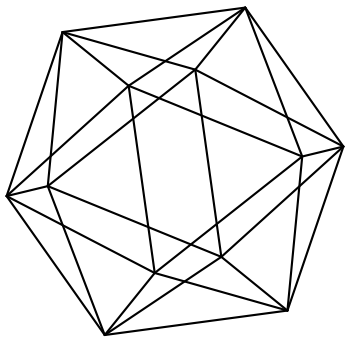
Max-Planck-Institut für Mathematik
Bonn

Smooth numbers with few non-zero binary digits

by

Maximilian Hauck
Igor E. Shparlinski



Max-Planck-Institut für Mathematik
Preprint Series 2023 (11)

Date of submission: June 14, 2023

Smooth numbers with few non-zero binary digits

by

Maximilian Hauck
Igor E. Shparlinski

Max-Planck-Institut für Mathematik
Vivatsgasse 7
53111 Bonn
Germany

Universität Bonn
Endenicher Allee 60
53115 Bonn
Germany

Department of Pure Mathematics
University of New South Wales
Sidney, NSW 2052
Australia

MPIM 23-11

SMOOTH NUMBERS WITH FEW NON-ZERO BINARY DIGITS

MAXIMILIAN HAUCK AND IGOR E. SHPARLINSKI

Abstract. We use bounds of character sums and some combinatorial arguments to show the abundance of very smooth numbers which also have very few non-zero binary digits.

1. Introduction

1.1. **Motivation and background.** Since recently there has been a lot of interest in arithmetic properties of integers with various digit restrictions in a given integer base g. This includes the work of Bourgain [Bou13, Bou15] and Swaenepoel [Swa20] on primes with prescribed digits on a positive proportion of positions in their digital expansion, the resolution of the Gelfond conjecture by Mauduit and Rivat [MR10], and the results of Maynard [May19, May22] on primes with missing digits, see also [Bug18, BK18, Col09, DES17, DMR20, Kar22, Nas15, Pratt20] and references therein.

Prime divisors of integers with very few non-zero g-ary digits have been studied in [Bou05, CKS18, Shp08].

A variety of results on primitive roots and quadratic nonresidues modulo a prime p, which satisfy various digit restrictions can be found in the series of work [DES13a, DES13b, DES17].

Furthermore, Mauduit and Rivat [MR09] and Maynard [May22] have also studied values of integral polynomials with various digital restrictions.

We also note that special integers with restricted digits appear in the context of cryptography [GS08, Meng13, Shp06].

Here we consider some digital problems for smooth integers. We recall that by a result of Bugeaud and Kaneko [BK18, Theorem 1.1] any integer N with at most $k \geqslant 3$ nonzero digits in any fixed basis $g \geqslant 2$ not dividing N has a prime divisor $p \mid N$ with

$$p \geqslant \left(\frac{1}{k-2} + o(1)\right) \frac{\log\log N \, \log\log\log N}{\log\log\log\log N} \tag{1.1}$$

as $N \to \infty$, see also [Bug21]. The proof of (1.1) uses some classical methods of Diophantine analysis, such as the bounds of linear forms in logarithms. Our technique is different and shows that there are many both reasonably smooth and sparse binary integers.

1.2. **Smooth sparse integers.** We recall that an integer s is called y-smooth if it has no prime divisors $p > y$, see [Gra08, HT86] for some background.

2020 *Mathematics Subject Classification.* 11A63, 11L40, 11N25.

Key words and phrases. Smooth integers, sparse binary representations, character sums.

For a fixed absolute constant $\zeta \geqslant 1$, given a real number $A > \zeta^3$ we define

$$\mu_0(A) = (1/2 - 1/2A)\left(1 - \zeta A^{-1/3}\right) \tag{1.2}$$

and define $\vartheta_0(A) < 1/2$ by the equation

$$H\left(\vartheta_0(A)\right) = 1 - \frac{1}{(1-\mu_0(A))A}, \tag{1.3}$$

where $H(\rho)$ is the *binary entropy* function defined by

$$H(\rho) = -\rho\frac{\log \rho}{\log 2} - (1-\rho)\frac{\log(1-\rho)}{\log 2}\ . \tag{1.4}$$

In particular, notice that $\vartheta_0(A) \to 1/2$ as $A \to \infty$.

We are interested in the existence of smooth integers with only few non-zero binary digits. Clearly, this question makes sense only for odd integers as otherwise powers of 2 give a perfect solution.

Theorem 1.1. *There is an absolute constant $\zeta \geqslant 1$ such that for $A > \zeta^3$ the following holds: For any $\vartheta < \vartheta_0(A)$ and sufficiently large n, there exists an odd n^A-smooth n-bit integer with at least*

$$\left(\mu_0(A) + (1-\mu_0(A))\,\vartheta\right) n$$

zeros in its binary expansion.

The proof of Theorem 1.1 is based on a refinement of the argument of [Shp06, Theorem 6] combined with the approach of [GS08], which in turn relies on bounds for short multiplicative character sums from [Iwa74] (see also [BS19]). It also uses a combinatorial argument, which originates from [Shp87] and has been used in [DES13a, DES13b, Nas15].

The constant ζ in (1.2) is directly related to the constant in the bound on short character sums in [BS19], see Section 3; it is effective and can be explicitly evaluated.

Using another approach based on the observation that $2^k + 1$ is quite smooth if we take k as the product of the first r odd primes, we obtain the following further result in the same direction:

Theorem 1.2. *Let $0 < \alpha < 1$ be fixed. Then there exist infinitely many n-bit integers $N \geqslant 1$ which are $Y^{1+o(1)}$-smooth, where*

$$Y = \exp\left(2^{1/2}e^{-\gamma}\alpha^{-1/2}\left(\log N\right)^{1/2}\left(\log\log\log N\right)^{-1}\right)$$

and γ denotes the Euler-Mascheroni constant, which have at most

$$\alpha n + O(n^{1/2}\log n)$$

non-zero digits in their binary expansion.

It is easy to see from the proof of Theorem 1.2 that one can obtain a more uniform version of this result when both α and N vary in such a way that $\alpha^{-1} = o(\log N)$. We now make this observation more concrete and use a similar argument to produce another construction in a different regime of smoothness and sparseness.

Theorem 1.3. *Let $0 \leqslant \alpha \leqslant 1$ be fixed. Then there exist infinitely many n-bit integers $N \geqslant 1$ which are $Y^{1+o(1)}$-smooth where*

$$Y = \exp\left(2e^{-\gamma}\left(\log 2\right)^{\alpha/2}(\log N)^{1-\alpha/2}\left(\log\log\log N\right)^{-1}\right)$$

and γ denotes the Euler-Mascheroni constant, which have at most

$$\frac{1}{2\log 2}n^{\alpha} + O(n^{\alpha/2}\log n)$$

non-zero digits in their binary expansion.

1.3. **Notation.** Throughout the paper, the notations $U = O(V)$, $U \ll V$ and $V \gg U$ are equivalent to $|U| \leqslant cV$ for some positive constant c, which throughout the paper is absolute. If $U \ll V$ and $V \gg U$, we write $U \asymp V$.

Moreover, for any quantity $V > 1$ we write $U = V^{o(1)}$ (as $V \to \infty$) to indicate a function of V which satisfies $V^{-\varepsilon} \leqslant |U| \leqslant V^{\varepsilon}$ for any $\varepsilon > 0$, provided V is large enough. One additional advantage of using $V^{o(1)}$ is that it absorbs $\log V$ and other similar quantities without changing the whole expression.

We also write $U \sim V$ as an equivalent of $(1-\varepsilon)V \leqslant U \leqslant (1+\varepsilon)V$ for any $\varepsilon > 0$, provided V is large enough.

For a finite set $\mathcal{S}$ we denote its cardinality by $\#\mathcal{S}$.

For $n \in \mathbb{N}$, we denote the n-th cyclotomic polynomial by Φ_n and the Euler function by φ. Furthermore, we write A_n for the largest absolute value of one of the coefficients of Φ_n.

For a natural number n, we use $s_2(n)$ to denote the sum of the digits of n in its binary representation.

We write $2 = p_1 < p_2 < \dots$ for the prime numbers.

We denote by $\gamma \approx 0.577$ the Euler-Mascheroni constant.

Finally, the expressions $1/ab$ mean $1/(ab)$ (which deviates from the canonical interpretation b/a).

2. Preparations

2.1. **Arithmetic functions.** We make use of the following well-known upper bound on the number $\tau(w)$ of divisors of a natural number w, see, for example, in [IK04, Equation (1.81)].

Lemma 2.1. *We have*

$$\tau(w) \leqslant w^{o(1)}\,.$$

as $w \to \infty$.

Bateman [Bat49] gives the following upper bound on the coefficients of cyclotomic polynomials:

Lemma 2.2. *We have*

$$A_n \leqslant \exp\left(\frac{1}{2}\tau(n)\log n\right)$$

for all $n \geq 1$.

We also need to bound the Euler function of the product of the first odd primes.

Lemma 2.3. *Let $k = p_2 \cdots p_r$ for some integer $r \geq 2$. Then*

$$\varphi(p_2 \cdots p_r) = 2e^{-\gamma} \frac{k}{\log\log k}(1 + o(1))$$

as $r \to \infty$.

Proof. This is a direct application of Mertens' so-called third theorem, see [Mer74]:

$$\frac{\varphi(p_2 \cdots p_r)}{p_2 \cdots p_r} = \prod_{i=2}^{r} \left(1 - \frac{1}{p_i}\right) = 2 \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right) = \frac{2e^{-\gamma}}{\log p_r}(1 + o(1)).$$

Using that by the prime number theorem

$$k = \exp\left((1 + o(1))\, p_r\right)$$

we conclude the proof. □

Finally, we need the following elementary result, see [HLS11, Proposition 2.2], which asserts that s_2 is subadditive,

Lemma 2.4. *For any $m, n \in \mathbb{N}$, we have*

$$s_2(m + n) \leqslant s_2(m) + s_2(n).$$

2.2. **Counting smooth numbers.** Let $\Psi(x, y)$ denote the number of y-smooth numbers less than or equal to x. We have the following result obtained. after simple manipulations, by combining [HT86, Theorem 3] and [HT86, Equation (2.4)]:

Lemma 2.5. *For $x \geq y \geq 2$ and $1 \leqslant c \leqslant y$, we have*

$$\begin{aligned}\Psi(cx, y) = \Psi(x, y) \left(1 + \frac{y}{\log x}\right)^{\log c/\log y} \\ \times \left(1 + O\left(\frac{\log\log(1+y)}{\log^2 y} \cdot \log\left(1 + \frac{y}{\log x}\right) \cdot \log c\right)\right) \\ \times \left(1 + O\left(\frac{1}{u} + \frac{\log y}{y}\right)\right)\end{aligned}$$

uniformly, where $u = \log x/\log y$.

We note that, for completeness, we have presented Lemma 2.5 in full generality, while we only use it for a fixed $c > 1$ and very small (compared to x) values of y. More precisely, we use Lemma 2.5 in the following form:

Lemma 2.6. *Fix real numbers $\alpha, \beta > 0$, $c \geq 1$ and $A > 1$. Then*

$$\Psi(c\alpha x^{\beta}, \log^A x) \sim c^{1-1/A} \Psi(\alpha x^{\beta}, \log^A x)$$

as $x \to \infty$.

Proof. Using Lemma 2.5 with αx^β in place of x and $y = \log^A x$, we obtain

$$\Psi(c\alpha x^\beta, \log^A x) = (1+o(1))\Psi(\alpha x^\beta, \log^A x)\left(1 + \frac{\log^A x}{\beta \log x + \log \alpha}\right)^{\log c/A \log\log x}.$$

Finally,

$$\begin{aligned}\left(1 + \frac{\log^A x}{\beta \log x + \log \alpha}\right)^{\log c/A \log\log x} &= \left((1/\beta + o(1)) \log^{A-1} x\right)^{\log c/A \log\log x} \\ &= (1+o(1))\left(\log^{A-1} x\right)^{\log c/A \log\log x}.\end{aligned}$$

Since $\left(\log^{A-1} x\right)^{\log c/A \log\log x} = c^{1-1/A}$, the result follows. □

We also need an asymptotic formula for $\Psi(x, \log^A x)$, see for example [Gra08, Equation (1.14)]:

Lemma 2.7. *For any fixed $A > 1$, we have*

$$\Psi(x, \log^A x) = x^{1-1/A+o(1)}$$

as $x \to \infty$.

Finally, we need an upper bound of Harper [Har16] on the number of smooth numbers in an arithmetic progression. In fact we present it in a very special case tailored to our applications. Namely, let $\Psi(x, y; q, a)$ denote the number of y-smooth numbers less than or equal to x in the residue class a modulo q. By [Har16, Smooth Number Result 3, Section 2.1] we have the following estimate.

Lemma 2.8. *For any fixed $A > 1$, $\beta > 0$ and $\varepsilon > 0$, we have*

$$\Psi(x^\beta, \log^A x; q, a) \leqslant (x^\beta/q)^{1-1/A} x^{o(1)}$$

for all a as $q \leqslant x^{\beta-\varepsilon}$ and $x \to \infty$.

2.3. **Sums involving binomial coefficients.** We recall the definition (1.4). We frequently use the following result from [MS77, Chapter 10, Corollary 9]:

Lemma 2.9. *For any natural number n and $0 < \rho \leqslant 1/2$, we have*

$$\sum_{0 \leqslant k \leqslant \rho n} \binom{n}{k} \leqslant 2^{nH(\rho)}.$$

Furthermore, we also need a bound on the product of binomial coefficient or, in other words, on the sum of their logarithms:

Lemma 2.10. *For any integer $n \geq 2$, we have*

$$\sum_{k=0}^{n} \log \binom{n}{k} = \frac{1}{2}n^2 + O(n \log n).$$

Proof. By the Stirling formula, we have $\log n! = n\log n - n + O(\log n)$ for $n \geq 2$ and therefore

$$\log \binom{n}{k} = n\log n - k\log k - (n-k)\log(n-k) + O(\log n)$$

whenever $2 \leqslant k \leqslant n-2$ and $n \geq 2$. Summing over all k, we see that

$$\sum_{k=0}^{n} \log \binom{n}{k} = n^2 \log n - 2\sum_{k=1}^{n} k\log k + O(n\log n).$$

For the remaining sum, partial summation yields

$$\begin{aligned}\sum_{k=1}^{n} k\log k &= \frac{n(n+1)}{2}\log n - \int_1^n \frac{\lfloor t\rfloor(\lfloor t\rfloor+1)}{2t}\,\mathrm{d}t \\ &= \frac{1}{2}n^2\log n - \int_1^n \frac{t}{2}\,\mathrm{d}t + O(n\log n) \\ &= \frac{1}{2}n^2\log n - \frac{1}{4}n^2 + O(n\log n)\end{aligned}$$

and hence we obtain

$$\sum_{k=0}^{n} \log \binom{n}{k} = \frac{1}{2}n^2 + O(n\log n),$$

as claimed. □

2.4. **Character sums modulo powers of 2.** From [BS19, Theorem 2.1], we deduce the following estimate for character sums modulo powers of 2:

Lemma 2.11. *There exists an effective constant $\xi > 0$ such that for all integers $k \geq 1$ and all non-principal characters χ modulo $q = 2^k$, we have*

$$\sum_{n=M+1}^{M+N} \chi(n) \ll N^{1-\xi\ell^2/k^2}$$

uniformly for all integers M and $N = 2^\ell$, where ℓ is an integer such that $\ell \leqslant k$, where implied constant is absolute.

Proof. Assume χ is induced by the primitive character $\widetilde{\chi}$ modulo $\widetilde{q} = 2^{k_0}$ with $k_0 \geq 1$ and let $1 \leqslant \widetilde{N} \leqslant \widetilde{q}$ such that $N \equiv \widetilde{N} \pmod{\widetilde{q}}$, that is, $\widetilde{N} = 2^{\ell_0}$ with $\ell_0 = \ell$ if $\ell \leqslant k_0$ and $\ell_0 = k_0$ otherwise. Then

$$\sum_{n=M+1}^{M+N} \chi(n) = \sum_{n=M+1}^{M+N} \widetilde{\chi}(n) = \sum_{n=M+1}^{M+\widetilde{N}} \widetilde{\chi}(n)$$

since complete sums of characters (of length $\widetilde{q}$ in this case) vanish. By [BS19, Theorem 2.1], there is an effective constant $\xi_0 > 0$ such that

$$\sum_{n=M+1}^{M+\widetilde{N}} \widetilde{\chi}(n) \ll \widetilde{N}^{1-\xi_0\ell_0{}^2/k_0^2}$$

and the implied constant is absolute. Put $\xi = \min\{1/2, \xi_0\}$.

Assuming first that $\ell_0 = \ell$, then clearly $\widetilde{N}^{1-\xi{\ell_0}^2/k_0^2} \leqslant N^{1-\xi\ell^2/k^2}$, so we are done. Now assume that $k_0 < \ell$ and $\ell_0 = k_0$. Then we have

$$k_0(1-\xi) \leqslant k_0(1-\xi\ell^2/k^2) < \ell(1-\xi\ell^2/k^2)$$

and therefore once again $\widetilde{N}^{1-\xi{\ell_0}^2/k_0^2} \leqslant N^{1-\xi\ell^2/k^2}$. □

2.5. **Smooth numbers with some bits prescribed.** Using techniques from [GS08] and modifying the proof of [Shp06, Theorem 6], we prove the following

Lemma 2.12. *Let $A > 1$ and $\varepsilon > 0$ be fixed real numbers. with $2\varepsilon < 1 - 1/A$. Then for sufficiently large n and any binary string σ of length*

$$m \leqslant n_0 \left(1 - \left(\frac{(1-1/A)\,(1/A+2\varepsilon)}{\xi\,(1-1/A-2\varepsilon)}\right)^{1/3}\right),$$

where $n_0 = \lfloor (1/2 - 1/2A - \varepsilon)n \rfloor$ and ξ is the constant from Lemma 2.11, there exist at least $2^{n(1-1/A)-m+o(n)}$ odd n-bit integers which are n^A-smooth and have the bit pattern σ at the positions $n_0 - 1, \ldots, n_0 - m$.

Proof. Let $\mathcal{W}$ be the set of odd n^A-smooth numbers in the interval $(2^{(n-1)/2}, 2^{n/2}]$ and let s denote the number defined by the binary string σ. We count the number $T(k)$ of solutions $w_1, w_2 \in \mathcal{W}$ of $w_1 w_2 \equiv 2^{n_0-m}s + k \pmod{2^{n_0}}$ for $0 \leqslant k < 2^{n_0-m}$. Then the product $w_1 w_2$ is n^A-smooth and has n bits as well as the desired bit pattern.

Let $\mathcal{X}$ be the set of multiplicative characters modulo 2^{n_0}. As in [GS08], recalling the orthogonality relation

$$\sum_{\chi\in\mathcal{X}} \chi(u) = \begin{cases} 0 & \text{if } u \not\equiv 1 \pmod{2^{n_0}}, \\ 2^{n_0-1} & \text{if } u \equiv 1 \pmod{2^{n_0}}, \end{cases} \tag{2.1}$$

see, for example, [IK04, Section 3.2], we can express $T(k)$ as

$$T(k) = \frac{1}{2^{n_0-1}} \sum_{w_1,w_2\in\mathcal{W}} \sum_{\chi\in\mathcal{X}} \chi\left((2^{n_0-m}s + k) w_1^{-1} w_2^{-1}\right),$$

where the inverses are taken modulo 2^{n_0} (recall that $w_1, w_2 \in \mathcal{W}$ are odd). Also observe that $T(k) = 0$ if k is even.

If k is odd, separating the contribution from the principal character χ_0 and changing the order of summation, we obtain

$$\begin{aligned} T(k) &= \frac{(\#\mathcal{W})^2}{2^{n_0-1}} + \frac{1}{2^{n_0-1}} \sum_{\substack{\chi\in\mathcal{X}\\ \chi\neq\chi_0}} \chi(2^{n_0-m}s + k) \sum_{w_1,w_2\in\mathcal{W}} \chi\left(w_1^{-1} w_2^{-1}\right) \\ &= \frac{(\#\mathcal{W})^2}{2^{n_0-1}} + \frac{1}{2^{n_0-1}} \sum_{\substack{\chi\in\mathcal{X}\\ \chi\neq\chi_0}} \chi(2^{n_0-m}s + k) \left(\sum_{w\in\mathcal{W}} \chi\left(w^{-1}\right)\right)^2. \end{aligned}$$

This yields

$$\sum_{k=0}^{2^{n_0-m}-1} T(k) = \frac{(\#\mathcal{W})^2}{2^m} + \Delta, \tag{2.2}$$

where, using $\chi\left(w^{-1}\right) = \overline{\chi}(w)$, the error term Δ is given by

$$\Delta = \frac{1}{2^{n_0-1}} \sum_{\substack{\chi \in \mathcal{X} \\ \chi \neq \chi_0}} \sum_{k=0}^{2^{n_0-m}-1} \chi(2^{n_0-m}s+k) \left(\sum_{w \in \mathcal{W}} \overline{\chi}(w) \right)^2 .$$

Now we estimate Δ, again following [GS08]. Namely, we have

$$|\Delta| \leqslant \frac{1}{2^{n_0-1}} \sum_{\substack{\chi \in \mathcal{X} \\ \chi \neq \chi_0}} \left| \sum_{k=0}^{2^{n_0-m}-1} \chi(2^{n_0-m}s+k) \right| \left| \sum_{w \in \mathcal{W}} \chi(w) \right|^2$$

by the triangle inequality. By Lemma 2.11, we have the estimate

$$\sum_{k=0}^{2^{n_0-m}-1} \chi(2^{n_0-m}s+k) \ll 2^{(n_0-m)(1-\xi(n_0-m)^2/n_0^2)}$$

for any non-principal character $\chi \in \mathcal{X}$ and n_0 sufficiently large. Moreover, by our choice of m and n_0, we may further estimate

$$\begin{aligned} n_0 - m &\geq \frac{n_0}{\xi^{1/3}} \left(\frac{(1-1/A)\,(1/A+2\varepsilon)}{1-1/A-2\varepsilon} \right)^{1/3} \\ &= \frac{n_0}{\xi^{1/3}} \left(\frac{1}{1-1/A-2\varepsilon} - 1 \right)^{1/3} (1-1/A)^{1/3} \\ &\geq \frac{n_0}{\xi^{1/3}} \left(\frac{n}{2n_0} - 1 + o(1) \right)^{1/3} (1-1/A)^{1/3} \\ &= \frac{n_0^{2/3}}{\xi^{1/3}} (n/2-n_0)^{1/3} (1-1/A)^{1/3} + o(n) \end{aligned}$$

and therefore the estimate above becomes

$$\sum_{k=0}^{2^{n_0-m}-1} \chi(2^{n_0-m}s+k) \ll 2^{n_0-m-(n/2-n_0)(1-1/A)+o(n)} .$$

Thus, we can further estimate

$$\begin{aligned} |\Delta| &\ll 2^{-m-(n/2-n_0)(1-1/A)+o(n)} \sum_{\substack{\chi \in \mathcal{X} \\ \chi \neq \chi_0}} \left| \sum_{w \in \mathcal{W}} \chi(w) \right|^2 \\ &\ll 2^{-m-(n/2-n_0)(1-1/A)+o(n)} \sum_{\chi \in \mathcal{X}} \left| \sum_{w \in \mathcal{W}} \chi(w) \right|^2 \\ &= 2^{-m-(n/2-n_0)(1-1/A)+o(n)} \sum_{\chi \in \mathcal{X}} \sum_{w_1, w_2 \in \mathcal{W}} \chi(w_1 w_2^{-1}) . \end{aligned}$$

To estimate the last sum, we change the order of summation and use the orthogonality relations (2.1). Namely, by Lemma 2.8, there are at most $2^{(n/2-n_0)(1-1/A)+o(n)}$ elements

of $\mathcal{W}$ in any given residue class modulo 2^{n_0}, so we obtain

$$\sum_{w_1,w_2\in\mathcal{W}}\sum_{\chi\in\mathcal{X}}\chi(w_1w_2^{-1}) = 2^{n_0-1}\cdot\#\{(w_1,w_2)\in\mathcal{W}^2 : w_1\equiv w_2 \pmod{2^{n_0}}\}$$
$$\leqslant 2^{n_0+(n/2-n_0)(1-1/A)+o(n)}\#\mathcal{W}\,.$$

Therefore, we overall get

$$|\Delta| \ll 2^{n_0-m+o(n)}\#\mathcal{W}\,.$$

To compare the error term with the main term, we need to determine $\#\mathcal{W}$. To do this, observe that for any x, there is a bijection

$$\{w\in[1,x/2] :\ w\ n^A\text{-smooth}\} \overset{1:1}{\longleftrightarrow} \{w\in[1,x] :\ w \text{ even},\ n^A\text{-smooth}\}$$
$$w \mapsto 2w$$

and thus the number of odd n^A-smooth numbers up to x is given by $\Psi(x,n^A)-\Psi(x/2,n^A)$. Therefore, Lemma 2.6 yields

$$\begin{aligned}\#\mathcal{W} &= \big(\Psi(2^{n/2},n^A)-\Psi(2^{n/2-1},n^A)\big)\\ &\qquad -\big(\Psi(2^{(n-1)/2},n^A)-\Psi(2^{(n-1)/2-1},n^A)\big)\\ &\sim \Psi(2^{(n-1)/2-1},n^A)\left(2^{3(1-1/A)/2}-2^{1-1/A}-2^{(1-1/A)/2}+1\right)\end{aligned}$$

and due to

$$2^{3(1-1/A)/2}-2^{1-1/A}-2^{(1-1/A)/2}+1 = (2^{(1-1/A)/2}-1)^2(2^{(1-1/A)/2}+1)\,,$$

this is a positive proportion of $\Psi(2^{(n-1)/2-1},n^A)$. Moreover, according to Lemma 2.7,

$$\begin{aligned}\Psi(2^{(n-1)/2-1},n^A) &\geq \Psi(2^{(n-3)/2},(n-3)^A\log^A 2/2^A)\\ &= 2^{(n-3)(1-1/A+o(1))/2} = 2^{n(1-1/A+o(1))/2}\,.\end{aligned}$$

Combining both results, we deduce that

$$\#\mathcal{W} \geq 2^{n(1-1/A+o(1))/2}\,. \tag{2.3}$$

Therefore, the ratio of the error term to the main term in (2.2) can be bounded by

$$\begin{aligned}\frac{|\Delta|}{(\#\mathcal{W})^2/2^m} &\ll \frac{2^{n_0-m+o(n)}\#\mathcal{W}}{(\#\mathcal{W})^2/2^m} = \frac{2^{n_0+o(n)}}{\#\mathcal{W}} \leqslant \frac{2^{n_0+o(n)}}{2^{n(1-1/A)/2}}\\ &\leqslant \frac{2^{n(1-1/A-2\varepsilon)/2+o(n)}}{2^{n(1-1/A)/2}} = 2^{-\varepsilon n+o(n)} = o(1)\,.\end{aligned}$$

Thus, putting this result back into (2.2), we obtain

$$\sum_{k=0}^{2^{n_0-m}-1} T(k) = \frac{(\#\mathcal{W})^2}{2^m}(1+o(1))\,.$$

Using (2.3) again, we see that this becomes

$$\sum_{k=0}^{2^{n_0-m}-1} T(k) \geq \frac{2^{n(1-1/A+o(1))}}{2^m}(1+o(1)) = \frac{2^{n(1-1/A+o(1))}}{2^m}\,.$$

To conclude, it still remains to check how many pairs (w_1, w_2) yield the same product $w = w_1 w_2$. However, any such product w satisfies $w \leqslant 2^n$ by construction and hence, by Lemma 2.1, it can occur at most $2^{o(n)}$ times. Therefore, the number of pairwise distinct products is at least

$$2^{o(n)} \sum_{k=0}^{2^{n_0-m}-1} T(k) \geq \frac{2^{n(1-1/A+o(1))}}{2^m},$$

as claimed. □

3. Proof of Theorem 1.1

Let $\varepsilon > 0$. We set

$$n_0 = \lfloor (1/2 - 1/2A - \varepsilon)n \rfloor \qquad \text{and} \qquad m = \left\lfloor n_0 \left(1 - \left(\frac{(1-1/A)\,(1/A+2\varepsilon)}{\xi\,(1-1/A-2\varepsilon)}\right)^{1/3}\right)\right\rfloor.$$

It is also convenient to define

$$\mu = m/n$$

and note that there are some constants $c_2 > c_1 > 0$ depending only on A, but not on ε such that for a sufficiently small ε,

$$(\xi A)^{-1/3} + c_1\varepsilon \leqslant \left(\frac{(1-1/A)\,(1/A+2\varepsilon)}{\xi\,(1-1/A-2\varepsilon)}\right)^{1/3} \leqslant (\xi A)^{-1/3} + c_2\varepsilon.$$

Note that the positivity of c_1 is crucial for us and in particular, this implies that for some constants $C_2 > C_1 > 0$ depending only on A, we have

$$\mu_0(A) - C_2\varepsilon \leqslant \mu \leqslant \mu_0(A) - C_1\varepsilon, \tag{3.1}$$

where $\mu_0(A)$ is given by (1.2) with $\zeta = \xi^{-1/3}$.

Applying Lemma 2.12 with the above values of n_0 and m, we see that the number T of odd n-bit n^A-smooth numbers with a string of m zeros at the positions $n_0-1, \ldots, n_0-m$ is at least

$$T \geq \frac{2^{n(1-1/A+o(1))}}{2^m} = 2^{(n-m)(1-1/(1-\mu)A)+o(n)}. \tag{3.2}$$

Now consider the $n-m$ bits we have not prescribed yet. If at most a proportion $\rho < \vartheta_0(A)$ of them are zeros, where $\vartheta_0(A)$ is given by (1.3), then, by Lemma 2.9, we can have at most

$$\sum_{0 \leqslant k \leqslant \rho(n-m)} \binom{n-m}{k} \leqslant 2^{(n-m)H(\rho)}$$

different integers. However, since $\rho < \vartheta_0(A)$, we know that

$$H(\rho) < 1 - \frac{1}{(1-\mu_0(A))A} < 1 - \frac{1}{(1-\mu)A}.$$

Recalling (3.2), we conclude that

$$2^{(n-m)H(\rho)} < T$$

if n is sufficiently large.

Thus, at least one of the T odd n-bit n^A-smooth numbers not only has a string of m consecutive zeros, which is by construction, but also has at least a proportion of ρ zeros among the remaining $n-m$ digits. That is, we see from (3.1) that its binary expansion contains at least

$$\begin{aligned} m+\rho(n-m) &= (1-\rho)m+\rho n = n\left(\mu(1-\rho)+\rho\right) \\ &\geqslant n\left(\mu_0(A)(1-\rho)+\rho-C_2(1-\rho)\varepsilon\right) \\ &\geqslant n\left(\mu_0(A)+\rho(1-\mu_0(A))-C_2(1-\rho)\varepsilon\right) \end{aligned}$$

zeros. Choosing $\rho>\vartheta$ concludes the proof since ε is arbitrarily small.

4. Proof of Theorem 1.2

Put $k=p_2\cdots p_r$ for some $r>2$ and consider $m=2^k+1$. We assume that $r\to\infty$, so we also have $k\to\infty$.

We have the following factorisation into cyclotomic polynomials:

$$m=2^k+1=-\prod_{d\mid k}\Phi_d(-2)$$

and we can bound the factors using Lemma 2.2 and Lemma 2.1 as

$$\begin{aligned} |\Phi_d(-2)| &\ll A_d 2^{\varphi(d)} \leqslant \exp\left(\frac{1}{2}\tau(d)\log d\right)2^{\varphi(d)} \\ &\ll \exp\left(k^{o(1)}\log k\right)2^{\varphi(d)} \ll 2^{\varphi(d)+k^{o(1)}} \leqslant 2^{\varphi(k)+k^{o(1)}}. \end{aligned}$$

Moreover, due to our choice of k, Lemma 2.3 implies

$$|\Phi_d(-2)| \ll 2^{(2e^{-\gamma}+o(1))k/\log\log k} < m^{(2e^{-\gamma}+o(1))/\log\log k} = m^{(2e^{-\gamma}+o(1))/\log\log\log m}.$$

In particular, m is $m^{(2e^{-\gamma}+o(1))/\log\log\log m}$-smooth.

Now, we consider the number

$$N=m^{\ell}, \tag{4.1}$$

where $\ell=\lfloor\beta k\rfloor$ for $\beta=2\alpha\log 2$. Without loss of generality we assume that k is large enough so $\ell\geqslant 2$. Thus N has

$$n=k\ell+O(\ell)=\beta^{-1}\ell^2+O(\ell) \tag{4.2}$$

bits, provided that $k\to\infty$. Due to

$$\ell\sim\beta k\sim\frac{\beta\log m}{\log 2}=\frac{2\alpha\log N}{\ell}$$

we also have $\ell\sim(2\alpha\log N)^{1/2}$. In particular,

$$\begin{aligned} \log\log\log N &= \log\log\left(\ell\log m\right) \leqslant \log\log\left((2\log m)^2\right) \\ &= \log\left(2\left(\log\log m+O(1)\right)\right) = (1+o(1))\log\log\log m. \end{aligned}$$

Therefore, we see that

$$\begin{aligned} m^{(2e^{-\gamma}+o(1))/\log\log\log m} &= N^{(2e^{-\gamma}+o(1))/(\ell\log\log\log N)} \\ &= N^{\left(2e^{-\gamma}+o(1)\right)/\left((2\alpha\log N)^{1/2}\log\log\log N\right)} = Y^{1+o(1)}, \end{aligned}$$

that is, N is $Y^{1+o(1)}$-smooth.

We now estimate the sparsity of N. By the binomial theorem, we know that

$$N = (2^k + 1)^\ell = \sum_{j=0}^{\ell} \binom{\ell}{j} 2^{kj}.$$

Noting that $\binom{\ell}{j}$ can have at most $\log \binom{\ell}{j} / \log 2 + 1$ binary digits and using the subadditivity of the sum of binary digits guaranteed by Lemma 2.4, we see that Lemma 2.10 shows that the total number of non-zero digits of N is at most

$$\begin{aligned} \frac{1}{\log 2} \sum_{j=0}^{\ell} \log \binom{\ell}{j} + (\ell + 1) &= \frac{1}{2 \log 2} \ell^2 + O(\ell \log \ell) \\ &= \alpha n + O\left(n^{1/2} \log n\right), \end{aligned}$$

where we used that $\ell^2 = \beta n + O\left(n^{1/2}\right)$ due to (4.2) in the last step. This proves the claim.

5. Proof of Theorem 1.3

We proceed exactly as in the proof of Theorem 1.2 up to the point where we defined N in (4.1). If $\alpha = 0$, we choose $\ell = 1$ and are done; otherwise, let $\ell = \lfloor k^\beta \rfloor$ for

$$\beta = \frac{\alpha}{2 - \alpha}.$$

Without loss of generality we assume that k is large enough so $\ell \geqslant 2$. Thus N has

$$n = k\ell + O(\ell) = \ell^{1+1/\beta} + O\left(\ell^{1/\beta}\right) \tag{5.1}$$

bits, provided that $k \to \infty$. Due to

$$\ell \sim k^\beta \sim \left(\frac{\log m}{\log 2}\right)^\beta = \left(\frac{\log N}{\ell \log 2}\right)^\beta$$

we also have $\ell \sim (\log N / \log 2)^{\beta/(1+\beta)}$. In particular,

$$\begin{aligned} \log\log\log N &= \log\log(\ell \log m) \leqslant \log\log\left((2 \log m)^{1+\beta}\right) \\ &= \log\left((1+\beta)(\log\log m + O(1))\right) \\ &= (1 + o(1)) \log\log\log m. \end{aligned}$$

Therefore, we see that

$$\begin{aligned} m^{(2e^{-\gamma}+o(1))/\log\log\log m} &= N^{(2e^{-\gamma}+o(1))/(\ell \log\log\log N)} \\ &= N^{\left(2e^{-\gamma}+o(1)\right)/\left((\log N/\log 2)^{\beta/(1+\beta)} \log\log\log N\right)} = Y^{1+o(1)} \end{aligned}$$

that is, N is $Y^{1+o(1)}$-smooth.

As before, we see that the number of non-zero digits of N is at most

$$\frac{1}{2\log 2}\ell^2 + O(\ell \log \ell) = \frac{1}{2\log 2} n^{2\beta/(1+\beta)} + O(n^{\beta/(1+\beta)} \log n)$$
$$= \frac{1}{2\log 2} n^{\alpha} + O(n^{\alpha/2} \log n)$$

due to $\ell = n^{\beta/(1+\beta)} + O(1)$ by (5.1) and this concludes the proof.

6. Comments

It is also interesting to study smooth values amongst balanced integers, that is, positive integers with equally many ones and zeros in their binary expansion.

Using simple counting arguments, one can show that for any integer n, there exist $2n$-bit balanced integers which are Y-smooth, where

$$Y = 2^{2n - \left(1/\sqrt{\pi} + o(1)\right) n^{1/2}}. \tag{6.1}$$

Indeed, first we observe that by the Stirling formula, the number of $2n$-bit balanced integers is given by

$$\binom{2n-1}{n-1} = \left(1/2\sqrt{\pi} + o(1)\right) \frac{4^n}{n^{1/2}}. \tag{6.2}$$

However, for any $u \in [1, 2]$,

$$\Psi(x, x^{1/u}) = (1 - \log u)\, x + O(x/\log x),$$

see [Gra08, Equations (1.3) and (1.8)]. Thus, if $u = 1 + \eta$ with $\eta \asymp (\log x)^{-1/2}$, then the number of non-$x^{1/u}$-smooth numbers $N \leqslant x$ can be estimated as

$$x - \Psi(x, x^{1/u}) = (1 + o(1))\eta x. \tag{6.3}$$

Hence, for a sufficiently small $\varepsilon > 0$ with $\eta = (1/2\sqrt{\pi} - \varepsilon)\, n^{-1/2}$ and sufficiently large n, we obtain from (6.2) and (6.3) that

$$4^n - \Psi\left(4^n, 4^{n/\left(1 + (1/2\sqrt{\pi} - \varepsilon) n^{-1/2}\right)}\right) < \binom{2n-1}{n-1},$$

and since $\varepsilon > 0$ is arbitrary, we see that we can take Y as in (6.1).

Furthermore, with a similar technique as in the proofs of Theorem 1.2 and Theorem 1.3, one can give an explicit construction of infinitely many rather smooth balanced numbers. Namely, taking

$$k_1 = p_2 \cdots p_r \qquad \text{and} \qquad k_2 = p_3 \cdots p_r = k_1/3,$$

the same argument as above shows that $2^{k_1} + 1$ is $2^{(2e^{-\gamma} + o(1))k_1/\log\log k_1}$-smooth and $2^{k_2} - 1$ is $2^{(2e^{-\gamma} + o(1))k_2/\log\log k_2}$-smooth. Thus, the number $N = (2^{k_1} + 1)(2^{k_2} - 1)$ is $2^{(2e^{-\gamma} + o(1))k_1/\log\log k_1} = N^{(3e^{-\gamma}/2 + o(1))/\log\log\log N}$-smooth and it has $n = k_1 + k_2 = 4k_2$ total digits, exactly $n/2$ of which are ones.

Acknowledgements

The authors are very grateful to Regis de la Bretèche for suggesting to use results from [Har16] and to the referee for the very careful reading of the manuscript.

This work started during a very enjoyable visit by the authors to the Max Planck Institute for Mathematics, Bonn, whose hospitality and support is very much appreciated.

During the preparation of this work I.S. was also supported in part by the Australian Research Council Grant DP200100355.

References

[Bat49] P. T. Bateman, 'Note on the coefficients of the cyclotomic polynomial', *Bull. Amer. Math. Soc.* **55** (1949), 1180–1181. 3

[BS19] W. Banks and I. E. Shparlinski, 'Bounds on short character sums and L-functions with characters to a powerful modulus', *J. d'Anal. Math.* **139** (2019), 239–263. 2, 6

[Bou05] J. Bourgain, 'Estimates on exponential sums related to Diffie–Hellman distributions', *Geom. Funct. Anal.* **15** (2005), 1–34. 1

[Bou13] J. Bourgain, 'Prescribing the binary digits of primes', *Israel J. Math.* **194** (2013), 935–955. 1

[Bou15] J. Bourgain, 'Prescribing the binary digits of primes, II', *Israel J. Math.* **206** (2015), 165–182. 1

[Bug18] Y. Bugeaud, 'On the digital representation of integers with bounded prime factors', *Osaka J. Math.* **55** (2018), 315–324. 1

[Bug21] Y. Bugeaud, 'On the Zeckendorf representation of smooth numbers', *Mosc. Math. J.* **21** (2021), 31–42. 1

[BK18] Y. Bugeaud and H. Kaneko, 'On the digital representation of smooth numbers, *Math. Proc. Cambridge Philos. Soc.* **165** (2018), 533–540. 1

[CKS18] M.-C. Chang, B. Kerr and I. E. Shparlinski, 'On the exponential large sieve inequality for sparse sequences modulo primes, *J. Math. Anal. Appl.* **459** (2018) 53–81. 1

[Col09] S. Col, 'Palindromes dans les progressions arithmétiques, *Acta Arith.* **137** (2009), 1–41. 1

[DES13a] R. Dietmann, C. Elsholtz and I. E. Shparlinski, 'On gaps between primitive roots in the Hamming metric', *Quart. J. Math.* **64** (2013), 1043–1055. 1, 2

[DES13b] R. Dietmann, C. Elsholtz and I. E. Shparlinski, 'On gaps between quadratic non-residues in the Euclidean and Hamming metrics', *Indag. Math.* **24** (2013), 930–938. 1, 2

[DES17] R. Dietmann, C. Elsholtz and I. E. Shparlinski, 'Prescribing the binary digits of squarefree numbers and quadratic residues', *Trans. Amer. Math. Soc.* **369** (2017), 8369–8388. 1

[DMR20] M. Drmota, C. Mauduit and J. Rivat, 'Prime numbers in two bases', *Duke Math. J.* **169** (2020), 1809–1876. 1

[Gra08] A. Granville, 'Smooth numbers: Computational number theory and beyond', *Proc. MSRI Conf. Algorithmic Number Theory: Lattices, Number Fields, Curves, and Cryptography, Berkeley 2000*, Cambridge Univ. Press, 267–323. 1, 5, 13

[GS08] S. W. Graham and I. E. Shparlinski, 'On RSA moduli with almost half of the bits prescribed', *Discr. Appl. Math.* **56** (2008), 3150–3154. 1, 2, 7, 8

[HLS11] K. G. Hare, S. Laishram and T. Stoll, 'Stolarsky's conjecture and the sum of digits of polynomial values', *Proc. Amer. Math. Soc.* **139** (2011), 39–49. 4

[Har16] A. J. Harper, 'Minor arcs, mean values, and restriction theory for exponential sums over smooth numbers', *Compos. Math.* **152** (2016), 1121–1158. 5, 14

[HT86] A. Hildebrand and G. Tenenbaum, 'On integers free of large prime factors', *Trans. Amer. Math. Soc.* **296** (1986), 265–290. 1, 4

[Iwa74] H. Iwaniec, 'On zeros of Dirichlet's L series', *Invent. Math.* **23** (1974), 97–104. 2

[IK04] H. Iwaniec and E. Kowalski, *Analytic number theory* Amer. Math. Soc., Providence, RI, 2004. 3, 7

[Kar22] F. Karwatowski, 'Primes with one excluded digit', *Acta Arith.* **202** (2022), 105–121. 1

[Mer74] F. Mertens, 'Ein Beitrag zur analytischen Zahlentheorie', *J. Reine Angew. Math.* **78** (1874), 46–62. 4

[MS77] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*, North-Holland, Amsterdam, 1977. 5

[MR09] C. Mauduit and J. Rivat, 'La somme des chiffres des carrés', *Acta Math.* **203** (2009), 107–148. 1

[MR10] C. Mauduit and J. Rivat, 'Sur un problème de Gelfond: la somme es chiffres des nombres premiers', *Ann. of Math.* **171** (2010), 1591–1646. 1

[May19] J. Maynard, 'Primes with restricted digits', *Invent. Math.* **217** (2019), 127–218. 1

[May22] J. Maynard, 'Primes and polynomials with restricted digits', *Int. Math. Res. Not.* **2022** (2022), 10626–10648. 1

[Meng13] X. Meng, 'On RSA moduli with half of the bits prescribed', *J. Number Theory* **133** (2013), 105–109. 1

[Nas15] E. Naslund, 'The tail distribution of the sum of digits of prime numbers', *Unif. Distrib. Theor.* **10** (2015), 63–68. 1, 2

[Pratt20] K. Pratt, 'Primes from sums of two squares and missing digits,*Proc. Lond. Math. Soc.* **120** (2020), 770–830. 1

[Shp87] I. E. Shparlinski, 'On primitive polynomials', *Problemy Peredachi Inform.* **23** (3) (1987), 100–103 (in Russian). 2

[Shp06] I. E. Shparlinski, 'On RSA moduli with prescribed bit patterns', *Designs, Codes and Cryptography* **39** (2006), 113–122. 1, 2, 7

[Shp08] I. E. Shparlinski, 'Exponential sums and prime divisors of sparse integers', *Period. Math. Hungar.* **57** (2008), 93–99. 1

[Swa20] C. Swaenepoel, 'Prime numbers with a positive proportion of preassigned digits, *Proc. Lond. Math. Soc.* **121** (2020), 83–151. 1

MH: Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany
Email address: max.hauck01@gmail.com

IES: Department of Pure Mathematics, University of New South Wales, Sydney, NSW 2052, Australia
Email address: igor.shparlinski@unsw.edu.au