Sums of the Thue–Morse sequence over arithmetic progressions

T.W. Cusick[1], P. Stănică[2]

[1]State University of New York, Department of Mathematics
Buffalo, NY 14260; Email: cusick@buffalo.edu
[2]Naval Postgraduate School, Applied Mathematics Department
Monterey, CA 93943; Email: pstanica@nps.edu

May 29, 2009

Abstract

In this note we use the theory of Boolean functions to find a new elementary proof for Moser's conjecture that states that in the bounded sequence of nonnegative integers divisible by 3 there are more integers with an even number of 1s in their base-2 representation. This proof is simpler than the original proof by D. J. Newman in 1969. We further apply the method to prove a similar result for $p = 5$, which was also done by Grabner in 1993. The methods seems to be extendable to other primes, but the computations for the relevant constants will be quite complex.

MSC: 06E30, 11A63, 11B85, 94C10

Keywords: Boolean functions; Thue-Morse sequence; Binary representation

1 Introduction

The well-known Thue-Morse sequence $T = \{t(n) : n = 0, 1, 2, \ldots\}$ is defined by $t(n) =$ parity of the sum of the bits occurring in the binary representation of the nonnegative integer n. Thus the sequence T, grouped in blocks of 4, is

$$T = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \cdots \tag{1}$$

For brevity, we call T the "TM sequence". We denote the 2^n-length initial segment of the TM sequence by T_{2^n}. It is easy to see from the definition that the TM sequence can also be generated by

$$\begin{aligned} &T_1 = t_0 = 0, \\ &T_{2^n} = T_{2^{n-1}}\overline{T_{2^{n-1}}},\ n \geq 1, \end{aligned} \tag{2}$$

where $\overline{B}$ denotes the complement of B.

Let $\mathbb{F}_2^n$ be the vector space of dimension n over the two-element field $\mathbb{F}_2$. Let us denote the addition operator over $\mathbb{F}_2$ by $\oplus$ (this is just addition modulo 2). A *Boolean function* on n variables may be viewed as a mapping from $\mathbb{F}_2^n$ into $\mathbb{F}_2$. We order $\mathbb{F}_2^n$ lexicographically, and denote $\mathbf{v}_0 = (0, \ldots, 0, 0)$, $\mathbf{v}_1 = (0, \ldots, 0, 1)$, $\mathbf{v}_{2^n-1} = (1, \ldots, 1, 1)$. We interpret a Boolean function $f(x_1, \ldots, x_n)$ as the outputs of the function obtained from all its inputs in lexicographic order, i.e., a binary string of length 2^n, $f = [f(\mathbf{v}_0),\ f(\mathbf{v}_1), f(\mathbf{v}_2), \ldots, f(\mathbf{v}_{2^n-1})]$. We will sometimes omit the commas in this representation of a Boolean function, and group the outputs in convenient blocks of size 4.

The purpose of this paper is to show how ideas from the theory of Boolean functions can be used to prove results concerning the TM sequence. The first part of the paper is devoted to a new proof of a conjecture of L. Moser [5] (see also [6]). To state this, we define for integers $m > 0$ and i,

$$S_{m,i}(n) = \sum_{\substack{0 \leq j \leq n, \\ j \equiv i \pmod{m}}} (-1)^{t(j)}. \tag{3}$$

Now, the conjecture of Moser says that $S_{3,0}(n) > 0$ for any $n \geq 1$, that is, in any bounded sequence of the arithmetic progression $\equiv 0 \mod 3$ of nonnegative integer written in base-2, there are always more such integers with an even number of 1's in their binary expansion. The conjecture was first proved by D. J. Newman [5] (with a different notation from ours) using generating functions. We give a simpler proof using Boolean functions, and our method also gives some better estimates for $S_{3,0}(n)$. We use the same method to show a similar result for $p = 5$, as well.

2 An Improvement of Newman's Proof of Moser's Conjecture

We will always have $m = 3$ in (3) in this section, so we define

$$S_i(n) = S_{3,i}(n) \text{ for } i = 0, 1, 2.$$

We also define the following set $\mathcal{B}$ of 4-bit strings:

$$\begin{aligned} \mathcal{B} = \{A = 0011,\ \bar{A} = 1100,\ B = 0101,\ \bar{B} = 1010, \\ C = 0110,\ \bar{C} = 1001,\ D = 0000,\ \bar{D} = 1111\}. \end{aligned} \tag{4}$$

These strings are needed in the following lemma, which characterizes all the Boolean functions in $n \geq 2$ variables that are affine (that is, linear in all the variables and with a constant term 0 or 1).

Lemma 1. (Folklore Lemma [7, Lemma 3.7.2]) *Any affine function $f = [t_1, \ldots, t_{2^n}]$ on n variables, $n \geq 2$, is a linear string of length 2^n made up of 4-bit blocks $I_1, \ldots, I_{2^{n-2}}$ given as follows:*

1. *The first block I_1 is one of $A, B, C, D, \bar{A}, \bar{B}, \bar{C}$ or $\bar{D}$.*
2. *The second block I_2 is I_1 or $\bar{I}_1$.*
3. *The next two blocks I_3, I_4 are I_1, I_2 or $\bar{I}_1$, $\bar{I}_2$.*

....

$n-1$. *The 2^{n-3} blocks $I_{2^{n-3}+1}, \ldots, I_{2^{n-2}}$ are $I_1, \ldots, I_{2^{n-3}}$ or $\bar{I}_1, ..., \bar{I}_{2^{n-3}}$.*

Our next lemma (which also appeared in [2]) shows that initial strings of the TM sequence can be simply characterized in terms of linear Boolean functions.

Lemma 2. *The initial segment of length 2^n, $n \geq 2$, of the TM sequence is the truth table of the Boolean function*

$$f(x_1, x_2, \ldots, x_n) = x_1 \oplus x_2 \oplus \cdots \oplus x_n,$$

defined on $\mathbb{F}_2^n$ (ordered lexicographically).

Proof. By the Folklore Lemma (or directly using (2)) it is easy to see that $x_1 \oplus \cdots \oplus x_n = C\bar{C} \cdots$, which by (2) is exactly the initial segment of length 2^n of the TM sequence. $\square$

By Lemma 2, we can write (1) as

$$T = C\bar{C}\bar{C}C\bar{C}CC\bar{C}\ldots \tag{5}$$

A simple induction argument using (5) gives Table 1. More precisely, to compute $S_0(2^k)$, we partition the multiples of 3 from 0 to $2^k - 1$ into two subsets, the odd, respectively, even parity numbers. Since the even number

Table 1: Values of $S_i(2^m - 1)$ for $m \geq 2$ (so $k \geq 1$)

i	$S_i(2^{2k} - 1)$	$S_i(2^{2k+1} - 1)$
0	$2 \cdot 3^{k-1}$	3^k
1	-3^{k-1}	-3^k
2	-3^{k-1}	0

are of the form $6n$, and $t(6n) = t(3n)$, this corresponds to $S_0(2^{k-1})$. The odd numbers are of the form $6n + 3$, and since $t(6n + 3) = 1 - t(3n + 1)$, this corresponds to $-S_1(2^{k-1})$. Therefore, we obtain $S_0(2^k) = S_0(2^{k-1}) - S_1(2^{k-1})$. In a similar way, $S_1(2^k) = -S_0(2^{k-1}) + S_2(2^{k-1})$ and $S_2(2^k) = S_1(2^{k-1}) - S_2(2^{k-1})$. By induction, we get Table 1 for all $k \geq 1$.

Our next lemma will enable the evaluation of $S_i(n)$ from the binary expansion of n, by iterative use of the formulas from Table 1.

Lemma 3. *For any positive integers k and $r \leq 2^{2k}-1$, we have for $t = 0, 1, 2$*

$$S_t(2^{2k} - 1 + r) = S_t(2^{2k} - 1) - S_{u(t)}(r - 1), \tag{6}$$

where $u(t)$ in $\{0, 1, 2\}$ is defined by $u(t) \equiv t + 2 \pmod 3$. For any positive integers k and $r \leq 2^{2k+1} - 1$, we have for $t = 0, 1, 2$

$$S_t(2^{2k+1} - 1 + r) = S_t(2^{2k+1} - 1) - S_{v(t)}(r - 1), \tag{7}$$

where $v(t)$ in $\{0, 1, 2\}$ is defined by $v(t) \equiv t + 1 \pmod 3$.

Proof. To establish (6), let $T(2^{2k+1})$ denote the first 2^{2k+1} elements $t(n)$, $0 \leq n \leq 2^{2k+1} - 1$, and let T_R and T_L denote the right and left halves of the string $T(2^{2k+1})$, respectively. By Lemmas 1 and 2, we have $T_R = \overline{T_L}$ and plainly

$$S_t(2^{2k} - 1 + r) = \sum_{\substack{t(j) \in T_L, \\ j \equiv t \pmod 3}} (-1)^{t(j)} + \sum_{\substack{t(j) \in T_R, j \leq 2^{2k}-1+r, \\ j \equiv t \pmod 3}} (-1)^{t(j)}. \tag{8}$$

The first sum on the right-hand side of (8) is $S_t(2^{2k}-1)$ and since the number of entries in T_L is $\equiv 1 \pmod 3$, it follows from the definition (3), using $T_R = \overline{T_L}$, that the second sum on the right-hand side of (8) is $-S_{u(t)}(r - 1)$. This proves (6).

An analogous argument proves (7); in this case the number of entries in T_L is $\equiv 2 \pmod 3$. $\square$

Now we turn to giving a lower bound for $S_0(n)$. Suppose the binary expansion of n is

$$n = 2^{k(1)} + 2^{k(2)} + \cdots + 2^{k(j)}, k(1) > k(2) > \ldots > k(j) \geq 0. \tag{9}$$

Then by Lemma 3

$$S_0(n) = S_0(2^{k(1)} - 1) - S_{p(2)}(2^{k(2)} - 1) + \cdots + (-1)^{j-1} S_{p(j)}(2^{k(j)} - 1), \tag{10}$$

where the subscripts $p(i)$, $2 \leq i \leq j$ are determined by (6) or (7).

Suppose the binary expansion of n is given by (9), so we can expand $S_0(n)$ in the form (10). Now we can follow Newman's argument in our notation. From Table 1 we find that the first two terms in (10) satisfy

$$S_0(2^{k(1)} - 1) \geq 3^{(k(1)-1)/2} \text{ and } S_{p(2)}(2^{k(2)} - 1) \leq 0 \tag{11}$$

and that all of the other terms satisfy

$$|S_{p(i)}(2^{k(i)} - 1)| \leq (2/3)3^{k(i)/2}. \tag{12}$$

Using (11) and (12) in (10) gives

$$\begin{aligned} S_0(n) &\geq 3^{(k(1)-1)/2} - \frac{2}{3}\sum_{i=3}^{j} 3^{k(i)/2} \geq 3^{(k(1)-1)/2}\left(1 - \frac{2}{3}\sum_{i=1}^{\infty}(\sqrt{3})^{-i}\right) \\ &= 3^{(k(1)-1)/2}\left(1 - \frac{2}{3}(\sqrt{3}-1)^{-1}\right) > 3^{k(1)/2}/20 > (n/3)^{\alpha}/20, \end{aligned}$$

where $\alpha = \log 3/\log 4$, thus obtaining Newman's bound.

However, from (10) we easily get a stronger lower bound than Newman's:

Theorem 1. *For any positive integer n, we have $S_0(n) > 0$, and in fact $S_0(n) \geq \frac{\sqrt{3}}{9} n^{\alpha}$, where $\alpha = \log 3/\log 4 = .79\ldots$.*

Proof. First suppose the binary expansion of n is given by (9) with $k(1) = 2m$ for some integer m. Then, from Table 1 and Lemma 3, the first two terms in (10) are are bounded below by $2 \cdot 3^{m-1}$ and 0. Now it follows trivially from Table 1 that for $m \geq 2$ (note we need to consider the cases j odd and j even separately, but the trivial lower bounds are the same) we have

$$j \leq 2m \Longrightarrow S_i(2^j - 1) \geq -3^{m-1} \tag{13}$$

and

$$j \leq 2m - 1 \Longrightarrow -S_i(2^j - 1) \geq -3^{m-1} \tag{14}$$

Using these bounds for the remaining terms in (10), we obtain

$$\begin{aligned} S_0(n) &\geq 2\cdot 3^{m-1} - 3^{m-2} - 3^{m-2} - 3^{m-3} - 3^{m-3}\cdots \\ &= 2\cdot 3^{m-1} - 3^{m-1} + 1 > 3^{m-1} = 2^{2m\alpha}/3 \\ &> n^{\alpha}/(3\cdot 2^{\alpha}) = (\sqrt{3}/9)n^{\alpha} = .19\ldots n^{\alpha}, \end{aligned}$$

which is stronger than Newman's bound.

If the binary expansion of n is given by (9) with $k(1) = 2m+1$, then the estimate along the same lines as above is

$$\begin{aligned} S_0(n) &\geq 3^m + 3^{m-1} - 3^{m-1} - 3^{m-2} - 3^{m-2}\cdots \\ &= 3^m - 3^{m-1} + 1 > (2/3)3^m = (2/3)2^{2m\alpha} \\ &> 2n^{\alpha}/(3\cdot 4^{\alpha}) = (2/9)n^{\alpha} = .22\ldots n^{\alpha}, \end{aligned}$$

which is even better. □

The exact lim inf of $S_0(n)/n^{\alpha}$ was calculated by Coquet [1, Theorem 1]. It is $2\sqrt{3}/3^{\alpha+1} = .48\ldots$. He also found the exact lim sup, which is $55/(3\cdot 65^{\alpha}) = .67\ldots$.

3 The case $p = 5$

Some other primes p also satisfy $S_{p,0}(n) > 0$ for any $n \geq 1$, or more generally $S_{p,0}(n) > 0$ for all but finitely many $n \geq 1$. The complete list of all primes $p < 1000$ with the latter property was obtained by Drmota and Skałba [3]; the list is $3, 5, 17, 43, 257, 463$. It is no surprise that extensive calculations were required to prove this.

Our Boolean functions method can be used to obtain such results for individual primes p, and here we explain the case $p = 5$. The method for any p imitates the proof we already gave for $p = 3$, but of course the complexity of the calculations increase as p increases. We should mention here that the result for $p = 5$ was proven by Grabner [4] who showed that

$$S_{5,0}(n)' = \sum_{\substack{0\leq j<n,\\ j\equiv 0 \pmod 5}} (-1)^{t(j)} = n^{\beta}\Phi(\log_{16} n) + \frac{\eta_5(n)}{5},$$

where Φ is a continuous nowhere differentiable periodic function of period 1, $\beta = \log 5/\log 16$, and $\eta_5 = 0$ for n even, and $(-1)^{t_{5n-1}}$ for n odd. Therefore, the sum $S_{5,0}(n)' > 0$.

Table 2: Values of $S_{5,i}(2^m - 1)$ for $m \geq 3$ (so $k \geq 1$)

i	$S_{5,i}(2^{4k-1} - 1)$	$S_{5,i}(2^{4k} - 1)$	$S_{5,i}(2^{4k+1} - 1)$	$S_{5,i}(2^{4k+2} - 1)$
0	$2 \cdot 5^{k-1}$	$4 \cdot 5^{k-1}$	5^k	5^k
1	0	-5^{k-1}	-5^k	-5^k
2	$-2 \cdot 5^{k-1}$	-5^{k-1}	0	-5^k
3	5^{k-1}	-5^{k-1}	0	5^k
4	-5^{k-1}	-5^{k-1}	0	0

An induction argument analogous to the one used to obtain Table 1 gives Table 2. The only difference is that we must run the induction over four columns (in general, we would expect $p - 1$ columns) instead of 2. We omit the routine details.

Our next lemma enables the evaluation of $S_{5,0}(n)$ from the binary expansion of n, by iterative use of the formulas from Table 2.

Lemma 4. *For each i, $1 \leq i \leq 4$, and for any positive integers k and $r \leq 2^{4k-2+i} - 1$, we have for $0 \leq t \leq 4$*

$$S_{5,t}(2^{4k-2+i} - 1 + r) = S_{5,t}(2^{4k-2+i} - 1) - S_{5,u_i(t)}(r - 1), 1 \leq i \leq 4, \quad (15)$$

where the values of $u_i(t) \in \{0, 1, 2, 3, 4\}$ are given by

$$u_i(t) \equiv t + 2^i \pmod 5, \ 1 \leq i \leq 4.$$

Proof. The argument is exactly like the proof of Lemma 3, but here we need to consider four cases instead of two, corresponding to the four columns in Table 2. We omit the details. $\square$

Following the argument in the proof of Theorem 1, we obtain:

Theorem 2. *For any positive integer n, we have $S_{5,0}(n) > cn^\beta$, where $\beta = \log 5/\log 16 = .58\ldots$ and $c > 0$ is an absolute constant (one can take $c = 0.066$).*

The constant c is very far from the best possible one, so we omit the tedious computations needed to get it. We are grateful to Ms. Thanh Nguyen for carrying out these computations.

References

[1] J. Coquet, "A summation formula related to the binary digits", *Invent. Math.* **73** (1983), 107–115.

[2] T.W. Cusick, H. Fredricksen, P. Stănică, "On the delta sequence of the Thue-Morse sequence", *Austr. J. Combin.* **39** (2007), 293–300.

[3] M. Drmota, M. Skałba, "Rarified sums of the Thue-Morse sequence", *Trans. Amer. Math. Soc.* **352** (2000), 609–642.

[4] P. Grabner, "A note on the parity of the sum-of-digits function", Séminaire Lotharingien de Combinatoire (Gerolfingen, 1993), 35–42, *Prépubl. Inst. Rech. Math. Av.*, 1993/34, Univ. Louis Pasteur, Strasbourg, 1993.

[5] D. J. Newman, "On the number of binary digits in a multiple of three", *Proc. Amer. Math. Soc.* **21** (1969), 719–721.

[6] D.J. Newman, M. Slater, "Binary digit distribution over naturally defined sequences", Trans. Amer. Math. Soc. **213** (1975), 71-78.

[7] P. Stănică, "Chromos, Boolean Functions and Avalanche Characteristics", Ph.D. Thesis, State University of New York at Buffalo, 1998.