $\mathbf{t}$, a contradiction, hence $\mathbf{t}$ has subwords of the form $0(n)0$ and $1(n)1$ for each $n \geqslant 0$. Thus $\mathbf{t}$ has a subword of the form $x(n)y$ for any choice of x, y and $n \geqslant 0$ and entry (f) is established.

Entry (g) need only be established for $x = 0$ due to the symmetries of $\mathbf{t}$. For brevity let $X \overset{\varphi}{\to} Y$ mean $\varphi(X) = Y$. From (f) and the recursions $0(k)0 \overset{\varphi}{\to} 01(2k)01 = 0(2k+1)01$ and $1(k)0 \overset{\varphi}{\to} 10(2k)01 \rhd 0(2k)01$ entry (g) holds for $yz = 01$. If $yz = 10$ then (g) follows similarly by applying φ to $1(k)1$ and $1(k)0$. If $yz = 00$ then we use what we have proved so far together with the recursions $0(n)10 \overset{\varphi}{\to} 01(2n)1001 \rhd 0(2n+2)00$, $1(n)10 \overset{\varphi}{\to} 10(2n)1001 \rhd 0(2n+1)00$. For the last case where $yz = 11$ let $N = \{n | 0(n)11 \lhd \mathbf{t}\}$. Then the recursions $0(n)01 \overset{\varphi}{\to}$

	Structure in $\mathbf{t}$	Restrictions
(a)	x	$\in O(\{0\})$
(b)	xy	$\in O(\{00,01\})$
(c)	xyz	$\in O(\{001,010\})$
(d)	$xyzw$	$\in O(\{0010,0011,0101,0110\})$
(e)	$xyzwt$	$\in O(\{00101,00110,01001\})$
(f)	$x(n)y$	–
(g)	$x(n)yz$	$n \neq 0$ if $x = y = z$
(h)	$01(n)01$	–
(i)	$01(n)10$	$n \neq 1$
(j)	$00(n)01$	$n \neq 2\cdot(4^k - 1),\ k \geqslant 0$
(k)	$00(n)10$	$n \neq 4^k - 2,\ k > 0$
(l)	$00(n)11$	$n = 2k,\ k \neq 2$
(m)	$00(n)00$	$n = 2k,\ k \neq 0$
(n)	X^{2+}	$\nexists$
(o)	X^2	$\lvert X\rvert = 2^n, 3\cdot 2^n$
(p)	$x(n-1)x$	$n \geqslant 1$
(q)	$xy(n-2)xy$	$n \geqslant 2$
(r)	$xyz(n-3)xyz$	$n \geqslant 3,\ n \neq 11, 13$

Table 1: Some common structures in $\mathbf{t}$

$01(2n)0110 \rhd 0(2n+2)11$, $0(n)01 \overset{\varphi}{\to} 0110(4n)01101001 \rhd 0(4n+1)11$ and $0(n)11 \overset{\varphi}{\to} 0110(4n)10011001 \rhd 0(4n+3)11$ show that if $n \in N$ then $2n+2, 4n+1, 4n+3 \in N$, so applying induction completes this last case of (g).

Entry (h) follows from entry (g) and the recursions $0(k)01 \overset{\varphi}{\to} 01(2k)0110 \rhd 01(2k)01$, $0(k)11 \overset{\varphi}{\to} 01(2k)1010 \rhd 01(2k+1)01$.

Entry (i) follows from (f) and the recursions $0(k)1 \overset{\varphi}{\to} 01(2k)10$, $0(k)00 \overset{\varphi}{\to} 01(2k)0101 \rhd 01(2k+1)10$, and we note that by (e), $01(1)10 \ntriangleleft \mathbf{t}$.

For (j), (k) put $N_0 = \{n \mid 00(n)01 \lhd \mathbf{t}\}$ and $N_1 = \{n \mid 00(n)10 \lhd \mathbf{t}\}$. Let $X \overset{\varphi}{\leftarrow} Y$ mean that X is of even length and is in phase, and that $\varphi(Y) = X$. In this case Y is a subword of $\mathbf{t}$. If $00(2k)10 \lhd \mathbf{t}$ and $k > 0$ then we have $10, 01, (2k-2), 01, 01 \overset{\varphi}{\leftarrow} 10(k-1)00$, so $0 \neq 2k \in N_1$ implies $k-1 \in N_0$. If $00(2k)01 \lhd \mathbf{t}$ and $k > 0$ then similarly $10, 01, (2k-2), 10, 10 \overset{\varphi}{\leftarrow} 10(k-1)11$, hence $00(k-1)10 \lhd \mathbf{t}$, so $2k \in N_0$ implies $k-1 \in N_1$. Applying φ to $00(k-1)xy$ gives the converse(s), so $2k \in N_i$ if and only if $k-1 \in N_{1-i}$. Now N_0, N_1 both contain all odd positive integers since we may reverse the subwords in (g) and for any $k \geqslant 0$ we have $10(k)0 \overset{\varphi}{\to} 1001(2k)01 \rhd 00(2k+1)01$ and $10(k)1 \overset{\varphi}{\to} 1001(2k)10 \rhd 00(2k+1)10$. Finally $0 \notin N_0$ and $2 \notin N_1$, so (j), (k) follow by induction.

For (l) note that $00(n)11 \lhd \mathbf{t}$ implies that the $0(n)1$ portion is in phase, hence $n = 2k$ is necessary. However if $00(4)11 \lhd \mathbf{t}$ then $1001(2)0110 \overset{\varphi}{\leftarrow} 10(1)01 \lhd \mathbf{t}$, contrary to (i) and the symmetries of $\mathbf{t}$. Finally using (g) we get $10(k)01 \overset{\varphi}{\to} 1001(2k)0110 \rhd$

$00(2k+2)11$, $k \neq 1$.

For (m) note that Lemma 2 implies $00(2k+1)00 \not\triangleleft \mathbf{t}$. Now (m) follows from the complement of (h) and the recursion $10(k)10 \xrightarrow{\varphi} 1001(2k)1001 \triangleright 00(2k+2)00$, $k \geqslant 0$.

Entry (n) says that $\mathbf{t}$ contains no power $\frac{p}{q} > 2$. Both (n) and (o) were known to Thue [11]. In entries (p) and (q) we mean that the structure exists for *some* choice of symbols x, y, so these entries follow immediately from (f) and (g) respectively.

For (r) let $n \in N$ if and only if $xyz(n-3)xyz \triangleleft \mathbf{t}$ for some xyz. By (h) and $01(k)01 \xrightarrow{\varphi} 0110(2k)0110 \triangleright 011(2k+1)011$ we have $2k+4 \in N$ for $k \geqslant 0$. Also $00(2k)11 \xrightarrow{\varphi} 0101(4k)1010$, the latter has $101(4k)101$ and $010(4k+2)010$, so (l) implies $4k+3, 4k+5 \in N$ for $k \geqslant 0$, $k \neq 2$.

Now if $w = xyz(10)xyz \triangleleft \mathbf{t}$ then xyz is not a synchronizing subword, so by Lemma 2 we may assume w.l.o.g. that $xyz = 010$. Should w begin in phase then $01, 01(8)10, 10 \xleftarrow{\varphi} 00(4)11 \triangleleft \mathbf{t}$ contrary to (l). Should w begin out of phase then $10, 10(10)01, 01 \xleftarrow{\varphi} 11(5)00 \triangleleft \mathbf{t}$, again contrary to (l). Therefore $n \neq 13$.

Similarly if $w = xyz(8)xyz \triangleleft \mathbf{t}$ then xyz is not a synchronizing subword, so again we may assume w.l.o.g. that $xyz = 010$. Should w begin in phase then $01, 01(6)10, 10 \xleftarrow{\varphi} 00(3)11 \triangleleft \mathbf{t}$ contrary to (l). Should w begin out of phase then $10, 10(8)01, 01 \xleftarrow{\varphi} 11(4)00 \triangleleft \mathbf{t}$, again contrary to (l). Therefore $n \neq 11$, and the proof of the last entry is complete.

3 Finding the powers

It is our goal to find a description of the exponent set $\{\frac{p}{q} : \mathrm{Mult}_{\mathbf{t}}(\frac{p}{q}) \cap \mathbb{N} \neq \emptyset\}$. Closely related to this set is

$$L(\mathbf{t}) = \{(n,k) | n \geqslant k \geqslant 0 \text{ and } XYX \triangleleft \mathbf{t},\ |X| = k, |XY| = n\}.$$

It is useful to call a subword XYX of $\mathbf{t}$ with $|X| = k$ and $|XY| = n$ an (n,k)-word. Given a pair (n,k) we ask, "is $(n,k) \in L(\mathbf{t})$?" In Lemma 3 below part (Y) gives some pairs for which the answer is "yes," parts (N1),(N2) give some pairs for which the answer is "no," and for the other pairs parts (E1), (E2) show how to replace (n,k) by an "equivalent" pair for which the answer to the question is the same as for (n,k).

Lemma 3 *Let $n \geqslant k \geqslant 0$, then*
(Y) $(n,k) \in L(\mathbf{t})$ if $k \leqslant 3$ and $(n,k) \neq (11,3), (13,3)$.
(N1) $(n,k) \notin L(\mathbf{t})$ if $(n,k) = (11,3), (13,3)$.
(N2) $(n,k) \notin L(\mathbf{t})$ if n is odd and $k \geqslant 4$.
(E1) If $n \equiv k \equiv 0 \pmod 2$ then $(n,k) \in L(\mathbf{t})$ if and only if $(n/2, k/2) \in L(\mathbf{t})$.
(E2) If $n \equiv 0 \pmod 2$ and $k \equiv 1 \pmod 2$ then $(n,k) \in L(\mathbf{t})$ if and only if $(n, k+1) \in L(\mathbf{t})$.

Proof: (Y), (N1): These follow from the entries (p)-(r) of Table 1.
(N2): By Lemma 2 subwords of $\mathbf{t}$ of length 4 or more are synchronizing so n must

be even for such k.
For *(E1)* both n, k are even, so we can bring into phase any out of phase (n,k)-word in $\mathbf{t}$ by a left shift. Applying φ^{-1} then gives a $(n/2, k/2)$-word in $\mathbf{t}$. Clearly the construction is reversible.
For *(E2)* let $X_1 Y X_2 \lhd \mathbf{t}$ be an (n,k)-word, $|X_i| = k$. W.l.o.g. take $X_1 Y X_2$ to be in phase (perform a left shift if necessary). Since n is even both X_i end out of phase, hence at their ends they split the same 2-block, (image under φ of a single symbol) so the X_i are followed by the same letter in $\mathbf{t}$. Therefore $X_1 a Y' X_2 a \lhd \mathbf{t}$ where $Y = aY'$, so $(n, k+1) \in L(\mathbf{t})$. Conversely it is clear that $(n, k+1) \in L(\mathbf{t})$ implies $(n,k) \in L(\mathbf{t})$.
□

Lemma 3 immediately gives an algorithm for deciding when n and k are such that $\mathbf{t}$ contains an (n,k)-word.

Preamble:
$Y = \{(n,k) | n \geqslant k \text{ and } k = 0,1,2\} \cup$
$\quad \{(n,3) | n \geqslant 3 \text{ and } n \neq 11, 13\}$

$N = \{(n,k) | n \text{ odd and } n \geqslant k \geqslant 4\} \cup$
$\quad \{(n,3) | n = 11, 13\}$

$f(n,k) = (n/2, k/2)$ for n even, k even,
$f(n,k) = (n/2, (k+1)/2)$ for n even, k odd.

```
Algorithm A:
 (1) Input (u,v) ← (n,k)
 (2) If (u,v) ∈ Y return '(n,k) ∈ L(t)' and HALT
 (3) If (u,v) ∈ N return '(n,k) ∉ L(t)' and HALT
 (4) (u,v) ← f(u,v)
 (5) GOTO (2)
```

We can now classify the set $L(\mathbf{t})$.

Lemma 4 *Let $a_\nu = 3 \cdot 2^\nu + 1$ for $\nu \geqslant 0$, and let n_1 be odd. Then*
(i) $(2^\nu n_1, k) \notin L(\boldsymbol{t})$ whenever $k \geqslant a_\nu$,
(ii) $(2^\nu n_1, k) \in L(\boldsymbol{t})$ whenever $k \leqslant 2^\nu n_1$ and $k < a_\nu$,
unless *$n_1 = 11, 13$ and $a_\nu - 2^\nu \leqslant k < a_\nu$.*

Proof: Write $f^j(n,k) = f(\ldots f(n,k)\ldots)$ for f applied j times to (n,k), $j \geqslant 0$, where f is as in Algorithm A.

If $k \geqslant a_\nu$ then $f^\nu(2^\nu n_1, k) = (n_1, k_1)$ and $k_1 \geqslant 4$, hence $(2^\nu n_1, k) \notin L(\mathbf{t})$ by Algorithm A and (i) is proved.

Now let $k \leqslant 2^\nu n_1$ and $k < a_\nu$, and let $f^\nu(2^\nu n_1, k) = (n_1, k_1)$. If $k < a_\nu - 2^\nu$ then $0 \leqslant k_1 \leqslant 2$, and then $(2^\nu n_1, k) \in L(\mathbf{t})$ by Algorithm A. If on the other hand

$a_\nu - 2^\nu \leqslant k < a_\nu$ then $k_1 = 3$, and then Algorithm A gives $(2^\nu n_1, k) \in L(\mathbf{t})$ unless $n_1 = 11$ or $n_1 = 13$. This proves (ii). □

As we observed in the introduction, a power $\frac{p}{q} \in E(\mathbf{w})$ need not occur in $\mathbf{w}$ with multiplier 1, for example the Kolakoski word has a $\frac{7}{4}$ power but not with multiplier 1. However, it *is* true that all powers that occur in the Thue-Morse word with a positive integer multiplier have an occurrence with multiplier 1. This is the key point of the next lemma.

Lemma 5 *If* $\frac{m}{n} \in E(\boldsymbol{t})$ *then* $1 \in \mathrm{Mult}_{\boldsymbol{t}}(\frac{m}{n})$.

Proof: By entry (n) of Table 1 we know that $m \leqslant 2n$, so write $m = n + k$, where $k \leqslant n$. First we will show that there is an odd integer $\ell \in \mathrm{Mult}_{\mathbf{t}}(\frac{m}{n})$. Since $\frac{m}{n} \in E(\mathbf{t})$ there is a positive integer ℓ such that $(n\ell, k\ell) \in L(\mathbf{t})$. Now if ℓ is even then $(n(\frac{\ell}{2}), k(\frac{\ell}{2})) \in L(\mathbf{t})$ by Lemma 3, part *(E1)*, and then $\frac{\ell}{2}$ is also a multiplier for $\frac{m}{n}$, and this is how ℓ can be reduced to an odd number.

Write $n = 2^\nu n_1$, where n_1 is odd, and as in Lemma 4 let $a_\nu = 3\cdot 2^\nu + 1$. We consider three cases depending on how large k is relative to a_ν.

(1) We have $k < a_\nu - 2^\nu$. Then $(n, k) \in L(\mathbf{t})$ by Lemma 4, and in this case $\frac{m}{n}$ clearly has 1 as a multiplier.

(2) We have $a_\nu - 2^\nu \leqslant k < a_\nu$. Then there is an odd ℓ such that $(2^\nu \cdot n_1\ell, k\ell) \in L(\mathbf{t})$. Here $n_1\ell$ is odd, so we cannot have $\ell \geqslant 3$ otherwise $k\ell \geqslant a_\nu$ and this contradicts Lemma $4(i)$. Therefore $\ell = 1$, as desired.

(3) We have $k \geqslant a_\nu$. In fact this case cannot happen, otherwise taking ℓ to be an odd multiplier we get $(2^\nu \cdot n_1\ell, k\ell) \in L(\mathbf{t})$ with $k\ell \geqslant a_\nu$ and $n_1\ell$ odd, again contradicting Lemma $4(i)$.

In all cases we find a multiplier of 1, and the proof is complete. □

By Lemma 5 we see that a $(\frac{m}{n})$-power occurs in $\mathbf{t}$ with positive integer multiplier if and only if $(n, m - n) \in L(\mathbf{t})$, and then Lemma 3 can be applied to give an explicit description of the exponent set of the Thue-Morse word.

Theorem 6 *Let* $1 < \frac{m}{n} < 2$ *be rational. Let* $m = n + k$ *and* $n = 2^{\nu(n)} n_1$ *with* n_1 *odd. A* $(\frac{m}{n})$*-power occurs in the Thue-Morse word with a positive integer multiplier if and only if* $k < 3\cdot 2^{\nu(n)} + 1$ *and*

$$\frac{m}{n} \notin \{\frac{13\cdot 2^j + a}{11\cdot 2^j} : 1 \leqslant a \leqslant 2^j, j \geqslant 0\} \cup \{\frac{15\cdot 2^j + a}{13\cdot 2^j} : 1 \leqslant a \leqslant 2^j, j \geqslant 0\}$$

Moreover, if a $(\frac{m}{n})$*-power occurs in the Thue-Morse word with a positive integer multiplier, then it occurs in the Thue-Morse word with multiplier 1.*

To understand Theorem 6 intuitively we imagine $\mathbf{t}$ being shifted 2^ν bits at a time. There are then only two possible incoming blocks of this length, which we think of as waves. Once a particular binary word of length k is long enough then its position among the waves, in other words its phase, becomes fixed. However, because the powers $\frac{14}{11}$ and $\frac{16}{13}$ are absent from $\mathbf{t}$, this phenomenon is not the only factor determining $E(\mathbf{t})$. In Figure 1 some other (absent) powers arising from $\frac{14}{11}$ are indicated, which we find by following Algorithm A backwards.

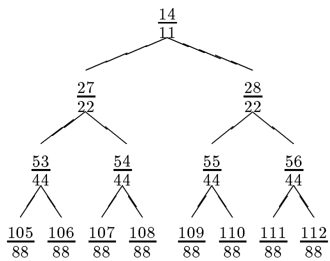

Figure 1: Powers not in $\mathbf{t}$ originating from $\frac{14}{11}$

The powers in Figure 1 are of the form $\frac{13\cdot 2^j+a}{11\cdot 2^j}$, $1 \leqslant a \leqslant 2^j$, $j \geqslant 0$, and they lie in the interval $(\frac{13}{11}, \frac{14}{11}]$. Similarly $\frac{16}{13}$ gives rise to absent powers of the form $\frac{15\cdot 2^j+a}{13\cdot 2^j}$, $1 \leqslant a \leqslant 2^j$, $j \geqslant 0$, that lie in the interval $(\frac{15}{13}, \frac{16}{13}]$. Incidentally $E(\mathbf{t})$ is dense in $[1, 2]$, since a square of length $2n$ in $\mathbf{t}$ gives $1 + \frac{a}{n} \in E(\mathbf{t})$ for each $0 < a \leqslant n$ and $\mathbf{t}$ contains arbitrarily long squares.

We remark that for *abelian* powers the problem is trivial, where a word XY is an abelian $\frac{|X|+|Y|}{|X|}$ power if $X = Y'X'$ and Y' is a permutation of Y. Note that abelian powers are rationals r with $1 < r \leqslant 2$. Given $r = p/q$ in this range, it is not hard to see that the prefix of $\mathbf{t}$ of length $2p$ is an abelian r-power.

4 Conclusion

It seems that the calculation of $E(\mathbf{w})$ for a $\mathbf{w}$ that is a fixed point of a "nice" morphism could be carried out along similar lines to our calculation of $E(\mathbf{t})$. Certainly if the morphism is uniform (images of all letters have the same length) one can look for synchronizing subwords and an algorithm like algorithm A, leading to a "phase inequality" which powers in $E(\mathbf{w})$ must satisfy. One would then hope to find a short list of absent powers which satisfy the phase inequality and such that any other absent power that satisfies the phase inequality could be traced to one of these powers by following the algorithm backwards.

More challenging is to find a good description of $E(\mathbf{f})$, where $\mathbf{f}$ is the Fibonacci infinite word. The morphism describing $\mathbf{f}$ is not uniform, so phase considerations may not be useful, but $\mathbf{f}$ is also a Sturmian word so perhaps this fact can be exploited.

Most challenging of all would be to calculate $E(\mathbf{k})$, where $\mathbf{k}$ is the Kolakoski word. Unlike $\mathbf{t}$ or $\mathbf{f}$, the word $\mathbf{k}$ is not the (encoded or unencoded) fixed point of a morphism. Multipliers for $\mathbf{k}$ are a mystery, e.g., 3 is the smallest integer multiplier in $\text{Mult}_{\mathbf{t}}(\frac{7}{4})$. One wonders if Carpi's [4] demonstration of the finiteness of $\text{Mult}_{\mathbf{k}}(\frac{2}{1})$ indicates that all fractional powers occurring in $\mathbf{k}$ have finitely many multipliers.

We leave these as open problems for the reader to solve.

References

[1] J. P. Allouche and J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, *Sequences and their applications (Singapore, 1998)* Springer, London, 1999.

[2] J. P. Allouche and J. Shallit, Automatic Sequences, Theory, Applications, Generalizations, Cambridge University Press, New York, 2003.

[3] J. Berstel, Axel Thue's papers on repetitions in words: a translation, Publications du Laboratoire de Combinatoire et d' Informatique Mathématique, 1994.

[4] On repeated factors in C^∞-words, Inform. Process. Lett. **52** (1994), 151–167.

[5] F. Dejean, Sur un théorème de Thue, *J. Combin. Theory Ser. A* **13** (1972), 90–99.

[6] D. S. Fitzpatrick, There are binary circular cube-free words of length n contained within the Thue-Morse word for all positive integers n, *Ars Combin.* **74** (2005), 323–329.

[7] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.

[8] F. Mignosi and G. Pirillo, Repetitions in the Fibonacci infinite word, Inform. Théorique Appl. **26** (1992), 199–204.

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

[10] E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres, *C. R. Acad. Sci. Paris Sér. I* **33** (1851), 225.

[11] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in "Selected mathematical papers of Axel Thue", T. Nagell, ed., Universitetsforlaget, Oslo, 1977, pp. 413–478.

(Received 4 Oct 2004; revised 8 June 2005)