FRACTAL DIGITAL SUMS AND CODES

Peter J. Grabner†, Tamás Herendi‡ and Robert F. Tichy†

†Institut für Mathematik
Technische Universität Graz
Steyrergasse 30
8010 Graz, Austria

‡Department of Mathematics
Kossuth Lajos University
Egyetem tér 1
4010 Debrecen, Hungary

Abstract. The positivity of a special digital sum is proved and its fractal nature is discussed. The result is interpreted in terms of running digital sums of a special code.

1. Introduction

In the area of the average case analysis of algorithms digital sums often play an important rôle. For instance, they can be used to analyze register allocation strategies, or, equivalently, the order of random channel networks (cf. [5]). Another application of digital sums appears in the analysis of Batcher's odd-even merge (cf. [4]). Further applications of digital sums are known in coding theory: the so-called running digital sum and run length of codes (cf. [11]).

Of course, these sums are also of number-theoretic interest. A survey on the asymptotic analysis of digital sums by means of the Mellin transform is given in [3]. One classical problem in that area is the study of sums of the type

$$S_p(N) = \sum_{n<N} (-1)^{\nu(pn)},$$

where $\nu(k)$ is the sum of the binary digits of k and p is a given odd number. We note here that $(-1)^{\nu(n)}$ is the classical Thue-Morse sequence. The first non-trivial case $p = 3$ was investigated by Newman [10] and Coquet [1], where they prove an old conjecture of Moser stating the positivity of $S_3(N)$. Furthermore, Coquet proved:

$$S_3(N) = N^{\log_4 3} F(\log_4 N) + \frac{\eta(N)}{3},$$

1991 *Mathematics Subject Classification.* Primary 11T23, 11Y55 Secondary 94B35, 11A63.
Key words and phrases. Thue-Morse sequence, digital expansions, exponential sums, codes.
†These authors were supported by the Austrian National Bank, project Nr. 4995

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

where F is a continuous, nowhere differentiable function of period 1 and $\eta(N)$ only attains the values $0, \pm 1$.

The case of arbitrary prime numbers p is studied in [6], where a similar fractal summation formula is given for the general case. An extension of Moser's problem to $p = 5$ was studied in [8]. Furthermore this was extended to numbers $p = 3^r 5^s$; the methods used there were developed in [7]. Recently Drmota and Skałba [2] proved that if 2 is a primitive root modulo p, then $p = 3$ and $p = 5$ are the only primes, such that $S_p(N)$ is always positive. If 2 generates the squares modulo p the only two primes, such that the asymptotic main term of $S_p(N)$ may be positive are 17 and 41. In Section 2 of the present paper we solve this problem in the case $p = 17$ and prove that $S_{17}(N) > 0$ for all N. Numerical experiments up to $N = 2^{20}$ show that $S_{41}(N) > 0$ for $N > 6903$. We remark here that the proof for $S_{41}(N) > 0$ for $N > N_0$ along the same lines as the proof for $S_{17}(N) > 0$ below would involve a $2^{20} \times 41$-matrix.

In Section 3 we analyze a special code, which is defined via digital expressions of the above type.

2. Digital Sums

We want to investigate

$$T(N) = S_{17}(N) = \sum_{n<N} (-1)^{\nu(17n)}$$

and prove

Theorem 1. *The function $T(N)$ satisfies*

$$T(N) = N^{\alpha} \Phi(\log_{256} N) + R_N \tag{2.1}$$

with a continuous nowhere differentiable periodic function Φ of period 1, $\alpha = \frac{\log(17+4\sqrt{17})}{\log 256}$. The function Φ is always greater than 0.08 and $|R_N| < 1.97$; $T(N)$ is always positive.

Proof. Let $\zeta_k = \exp(\frac{2k\pi i}{17})$ for $k = 0, \dots, 16$. Then it is an immediate consequence of $256^n \equiv 1 \mod 17$ that

$$g_k(n) = \zeta_k^n (-1)^{\nu(n)} \tag{2.2}$$

satisfies

$$g_k(256n + b) = g_k(n) g_k(b) \quad \text{for } 0 \le b \le 255.$$

This property is called "complete 256-multiplicativity". Thus we obtain

$$g_k\left(\sum_{l=0}^{L} a_l \cdot 256^l\right) = \prod_{l=0}^{L} g_k(a_l). \tag{2.3}$$

The value of $g_k(n)$ only depends on the 256-ary expansion of n.

Setting $G_k(M) = \sum_{n<M} g_k(n)$ we have

$$T(N) = \frac{1}{17} G_0(17N) + \frac{1}{17} \sum_{k \in \langle 2 \rangle} G_k(17N) + \frac{1}{17} \sum_{k \in 3\langle 2 \rangle} G_k(17N), \tag{2.4}$$

where $\langle 2\rangle$ denotes the subgroup of the multiplicative residue group generated by 2. We split up the summation as above, since the second term contributes the asymptotic main term, the evaluation of the first is trivial ($G_0(N) = 0, \pm 1$) and third one is $O(1)$.

We will now investigate the asymptotic behaviour of $G_k(M)$, $k \in \langle 2\rangle$:
Let $M = \sum_{l=0}^{L} a_l \cdot 256^l$ be the 256-adic expansion of M and set $M_p = \sum_{l=p}^{L} a_l \cdot 256^l$. Then we have
(2.5)

$$G_k(M) = \sum_{n<M_L} g_k(n) + \sum_{p=0}^{L-1} \sum_{n=M_{p+1}}^{M_p-1} g_k(n) = G_k(a_L \cdot 256^L) + \sum_{p=0}^{L} g_k(M_{p+1}) G_k(a_p \cdot 256^p).$$

Thus we have reduced the problem to the computation of $G_k(a \cdot 256^l)$:

$$G_k(a \cdot 256^l) = \sum_{\varepsilon<a} g_k(\varepsilon) G_k(256^l) = G_k(a) G_k(256)^l.$$

Notice that

$$G_k(256) = \sum_{n=0}^{255} \zeta_k^n (-1)^{\nu(n)} = \prod_{l=0}^{7} \left(1 - \zeta_k^{2^l}\right) = \begin{cases} 17 + 4\sqrt{17} & \text{for } k \in 3\langle 2\rangle \\ 17 - 4\sqrt{17} & \text{for } k \in \langle 2\rangle. \end{cases} \tag{2.6}$$

This holds because 2 generates the squares mod 17 and 3 is a quadratic non-residue.

We rewrite (2.5)

$$G_k(M) = (17 + 4\sqrt{17})^L \sum_{p=0}^{L} (17 + 4\sqrt{17})^{p-L} G_k(a_p) \prod_{l=p+1}^{L} g_k(a_l) \tag{2.7}$$

and set

$$\varphi_k \left(\sum_{l=0}^{\infty} a_l \cdot 256^{-l} \right) = \sum_{l=0}^{\infty} \prod_{p=0}^{l-1} g_k(a_p) G_k(a_l) (17 + 4\sqrt{17})^{-l}. \tag{2.8}$$

Notice that these functions are well-defined and continuous (this is proved in a more general setting in [7]) and $\varphi_k(1) = 1$, $\varphi_k(256) = 17 + 4\sqrt{17}$.

Inserting the definition of φ_k into (2.7) yields

$$\begin{aligned} G_k(M) &= (17 + 4\sqrt{17})^{\lfloor \log_{256} M \rfloor} \varphi_k \left(\frac{M}{256^{\lfloor \log_{256} M \rfloor}} \right) \\ &= M^{\alpha} (17 + 4\sqrt{17})^{-\{\log_{256} M\}} \varphi_k \left(256^{\{\log_{256} M\}} \right), \end{aligned} \tag{2.9}$$

where $\lfloor x \rfloor$ and $\{x\}$ denote respectively the integer and the fractional part of x as usual. We set now $\psi_k(x) = \varphi_k(x) x^{-\alpha}$ for $1 \leq x \leq 256$ and observe that

$$\Psi(x) = \frac{1}{17} \sum_{k \in 3\langle 2\rangle} \psi_k(x)$$

is a continuous function which can be continued periodically (with period 1). The proof that Ψ and therefore Φ are nowhere differentiable would run along the same lines as the proof in Coquet's paper [1] and is omitted here. We do not make any use of this fact. Then we have

$$T(N)=\frac{1}{17}G_0(17N)+(17N)^\alpha\Psi(17N)+\frac{1}{17}\sum_{k\in\langle 2\rangle}G_k(17N)$$

and

$$\Phi(y)=(17+4\sqrt{17})^\alpha\Psi(17\cdot 256^y).$$

In order to estimate Φ from below we derive an explicit formula for $\varphi(x)=\frac{1}{17}\sum_{k\in 3\langle 2\rangle}\varphi_k(x)$:

$$\begin{aligned}\varphi(x)&=\frac{1}{17}\sum_{k\in 3\langle 2\rangle}\sum_{\ell=0}^{\infty}\prod_{p=0}^{\ell-1}g_k(a_p)G_k(a_\ell)(17+4\sqrt{17})^{-\ell}\\&=\frac{1}{17}\sum_{\ell=0}^{\infty}(17+4\sqrt{17})^{-\ell}(-1)^{\nu(a_0)+\cdots+\nu(a_{\ell-1})}\sum_{k\in 3\langle 2\rangle}\zeta_k^{a_0+\cdots+a_{\ell-1}}G_k(a_\ell).\end{aligned}$$

We introduce some notations:

$$\beta(r,s)=\frac{1}{17}\sum_{k\in 3\langle 2\rangle}\zeta_k^r G_k(s)$$

$$A_\ell(x)=\sum_{p<\ell}a_p(x),\quad B_\ell(x)=\sum_{p<\ell}\nu(a_p)\quad\text{for } x=\sum_{p=0}^{\infty}\frac{a_p}{256^p}, 0\le a_p\le 255.$$

Then we can rewrite

$$\varphi(x)=\sum_{\ell=0}^{\infty}(-1)^{B_\ell(x)}\frac{\beta(A_\ell(x)\mod 17,a_\ell)}{(17+4\sqrt{17})^\ell}.$$

Clearly, the values of $\varphi(x)$ are determined by the entries of the matrix $\beta(r,s)$, $r=0,\dots,16$, $s=0,\dots,255$. Numerical computations with MAPLE show that the maximal entry of this matrix is $\beta(0,255)=128+32\sqrt{17}$. Furthermore the minimal entry in the first column is $\beta(0,1)=8$. In order to give a lower bound for $\varphi(x)$ we compute the minimal value of

$$\beta(0,a_0)+(-1)^{\nu(a_0)}\frac{\beta(a_0\mod 17,a_1)}{17+4\sqrt{17}}-\beta(0,255)\sum_{\ell=2}^{\infty}(17+4\sqrt{17})^{-\ell},$$

which is attained for $a_0=1$ and $a_1=11$, $(\beta(1,11)=-\frac{1}{2}+\frac{3}{2}\sqrt{17})$. This yields the lower bounds

$$\begin{aligned}&\varphi(x)\ge\frac{13}{34}+\frac{9}{578}\sqrt{17},\\ &\Phi(x)\ge 17^\alpha\left(\frac{13}{34}+\frac{9}{578}\right)\frac{1}{17+4\sqrt{17}}>0.08.\end{aligned}\tag{2.10}$$

Starting from (2.5) and performing the same calculations as above, we derive

$$\begin{aligned}R_N &= \bar{\beta}(0,a_L)(17-4\sqrt{17})^L\\ &\quad+\sum_{p=0}^{L}(-1)^{\nu(a_{p+1})+\cdots+\nu(a_L)}\bar{\beta}(a_{p+1}+\cdots+a_L \bmod 17, a_p)(17-4\sqrt{17})^p\end{aligned}$$

for $N=\sum_{p=0}^{L}a_p\cdot 256^p$, where $\bar{\beta}$ is the conjugate of β in the field $\mathbb{Q}(\sqrt{17})$. Again numerical computations with MAPLE provide the entry of $\bar{\beta}$ of maximal modulus: $\bar{\beta}(10,198)=17-8\sqrt{17}$. This yields an estimate for $|R_N|$:

$$|R_N|\le\frac{1}{17}\left(\left(8\sqrt{17}-17\right)\sum_{\ell=0}^{\infty}(17-4\sqrt{17})^{\ell}+1\right)=\frac{18}{17}+\frac{15}{68}\sqrt{17}<1.97. \tag{2.11}$$

Combining (2.10) and (2.11) implies the lower bound

$$T(N)\ge 0.08\cdot N^{\alpha}-1.97.$$

From this we obtain $T(N)>0$ for $N\ge 158$. Checking the remaining values $N=1,\dots,157$ directly with MAPLE yields $T(N)>0$ for all N. This completes the proof of Theorem 1. □

Remark 1. *The graph of the function Φ is of "fractal nature". As the following picture and numerical experiments show, the range of the function Φ is between* 1.105 *and* 2.892*, which means that our lower bound is quite weak. For more details on functions of that type see* [3] *and* [8].

The graph of Φ

3. Encoding with Digital Sums

As an application, we can use the sum-of-digits function for generating a code. As above, let $\nu(n) = \sum_{i=0}^{L} b_i$ be the binary sum-of-digits function, where $n = \sum_{i=0}^{L} b_i 2^i$.

Let

$$A = A(L,k) = \{a(p) = (\nu(p), \nu(2p), \dots, \nu(k \cdot p)) | p \in 1, 3, 5, \dots, 2^L - 1\}$$

be the language of the code, where $L, k \in \mathbb{N}$. We will show that if k is big enough, then the code is uniquely decodable.

As usual, we define the running digital sum (RDS) and runlength (RL) for A. For a codeword $a = a(p) = (a_1, a_2, \dots, a_k)$ its RDS at time t $1 \leq t \leq k$ is given by $S_t(a) = \left|\sum_{j=1}^{t} (-1)^{a_j}\right|$. Furthermore define $\rho(a) = \max\limits_{1 \leq t \leq k} S_t(a)$ and the RDS $\rho = \max\limits_a \rho(a)$. The maximum number of consecutive identical values in a is denoted by $\kappa(a)$. Then the RL is given by $\kappa = \max\limits_a \kappa(a)$. For more details concerning RL and RDS we refer to [9], [12] and [11] .

Remark 2. *As it is proved in* [1] *for* $p = 3$ *or in* [8] *for* $p = 5$*, or in Section 1 for* $p = 17$*, there exists some number* p *for which* $\rho(a(p)) > c \cdot k^{\alpha}$*, with some* $0 < \alpha < 1$ *and* $c > 0$*. Thus an upper bound* ρ_0 *for the RDS implies an upper bound also for the code length* k*.*

Remark 3. *If* $p = \sum_{i=0}^{l} 2^i$ *and* $k \leq 2^{l+1}$ *, then* $\kappa(a(p)) = k$ *. (It is easy to prove that if* $1 \leq k \leq 2^{l+1}$ *then* $\nu(p \cdot k) = l+1$*, and* $\nu(p \cdot (2^{l+1}+1)) = 2l+2$*.) This means that* $\kappa = k$ *if* $k \leq 2^{l+1}$*. One can also prove that if* $p \in \mathbb{N}$ *and* $l = \lfloor \log(p) \rfloor$ *, then among any* $2^{l+1}+1$ *consecutive elements of* $a(p)$ *there exist two different ones.*

Proof. Let $p \in \mathbb{N}$ and let $(a_1, a_2, \dots, a_k)$ be its codeword $a(p)$. Let furthermore $1 \leq t \leq k$ and let us consider the subword $(a_t, a_{t+1}, \dots, a_{t+2^{l+1}})$. By the Euler-Fermat theorem there exists a number u such that $p \cdot u = 1 \mod 2^{l+1}$. Let $q = \min\{ i \,| t \leq i \leq t + 2^{l+1}$ and $i = u \cdot (-p-1) \mod 2^{l+1}\}$. Then $t \leq q < t + 2^{l+1}$ and thus q and $q+1$ are also contained in the above subword. But then $\nu((q+1) \cdot p) = \nu(q \cdot p) + \nu(p) > \nu(q \cdot p)$, which proves the statement. □

Let $p = \sum_{i=0}^{L} b_i 2^i$, $k \geq 2^{L+1}+1$ and let $a(p) = (\nu(p), \nu(2p), \dots, \nu(k \cdot p))$ be the corresponding codeword. Suppose we know k and $a(p)$. Define further $\mathrm{inv}_j(b_0, b_1, \dots, b_j) = (c_0, c_1, \dots, c_j)$ such that

$$1 \equiv \left(\sum_{i=0}^{j} b_i 2^i\right) \left(\sum_{i=0}^{j} c_i 2^i\right) \mod 2^{j+1}.$$

Then we can define the following (decoding) algorithm:

Decoding Algorithm.

```
1. Let $l = \max\{ i \mid 0 \leq i \leq L$ and $\nu(p \cdot (2^i + 1)) < 2\nu(p)\}$;
2. For $i > l$ Let $b_i = 0$;
```

```
3.      Let $b_0 = b_l = 1$;
4.      For $i$ from $0$ to $\lfloor l/2 \rfloor - 1$ do
4.1.        Let $(c_0, c_1, \dots, c_i) = \mathrm{inv}_i(b_0, b_1, \dots, b_i)$;
4.2.        Let $c = \sum_{j=0}^{i} c_j 2^j$;
            Let $d = c + 2^{i+1}$;
            Let$\eta = \nu(p) + \nu(p \cdot c) - \nu(p + 2^{i+1} p \cdot c)$;
            $\vartheta = \nu(p) + \nu(p \cdot d) - \nu(p + 2^{i+1} p \cdot d)$;
4.3.        If $\eta < \vartheta$ then
4.3.1.          Let $c_{i+1} = 0$;
4.3.2.          Let $(b_0, b_1, \dots, b_{i+1}) = \mathrm{inv}_{i+1}(c_0, c_1, \dots, c_{i+1})$;
4.3.3.          If $\eta = 0$ then $b_{i+1} = 0$ else $b_{i+1} = 1$;
4.4.        If $\eta > \vartheta$ then
4.4.1.          Let $c_{i+1} = 1$;
4.4.2.          Let $(b_0, b_1, \dots, b_{i+1}) = \mathrm{inv}_{i+1}(c_0, c_1, \dots, c_{i+1})$;
4.4.3.          If $\vartheta = 0$ then $b_{i+1} = 0$ else $b_{i+1} = 1$;
4.5.        If $\eta = \vartheta$ then
4.5.1.          Let $b_{l-i-1} = 1$
4.5.2.          If $\nu\left(\sum_{j=0}^{i} b_j 2^j\right) + \nu\left(\sum_{j=0}^{i} 2^j\right) - \nu\left(\sum_{j=0}^{i} b_j 2^j + \sum_{j=0}^{i} 2^j\right) =$
            $2\nu(p) - \nu\left(p + 2^{l-i+1}\right)$ then Let $b_{i+1} = 0$ else Let $b_{i+1} = 1$.
```

Theorem 2. *Let $a(p)$ be an error free codeword and let b_i $(0 \le i \le L)$ be obtained by the above Decoding Algorithm. Then $p = \sum_{i=0}^{L} b_i 2^i$.*

Proof. Steps 1,2 and 3 give the correct l because if $l < L$ then $\nu(p \cdot (2^{l+1} + 1)) = 2\nu(p)$ and $\nu(p \cdot (2^l + 1)) < 2\nu(p)$. (This is because in the addition $p + p \cdot 2^{l+1}$ there are no carries and so there are no “lost digits”, while in the addition $p + p \cdot 2^l$ there is at least one carry and so there is a “lost digit”.)

Suppose now, that $b_0, b_1, \dots, b_i$ and $b_{l-i}, b_{l-i+1}, \dots, b_l$ are given. Then by the Euler-Fermat theorem there exist uniquely determined $c_0, c_1, \dots, c_i$ such that

$$\sum_{j=0}^{m} d_j 2^j = \left(\sum_{j=0}^{i} b_j 2^j\right)\left(\sum_{j=0}^{i} c_j 2^j\right)$$

with some $m \in \mathbb{N}$ and $d_j \in \{0, 1\}$, $0 \le j \le m$, and $d_0 = 1, d_1 = d_2 = \dots = d_i = 0$.

Let $c = \sum_{j=0}^{i} c_j 2^j$. If $d = \sum_{j=0}^{m} d_j 2^j = c \cdot p$ then $d_0 = 1$ and $d_1 = d_2 = \dots = d_i = 0$.

Furthermore, if $e = \sum_{j=0}^{m'} e_j 2^j = (c + 2^{i+1}) \cdot p$, then $d_{i+1} + e_{i+1} = 1$ and $e_j = f_j$ if $0 \le j \le i$.

Computing the values $\eta = \nu(p) + \nu(d) - \nu(p + 2^{i+1} \cdot d)$ and $\vartheta = \nu(p) + \nu(e) - \nu(p + 2^i \cdot e)$ yields either (a) $\eta < \vartheta$ (b) $\vartheta < \eta$ or (c) $\eta = \vartheta$. Here η (and ϑ) represents the numbers of carries in the additions $p + 2^{i+1} \cdot e$ (and $p + 2^{i+1} \cdot f$). In case (a) we have $f_{i+1} = 1$ since there are more carries. Similarly, in case (b) we obtain $e_{i+1} = 1$. Hence the value of b_{i+1} can be computed by the function inv_{i+1}. Furthermore, in case (a) we have $b_{l-i-1} = 0$ for $\eta = 0$ and $b_{l-i-1} = 1$, otherwise. Case (b) can be handled similarly. Case (c) appears if and only if $b_{l-i-1} = b_{l-i} = \dots = b_l = 1$. But then let $\gamma = \nu\left(\sum_{j=0}^{i} b_j 2^j\right) + \nu\left(\sum_{j=0}^{i} 2^j\right) -$

$\nu\left(\sum_{j=0}^{i} b_j 2^j + \sum_{j=0}^{i} 2^j\right)$. Since the $i+1$-st digit of the last sum of the expression has value 1, $\gamma = 2\nu(p) - \nu\left(p + p\cdot 2^{l-i+1}\right)$ if and only if $b_{i+1} = 0$.

By repeating this step we can compute p . This implies that the coding system is uniquely decodable. □

Remark 4. *From the digital construction it is clear that the length of the codewords is exponentially growing with the length of the original words. Thus the code is exponential in contrast to many well-known and standard other codes.*

Remark 5. *The code defined above is extremely redundant, because all parts of the codeword depend on the original word.*

Remark 6. *Because of the previous remarks the above code is strongly error correcting, and the decoding procedure is linear.*

Acknowledgement. The authors are indebted to Brian Ward for preparing the numerical experiments which were necessary for this work.

References

[1] J. Coquet, *A Summation Formula Related to the Binary Digits*, Invent. math. **73** (1983), 107–115.

[2] M. Drmota and M. Skałba, *Sign-changes of the Thue-Morse Fractal Function and Dirichlet L-Series*, manuscripta math. **86** (1995), 519–541.

[3] P. Flajolet, P.J. Grabner, P. Kirschenhofer, H. Prodinger and R.F. Tichy, *Mellin transforms and asymptotics: digital sums*, Theoret. Comput. Sci. **123** (1994), 291–314.

[4] P. Flajolet and L. Ramshaw, *A note on Gray code and odd-even merge*, SIAM J. Comput. **9** (1980), 142–158.

[5] P. Flajolet, J.C. Raoult and J. Vuillemin, *The number of registers required for evaluating arithmetic expressions*, Theoret. Comput. Sci. **9** (1979), 99–125.

[6] S. Goldstein, K.A. Kelly and E.R. Speer, *The fractal structure of rarefied sums of the Thue-Morse sequence*, J. Number Th. **42** (1992), 1–19.

[7] P.J. Grabner, *Completely q-Multiplicative Functions: the Mellin Transform Approach*, Acta Arith. **65** (1993), 85–96.

[8] P.J. Grabner, *A Note on the Parity of the Sum-of-Digits Function*, Actes 30$^{\text{e}}$ Séminaire Lotharingien, 1993, pp. 35–42.

[9] S.Litsyn and A.Tietäväinen, *Character Sum Constructions of Constrained Error-Correcting Codes*, Appl. Algebra in Eng., Comm. and Comp. **5** (1994), 45–51.

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

[11] K.A.Schouhamer-Immink, *Coding Techniques for Digital Recorders*, Englewood Cliffs: Prentice-Hall, 1991.

[12] IEEE Trans.Inform. Theory, Special Issue on Coding for Storage Devices, Part 2 (May 1991).