THE SUM OF DIGITS FUNCTION OF SQUARES

MICHAEL DRMOTA* AND JOËL RIVAT**

Abstract. We consider the set of squares n^2, $n < 2^k$, and split up the sum of binary digits $s(n^2)$ into two parts $s_{[<k]}(n^2) + s_{[\ge k]}(n^2)$, where $s_{[<k]}(n^2) = s(n^2 \bmod 2^k)$ collects the first k digits and $s_{[\ge k]}(n^2) = s(\lfloor n^2/2^k \rfloor)$ collects the remaining digits. We present very precise results on the distribution on $s_{[<k]}(n^2)$ and $s_{[\ge k]}(n^2)$. For example, we provide asymptotic formulas for the numbers $\#\{n < 2^k : s_{[<k]}(n^2) = m\}$ and $\#\{n < 2^k : s_{[\ge k]}(n^2) = m\}$ and show that these partial sum of digits functions are asymptotically equidistributed in residue classes. These results are motivated by a conjecture by Gelfond [11] saying that the (total) sum of digits function $s(n^2)$ is asymptotically equidistributed in residue classes.

1. Introduction

Let $s(n)$ denote the binary sum of digits function, that is,

$$s(n) = \sum_{j \ge 0} \varepsilon_j(n),$$

where $\varepsilon_j(n) \in \{0, 1\}$ $(j \ge 0)$ are the digits in the binary digital expansion

$$n = \sum_{j \ge 0} \varepsilon_j(n) 2^j$$

of $n \ge 0$.

The main purpose of this paper is to analyze the distribution of the sum of digits function of squares $s(n^2)$. There are some known facts, for example, Peter [17] has proved that

$$\frac{1}{N} \sum_{n<N} s(n^2) = \log_2 N + \gamma(\log_2 N) + O(N^{-\eta}) \qquad (1.1)$$

where $\log_2 N = (\log N)/(\log 2)$, γ is a continuous periodic function and $\eta > 0$. Furthermore, Bassily and Kátai [1] studied the distribution of q-additive functions on polynomial sequences $P(n)$. In particular for $s(n^2)$ one gets

$$\frac{1}{N} \# \left\{ n < N : s(n^2) \le \log_2 N + y \sqrt{\frac{1}{2} \log_2 N} \right\} = \Phi(y) + o(1), \qquad (1.2)$$

where $\Phi(y)$ denotes the normal distribution function.

These results show that $s(n^2)$, $n < N$, behaves (asymptotically) like the sum of $2 \log_2 N$ independent random variables X_j (the digits) with $\mathbf{Pr}\{X_j = 0\} = \mathbf{Pr}\{X_j = 1\} = 1/2$. However, (1.1) and (1.2) give only information on the overall distribution for $s(n^2)$ $(n < N)$ and they do not provide asymptotic relations for the numbers $\#\{n < N : s(n^2) = m\}$. In particular it is an open problem (see Gelfond

Date: May 8, 2004.

1991 *Mathematics Subject Classification.* Primary: 11A63, Secondary: 11N60.

[1] This research was supported by the Austrian Science Foundation FWF, grant S8302-MAT.

*Institute of Discrete Mathematics and Geometry, Technische Universität Wien, Wiedner Hauptstraße 8-10/113, A-1040 Wien, Austria.

**Institut de Mathématiques de Luminy, CNRS UMR 6206, Université de la Méditerranée, Campus de Luminy, Case 907, 13288 MARSEILLE Cedex 9, France.

[11]) whether $\#\{n < N : s(n^2) \equiv 0 \bmod 2\} \sim N/2$ or not. Quite recently Dartyge and Tenenbaum [4] could show that

$$\#\{n < N : s(n^2) \equiv 0 \bmod 2\} \gg N \quad \text{and} \quad \#\{n < N : s(n^2) \equiv 1 \bmod 2\} \gg N.$$

It seems to be a very difficult problem to obtain precise information on the exact distribution of $s(n^2)$. In what follows we present a completely new approach to this kind of problems. We consider the set of squares n^2, $n < 2^k$, and split up the sum of binary digits $s(n^2)$ into two parts $s_{[<k]}(n^2) + s_{[\ge k]}(n^2)$, where

$$s_{[<k]}(n^2) = s(n^2 \bmod 2^k) = \sum_{j<k} \varepsilon_j(n^2)$$

collects the first k digits and

$$s_{[\ge k]}(n^2) = s\left(\left\lfloor \frac{n^2}{2^k} \right\rfloor\right) = \sum_{j\ge k} \varepsilon_j(n^2)$$

collects the remaining digits. Interestingly, we obtain very precise results on the distribution on $s_{[<k]}(n^2)$ and $s_{[\ge k]}(n^2)$. For example, we provide asymptotic formulas for the numbers $\#\{n < 2^k : s_{[<k]}(n^2) = m\}$ and $\#\{n < 2^k : s_{[<k]}(n^2) = m\}$ and show that these partial sum of digits functions are asymptotically equidistributed in residue classes. Unfortunately these results cannot be applied to obtain corresponding results for $s(n^2) = s_{[<k]}(n^2) + s_{[\ge k]}(n^2)$ (we only get upper bounds). Our methods rely on generating functions and estimates on multivariate exponential sums.

In section 2 we collect and discuss the main results of this paper. In section 3 and 4 we prove Theorem 1, in particular we derive representations for the generating function of $s_{[<k]}(n^2)$ and $s_{[\ge k]}(n^2)$ respectively. In section 5 we indicate how Theorems 2 and 3 can be derived from Theorem 1. Finally, in section 6 we present some results on squares with a large sum-of-digits function.

2. Results

The basic result is the following one.

Theorem 1. *Let* $s_{[<k]}(n) := \sum_{j<k} \varepsilon_j = s(n \bmod 2^k)$ *and* $s_{[\ge k]}(n) := \sum_{j\ge k} \varepsilon_j = s(\lfloor n/2^k \rfloor)$ *denote the partial sum of digits funtions that collect the first k resp. the remaining digits ε_j of n. Then we have (for complex $x \ne \pm\sqrt{2} - 1$)*

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} = \begin{cases} \dfrac{4x(1+x)^{k-1}}{x^2+2x-1} - 2^{\frac{k}{2}} \dfrac{(1+x)^2(1-x)}{x^2+2x-1} & \text{if } k \text{ is even,} \\ \dfrac{4x(1+x)^{k-1}}{x^2+2x-1} + 2^{\frac{k-1}{2}} \dfrac{x^3+3x^2-4x-2}{x^2+2x-1} & \text{if } k \text{ is odd.} \end{cases} \quad (2.1)$$

Set $\xi_1 = 2^{5/6} - 1 = 0.78179\ldots$ *and* $\xi_2 = 1 + \sqrt{2} + \sqrt{2(1+\sqrt{2})} = 4.61158\ldots$ *and suppose that* $\xi_1 + \varepsilon \le |x| \le \xi_2 - \varepsilon$ *for some* $\varepsilon > 0$. *Then there exists* $\eta > 0$ *such that uniformly in that range*

$$\sum_{n<2^k} x^{s_{[\ge k]}(n^2)} = C(x)(1+x)^k + O\left((1+|x|)^{k(1-\eta)}\right), \quad (2.2)$$

where $C(x)$ is a continuous function (defined in (4.7)) that is also analytic in the range $|x| < |1+x|$.

Remark 1. Formula (2.2) can be stated in a litte bit more precise form, compare with Proposition 2. In particular, for $x = -1$ one gets

$$\sum_{n<2^k} (-1)^{s_{[\ge k]}(n^2)} = O\left(k^4 2^{5k/6}\right).$$

Furthermore, it is also possible to extend (2.2) to

$$\sum_{n<N} x^{s_{[\geq k]}(n^2)} = C(x, \log_2 N)(1+x)^{\log_2 N} + O\left((1+|x|)^{\log_2 N(1-\eta)}\right)$$

for k close to $\log_2 N$. Here $C(x,t)$ denotes a continuous function that is periodic in t: $C(x,t+1) = C(x,t)$. However, for the sake of shortness we just prove (2.2).

Remark 2. We also want to note that it seems that there are extensions of Theorem 1 to $s(n^d)$ for $d \geq 2$. In particular, it is possible to extend the methods of section 3 to provide an explicit representation for

$$\sum_{n<2^k} x^{s_{[<k]}(n^d)},$$

and the double large sieve methods of section 4 seem to work even for $s_{[\geq k]}(n^d) = s(\lfloor n^d/2^k \rfloor)$, at least up to some d_0.

From (2.1) and (2.2) the following distributional properties of $s_{[<k]}(n^2)$ and $s_{[\geq k]}(n^2)$ follow almost directly.

Theorem 2. *Let $\varepsilon > 0$. Then for $(1-1/\sqrt{2}+\varepsilon)k \leq m \leq (1-\varepsilon)k$ we have uniformly*

$$\#\{n < 2^k : s_{[<k]}(n^2) = m\} = \frac{4m(k-m)^2}{k(4km - 2m^2 - k^2)} \binom{k}{m} \left(1 + \mathcal{O}\left(k^{-1}\right)\right). \tag{2.3}$$

For $\varepsilon \leq m \leq (1-1/\sqrt{2}-\varepsilon)k$ we uniformly have

$$\#\{n < 2^k : s_{[<k]}(n^2) = m\} = \begin{cases} 2^{k/2}(1+\sqrt{2})^m \left(1 + \mathcal{O}\left(e^{-\eta m}\right)\right) & \text{for even } k, \\ \frac{6+\sqrt{2}}{4} 2^{k/2}(1+\sqrt{2})^m \left(1 + \mathcal{O}\left(e^{-\eta m}\right)\right) & \text{for odd } k, \end{cases} \tag{2.4}$$

and for some $\eta > 0$. Furthermore, for $(\xi_1/(1+\xi_1)+\varepsilon)k \leq m \leq (\xi_2/(1+\xi_2)-\varepsilon)k$ (where ξ_1 and ξ_2 are defined in Theorem 1) we uniformly obtain

$$\#\{n < 2^k : s_{[\geq k]}(n^2) = m\} = C\left(\frac{m}{k-m}\right) \cdot \binom{k}{m} \left(1 + \mathcal{O}\left(k^{-1}\right)\right), \tag{2.5}$$

where $C(x)$ is the continuous function of Theorem 1.

Remark 3. Note that $s_{[<k]}(n^2)$ and $s_{[\geq k]}(n^2)$ also satisfy *central limit theorems* of the forms

$$\frac{1}{2^k} \# \left\{ n < 2^k : s_{[<k]}(n^2) \leq \frac{k}{2} + y\sqrt{\frac{k}{4}} \right\} = \Phi(y) + o(1),$$

and

$$\frac{1}{2^k} \# \left\{ n < 2^k : s_{[\geq k]}(n^2) \leq \frac{k}{2} + y\sqrt{\frac{k}{4}} \right\} = \Phi(y) + o(1).$$

This follows (almost) directly from the methods of [1] and is in accordance to Theorem 2, that might be interpreted as a *local central limit theorem*. By adapting the methods of [1] we also get a joint central limit theorem of the form

$$\frac{1}{2^k} \# \left\{ n < 2^k : s_{[<k]}(n^2) \leq \frac{k}{2} + y_1\sqrt{\frac{k}{4}},\ s_{[\geq k]}(n^2) \leq \frac{k}{2} + y_2\sqrt{\frac{k}{4}} \right\}$$
$$= \Phi(y_1)\Phi(y_2) + o(1).$$

This means that in an *overall sense* $s_{[<k]}(n^2)$ and $s_{[\geq k]}(n^2)$ are asymptotically independent.

Theorem 3. *Let $M \ge 2$ be an integer. Then there exists $\eta > 0$ such that for all integers c we have*

$$\frac{1}{2^k}\cdot \#\{n < 2^k : s_{[<k]}(n^2) \equiv c \bmod M\} = \frac{1}{M} + O(2^{-\eta k}) \qquad (2.6)$$

and

$$\frac{1}{2^k}\cdot \#\{n < 2^k : s_{[\ge k]}(n^2) \equiv c \bmod M\} = \frac{1}{M} + O(2^{-\eta k}) \qquad (2.7)$$

3. Generating Function for $s_{[<k]}(n^2)$

In this section we prove the first part of Theorem 1, the explicit formula (2.1). for the generating function of $s_{[<k]}(n^2)$.

Lemma 1. *Set*

$$S_k^{(0)}(x) = \sum_{n<2^k,\, n\equiv 1 \bmod 2} x^{s_{[<k]}(n^2)}$$

and

$$S_k^{(1)}(x) = \sum_{n<2^k,\, n\equiv 1 \bmod 2} x^{s_{[<k+1]}(n^2)}.$$

Then we have $S_1^{(0)}(x) = x$, $S_2^{(0)}(x) = 2x$, and

$$S_k^{(0)}(x) = 4x(1+x)^{k-3} \quad \text{for } k \ge 3, \qquad (3.1)$$

and $S_1^{(1)}(x) = x$ and

$$S_k^{(1)}(x) = 2x(1+x)^{k-2} \quad \text{for } k \ge 2. \qquad (3.2)$$

Proof. The representations for $S_1^{(0)}(x)$, $S_2^{(0)}(x)$, and $S_1^{(1)}(x)$ are trivial. We next show that

$$S_k^{(0)}(x) = 2S_{k-1}^{(1)}(x) \quad \text{for } k \ge 2 \qquad (3.3)$$

and

$$S_k^{(1)}(x) = (1+x)S_{k-1}^{(1)}(x) \quad \text{for } k \ge 2. \qquad (3.4)$$

Obviously, (3.3) and (3.4) prove the lemma.

First, we have

$$\begin{aligned} S_k^{(0)}(x) &= \sum_{n<2^k,\, n\equiv 1 \bmod 2} x^{s(n^2) \bmod 2^k} \\ &= \sum_{n<2^{k-1},\, n\equiv 1 \bmod 2} x^{s(n^2 \bmod 2^k)} + \sum_{n<2^{k-1},\, n\equiv 1 \bmod 2} x^{s((n+2^{k-1})^2 \bmod 2^k)}. \end{aligned}$$

Since $k \ge 2$ it follows that

$$(n+2^{k-1})^2 = n^2 + 2^k n + 2^{2k-2} \equiv n^2 \bmod 2^k.$$

Hence, $S_k^{(0)}(x) = 2S_{k-1}^{(1)}(x)$, that is, we have proved (3.3).

Next, we get

$$\begin{aligned} S_k^{(1)}(x) &= \sum_{n<2^k,\, n\equiv 1 \bmod 2} x^{s(n^2 \bmod 2^{k+1})} \\ &= \sum_{n<2^{k-1},\, n\equiv 1 \bmod 2} x^{s(n^2 \bmod 2^{k+1})} + \sum_{n<2^{k-1},\, n\equiv 1 \bmod 2} x^{s((n+2^{k-1})^2 \bmod 2^{k+1})}. \end{aligned}$$

Since $n \equiv 1 \bmod 2$ and $k \ge 2$ we now have

$$(n+2^{k-1})^2 = n^2 + 2^k n + 2^{2k-2} \equiv n^2 + 2^k \bmod 2^{k+1}.$$

Hence if $\varepsilon_k(n^2) = 0$ then $s(n^2 \bmod 2^{k+1}) = s(n^2 \bmod 2^k)$ and $s((n+2^{k-1})^2 \bmod 2^{k+1}) = s(n^2 \bmod 2^k) + 1$. On the other hand, if $\varepsilon_k(n^2) = 1$ then $s(n^2 \bmod 2^{k+1}) =$

$s(n^2 \bmod 2^k)+1$ and $s((n+2^{k-1})^2 \bmod 2^{k+1}) = s(n^2 \bmod 2^k)$. So after all we get $S_k^{(1)}(x) = (1+x)S_{k-1}^{(1)}(x)$. This completes the proof of the lemma. □

Proposition 1. *Suppose that $k \ge 4$ is even. If $x \ne \pm\sqrt{2}-1$ then*

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} = \frac{4x(1+x)^{k-1}}{x^2+2x-1} - 2^{\frac{k}{2}}\frac{(1+x)^2(1-x)}{x^2+2x-1}. \tag{3.5}$$

If $x = \pm\sqrt{2}-1$ then

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} = k2^{\frac{k}{2}-1}x(1+x) + 2^{\frac{k}{2}}(1+x)(1-x). \tag{3.6}$$

If $k \ge 3$ is odd and $x \ne \pm\sqrt{2}-1$ then

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} = \frac{4x(1+x)^{k-1}}{x^2+2x-1} + 2^{\frac{k-1}{2}}\frac{x^3+3x^2-4x-2}{x^2+2x-1}. \tag{3.7}$$

and if $x = \pm\sqrt{2}-1$ then

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} = k2^{\frac{k-1}{2}}x + 2^{\frac{k-1}{2}}(1-x). \tag{3.8}$$

Proof. By splitting between odd and even $n < 2^k$ we directly get the recurrence

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} = S_k^{(0)}(x) + 2\sum_{n<2^{k-2}} x^{s_{[<k-2]}(n^2)}$$

and thus (if $k \ge 4$ is even)

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} = S_k^{(0)}(x) + 2S_{k-2}^{(0)}(x) + \cdots + 2^{\frac{k}{2}-1}S_2^{(0)}(x) + 2^{\frac{k}{2}}.$$

Hence, by using (3.1) we directly obtain (3.5) and (3.6). In the same way we derive (3.7) and (3.8). □

4. Generating Function for $s_{[\ge k]}(n^2)$

We now turn to the generating function (2.2) for $s_{[\ge k]}(n^2)$ that is much more difficult to handle than that for $s_{[<k]}(n^2)$.

Lemma 2. *Set $a_j = \#\left\{n < 2^k : \sqrt{2^k j} \le n < \sqrt{2^k(j+1)}\right\}$. Then*

$$\sum_{n<2^k} x^{s_{[\ge k]}(n^2)} = \sum_{j<2^k} a_j x^{s(j)}.$$

Proof. The proof is obvious by observing that $s_{[\ge k]}(n^2) = s(\lfloor n^2/2^k \rfloor)$. □

Note that with help of $\Psi(x) = x - \lfloor x \rfloor - \frac{1}{2}$ we can rewrite a_j to

$$\begin{aligned} a_j &= \#\left\{n < 2^k : \sqrt{2^k j} \le n < \sqrt{2^k(j+1)}\right\} \\ &= \left\lceil \sqrt{2^k(j+1)} \right\rceil - \left\lceil \sqrt{2^k j} \right\rceil \\ &= 2^{\frac{k}{2}}\left(\sqrt{j+1} - \sqrt{j}\right) + \Psi\left(-\sqrt{2^k(j+1)}\right) - \Psi\left(-\sqrt{2^k j}\right). \end{aligned}$$

Hence, we have to deal with three sums:

$$S_1 = 2^{\frac{k}{2}} \sum_{j<2^k} \left(\sqrt{j+1}-\sqrt{j}\right) x^{s(j)},$$

$$S_2 = \sum_{j<2^k} \Psi\left(-\sqrt{2^k j}\right) x^{s(j)},$$

$$S_3 = \sum_{j<2^k} \Psi\left(-\sqrt{2^k (j+1)}\right) x^{s(j)}.$$

It turns out that the first sum is easy to handle, whereas the other two require non-trivial tools from multivariate exponential sums. (We use the double large sieve by Bombieri and Iwanies [2].)

Our goal is to prove the following representation (that follows from a combination of Lemma 3, 8 and 9.)

Proposition 2. *Set $\xi_1 = 2^{5/6}-1 = 0.78179\ldots$ and $\xi_2 = 1+\sqrt{2}+\sqrt{2(1+\sqrt{2})} = 4.61158\ldots$ and suppose that $\xi_1 < x_1 \le x_2 < \xi_2$. Then for every*

$$\eta > \min\left(1-\frac{6}{5}\log(2)/\log(1+x_1), \frac{1}{2}-\left(\frac{1}{4}\log(1+x_2^2)+\frac{1}{8}\log 2\right)/\log(1+x_2)\right).$$

we have uniformly for $x_1 \le |x| \le x_2$

$$\sum_{n<2^k} x^{s_{[\ge k]}(n^2)} = C(x)(1+x)^k + O\left((1+|x|)^{k(1-\eta)}\right), \tag{4.1}$$

where $C(x)$ is a continuous function (defined in (4.7)).

4.1. **Evaluation of S_1.**

Lemma 3. *Suppose that N is represented in its digital expansion of the form $N = 2^{k_1}+2^{k_2}+\cdots+2^{k_L}$ with $k_1 > k_2 > \cdots > k_L \ge 0$ then we have*

$$\sum_{n<N} x^{s(n)} = (1+x)^{k_1} + x(1+x)^{k_2} + \cdots + x^{L-1}(1+x)^{k_L} \tag{4.2}$$

In particular, if $|x| < |1+x|$ and $|1+x| \ge 1$ this can be written as

$$\sum_{n<N} x^{s(n)} = \gamma(x, \log_2 N)\,(1+x)^{\log_2 N}, \tag{4.3}$$

where $\gamma(x,t)$, $x \in \mathbb{C}$ is analytic in x and periodic in t with peroid 1. Furthermore, $\gamma(x,t)$ is Lipschitz continuous in t of the form

$$|\gamma(x,t_1)-\gamma(x,t_2)| \ll |t_1-t_2|^{\log_2 \frac{|1+x|}{|x|}}.$$

Furthermore, we have

$$\left|\sum_{n<N} x^{s(n)}\right| \le \begin{cases} \frac{1-|1+x|^{k_1}}{1-|1+x|} & \text{if } |x+1|<1 \text{ and } |x| \le |1+x|, \\ \frac{1-|x|^{k_1}}{1-|x|} & \text{if } |x+1|<1 \text{ and } |x| > |1+x|, \\ k_1|x|^{k_1} & \text{if } |x+1| \ge 1 \text{ and } |x| > |1+x|, \end{cases} \tag{4.4}$$

where $k_1 = \lfloor \log_2 N \rfloor$ (as above).

Proof. First note that

$$\sum_{n<N} x^{s(n)} = \sum_{n<2^k} x^{s(n)} + x \sum_{n<N'} x^{s(n)},$$

if $N = 2^k + N'$ with $N' < 2^k$. Furthermore, we have

$$\sum_{n<2^k} x^{s(n)} = (1+x)^k.$$

Hence, (4.2) follows.

For $0 \le t < 1$ let the binary expansion of 2^t be written as

$$2^t = 1 + \sum_{j\ge 1} 2^{-\ell_j},$$

where $1 \le \ell_1 < \ell_2 < \cdots$, and set

$$\gamma(x,t) = (1+x)^{-t}\left(1 + \sum_{j\ge 1} x^j (1+x)^{-\ell_j}\right),$$

where we have to assume that $|x| < |1+x|$ and $|1+x| \ge 1$. Then it is easy to verify that γ is analytic in x and Lipschitz continuous in t (of the above form). Furthermore $\gamma(x,1) = \gamma(x,0) = 1$. Thus, we can extend it to a continuous periodic function in t. Finally with help of (4.2) and we also get (4.3) in a direct way.

The estimates (4.4) follow immediately from (4.2). □

Lemma 4. *There exists a continuous function $C(x)$, $x \in \mathbb{C}$ with $|x| < |1+x|$ and $|1+x| \ge 1$, such that*

$$S_1 = 2^{\frac{k}{2}} \sum_{j<2^k} \left(\sqrt{j+1} - \sqrt{j}\right) x^{s(j)} = C(x)(1+x)^k + O\left(|1+x|^{k(1-\beta)}\right) \tag{4.5}$$

uniformly for $|x| \le (1-\eta_1)|1+x|$, where $0 < \eta_1 < 1$ and $\beta = \log_2((|1+x|)/|x|) > 0$. Furthermore, if x varies in a compact set K of the complex plane that does not contain the positive real line then we have

$$S_1 \ll (1+|x|)^{(1-\eta)k} \tag{4.6}$$

uniformly for $x \in K$, where

$$\eta = \max_{x\in K}\left(\frac{\log(|1+x|)}{\log(1+|x|)}, \frac{\log(|x|)}{\log(1+|x|)}\right).$$

Proof. Set $c_j = \sqrt{j+1} - \sqrt{j}$. Then we have

$$c_j = \frac{1}{2j^{1/2}} + O\left(j^{-3/2}\right), \quad c_j - c_{j+1} = \frac{1}{4j^{3/2}} + O\left(j^{-5/2}\right),$$

and by partial summation we obtain

$$\begin{aligned} 2^{-\frac{k}{2}} S_1 &= \sum_{j<2^k} c_j\, x^{s(j)} \\ &= c_{2^k-1} \sum_{j<2^k} x^{s(j)} + \sum_{j<2^k-1} (c_j - c_{j+1}) \sum_{m\le j} x^{s(m)} \\ &= \left(2^{-\frac{k}{2}} + O\left(2^{-\frac{3k}{2}}\right)\right) \sum_{j<2^k} x^{s(j)} \\ &\quad + \sum_{j<2^k-1}\left(\frac{1}{4j^{3/2}} + O\left(j^{-5/2}\right)\right) \sum_{m\le j} x^{s(m)}. \end{aligned}$$

Now suppose that $|x| < |1+x|$ and $|1+x| \ge 1$. Then we can use (4.3) from Lemma 3 to proceed further. For this purpose we will use the (easy to derive) asymptotic formula

$$\sum_{n<N} n^\alpha \gamma(\log_2 n) = N^{\alpha+1} \int_0^1 t^\alpha \gamma(\log_2 t - \log_2 N)\, dt + O\left(N^{\max(\Re\alpha+1-\beta, \Re\alpha)}\right),$$

where α has real part $\Re\alpha > -1$ and $\gamma(t)$ is a continuous periodic function (with period 1 and Lipschitz exponent $\beta > 0$). In particular we use $\beta = \log_2((|1+$

$x|)/|x|) > 0$ and $\alpha = \log_2(1+x) - \frac{3}{2}$ that has real part $\Re\alpha = \log_2|1+x| - \frac{3}{2} \ge -\frac{1}{2}$ and obtain

$$\begin{aligned}\sum_{j<2^k-1} \frac{1}{j^{3/2}} \sum_{m\le j} x^{s(m)} &= \sum_{j<2^k-1} j^{\log_2(1+x)-\frac{3}{2}}\gamma(x,\log_2 j)\\ &= 2^{k(\alpha+1)}\int_0^1 t^\alpha \gamma(x,\log_2 t)\,dt + O\left(2^{k(\alpha+1-\beta)}\right)\\ &= 2^{-\frac{k}{2}}C(x)(1+x)^k + O\left(2^{-\frac{k}{2}}|1+x|^{k(1-\beta)}\right),\end{aligned}$$

where

$$C(x) = \int_0^1 t^{\log_2(1+x)-\frac{3}{2}}\gamma(x,\log_2 t)\,dt. \tag{4.7}$$

Of course, the remaining part can be treated in a similar way so that after all we get (4.5).

For the proof of (4.6) we just have to use the upper bounds (4.4) instead of (4.3) and proceed along similar lines. It is also a little bit easier since we only have to provide upper bounds. □

Note that the proof also shows that for real $x > 0$ one always gets

$$\sum_{n<2^k} x^{s_{[\ge k]}(n)} \le C(1+x)^k$$

for a certain constant $C > 0$.

4.2. **Reduction of S_2 and S_3.** Let $x\in\mathbb{C}$, $k\ge 2$, $u\in[0,1]$. We consider

$$S(x) := \sum_{\ell<2^k} \Psi\left(-\sqrt{2^k(\ell+u)}\right) x^{s(\ell)}.$$

Let r be an integer with $0<r<k$. We write $\ell = 2^r m + n$ and obtain

$$\begin{aligned}S(x) &= \sum_{0\le i<k-r}\sum_{0\le j<r} S(2^i,2^j) + O\left(\sum_{m<2^{k-r}} |x|^{s(m)}\right) + O\left(\sum_{n<2^r}|x|^{s(n)}\right),\\ S(x) &= \sum_{0\le i<k-r}\sum_{0\le j<r} S(2^i,2^j) + O\left((1+|x|)^{k-r}\right) + O\left((1+|x|)^r\right)\end{aligned} \tag{4.8}$$

where

$$S(M,N,x) := \sum_{M\le m<2M}\sum_{N\le n<2N} \Psi\left(-\sqrt{2^k(2^r m+n+u)}\right) x^{s(m)}x^{s(n)}. \tag{4.9}$$

Lemma 5 (Vaaler, 1985)**.** *For $H\in\mathbb{N}$, $h\in\mathbb{Z}$, $1\le|h|\le H$, let*

$$0<\theta_H(h) := \pi\frac{|h|}{H+1}\left(1-\frac{|h|}{H+1}\right)\cot\left(\pi\frac{|h|}{H+1}\right) + \frac{|h|}{H+1} < 1.$$

Then, the trigonometric polynomial

$$\Psi_H^*(x) = -\frac{1}{2i\pi}\sum_{1\le|h|\le H}\frac{\theta_H(h)}{h}\,\mathrm{e}(hx)$$

satisfies

$$|\Psi(x)-\Psi_H^*(x)| \le \frac{1}{2H+2}\sum_{|h|\le H}\left(1-\frac{|h|}{H+1}\right)\mathrm{e}(hx) \qquad (x\in\mathbb{R}).$$

Proof. For $x \notin \mathbb{Z}$ this is inequality (7.14) of Vaaler [18]. For $x \in \mathbb{Z}$, both sides are equal to 1/2, so the result remains true. □

Let $H_0 > 0$. Using this lemma with H_0 and splitting the summation over h we get

$$\begin{aligned} &|S(M,N,x)| && (4.10)\\ &\ll \frac{1}{H_0} \sum_{M\le m<2M} \sum_{N\le n<2N} |x|^{s(m)+s(n)} \\ &\quad + \sum_{q\ll \log 2H_0} |S_1(H_0/2^q, M, N)| + \sum_{q\ll \log 2H_0} |S_2(H_0/2^q, M, N)|, && (4.11) \end{aligned}$$

where $S_1(H,M,N,x)$ is the sum

$$\sum_{H\le h<2H} \sum_{M\le m<2M} \sum_{N\le n<2N} \frac{\theta_{H_0}(h)}{h} x^{s(m)} x^{s(n)} \ \mathrm{e}\left(h\sqrt{2^k(2^r m+n+u)}\right)$$

and $S_2(H,M,N,x)$ is the sum

$$\sum_{M\le m<2M} \sum_{N\le n<2N} |x|^{s(m)} |x|^{s(n)} \sum_{H\le h<2H} \frac{c_{H_0}(h)}{h} \ \mathrm{e}\left(h\sqrt{2^k(2^r m+n+u)}\right)$$

with

$$c_{H_0}(h) := \frac{h}{H_0+1}\left(1-\frac{h}{H_0+1}\right).$$

Hence

$$|S_1(H,M,N,x)| \ll H^{-1} S_3(H,M,N,x), \qquad (4.12)$$
$$|S_2(H,M,N,x)| \ll H^{-1} S_3(H,M,N,|x|), \qquad (4.13)$$

where $S_3(H,M,N,x)$ is the sum

$$\sum_{H\le h<2H} \left| \sum_{M\le m<2M} \sum_{N\le n<2N} x^{s(m)} x^{s(n)} \ \mathrm{e}\left(h\sqrt{2^k(2^r m+n+u)}\right) \right|$$

By Taylor expansion we can write

$$\begin{aligned} &h\sqrt{2^k(2^r m+n+u)} \\ &= h2^{(k+r)/2} m^{1/2} + \tfrac{1}{2} h2^{(k-r)/2} m^{-1/2}(n+u) - \tfrac{1}{8} h2^{(k-3r)/2} m^{-3/2}(n+u)^2 \\ &\quad + O(H2^{(k-5r)/2} M^{-5/2} N^3). \end{aligned}$$

We introduce

$$\phi(h,m,n+u) = \tfrac{1}{2} h2^{(k-r)/2} m^{-1/2}(n+u) - \tfrac{1}{8} h2^{(k-3r)/2} m^{-3/2}(n+u)^2$$

Then

$$S_3(H,M,N,x) \le S_4(H,M,N,x) + O(H^2 2^{(k-5r)/2} M^{-3/2} N^4), \qquad (4.14)$$

where $S_4(H,M,N,x)$ is the sum

$$\sum_{H\le h<2H} \sum_{M\le m<2M} |x|^{s(m)} \left| \sum_{N\le n<2N} x^{s(n)} \ \mathrm{e}\left(\phi(h,m,n+u)\right) \right| \qquad (4.15)$$

4.3. **Exponential sums estimates.**

Lemma 6. *Let $X > 1$ be a real number and M, N, H be integers with $0 < M < X^2$, $0 < N < X^2$, $H > 0$. Let a_m, b_n, $\rho_{h,m}$ be complex numbers for $m = M, \dots, 2M-1$, $n = N, \dots, 2N-1$, $h = H, \dots, 2H-1$. We suppose $|\rho_{h,m}| \le 1$. Then uniformly for any $\alpha \in [1/16, 1]$ and $u \in [0,1]$, writing*

$$\phi(h,m,n) = Xhm^{-1/2}n - \alpha X^{-1}hm^{-3/2}n^2,$$

we have

$$\left| \sum_{H \le h < 2H} \sum_{M \le m < 2M} \sum_{N \le n < 2N} \rho_{h,m} a_m b_n \ \mathrm{e}(\phi(h,m,n+u)) \right|^2 \ll (M + \Delta_1^{-1})(1 + \Delta_2^{-1}) \ H^2 \sum_{M \le m < 2M} |a_m|^2 \sum_{N \le n < 2N} |b_n|^2 .$$

where

$$\Delta_1^{-1} = 2^9 XHM^{-1/2}N, \quad \Delta_2^{-1} = 2^9 X^{-1}HM^{-3/2}N^2.$$

Proof. By Lemma 2.4 of [2], we have

$$\left| \sum_{H \le h < 2H} \sum_{M \le m < 2M} \sum_{N \le n < 2N} \rho_{h,m} a_m b_n \ \mathrm{e}(\phi(h,m,n+u)) \right|^2 \ll (1 + \Delta_1^{-1})(1 + \Delta_2^{-1}) \sum_{(h,h',m,m') \in E_1} |\rho_{h,m}\rho_{h',m'} a_m a_{m'}| \sum_{(n,n') \in E_2} |b_n b_{n'}|$$

where E_1 is the set of quadruples (h, h', m, m') such that

$$\left| hm^{-1/2} - h'm'^{-1/2} \right| \le \Delta_1 HM^{-1/2}, \quad \left| hm^{-3/2} - h'm'^{-3/2} \right| \le \Delta_2 HM^{-3/2},$$

and E_2 is the set of pairs $(n, n') \in \{N, \dots, 2N-1\}^2$ such that

$$|n - n'| \le \Delta_1 N, \quad \left| n^2 - n'^2 \right| \le \Delta_2 N^2.$$

We observe that $M < X^2$ implies

$$\Delta_1 N = 2^{-9} X^{-1} H^{-1} M^{1/2} < 1.$$

Therefore all pairs (n, n') in E_2 satisfy $n = n'$, *i.e.* E_2 is exactly the diagonal of $\{N, \dots, 2N-1\}^2$ and we get

$$\sum_{(n,n') \in E_2} |b_n b_{n'}| = \sum_{N \le n < 2N} |b_n|^2 .$$

Now we write $|a_m a_{m'}| \le (|a_m|^2 + |a_{m'}|^2)/2$. Hence by symmetry of the roles of m and m', we have

$$\sum_{(h,h',m,m') \in E_1} |\rho_{h,m}\rho_{h',m'} a_m a_{m'}| \le \sum_{(h,h',m,m') \in E_1} |a_m a_{m'}| \le \sum_{(h,h',m,m') \in E_1} |a_m|^2 .$$

If we fix h, h' and m, we observe that m' must be in an interval of length $\ll \Delta_1 M$, thus there are $\ll 1 + \Delta_1 M$ such m'. We obtain

$$\begin{aligned} \sum_{(h,h',m,m') \in E_1} |a_m|^2 &\ll \sum_{H \le h < 2H} \sum_{H \le h' < 2H} \sum_{M \le m < 2M} |a_m|^2 (1 + \Delta_1 M) \\ &= H^2 (1 + \Delta_1 M) \sum_{M \le m < 2M} |a_m|^2 , \end{aligned}$$

and the inequality of Lemma 6 follows since $\Delta_1^{-1} \gg 1$ and $(1 + \Delta_1^{-1})(1 + \Delta_1 M) \ll M + \Delta_1^{-1}$. □

Lemma 7. *Suppose that the assumptions of Lemma 6 are satisfied and that* $|a_m| \leq 1$*. Then*

$$\left|\sum_{H\leq h<2H}\sum_{M\leq m<2M}\sum_{N\leq n<2N}\rho_{h,m}a_m b_n\ \mathrm{e}(\phi(h,m,n+u))\right|^2$$
$$\ll\ \Delta_1^{-1}(1+\Delta_2^{-1})\ (HM\log(2HM)+\Delta_1H^2M^2)\sum_{N\leq n<2N}|b_n|^2.$$

Proof. We follow the proof of Lemma 6, except for the sum

$$\sum_{(h,h',m,m')\in E_1}|\rho_{h,m}\rho_{h',m'}a_m a_{m'}|$$

for which we use $|\rho_{h,m}\rho_{h',m'}a_m a_{m'}|\leq 1$, and Lemma 1 of [8], that asserts

$$\sum_{(h,h',m,m')\in E_1}1\ll HM\log(2HM)+\Delta_1H^2M^2.$$

Of course, this completes the proof. □

4.4. **Application of the exponential sums estimates.**

Lemma 8. *Uniformly for* $x\in\mathbb{C}$, $0.84\leq|x|\leq\xi_2$, $k\geq 2$, $u\in[0,1]$, *we have*

$$\left|\sum_{\ell<2^k}\Psi\left(-\sqrt{2^k(\ell+u)}\right)x^{s(\ell)}\right|\ll k^2 2^{k/8}(1+|x|)^{k/2}(1+|x|^2)^{k/4}.$$

Note that in the full range $0.84\leq|x|\leq\xi_2-\varepsilon$ this upper bound is better than the trivial estimate $(1+|x|)^k$.

Proof. We write $X=2^{(k-r)/2-1}$. By Lemma 6 we have

$$H^{-2}\left|S_4(H,M,N)\right|^2$$
$$\ll\ (M+XHM^{-1/2}N)(1+X^{-1}HM^{-3/2}N^2)\sum_{M\leq m<2M}|x|^{2s(m)}\sum_{N\leq n<2N}|x|^{2s(n)}$$

To simplify the computations, we choose

$$r=\lfloor k/2\rfloor.$$

Then $X\asymp 2^{k/4}$, $M<2^{k/2}$ and $N<2^{k/2}$. Hence

$$M+XHM^{-1/2}N\ll 2^{k/2}+HM^{-1/2}2^{3k/4}\ll HM^{-1/2}2^{3k/4},$$

and

$$1+X^{-1}HM^{-3/2}N^2\ll 1+HM^{-3/2}2^{3k/4}\ll HM^{-3/2}2^{3k/4}.$$

Therefore

$$H^{-2}\left|S_4(H,M,N)\right|^2\ll H^2M^{-2}2^{3k/2}\sum_{M\leq m<2M}|x|^{2s(m)}\sum_{N\leq n<2N}|x|^{2s(n)}.$$

Using

$$\sum_{2^i\leq n<2^{i+1}}|x|^{2s(n)}=|x|^2\left(1+|x|^2\right)^i,$$

we get

$$H^{-1}\left|S_4(H,2^i,2^j)\right|\ll H2^{3k/4-i}\left|x\right|^2\left(1+|x|^2\right)^{(i+j)/2}.$$

We report this upperbound in (4.14) and get for $i\leq\lfloor k/2\rfloor$, $j\leq\lceil k/2\rceil$:

$$H^{-1}\left|S_3(H,2^i,2^j)\right|\ \ll\ H2^{3k/4-i}\left|x\right|^2\left(1+|x|^2\right)^{(i+j)/2}+H2^{-3k/4-3i/2+4j}$$
$$\ll\ H2^{3k/4-i}\left|x\right|^2\left(1+|x|^2\right)^{i/2+k/4}+H2^{5k/4-3i/2}.$$

We choose

$$H_0 = \left\lceil 2^{i/2-3k/8}(1+|x|)^{i/2+k/4}(1+|x|^2)^{-i/4-k/8} \right\rceil$$

and by (4.12), (4.13), and (4.11) we get for $i \le \lfloor k/2 \rfloor$, $j \le \lceil k/2 \rceil$:

$$\begin{aligned} \left|S(2^i,2^j,x)\right| \ll\ & 2^{3k/8-i/2}\left|x\right|^2(1+|x|)^{i/2+k/4}(1+|x|^2)^{i/4+k/8} \\ & + 2^{3k/4-i}\left|x\right|^2(1+|x|^2)^{i/2+k/4} \\ & + 2^{7k/8-i}(1+|x|)^{i/2+k/4}(1+|x|^2)^{-i/4-k/8} + 2^{5k/4-3i/2}. \end{aligned}$$

In order to simplify the computations, we assume that $|x| \ge 2/3$. This implies that

$$(1+|x|)\sqrt{1+|x|^2} > 2,$$

and we get an upper bound for the first term by replacing i by $k/2$ in it. For the remaining terms we will assume first that $i \ge k/3$. We get

$$\begin{aligned} \left|S(2^i,2^j,x)\right| \ll\ & 2^{k/8}\left|x\right|^2(1+|x|)^{k/2}(1+|x|^2)^{k/4} \\ & + 2^{5k/12}\left|x\right|^2(1+|x|^2)^{5k/12} + 2^{k/4}\left|x\right|^2(1+|x|^2)^{k/2} \\ & + 2^{13k/24}(1+|x|)^{5k/12}(1+|x|^2)^{-5k/24} + 2^{3k/4}. \end{aligned}$$

When $i < k/3$, we may use the trivial upper bound:

$$\left|S(2^i,2^j,x)\right| \ll |x|^2(1+|x|)^{i+k/2} \ll |x|^2(1+|x|)^{5k/6}.$$

For $0.84 \le |x| \le \xi_2$ we can exhibit numerically a dominating term, and in that range we get for all $i \le \lfloor k/2 \rfloor$, $j \le \lceil k/2 \rceil$:

$$\left|S(2^i,2^j,x)\right| \ll 2^{k/8}(1+|x|)^{k/2}(1+|x|^2)^{k/4}.$$

Finally by (4.8) we obtain the result of Lemma 8. □

Lemma 9. *Uniformly for $x \in \mathbb{C}$, $|x| \le 1$, $k \ge 2$, $u \in [0,1]$, we have*

$$\left|\sum_{\ell < 2^k} \Psi\left(-\sqrt{2^k(\ell+u)}\right) x^{s(\ell)}\right| \ll k^4 2^{5k/6}.$$

Note that this upper bound is only significant if $k^4 2^{5k/6} \le (1+|x|)^k$, that is, if $|x| > 2^{5/6} - 1 = 0.781\ldots$.

Proof. We write $X = 2^{(k-r)/2-1}$. By Lemma 6 we have

$$\begin{aligned} & |S_4(H,M,N)|^2 \\ & \ll\ XHM^{-1/2}N(1+X^{-1}HM^{-3/2}N^2)(HM\log(2HM)+X^{-1}H^{-1}M^{3/2})N. \end{aligned}$$

Again, we choose

$$r = \lfloor k/2 \rfloor.$$

Then $X \asymp 2^{k/4}$, $M < 2^{k/2}$ and $N < 2^{k/2}$. Hence

$$XHM^{-1/2}N \ll 2^{k/2} + HM^{-1/2}2^{3k/4} \ll HM^{-1/2}2^{3k/4},$$

and

$$1 + X^{-1}HM^{-3/2}N^2 \ll 1 + HM^{-3/2}2^{3k/4} \ll HM^{-3/2}2^{3k/4}.$$

Furthermore

$$HM + X^{-1}H^{-1}M^{3/2} \ll HM(1+X^{-1}H^{-2}M^{1/2}) \ll HM(1+H^{-2}) \ll HM.$$

Therefore

$$|S_4(H,M,N)|^2 \ll H^2M^{-2}2^{3k/2}HMN\log(2HM) \ll H^3M^{-1}2^{2k}\log(2HM)$$

and

$$H^{-1}|S_4(H,M,N)| \ll H^{1/2}M^{-1/2}2^k\log(2HM).$$

We choose $H_0 = \lceil M 2^{-k/3} \rceil$. Then for $1 \le H \le H_0$,

$$H^{-1} |S_4(H,M,N)| \ll (2^{5k/6} + M^{-1/2} 2^k) \log(2M).$$

We report this upperbound in (4.14) and get for $H^{-1} |S_3(H,M,N)|$ the upperbound

$$\begin{aligned} &\ll (2^{5k/6} + M^{-1/2}2^k + H 2^{-3k/4} M^{-3/2} N^4) \log(2M) \\ &\ll (2^{5k/6} + M^{-1/2}2^k + M^{-1/2} 2^{11k/12} + M^{-3/2} 2^{5k/4}) \log(2M) \\ &\ll (2^{5k/6} + M^{-1/2}2^k + M^{-3/2} 2^{5k/4}) \log(2M). \end{aligned}$$

By (4.12), (4.13), and (4.11) we get

$$\begin{aligned} |S(M,N,x)| &\ll \frac{MN}{H_0} + (2^{5k/6} + M^{-1/2}2^k + M^{-3/2}2^{5k/4}) \log(2M) \log(2H_0) \\ &\ll (2^{5k/6} + M^{-1/2}2^k + M^{-3/2}2^{5k/4}) \log(2M) \log(2H_0). \end{aligned}$$

Hence for $M \gg 2^{k/3}$ we have

$$|S(M,N,x)| \ll k^2 2^{5k/6}. \tag{4.16}$$

For $M \ll 2^{k/3}$ the trivial upperbound $|S(M,N,x)| \le MN$ prove that (4.16) is indeed valid for all $M \ll 2^{k/2}$. Finally by (4.8) we obtain the result of Lemma 9. □

5. Proof of Theorems 2 and 3

5.1. **Proof of Theorem 2.** We start by stating the following lemma.

Lemma 10. *Let $A(x) = \sum_{m \ge 0} a_m x^m$ be a power series of non-negative numbers a_m. Suppose that for some $0 < r_1 < r_2$ and $\eta > 0$ we have*

$$A(x) = C(x)(1+x)^k + \mathcal{O}\left((1+|x|)^{k(1-\eta)}\right)$$

uniformly for $r_1 \le |x| \le r_2$, where $C(x)$ is a twice continuously differentible function. Then

$$a_m = C\left(\frac{j}{k-j}\right) \cdot \binom{k}{m} \left(1 + \mathcal{O}\left(k^{-1}\right)\right)$$

uniformly for $r_1/(1+r_1)\, k \le m \le r_2/(1+r_2)\, k$.

Proof. The proof is immediate by using Cauchy's formula

$$a_m = \frac{1}{2\pi i} \int_{|x|=x_0} \frac{A(x)}{x^{m+1}}\, dx,$$

where $x_0 = m/(k-m)$, and classical saddle point approximations, e.g. compare with [7]. □

Two parts of Theorem 2, that is (2.3) and (2.5), are a direct corollary from Propositions 1 and 2 (or from Theorem 1).

The proof of (2.4) is also almost direct. In the range $\varepsilon \le m \le (1 - 1/\sqrt{2} - \varepsilon)k$ the *second order terms* in the generating function of $s_{[<k]}(n^2)$

$$2^{\frac{k}{2}} \frac{(1+x)^2(1-x)}{x^2+2x-1} \quad \text{resp.} \quad 2^{\frac{k-1}{2}} \frac{x^3+3x^2-4x-2}{x^2+2x-1}$$

dominate the asymptotic behaviour of the coefficients. These terms are just rational function. Thus the asymptotic behaviour for the coefficients follows immediately (and it is easy to show that these terms are dominant).

5.2. **Proof of Theorem 3.** As above set $A(x) = \sum_{n<2^k} x^{s_{[<k]}(n^2)}$. Then

$$\#\{n < 2^k : s_{[<k]}(n^2) \equiv c \bmod M\} = \frac{1}{M}\sum_{m=0}^{M-1} e(-mc/M) A\left(e(m/M)\right).$$

Hence, (2.6) (and similarly (2.7)) follows from Proposition 1 (and from Proposition 2).

6. Squares with Large Sum of Digits Function

In this section we want to collect some results on squares n^2 with large sum-of-digits function $s(n^2)$.

First of all we have the following upper bound.

Theorem 4. *Then, for every $\varepsilon > 0$ exists a constant $C > 0$ such that*

$$\#\{n < 2^k : s(n^2) \ge L\} \le \frac{C}{\left(\frac{L}{2k}\right)^{L/2}\left(1-\frac{L}{2k}\right)^{k-L/2}} \tag{6.1}$$

for $k \le L \le (2-\varepsilon)k$.

Proof. Suppose that $A(x) = \sum_{m\ge 0} a_m x^m$ is a power series with non-negative coefficients a_m. Then we obviously have (for $x \ge 1$)

$$\sum_{m\ge L} a_m \le x^{-L} A(x). \tag{6.2}$$

The idea of the proof is to apply this (easy) method to

$$A(x) = \sum_{n<2^k} x^{s(n^2)} = \sum_{n<2^k} x^{s_{[<k]}(n^2)+s_{[\ge k]}(n^2)}.$$

By Propositions 1 and 2 we have for $x \ge 1$

$$\sum_{n<2^k} x^{s_{[<k]}(n^2)} \ll (1+x)^k$$

and

$$\sum_{n<2^k} x^{s_{[\ge k]}(n^2)} \ll (1+x)^k.$$

Hence, by Cauchy-Schwarz's inequality

$$\begin{aligned} |A(x)|^2 &= \left|\sum_{n<2^k} x^{s(n^2)}\right|^2 \\ &\le \sum_{n<2^k} x^{2s_{[<k]}(n^2)} \cdot \sum_{n<2^k} x^{2s_{[\ge k]}(n^2)} \\ &\ll (1+x^2)^{2k}. \end{aligned}$$

Hence, with $x = \sqrt{L/(2k-L)}$ and by the use of (6.2) we directly obtain (6.1). □

Unfortunately we do not have a corresponding lower bound. Nevertheless we can prove the existence of squares n^2 with very large sum-of-digits function.

Theorem 5. *For every $\varepsilon > 0$ there exist infinitely many positive integers n such that*

$$s(n^2) \ge (2-\varepsilon)\log_2 n.$$

Interestingly it seems to be a non-trivial problem to construct squares with a large sum-of-digits as the following two examples by G. Baron (TU Wien) and J. Cassaigne (Marseille) show. Nevertheless, if one combines these kinds of examples with a second *trick* then we obtain a constructive proof of Theorem 5.

First Example. (G. Baron) Set $N_1 = 2^m - 1$ and $L_1 = m$ (for some $m \geq 1$) and inductively $N_{n+1} = 2^{2L_n}(2^{L_n} - 1) - N_n, L_{n+1} = 3L_n$. Then N_n^2 has many 1-s in the binary expansion. For example, for $m = 8$ we get

$$
\begin{aligned}
N_2^2 =& 111111011111111100000011111111101111111000000001\\
N_3^2 =& 1111111111111111111111011111111111111111111111111000000100 0\\
& 0000011111111111111101111111100000001011111101111111110000\\
& 0011111111101111111000000001.
\end{aligned}
$$

The ratio $s(N_n^2)/\log_2(N_n^2)$ approaches 3/4 from below. (Other nice examples are $(2^{5k} - 2^{3k} - 2k + 1)^2$ with ratio 3/5 and $(2^{3k} - 2^{2k} - 2^k + 1)^2$ with ratio 2/3.)

Second Example. (J. Cassaigne) Set

$$N_n = 181(2^{14} + 1)(2^{42} + 1) \cdots (2^{14 \cdot 3^{n-1}} + 1).$$

Then N_n^2 has $14 \cdot 3^n + 1$ digits with $12 \cdot 3^n + 1$ ones. The ratio tends to 6/7:

$$
\begin{aligned}
N_0 =& 181 = 10110101\\
N_0^2 =& 32761 = 111111111111001\\
N_1 =& 2965685 = 1011010100000010110101\\
N_1^2 =& 8795287519225 = 1111111111111001111111111001111111111111001\\
N_2 =& 13043220567286431925\\
=& 1011010100000010110101000000000000000000001011010100000010110101\\
N_2^2 =& 170125602766883791039492846197659205625\\
=& 11111111111110011111111110011111111111111001111111111100\\
& 11111111110011111111111110011111111111110011111111110011\\
& 111111111111001.
\end{aligned}
$$

This second example is also interesting since it provides an infinite sequence of numbers N_n with very few a very small sum-of-digits function $s(N_n) = 5 {\cdot} 2^n = o(3^n)$. On the other hand the squares have many non-zero digits. Note also that this method is not restricted to the initial value $N_0 = 181$.

Proof. (Theorem 5) The idea of the proof is to repeat a two step procedure of the following kind. The first initial step is exactly Example 2. Recall that the numbers N_n satisfy $s(N_n) = 5 \cdot 2^n$ and $s(N_n^2) = 12 \cdot 3^n + 1$.

Now observe that if n is an odd positive integer with $n^2 < 2^{k+1}$ then

$$s((2^k - n)^2) = s(2^{k-1}((s^{k+1} - 1) - (n - 1)) + n^2) = k - s(n) + s(n^2).$$

This means that if we consider the numbers

$$\tilde{N}_n = 2^{28 \cdot 3^{n-1} + 2} - N_n$$

then the ratio $s(N_n^2)/\log_2(N_n^2)$ approaches 13/14 (as $n \to \infty$). This completes the second initial step.

We now repeat these two steps appropriately. First we go back to the construction principle of the second example. Fix some large n_0, set $L_0 := 56 \cdot 3^{n_0} + 4$ and define

$$N_{n,1} := N_{n_0}((2^{L_0} + 1)(2^{L_0 \cdot 3} + 1) \cdots (2^{L \cdot 3^{n-1}} + 1).$$

Then the ratio $s(N_{n,1})/\log_2(N_{n,1})$ approaches zero for large n whereas the ratio $s(N_{n,1}^2)/\log_2(N_{n,1}^2)$ is approximately 13/14. With help of the second step we can now construct a numbers $\tilde{N}_{n,1}$ for which the ratio $s(\tilde{N}_{n,1}^2)/\log_2(\tilde{N}_{n,1}^2)$ is approximately 27/28. In this way we can proceed further.

After k repetitions of the two step procedure we can approach the ratio $1-1/(7\cdot 2^k)$ to arbitrary precision. This completes the proof of Theorem 5. □

References

[1] N. L. Bassily and I. Kátai, Distribution of the values of q-additive functions on polynomial sequences, *Acta Math. Hung.* **68** (1995), 353–361.

[2] E. Bombieri and H. Iwaniec, On the Order of $\zeta(\frac{1}{2}+it)$, *Annali Scuola Norm. Sup. Pisa* **13** (1986), 449–472.

[3] J. Coquet, *Corrélation de suites arithmétiques*, Sémin. Delange-Pisot-Poitou, 20e Année 1978/79, Exp. 15, 12 p. (1980).

[4] C. Dartyge and G. Tenenbaum, *Congruences de sommes de chiffres de valeurs polynomiales,* manuscript.

[5] H. Delange, *Sur les fonctions q-additives ou q-multiplicatives*, Acta Arith. **21** (1972), 285–298.

[6] H. Delange, Sur la fonction sommatoire de la fonction "Somme de Chiffres", *L'Enseignement math.* **21** (1975), 31–77.

[7] M. Drmota, A bivariate asymptotic expansion of coefficients of powers of generating functions, *Europ. J. Combinatorics* **15** (1994), 139–152.

[8] E. Fouvry and H. Iwaniec, Exponential Sums with Monomials, *Journal of Number Theory* **33** (1989), 311–333.

[9] E. Fouvry and C. Mauduit, Sommes des chiffres et nombres presque premiers. *Math. Ann.* **305** (1996), 571-599.

[10] E. Fouvry and C. Mauduit, Methodes de crible et fonctions sommes des chiffres. *Acta Arith.* **77** (1996), 339-351.

[11] A. O. Gelfond, Sur les nombres qui ont des propriétés additives et multiplicatives données, *Acta Arith.* **13** (1968), 259–265.

[12] P. J. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy, On the moments of the sum-of-digits function, in: *Applications of Fibonacci Numbers* **5** (1993), 263–271

[13] L. K. Hua, *Additive Theory of Prime Numbers*, Translations of Mathematical Monographs Vol. 13, Am. Math. Soc., Providence, 1965.

[14] R. E. Kennedy and C. N. Cooper, An extension of a theorem by Cheo and Yien concerning digital sums, *Fibonacci Q.* **29** (1991), 145–149.

[15] D.-H. Kim, On the joint distribution of q-additive functions in residue classes , *J. Number Theory* **74** (1999), 307 – 336.

[16] P. Kirschenhofer, On the variance of the sum of digits function, *Lecture Notes Math.* **1452** (1990), 112–116.

[17] M. Peter, The summatory function of the sum-of-digits function on polynomial sequences, *Acta Arith.* **104** (2002), 85–96.

[18] J.D. Vaaler, Some extremal functions in Fourier analysis, *Bull. Amer. Math. Soc.* **12** (1985), 183–216.