Thue-Morse along two polynomial subsequences

Thomas Stoll

Thue–Morse along two polynomial subsequences

Thomas Stoll*

Institut Elie Cartan de Lorraine, UMR 7502,
Université de Lorraine / CNRS
thomas.stoll@univ-lorraine.fr
http://iecl.univ-lorraine.fr/~Thomas.Stoll/

Abstract. The aim of the present article is twofold. We first give a survey on recent developments on the distribution of symbols in polynomial subsequences of the Thue–Morse sequence $\mathbf{t} = (t(n))_{n\geq 0}$ by highlighting effective results. Secondly, we give explicit bounds on

$$\min\{n : (t(pn), t(qn)) = (\varepsilon_1, \varepsilon_2)\},$$

for odd integers p, q, and on

$$\min\{n : (t(n^{h_1}), t(n^{h_2})) = (\varepsilon_1, \varepsilon_2)\}$$

where $h_1, h_2 \geq 1$, and $(\varepsilon_1, \varepsilon_2)$ is one of $(0,0), (0,1), (1,0), (1,1)$.

Keywords: Thue–Morse sequence, sum of digits, polynomials

1 Introduction

The Thue–Morse sequence

$$\mathbf{t} = (t(n))_{n\geq 0} = 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, \ldots$$

can be defined via

$$t(n) = s_2(n) \bmod 2, \tag{1}$$

where $s_2(n)$ denotes the number of one bits in the binary expansion of n, or equivalently, the sum of digits of n in base 2. This sequence can be found in various fields of mathematics and computer science, such as combinatorics on words, number theory, harmonic analysis and differential geometry. We refer the reader to the survey articles of Allouche and Shallit [2], and of Mauduit [14] for a concise introduction to this sequence. As is well-known, Thue–Morse is 2-automatic and can be generated by the morphism $0 \mapsto 01$, $1 \mapsto 10$. It is also the prime example of an overlapfree sequence.

* Work supported by the ANR-FWF bilateral project MuDeRa "Multiplicativity: Determinism and Randomness" (France-Austria) and the joint project "Systèmes de numération : Propriétés arithmétiques, dynamiques et probabilistes" of the Université de Lorraine and the Conseil Régional de Lorraine.

The overall distribution of the symbols 0 and 1 in Thue–Morse is trivial since the sequence consists exclusively of consecutive blocks of the forms 01 and 10, thus there are "as many 0's as 1's" in the sequence. The investigation of Thue–Morse along subsequences can be said to have started with an influential paper by Gelfond in 1967/68 [9]. He proved, via exponential sums techniques, that $\mathbf{t}$ is uniformly distributed along arithmetic progressions, namely,

$$\#\{n < N : \quad t(an+b) = \varepsilon\} \sim \frac{N}{2}, \qquad (\varepsilon = 0 \text{ or } 1)$$

with an explicit error term. Gelfond's result shows that there are "as many 0's as 1's" in the sequence also regarding arithmetic progressions. His result, however, gives no information on how long one actually has to wait to "see" the first, say, "1" along a specific arithmetic progression. Newman [19] showed that there is a weak preponderance of the 0's over the 1's in the sequence of the multiples of three. More precisely, he showed that

$$\#\{n < N : \quad t(3n) = 0\} = \frac{N}{2} + C(N),$$

with $c_1' N^{\log_4 3} < C(N) < c_2' N^{\log_4 3}$ for all $N \geq 1$ and certain positive constants c_1', c_2'. For the multiples of three one has to wait for 7 terms to "see the first 1", i.e.,

$$\min\{n : \quad t(3n) = 1\} = 7.$$

Morgenbesser, Shallit and Stoll [18] proved that for $p \geq 1$,

$$\min\{n : \quad t(pn) = 1\} \leq p + 4,$$

and this becomes sharp for $p = 2^{2r} - 1$ for $r \geq 1$ (Note that $3 = 2^{2\cdot 1} - 1$ is exactly of that form). A huge literature is nowadays available for classes of arithmetic progressions where such Newman-type phenomena exist and many generalizations have been considered so far (see [3, 4, 8, 10, 12, 22] and the references given therein). Still, a full classification is not yet at our disposal.

Most of the results that hold true for Thue–Morse in the number-theoretic setting of (1) have been proven for the sum of digits function in base q, where q is an integer greater than or equal to 2, and where the reduction in (1) is done modulo an arbitrary integer $m \geq 2$. We refrain here from the general statements and refer the interested readers to the original research papers.

Following the historical line, Gelfond [9] posed two challenging questions concerning the distribution of the sum of digits function along primes and along polynomial values instead of looking at linear subsequences. A third question was concerned with the simultaneous distribution when the sum of digits is taken to different bases; this question has been settled by Kim [13]. In recent years, this area of research gained much momentum due to an article by Mauduit and Rivat [15] who answered Gelfond's question for primes with an explicit error

term. In a second paper [16], they also answered Gelfond's question for the sequence of squares. Their result implies that

$$\#\{n < N : \quad t(n^2) = \varepsilon\} \sim \frac{N}{2}.$$

In a very recent paper, Drmota, Mauduit and Rivat [6] showed that $\mathbf{t}$ along squares gives indeed a normal sequence in base 2 meaning that each binary block appears with the expected frequency. This quantifies a result of Moshe [17] who answered a question posed by Allouche and Shallit [1, Problem 10.12.7] about the complexity of Thue–Morse along polynomial extractions.

We are still very far from understanding

$$\#\{n < N : \quad t(P(n)) = \varepsilon\},$$

where $P(x) \in \mathbb{Z}[x]$ is a polynomial of degree ≥ 3. Drmota, Mauduit and Rivat [7] obtained an asymptotic formula for $\#\{n < N : \ s_q(P(n)) = \varepsilon \pmod m\}$ whenever q is sufficiently large in terms of the degree of P. The case of Thue–Morse is yet out of reach of current technology. The currently best result is due to the author [23], who showed that there exists a constant $c = c(P)$ depending only on the polynomial P such that

$$\#\{n < N : \quad t(P(n)) = \varepsilon\} \geq cN^{4/(3\deg P+1)}. \tag{2}$$

This improves on a result of Dartyge and Tenenbaum [5] who had $N^{2/(\deg P)!}$ for the lower bound. The method of proof for (2) is constructive and gives an explicit bound on the minimal non-trivial n such that $t(n^h) = \varepsilon$ for fixed $h \geq 1$. Since $t(n^h) = 1$ for all $n = 2^r$, and $t(0^h) = 0$, we restrict our attention to

$$\mathcal{A} = \{n : \quad n \neq 2^r, \ r \geq 0\} = \{3, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, \ldots\}.$$

From the proof of (2) follows that

$$\min(n \in \mathcal{A} : \quad t(n^h) = \varepsilon) \leq 64^{h+1} \left(8h \cdot 12^h\right)^{3h+1}.$$

Hence, there exists an absolute constant $c_1 > 0$ such that

$$\min(n \in \mathcal{A} : \quad t(n^h) = \varepsilon) \leq \exp(c_1 h^2). \tag{3}$$

With some extra work, a similar result can be obtained for a general polynomial $P(n)$ instead of n^h, where there corresponding constant will depend on the coefficients of P.

The joint distribution of the binary digits of integer multiples has been studied by J. Schmid [20] and in the more general setting of polynomials by Steiner [21]. The asymptotic formulas do not imply effective bounds on the first n that realizes such a system and it is the aim of this paper to prove effective bounds in the case of two equations for integer multiples and for monomials.

Our first result is as follows.

Theorem 1. *Let $p > q \geq 1$ be odd integers. Then there exists an absolute constant $c_2 > 0$ such that*

$$\min(n:\ t(pn) = \varepsilon_1,\ t(qn) = \varepsilon_2) \leq \exp(c_2 \log p) \qquad (\varepsilon_1, \varepsilon_2 \in \{0,1\}).$$

Remark 1. Note that for $p = 2^r + 1$ with $r \geq 1$ we have $t(pn) = 0$ for all $n < p-1$, so that there is no absolute bound for the minimal n.

There are examples that show that sometimes one has to "wait" quite some time to see all of the four possibilities for $(\varepsilon_1, \varepsilon_2)$ when the extraction is done along two monomial sequences. For instance, we have

$$\min\{n \in \mathcal{A}:\quad t(n^{130}) = \varepsilon_1)\} \leq 7 \quad \text{and} \quad \min\{n \in \mathcal{A}:\quad t(n^{53}) = \varepsilon_2)\} \leq 5$$

for $\varepsilon_1, \varepsilon_2 = 0, 1$ but

$$\min\left\{n:\quad (t(n^{130}), t(n^{53})) = (0,0)\right\} = 113.$$

The construction that we will use to prove Theorem 1 will not be useful to study the minimal n along polynomial subsequences since in this case we would need to keep track of the binary digits sum of various binomial coefficients. Instead, we will use ideas from work of Hare, Laishram and the author [11] to show the following result.

Theorem 2. *Let $h_1 > h_2 \geq 1$ be integers. Then there exists an absolute constant $c_3 > 0$ such that*

$$\min(n \in \mathcal{A}:\ t(n^{h_1}) = \varepsilon_1,\ t(n^{h_2}) = \varepsilon_2) \leq \exp(c_3 h_1^3) \qquad (\varepsilon_1, \varepsilon_2 \in \{0,1\}).$$

Remark 2. The method also allows to treat general *monic* polynomials $P_1(x)$, $P_2(x) \in \mathbb{Z}[x]$ of different degree h_1, h_2 in place of x^{h_1}, x^{h_2}. Even more generally, we can deal with *non-monic* polynomials $P_1(x), P_2(x) \in \mathbb{Z}[x]$ provided h_1 *is odd*. As we will see in the proof (compare with the remark after (14)), the latter condition relies on the fact that for odd h the congruence $x^h \equiv a \pmod{16}$ admits a solution mod 16 for each odd a, while this is not true in general if h is even.

We write $\log_2$ for the logarithm to base 2. Moreover, for $n = \sum_{j=0}^{\ell} n_j 2^j$ with $n_j \in \{0,1\}$ and $n_\ell \neq 0$ we write $(n_\ell, n_{\ell-1}, \cdots, n_1, n_0)_2$ for its digital representation in base 2 and set $\ell = \ell(n)$ for its length. To simplify our notation, we allow to fill up by a finite number of 0's to the left, i.e., $(n_\ell n_{\ell-1} \cdots n_1 n_0)_2 = (0 n_\ell n_{\ell-1} \cdots n_1 n_0)_2 = (0 \cdots 0 n_\ell n_{\ell-1} \cdots n_1 n_0)_2$.

The paper is structured as follows. In Section 2 we prove Theorem 1 and in Section 3 we show Theorem 2.

2 Thue–Morse at distinct multiples

The proof of Theorem 1 is based on the following lemma.

Lemma 1. *Let p, q be odd positive integers with $p > q \geq 1$ and let $(\varepsilon_1, \varepsilon_2)$ be one of $(0,0),(0,1),(1,0),(1,1)$. Then we have*

$$\min\{n : \ (t(pn), t(qn)) = (\varepsilon_1, \varepsilon_2)\} \leq C_{\varepsilon_1,\varepsilon_2}(p,q),$$

where

$$C_{0,0}(p,q) = C_{1,1}(p,q) = 4p, \quad C_{0,1}(p,q) = \frac{2^{23}p^{11}}{(p-q)^6} \quad C_{1,0}(p,q) = \frac{2^6 p^3}{(p-q)^2}.$$

Proof. Recall that for $1 \leq b < 2^k$ and $a, k \geq 1$, we have

$$s_2(a2^k + b) = s_2(a) + s_2(b), \qquad s_2(a2^k - b) = s_2(a-1) + k - s_2(b-1). \quad (4)$$

In the sequel we will make frequent use of these splitting formulas. We first deal with the two cases when $(\varepsilon_1, \varepsilon_2)$ is one of $(0,0),(1,1)$. If $2^k > p > q$ then $s_2(p(2^k-1)) = s_2(q(2^k-1)) = k$. Moreover, since $k \equiv 0$ or $1 \bmod 2$ and

$$2^k - 1 \leq 2^{\log p/\log 2 + 2} - 1 = 4p - 1 < 4p,$$

we get that $C_{0,0}(p,q), C_{1,1}(p,q) \leq 4p$. Finding explicit bounds for $C_{0,1}(p,q)$ and $C_{1,0}(p,q)$ is more involved. To begin with, we first claim that there exists $n_1 \geq 1$ with the following two properties:

(a) $\ell(pn_1) > \ell(qn_1)$,
(b) $pn_1 \equiv 1 \bmod 4$.

As for (a), we need to find two integers a, n_1 such that $2^a \leq pn_1$ and $2^a > qn_1$. This is equivalent to

$$\frac{2^a}{p} \leq n_1 < \frac{2^a}{q}. \quad (5)$$

For odd k, n either $kn \equiv 1 \pmod 4$ or $k(n+2) \equiv 1 \pmod 4$, so provided $2^a\left(\frac{1}{q} - \frac{1}{p}\right) \geq 4$, we can find an odd n_1 that satisfies both (a) and (b). By taking a to be the unique integer with

$$\frac{4pq}{p-q} \leq 2^a < 2 \cdot \frac{4pq}{p-q}, \quad (6)$$

we get an n_1 with

$$n_1 < \frac{8p}{p-q}.$$

Now, define $n_2 = 2^{\ell(pn_1)} + 1$. Since both p and n_1 are odd we have $n_2 \leq pn_1$. Then

$$s_2(pn_1n_2) = s_2(pn_1 2^{\ell(pn_1)} + pn_1) = 2s_2(pn_1) - 1 \equiv 1 \pmod 2, \quad (7)$$

since there is exactly one carry propagation from the most significant digit of pn_1 to the digit at digit place $\ell(pn_1)$ of $pn_1 2^{\ell(pn_1)}$ which stops immediately after one step because of property (b). On the other hand, (a) implies that

$$s_2(qn_1n_2) = s_2(qn_1 2^{\ell(pn_1)} + qn_1) = 2s_2(qn_1) \equiv 0 \pmod 2, \tag{8}$$

because the terms qn_1 and $qn_1 2^{\ell(pn_1)}$ do not interfere and there is therefore no carry propagation while adding these two terms. We therefore can set

$$n = n_1n_2 \le \left(\frac{8p}{p-q}\right)^2 p$$

to get that $C_{1,0}(p,q) \le 2^6 \cdot \frac{p^3}{(p-q)^2}$. For $(\varepsilon_1, \varepsilon_2) = (0,1)$, take m to be the unique odd integer with

$$pn_1n_2 < 2^m \le 4pn_1n_2$$

and put

$$n_3 = 2^{2m} + 2^m - 1 \le 2(4pn_1n_2)^2 \le 2^5 (pn_1)^4.$$

Then by (7),

$$\begin{aligned} s_2(pn_1n_2n_3) &= s_2\left((pn_1n_2 2^m + pn_1n_2)2^m - pn_1n_2\right) \\ &= s_2\left(pn_1n_2 2^m + pn_1n_2 - 1\right) + m - s_2\left(pn_1n_2 - 1\right) \\ &= s_2\left(pn_1n_2 - 1\right) + s\left(pn_1n_2\right) + m - s_2\left(pn_1n_2 - 1\right) \\ &\equiv m + 1 \equiv 0 \pmod 2. \end{aligned}$$

A similar calculation shows by (8) that

$$s\left(qn_1n_2n_3\right) \equiv m \equiv 1 \pmod 2.$$

We set $n = n_1n_2n_3$ and get

$$n \le \left(\frac{8p}{p-q}\right)^2 p \cdot 2^5 p^4 \cdot \frac{(8p)^4}{(p-q)^4} = 2^{23} \cdot \frac{p^{11}}{(p-q)^6}.$$

This completes the proof. □

Proof (Theorem 1). This follows directly from Lemma 1 and

$$2^{23} \cdot \frac{p^{11}}{(p-q)^6} \le 2^{17} p^{11}.$$

□

3 Thue–Morse at two polynomials

This section is devoted to the proof of a technical result which implies Theorem 2. Considering the extractions of Thue–Morse along n^{h_1} and n^{h_2}, it is simple to

get two and not too difficult to get three out of the four possibilities for $(\varepsilon_1, \varepsilon_2)$. However, to ensure that we see all of the four possibilities, we need a rather subtle construction. The difficulty is similar to that one to get $C_{0,0}(p,q)$ in the proof of Theorem 1. The idea is to shift two specific blocks against each other while all other terms in the expansions are non-interfering. Via this procedure, we will be able to keep track of the number of carry propagations. In the proof of Theorem 1 we have used the blocks pn_1 and qn_1. In the following, we will make use of $u_1 = 9 = (1001)_2$ and $u_2 = 1 = (0001)_2$. Then

$$\begin{aligned} s_2(u_1 + u_2) &= 2, \\ s_2(u_1 + 2u_2) &= 3, \\ s_2(u_1 + 2^2u_2) &= 3, \\ s_2(u_1 + 2^3u_2) &= 2, \end{aligned}$$

and mod 2 we get the sequence $(0,1,1,0)$. These particular expansions and additions will be of great importance in our argument (compare with (16)–(19)).

Lemma 2. *Let h_1, h_2 be positive integers with $h_1 > h_2 \geq 1$ and let $(\varepsilon_1, \varepsilon_2)$ be one of $(0,0),(0,1),(1,0),(1,1)$. Then we have*

$$\min\left\{n \in \mathcal{A} :\ (t(n^{h_1}), t(n^{h_2})) = (\varepsilon_1, \varepsilon_2)\right\} \leq C,$$

where

$$C = 32^{3h_1+5}\left(\frac{1}{15}\left(\frac{150}{h_1}\,8^{h_1}\right)^2\right)^{5h_1^2+h_1-5}.$$

Proof. Let $h \geq 1$ and put

$$\begin{aligned} l &= \left\lfloor \log_2\left(\frac{1}{15}\left(\frac{150}{h}\,8^h\right)^2\right)\right\rfloor + 1, \qquad (9)\\ a &= (2^{lh}-1)2^4 + 1, \\ b &= \left\lfloor \frac{150}{h}\,8^h a^{1-\frac{1}{h}}\right\rfloor, \\ M &= \left\lfloor 15a^{1-\frac{1}{h}}\right\rfloor + 1, \\ k &= lh^2 + 3h + 4 - l. \end{aligned}$$

It is straightforward to check that for all $h \geq 1$, we have

$$l \geq 12, \quad a \geq 65521, \quad b \geq 1200, \quad M \geq 16, \quad k \geq 7. \qquad (10)$$

Obviously, we have $b \geq M$. Moreover,

$$a = 16 \cdot 2^{lh} - 15 \geq 15 \cdot 2^{lh} \geq \left(\frac{150}{h}\,8^h\right)^h$$

and therefore $a \geq b$. Furthermore, for $h \geq 2$, we have $aM > b^2$ since by

$$a > \left(\frac{1}{15} \left(\frac{150}{h} \, 8^h \right)^2 \right)^h$$

we have

$$aM - b^2 > a \cdot 15 a^{1-\frac{1}{h}} - \left(\frac{150}{h} \, 8^h a^{1-\frac{1}{h}} \right)^2 > 0.$$

Let

$$T(x) = ax^5 + bx^4 + Mx^3 + Mx^2 - x + M$$

and write $T(x)^h = \sum_{i=0}^{5h} \alpha_i x^i$. Obviously, $\alpha_0 > 0$ and $\alpha_1 < 0$. We claim that for $h \geq 1$ we have $\alpha_i > 0$ for $2 \leq i \leq 5h$. To see this we write

$$T(x)^h = \left(ax^5 + bx^4 + Mx^3 + Mx^2 + M \right)^h + r(x), \tag{11}$$

with

$$r(x) = \sum_{j=1}^{h} \binom{h}{j} (-x)^j \left(ax^5 + bx^4 + Mx^3 + Mx^2 + M \right)^{h-j} = \sum_{j=1}^{5(h-1)} d_j x^j.$$

Since $a \geq b \geq M$ the coefficient of x^i in the first term in (11) is $\geq M^h$. On the other hand,

$$|d_j| < h 2^h (5a)^{h-1},$$

and

$$M^h \geq \left(15 a^{1-1/h} \right)^h > \left(3 \cdot 5 a^{1-1/h} \right)^h > \left(2 h^{1/h} (5a)^{1-1/h} \right)^h = |d_j|$$

which proves the claim. Next, we need a bound on he size of α_i. The coefficients α_i, $0 \leq i \leq 5h-2$, are bounded by the corresponding coefficients in the expansion of $(ax^5 + bx^4 + M(x^3 + x^2 + x + 1))^h$. Since $aM > b^2$ and $M \leq 16 a^{1-1/h}$, each of these coefficients is bounded by

$$\left(a^{h-1} M \right) \cdot 6^h < a^{h-1} M 8^h \leq a^{h-\frac{1}{h}} \cdot 16 \cdot 8^h,$$

and therefore

$$|\alpha_i| \leq a^{h-\frac{1}{h}} \cdot 16 \cdot 8^h, \qquad i = 0, 1, 2, \ldots, 5h-2. \tag{12}$$

Moreover, we have

$$\alpha_{5h-1} = hba^{h-1} = \left\lfloor \frac{150}{h} \, 8^h a^{1-\frac{1}{h}} \right\rfloor h a^{h-1} \tag{13}$$

and

$$149 \cdot 8^h a^{h-\frac{1}{h}} \leq \alpha_{5h-1} \leq 150 \cdot 8^h a^{h-\frac{1}{h}},$$

which is true for all $a \geq 1$ and $h \geq 1$. Note that both the bound in (12) and the coefficient α_{5h-1} are increasing functions in h. From now on, suppose that $h \geq 2$. We further claim that

$$a^{h-\frac{1}{h}} \cdot 16 \cdot 8^h < 2^k, \qquad 9 \cdot 2^k \leq \alpha_{5h-1} < 10 \cdot 2^k, \tag{14}$$

which will give us the wanted overlap for the digital blocks of α_{5h-1} and α_{5h}. By (14) the binary expansion of α_{5h-1} is $(1001\cdots)_2$ and interferes with the digital block coming from $\alpha_{5h} = a^h$ which is $(\cdots 0001)_2$ since $a \equiv 1 \pmod{16}$. To prove (14), we show a stronger inequality that in turn implies (14), namely,

$$144 \cdot 8^h a^{h-\frac{1}{h}} < 9 \cdot 2^k \leq 149 \cdot 8^h a^{h-\frac{1}{h}}. \tag{15}$$

Passing to logarithms, this is equivalent to

$$(k - 3h - 4) - \delta \leq \left(h - \frac{1}{h}\right) \log_2 a < k - 3h - 4$$

with

$$\delta = \log_2 149 - \log_2 9 - 4 = 0.04924\cdots > \frac{1}{25}.$$

We rewrite

$$\left(h - \frac{1}{h}\right) \log_2 a = \left(h - \frac{1}{h}\right) \left(lh + \log_2 \left(1 - \left(\frac{1}{2^{lh-4}} - \frac{1}{2^{lh}}\right)\right)\right),$$

which on the one hand shows that

$$\left(h - \frac{1}{h}\right) \log_2 a < \left(h - \frac{1}{h}\right) lh = lh^2 - l = k - 3h - 4,$$

and on the other hand by $-x/(1-x) < \log(1-x) < 0$ for $0 < x < 1$ that

$$\left(h - \frac{1}{h}\right) \log_2 a > lh^2 - l - \frac{1}{\log 2} \cdot \frac{(2^4 - 1) \cdot \left(h - \frac{1}{h}\right)}{2^{lh} - 2^4 + 1}.$$

Finally, we easily check that for all $h \geq 2$ and $l \geq 5$ we have

$$\frac{1}{\log 2} \cdot \frac{15\left(h - \frac{1}{h}\right)}{2^{lh} - 15} < \frac{1}{25},$$

which finishes the proof of (15) and thus of (14).

After this technical preliminaries we proceed to the evaluation of the sum of digits. First note that for all $h \geq 1$ by construction none of $n = T(2^k)$, $T(2^{k+1})$, $T(2^{k+2})$, $T(2^{k+3})$ is a power of two and therefore $n \in \mathcal{A}$ in these four cases. Let $h = h_1 \geq 2$ and define a, b, M, l, k according to (9). To begin with, by (12), (13), (14) and the splitting formulas (4), we calculate

$$s_2(T(2^k)^{h_1}) = s_2\left(\sum_{i=0}^{5h_1} \alpha_i 2^{ik}\right) = A_1 - 1 + A_2 + k + A_3, \tag{16}$$

where

$$\begin{aligned} A_1 &= s_2(\alpha_{5h_1}) + s_2(\alpha_{5h_1-1}), \\ A_2 &= s_2\left(\sum_{i=3}^{5h_1-2} \alpha_i\right) + s_2(\alpha_0), \\ A_3 &= s_2(\alpha_2 - 1) - s_2(\alpha_1 - 1). \end{aligned}$$

Note that the summand k in (16) comes from formula (4) due to the negative coefficient α_1, and the -1 comes from the addition of $(\cdots 0001)_2$ and $(1001)_2$ (the ending and starting blocks corresponding to α_{5h_1} and α_{5h_1-1}) which gives rise to exactly one carry. A similar calculation shows that

$$s_2(T(2^{k+1})^{h_1}) = A_1 + A_2 + (k+1) + A_3, \tag{17}$$

$$s_2(T(2^{k+2})^{h_1}) = A_1 + A_2 + (k+2) + A_3, \tag{18}$$

$$s_2(T(2^{k+3})^{h_1}) = A_1 + A_2 - 1 + (k+3) + A_3. \tag{19}$$

In (17) we add $(1001)_2$ to $(\cdots 00010)_2$ which gives no carry. The same happens for the addition of $(1001)_2$ to $(\cdots 000100)_2$ in (18). Finally, in (19) we again have exactly one carry in the addition of $(1001)_2$ to $(\cdots 0001000)_2$. If we look mod 2 this shows that

$$\big(t(T(2^k)^{h_1}), t(T(2^{k+1})^{h_1}), t(T(2^{k+2})^{h_1}), t(T(2^{k+3})^{h_1})\big)$$

is either $(0,0,1,1)$ or $(1,1,0,0)$. For $h_2 < h_1$ with $h_2 \geq 1$ all coefficients are non-interfering. To see this, consider the coefficients of $T(x)^{h_2} = \sum_{i=0}^{5h_2} \alpha'_i x^i$. By (12), they are clearly bounded in modulus by $a^{h_1 - \frac{1}{h_1}} \cdot 16 \cdot 8^{h_1} < 2^k$ for $i = 0, 1, 2, \ldots, 5h_2 - 2$. Also, by (10) and $h_1 \geq 2$,

$$\alpha'_{5h_2-1} \leq 150 \cdot 8^{h_2} a^{h_2 - \frac{1}{h_2}} < \frac{150}{8} \cdot 8^{h_1} a^{h_1 - 1} < a^{h_1 - \frac{1}{h_1}} \cdot 16 \cdot 8^{h_1} < 2^k,$$

and thus we don't have carry propagations in the addition of terms in the expansion of $T(2^k)^{h_2}$. Similarly, we show that

$$\big(t(T(2^k)^{h_2}), t(T(2^{k+1})^{h_2}), t(T(2^{k+2})^{h_2}), t(T(2^{k+3})^{h_2})\big)$$

is either $(0,1,0,1)$ or $(1,0,1,0)$. This yields in any case that

$$\{(t(T(2^{k+i})^{h_1}), t(T(2^{k+i})^{h_2})) : \quad i = 0,1,2,3\} = \{(0,0),(0,1),(1,0),(1,1)\},$$

which are the four desired values. Finally,

$$\begin{aligned} T(2^{k+3}) &\leq 2^{5(k+3)} \cdot 2^{lh_1+4} \\ &\leq 2^{l(5h_1^2+h_1-5)} \cdot 2^{5(3h_1+4)+4} \\ &\leq 32^{3h_1+5} \left(\frac{1}{15}\left(\frac{150}{h_1} 8^{h_1}\right)^2\right)^{5h_1^2+h_1-5}, \end{aligned}$$

which completes the proof. □

Proof (Theorem 2). This follows from Lemma 2 and

$$32^{3h_1+5}\left(\frac{1}{15}\left(\frac{150}{h_1}\,8^{h_1}\right)^2\right)^{5h_1^2+h_1-5} \le \exp\left(c_3 h_1^3\right)$$

for some suitable positive constant c_3. □

Acknowledgements

I am pleased to thank Jeff Shallit for bringing to my attention the question on bounding Thue–Morse on two multiples. I also thank him and E. Rowland for several discussions.

References

1. Allouche J.-P., Shallit J.: Automatic Sequences. Theory, Applications, Generalizations., Cambridge University Press, Cambridge, 2003.
2. Allouche J.-P., Shallit J.: The ubiquitous Prouhet-Thue-Morse sequence, in: Sequences and their applications (Singapore, 1998), 1–16, *Springer Ser. Discrete Math. Theor. Comput. Sci.*, London, 1999.
3. Boreico I., El-Baz D., T. Stoll: On a conjecture of Dekking: the sum of digits of even numbers, *J. Théor. Nombres Bordeaux* **26** (2014), 17–24.
4. Coquet J.: A summation formula related to the binary digits, *Invent. Math.* **73** (1983) 107–115.
5. Dartyge C., Tenenbaum G.: Congruences de sommes de chiffres de valeurs polynomiales, *Bull. London Math. Soc.* **38** (2006), no. 1, 61–69.
6. Drmota M., Mauduit C., Rivat J.: The Thue-Morse sequence along squares is normal, *manuscript.*
7. Drmota M., Mauduit C., Rivat J.: The sum of digits function of polynomial sequences, *J. London Math. Soc.* **84** (2011), 81–102.
8. Drmota M., Skałba M.: Rarified sums of the Thue-Morse sequence, *Trans. Amer. Math. Soc.* **352** (2000) 609–642.
9. Gelfond A. O.: Sur les nombres qui ont des propriétés additives et multiplicatives données, *Acta Arith.* **13** (1967/1968), 259–265.
10. Goldstein S., Kelly K. A., Speer E. R.: The fractal structure of rarefied sums of the Thue-Morse sequence, *J. Number Theory* **42** (1992) 1–19.
11. Hare K. G., Laishram S., Stoll T.: Stolarsky's conjecture and the sum of digits of polynomial values, *Proc. Amer. Math. Soc.* **139** (2011), 39–49.
12. Hofer R.: Coquet-type formulas for the rarefied weighted Thue–Morse sequence, *Discrete Math.* **311** (2011), 1724–1734.
13. Kim D.-H.: On the joint distribution of q-additive functions in residue classes, *J. Number Theory* **74** (1999), 307–336.
14. Mauduit C.: Multiplicative properties of the Thue–Morse sequence, *Period. Math. Hungar.* **43** (2001), 137–153.
15. Mauduit C., Rivat J.: Sur un problème de Gelfond: la somme des chiffres des nombres premiers, *Ann. of Math.* **171** (2010), 1591–1646.

16. Mauduit C., Rivat J.: La somme des chiffres des carrés, *Acta Math.* **203** (2009), 107–148.
17. Moshe Y.: On the subword complexity of Thue–Morse polynomial extractions, *Theoret. Comput. Sci.* **389** (2007), 318–329.
18. Morgenbesser J., Shallit J., Stoll T.: Thue–Morse at multiples of an integer, *J. Number Theory* **131** (2011), 1498–1512.
19. Newman D. J.: On the number of binary digits in a multiple of three, *Proc. Amer. Math. Soc.* **21** (1969) 719–721.
20. Schmid J.; The joint distribution of the binary digits of integer multiples, *Acta Arith.* **43** (1984), 391–415.
21. Steiner W.: On the joint distribution of q-additive functions on polynomial sequences. Proceedings of the Conference Dedicated to the 90th Anniversary of Boris Vladimirovich Gnedenko (Kyiv, 2002). *Theory Stoch. Process.* **8** (2002), 336–357.
22. Stoll T.: Multi-parametric extensions of Newman's phenomenon, *Integers* **5** (2005), A14, 14 pp. (electronic).
23. Stoll T.: The sum of digits of polynomial values in arithmetic progressions, *Funct. Approx. Comment. Math.* **47** (2012), part 2, 233–239.