COMPOSITIO MATHEMATICA

The level of distribution of the Thue–Morse sequence

Lukas Spiegelhofer

Compositio Math. **156** (2020), 2560–2587.

doi:10.1112/S0010437X20007563

FOUNDATION COMPOSITIO MATHEMATICA

LONDON MATHEMATICAL SOCIETY
EST. 1865

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

CrossMark

Compositio Math. 156 (2020) 2560–2587
doi:10.1112/S0010437X20007563

The level of distribution of the Thue–Morse sequence

Lukas Spiegelhofer

Abstract

The level of distribution of a complex-valued sequence b measures the quality of distribution of b along sparse arithmetic progressions $nd + a$. We prove that the Thue–Morse sequence has level of distribution 1, which is essentially best possible. More precisely, this sequence gives one of the first nontrivial examples of a sequence satisfying a Bombieri–Vinogradov-type theorem for each exponent $\theta < 1$. This result improves on the level of distribution 2/3 obtained by Müllner and the author. As an application of our method, we show that the subsequence of the Thue–Morse sequence indexed by $\lfloor n^c \rfloor$, where $1 < c < 2$, is *simply normal*. This result improves on the range $1 < c < 3/2$ obtained by Müllner and the author and closes the gap that appeared when Mauduit and Rivat proved (in particular) that the Thue–Morse sequence along the squares is simply normal.

1. Introduction

The Thue–Morse sequence $\mathbf{t}$ is one of the most easily defined automatic sequences. Like any automatic sequence, it can be defined using a *uniform morphism* over a finite alphabet: $\mathbf{t}$ is the unique fixed point of the substitution $\mathtt{0} \mapsto \mathtt{01}$, $\mathtt{1} \mapsto \mathtt{10}$ that starts with $\mathtt{0}$. Therefore, $\mathbf{t} = (\mathtt{0110100110010110}\ldots)$. Alternatively, this sequence can be defined using the *binary sum-of-digits function* s, which counts the number of $\mathtt{1}$s in the binary expansion of a nonnegative integer n: we have $\mathbf{t}(n) = \mathtt{0}$ if and only if $s(n) \equiv 0 \bmod 2$. The equivalence of these two definitions can be proved via a third description: start with the one-element sequence $\mathbf{t}^{(0)} := (\mathtt{0})$ and define $\mathbf{t}^{(n+1)}$ by concatenating $\mathbf{t}^{(n)}$ and the Boolean complement $\neg\mathbf{t}^{(n)}$. Then $\mathbf{t}$ is the (pointwise) limit of this sequence of finite words. In this work, we will adopt the second viewpoint. In fact, in the proofs we will work with the sequence $(-1)^{s(n)}$ instead of $\mathbf{t}$, and we also call this sequence the Thue–Morse sequence by slight abuse of notation. When working with exponential sums, we will always use the 'multiplicative version' $(-1)^{s(n)}$. For an overview on the Thue–Morse sequence, we refer the reader to the article by Allouche and Shallit [AS99], which points out occurrences of this sequence in different fields of mathematics and offers a good bibliography.

Received 31 October 2019, accepted in final form 2 June 2020, published online 25 January 2021.
2010 Mathematics Subject Classification 11N69, 11B25, 11B85 (primary), 11A63, 11K16, 11B83 (secondary).
Keywords: Thue–Morse sequence, level of distribution, Bombieri–Vinogradov theorem, Elliott–Halberstam conjecture, arithmetic progression, Piatetski-Shapiro sequence, normal sequence, Gel'fond problem.

The author acknowledges support by the Austrian Science Fund (FWF), Project F5502-N26, which is a part of the Special Research Program 'Quasi Monte Carlo Methods: Theory and Applications'. The author also wishes to acknowledge support by the project MuDeRa, which is a joint project between the FWF (I-1751-N26) and the ANR (Agence Nationale de la Recherche, France, ANR-14-CE34-0009).
This journal is © Foundation Compositio Mathematica 2021.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

We also wish to mention the survey paper [Mau01] by Mauduit on the Thue–Morse sequence. For a comprehensive treatment of automatic and morphic sequences, see the book [AS03] by Allouche and Shallit.

The main topic of this article is the study of $\mathbf{t}$ along arithmetic progressions and, more generally, along Beatty sequences $\lfloor n\alpha+\beta\rfloor$, where α and β are real numbers and $\alpha \geq 0$. This topic can be traced back at least to Gel'fond [Gel68], who proved the following theorem on the base-q sum-of-digits function s_q defined by $s_q(\varepsilon_\nu q^\nu + \cdots + \varepsilon_0 q^0) = \varepsilon_\nu + \cdots + \varepsilon_0$ for $\varepsilon_i \in \{0,\ldots,q-1\}$.

THEOREM A (Gel'fond). *Let* q, m, d, b, a *be integers and* $q, m, d \geq 2$. *Suppose that* $\gcd(m, q-1) = 1$. *Then*

$$|\{1 \leq n \leq x : n \equiv a \bmod d, s_q(n) \equiv b \bmod m\}| = \frac{x}{dm} + \mathcal{O}(x^\lambda)$$

for some $\lambda < 1$ *independent of* x, d, a *and* b.

We are particularly interested in the error term for *sparse* arithmetic progressions, having large common difference d. This leads us directly to the other main concept of this paper, the notion of *level of distribution*. (We use this term in the same way as Goldston *et al.* [GPY09]; the term is also used for a very similar concept by other authors. Moreover, the term *exponent of distribution* is also common.) Very roughly speaking, the level of distribution is a measure of how well a given sequence behaves on arithmetic progressions. A formal definition can be found in the article [FM96a] by Fouvry and Mauduit. We adapt this definition.

DEFINITION 1. Let $c = (c_n)_{n\geq 0}$ be a sequence of complex numbers and, for each integer $d \geq 1$, let $\mathcal{Q}(d)$ and $\mathcal{R}(d) \neq \emptyset$ be subsets of $\mathbb{Z}/d\mathbb{Z}$ such that $\mathcal{Q}(d) \subseteq \mathcal{R}(d)$. The sequence c has *level of distribution* θ with respect to $\mathcal{Q}$ and $\mathcal{R}$ if for all $\varepsilon > 0$ and $A > 0$ we have for all $x \geq 1$ that

$$\sum_{1\leq d\leq D} \max_{0\leq y\leq x} \max_{\substack{0\leq a<d\\ a+d\mathbb{Z}\in\mathcal{Q}(d)}} \left| \sum_{\substack{0\leq n<y\\ n\equiv a \bmod d}} c_n - \frac{1}{|\mathcal{R}(d)|} \sum_{\substack{0\leq n<y\\ n+d\mathbb{Z}\in\mathcal{R}(d)}} c_n \right| \ll (\log 2x)^{-A} \sum_{0\leq n<x} |c_n|,$$

where $D = x^{\theta-\varepsilon}$. The implied constant may depend on A and ε. In this definition, the maximum over the empty index set is defined to be 0.

The most well-known cases are $\mathcal{R}(d) = \mathbb{Z}/d\mathbb{Z}$ or $\mathcal{R}(d) = (\mathbb{Z}/d\mathbb{Z})^*$; the treatment of the main term

$$\frac{1}{|\mathcal{R}(d)|} \sum_{\substack{0\leq n<y\\ n+d\mathbb{Z}\in\mathcal{R}(d)}} c_n$$

is usually the easy part of an estimate as in the definition. In the case of the Bombieri–Vinogradov theorem, we use $\mathcal{R}(d) = (\mathbb{Z}/d\mathbb{Z})^*$, since the prime numbers are distributed evenly in the residue classes relatively prime to d. The summands in Definition 1 measure the maximal deviation of a sum over an arithmetic progression from the expected value, where the maximum is taken over a set $\mathcal{Q}(d)$ of residue classes and the length of the progression may also vary.

The level of distribution is an important concept in sieve theory. As a striking application, a variant of this concept was used in the paper by Zhang [Zha14] on bounded gaps between primes. For more information on this subject, we refer the reader to the survey by Kontorovich [Kon14].

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Moreover, we wish to draw the attention of the reader to the book [FI10] on sieve theory by Friedlander and Iwaniec, in particular Chapter 22 on the level of distribution.

We are ready to present our main result. Note that we use $\mathcal{O}_\varepsilon$ to indicate that the implied constant may depend on ε.

Theorem 1.1. *The Thue–Morse sequence has level of distribution* 1 *with respect to* $\mathcal{Q}$ *and* $\mathcal{R}$ *given by* $\mathcal{Q}(d)=\mathcal{R}(d)=\mathbb{Z}/d\mathbb{Z}$. *More precisely, for all* $\varepsilon>0$, *we have*

$$\sum_{1\le d\le D}\max_{\substack{y,z\ge 0\\ z-y\le x}}\max_{0\le a<d}\left|\sum_{\substack{y\le n<z\\ n\equiv a \bmod d}}(-1)^{s(n)}\right|=\mathcal{O}_\varepsilon(x^{1-\eta})$$

for some $\eta>0$ *depending on* ε, *where* $D=x^{1-\varepsilon}$.

Before presenting some history, we wish to say a word about the proof: we are going to reduce the problem to the estimation of a certain *Gowers uniformity norm* of the Thue–Morse sequence. These expressions appear by repeated application of Van der Corput's inequality and have the form

$$\sum_{\substack{0\le n<2^\rho\\ 0\le r_1,\ldots,r_k<2^\rho}}\prod_{\varepsilon\in\{0,1\}^k}(-1)^{s_\rho(n+\varepsilon\cdot r)},$$

where $\varepsilon\cdot r=\sum_{1\le i\le k}\varepsilon_i r_i$ and s_ρ is the truncated sum-of-digits function in base 2 defined by $s_\rho(n)=s(n \bmod 2^\rho)$. Note that, strictly speaking, this is not the Gowers norm of the Thue–Morse sequence, but the Gowers norm of order k of the projection of $(-1)^{s(n)}$ to $\mathbb{Z}/2^\rho\mathbb{Z}$. The proof of a very similar statement was given recently by Konieczny [Kon19], and we use the proof from that paper to prove our estimate.

Gowers norms are certain averaged multiple correlations and were introduced by Gowers [Gow98, Gow01], who used them to give a new proof of Szemerédi's theorem. These norms are a central tool in *higher order Fourier analysis* [Tao12]; this theory can be used to study questions in additive combinatorics, such as the behaviour of an arithmetic function f on arithmetic progressions $n,n+d,n+2d,\ldots,n+(\ell-1)d$. In the ground-breaking paper [GT08] by Green and Tao, Gowers norms were used to prove the existence of arbitrarily long arithmetic progressions in the primes. Our result is a statement on arithmetic progressions too; although it is different in nature, Gowers norms are applicable here.

In order to put Theorem 1.1 into context, we present some related theorems. The well-known Bombieri–Vinogradov theorem concerns the level of distribution of the von Mangoldt function Λ, which is defined by $\Lambda(n)=\log p$ if $n=p^\ell$ for some prime p and some $\ell\ge 1$ and $\Lambda(n)=0$ otherwise. This theorem states that Λ has level of distribution $1/2$ with respect to $\mathcal{Q}$ and $\mathcal{R}$ given by $\mathcal{Q}(d)=\mathcal{R}(d)=(\mathbb{Z}/d\mathbb{Z})^*$.

Theorem B (Bombieri–Vinogradov). *Let* $d\ge 1$ *and* a *be integers and define*

$$\psi(x;d,a)=\sum_{\substack{1\le n\le x\\ n\equiv a \bmod d}}\Lambda(n).$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

For all real numbers $A>0$ there exist $B>0$ and a constant C such that setting $D=x^{1/2}(\log x)^{-B}$ we have for all $x\geq 2$

$$\sum_{1\leq d\leq D}\max_{1\leq y\leq x}\max_{\substack{0\leq a<d\\ \gcd(a,d)=1}}\left|\psi(y;d,a)-\frac{y}{\varphi(d)}\right|\leq Cx(\log x)^{-A}.$$

Here φ denotes Euler's totient function.

No improvement on the level of distribution $1/2$ in this theorem is currently known. Meanwhile the Elliott–Halberstam conjecture [EH70] states that we can choose $D=x^{1-\varepsilon}$ for any $\varepsilon>0$. That is, it is conjectured that the primes have level of distribution 1. Improvements on the exponent $1/2$ exist for certain sequences of integers; we refer to the articles [Fou82, Fou84] by Fouvry, [FI80] by Fouvry and Iwaniec and [FI85] by Friedlander and Iwaniec. Moreover, we mention the series [BFI86, BFI87, BFI89] by Bombieri *et al.* concerning this topic. In this context, we also note the result of Goldston *et al.* [GPY09], who showed in particular the following conditional result: if the primes have level of distribution θ for some $\theta>1/2$, then there exists a constant C such that $p_{n+1}-p_n<C$ infinitely often, where p_n is the nth prime. In a ground-breaking paper we mentioned before, Zhang [Zha14] used the Goldston *et al.* method and a variant of the Bombieri–Vinogradov theorem to prove the above result unconditionally. Maynard [May15] later proved the bounded gaps result using only the classical Bombieri–Vinogradov theorem.

Improvements on the level $1/2$ are also known for the sum-of-digits function modulo m. Fouvry and Mauduit [FM96b] established 0.5924 as a level of distribution of the Thue–Morse sequence, with respect to $\mathcal{Q}$ and $\mathcal{R}$, where $\mathcal{Q}(d)=\mathcal{R}(d)=\mathbb{Z}/d\mathbb{Z}$.

Theorem C (Fouvry–Mauduit). *Set*

$$A(x;d,a)=|\{0\leq n<x:\mathbf{t}(n)=\mathtt{0},n\equiv a \bmod d\}|.$$

Then

$$\sum_{1\leq d\leq D}\max_{1\leq y\leq x}\max_{0\leq a<d}\left|A(y;d,a)-\frac{y}{2d}\right|\leq Cx(\log 2x)^{-A}\tag{1.1}$$

for all real A and $D=x^{0.5924}$, where C may depend on A.

More generally, for $m\geq 2$, they also studied the sum-of-digits function in base 2 modulo m, obtaining the weaker level of distribution 0.55711. Using sieve theory, they applied this result to the study of the sum of digits modulo m of numbers having at most two prime factors. Later, Mauduit and Rivat [MR10], in an important paper, managed to treat the sum of digits modulo m of prime numbers, thereby answering one of the questions posed by Gel'fond [Gel68].

Müllner and the author [MS17] improved the exponent 0.5924 to $2/3-\varepsilon$, thereby establishing $2/3$ as an admissible level of distribution of the Thue–Morse sequence.

Fouvry and Mauduit [FM96a] also considered, more generally, the sum-of-digits function s_q in base q modulo an integer m such that $\gcd(m,q-1)=1$. They obtained the result that the level of distribution approaches 1 as the base q gets larger.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Theorem D (Fouvry–Mauduit). *Let $q \geq 2$, $m \geq 1$ and b be integers such that* $\gcd(m, q-1) = 1$. *There exists $\theta_q > 0$ such that for all A and $\varepsilon > 0$ we have for all $x \geq 1$*

$$\sum_{1 \leq d \leq D} \max_{0 \leq y \leq x} \max_{0 \leq a < d} \left| \sum_{\substack{n < y, s_q(n) \equiv b \bmod m \\ n \equiv a \bmod d}} 1 - \frac{1}{d} \sum_{n < y, s_q(n) \equiv b \bmod m} 1 \right| = \mathcal{O}_{m,q,A,\varepsilon}(x (\log 2x)^{-A}),$$

where $D = x^{\theta_q - \varepsilon}$. The implied constant depends at most on m, q, A and ε. As $q \to \infty$, the value of θ_q tends to 1.

As an application of this theorem, they considered the sum of digits in base q of integers having at most two prime factors; moreover, they studied the sum $\sum_{n<x, s_q(n) \equiv b \bmod m} \Lambda_\ell(n)$, where Λ_ℓ is the generalized von Mangoldt function of order $\ell \geq 2$ [FM96a, Corollaire 2].

Theorem D motivates us to look for sequences having level of distribution equal to 1. In the paper by Fouvry and Mauduit [FM96a] cited above, for example, a list of sequences having this property is given. Also, we note [FI10, Chapter 22.3], which studies the level of distribution for additive convolutions, giving further examples. However, in these examples, other than the trivial example $c_n = 1$ for all n, the maximum over a does not play a rôle: the set $\mathcal{Q}(d)$ consists of at most one element.

We are interested in sequences c having level of distribution 1 and such that the set $\mathcal{Q}(d)$ contains 'many' residue classes. In other words, we want to find analogues of the Elliott–Halberstam conjecture. Requiring monotonicity of c, examples can be constructed easily: $c(n) = n$ is such an example and, more generally, increasing sequences c satisfying certain growth conditions have this property. Apart from such 'trivial' sequences, no other examples seem to be known. Our Theorem 1.1, giving such an example, might therefore be of interest.

We believe that our method can be adapted to $s_q(n) \bmod m$ for all $m \geq 1$ and general bases $q \geq 2$, which would yield $\theta_q = 1$ for all $q \geq 2$ in Theorem D.

The second focus of this paper concerns *Piatetski-Shapiro sequences*, which are sequences of the form $(\lfloor n^c \rfloor)_{n \geq 0}$ for some $c \geq 1$. In order to state the second main theorem, we do not need additional preparation.

Theorem 1.2. *Let $1 < c < 2$. The Thue–Morse sequence along $\lfloor n^c \rfloor$ is simply normal. That is, each of the letters* 0 *and* 1 *appears with asymptotic frequency $1/2$ in $n \mapsto \mathbf{t}(\lfloor n^c \rfloor)$.*

In our earlier paper [MS17] with Müllner, this theorem is proved via a Beatty sequence variant of Theorem 1.1. That theorem in turn is proved by arguments analogous to the arguments in the proof of Theorem 1.1 and reduces to the same estimate of the Gowers uniformity norm of Thue–Morse. Theorem 1.2 is therefore an application of the method of proof of Theorem 1.1.

Again, we present some historical background. Studying Piatetski-Shapiro subsequences of a given sequence can be seen as a step towards proving theorems on polynomial subsequences. For example, it is unknown whether there are infinitely many primes of the form $n^2 + 1$; therefore, it is of interest to consider primes of the form $\lfloor n^c \rfloor$ for $1 < c < 2$ and prove an asymptotic formula for the number of such primes. Piatetski-Shapiro [Pia53] proved such a formula for $1 < c < 12/11$, and the currently best known bound is $1 < c < 2817/2426$ due to Rivat and Sargos [RS01]. In a similar way, the study of the sum-of-digits function along $\lfloor n^c \rfloor$ can be justified. It is another problem posed by Gel'fond [Gel68] to study the distribution of the sum of digits of polynomial sequences in residue classes. Since this problem could not be solved at first, Mauduit and Rivat

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

[MR95, MR05] considered q-multiplicative functions along $\lfloor n^c \rfloor$ (where a q-multiplicative function $f : \mathbb{N} \to \{z \in \mathbb{C} : |z| = 1\}$ satisfies $f(aq^m + b) = f(aq^m)f(b)$ for nonnegative integers a, b, m such that $b < q^m$) and they obtained an asymptotic formula for $c < 7/5$.

Theorem E (Mauduit–Rivat). *Let $1 < c < 7/5$ and set $\gamma = 1/c$. For all $\delta \in (0, (7-5c)/9)$ there exists a constant $C > 0$ such that for all q-multiplicative functions $f : \mathbb{N} \to \{z \in \mathbb{C} : |z| = 1\}$ and all $x \geq 1$ we have*

$$\left| \sum_{1 \leq n \leq x} f(\lfloor n^c \rfloor) - \sum_{1 \leq m \leq x^c} \gamma m^{\gamma - 1} f(m) \right| \leq C x^{1-\delta}.$$

Since the Thue–Morse sequence is 2-multiplicative, it follows in particular that the subsequence indexed by $\lfloor n^c \rfloor$ assumes each of the two values $\mathtt{0}$, $\mathtt{1}$ with asymptotic frequency $1/2$, as long as $1 < c < 7/5$. This means that this subsequence is simply normal. In the paper [DDM12] by Deshouillers *et al.*, a statement as in Theorem E for arbitrary automatic sequences and $1 < c < 7/5$ is proved. Moreover, we wish to note the paper [Mor11] by Morgenbesser, who proved uniform distribution of $s_q(\lfloor n^c \rfloor)$ in residue classes for *all* noninteger $c > 0$, as long as the base q is large enough (depending on c).

Some progress on Gel'fond's question on polynomials was made by Drmota and Rivat [DR05] and by Dartyge and Tenenbaum [DT06]; finally, Mauduit and Rivat [MR09] managed to answer Gel'fond's question for the polynomial n^2. This latter paper was generalized by Drmota *et al.* [DMR19], who showed that in fact $\mathbf{t}(n^2)$ defines a *normal sequence*, by which we understand an infinite sequence on $\{\mathtt{0}, \mathtt{1}\}$ such that every finite sequence of length L occurs as a factor (contiguous finite subsequence) with asymptotic frequency 2^{-L}. This result also generalizes a paper by Moshe [Mos07], who showed that every finite word on $\{\mathtt{0}, \mathtt{1}\}$ occurs as a factor of $n \mapsto \mathbf{t}(n^2)$ at least once.

However, the distribution of the sum of digits of $\lfloor n^c \rfloor$ in residue classes, for $c \in [7/5, 2)$, remained an open problem. Progress in this direction was made by the author [Spi14], who improved the bound on c to $1 < c \leq 1.42$ for the Thue–Morse sequence. The key idea in that paper is to approximate $\lfloor n^c \rfloor$ by a Beatty sequence $\lfloor n\alpha + \beta \rfloor$ and thus reduce the problem to a linear one. Müllner and the author [MS17], using the same linearization argument and a Bombieri–Vinogradov-type theorem for the Thue–Morse sequence on Beatty sequences, were able to extend this range to $1 < c < 3/2$. In that paper, we also handled occurrences of factors in Piatetski-Shapiro subsequences of $\mathbf{t}$, thus showing that $\mathbf{t}(\lfloor n^c \rfloor)$ defines a normal sequence for $1 < c < 3/2$.

Theorem F (Müllner–Spiegelhofer). *Let $1 < c < 3/2$. Then the sequence $\mathbf{u} = (\mathbf{t}(\lfloor n^c \rfloor))_{n \geq 0}$ is normal. More precisely, for any $L \geq 1$ there exist an exponent $\eta > 0$ and a constant C such that*

$$\left| |\{n < N : \mathbf{u}(n+i) = \omega_i \text{ for } 0 \leq i < L\}| - N/2^L \right| \leq C N^{1-\eta}$$

for all $(\omega_0, \ldots, \omega_{L-1}) \in \{\mathtt{0}, \mathtt{1}\}^L$.

This theorem also improved on an earlier result by the author [Spi15], who obtained normality for $1 < c < 4/3$, using an estimate for Fourier coefficients related to the Thue–Morse sequence provided by Drmota *et al.* [DMR19].

Our Theorem 1.2 finally closes the gap in the set of exponents c such that we have an asymptotic formula for Thue–Morse on $\lfloor n^c \rfloor$. This gap appeared with the Mauduit–Rivat result

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

on squares [MR09]; at that time, the gap was $[7/5, 2)$: now, after our paper with Müllner [MS17], it was only left to close the smaller gap $[3/2, 2)$.

However, the case $c > 2$ remains open for now for $c \in \mathbb{Z}$ (which is contained in Gel'fond's problem on polynomial subsequences) as well as for Piatetski-Shapiro sequences. For example, it is a notorious open question to prove that $\mathtt{0}$ occurs with frequency $1/2$ in $n \mapsto \mathbf{t}(n^3)$.

Mauduit [Mau01, Conjecture 1] conjectured that

$$\lim_{N\to\infty} \frac{1}{N}\{1 \le n \le N : s_q(\lfloor n^c \rfloor) \equiv b \bmod m\} = \frac{1}{m}$$

for almost all $c > 1$ with respect to Lebesgue measure, where $q \ge 2$, $m \ge 1$ and b are integers. While this almost-all result is known for $1 < c < 2$, as he notes just before this conjecture, we believe (as we noted before) that our method can be adapted to generalize our results to general sequences $s_q(n) \bmod m$ and thus to prove the asymptotic identity for all $c \in (1, 2)$. However, while we are confident that the asymptotic identity in Mauduit's conjecture holds for *all* noninteger $c > 1$, the case $c > 2$ cannot yet be handled by our methods.

We note that it would definitely be interesting to generalize the normality result from Theorem F to all exponents $1 < c < 2$.

Notation. For a real number x, we write $e(x) = \exp(2\pi i x)$, $\{x\} = x - \lfloor x \rfloor$, $\|x\| = \min_{n\in\mathbb{Z}}|x - n|$ and $\langle\cdot\rangle = \lfloor x + 1/2 \rfloor$ (the 'nearest integer' to x). For a prime number p let $\nu_p(n)$ be the exponent of p in the prime factorization of n. We define the *truncated binary sum-of-digits function*

$$s_\lambda(n) \coloneqq s(n'),$$

where $0 \le n' < 2^\lambda$ and $n' \equiv n \bmod 2^\lambda$, which is the 2^λ-periodic extension of the restriction of s to $\{0, \ldots, 2^\lambda - 1\}$. For $\mu \le \lambda$ we define the *two-fold restricted binary sum-of-digits function*

$$s_{\mu,\lambda}(n) = s_\lambda(n) - s_\mu(n).$$

For a real number $x \ge 0$ we set

$$\log^+ x = \max\{1, \log x\}.$$

The symbol $\mathbb{N}$ denotes the set of nonnegative integers.

Constants implied by the symbols $\ll$ and $\mathcal{O}$ may depend on the variable k (which describes the number of times that we apply Van der Corput's inequality), but are otherwise absolute. Exceptions to this rule will be indicated in the text.

2. Results

In order to (re)state our main theorem, we introduce some notation. Let α, β, y and z be nonnegative real numbers such that $\alpha \ge 1$. We define

$$A(y, z; \alpha, \beta) = |\{y \le m < z : \mathbf{t}(m) = \mathtt{0} \text{ and } \exists n \in \mathbb{Z} \text{ such that } m = \lfloor n\alpha + \beta \rfloor\}|.$$

For integers $d = \alpha$ and $a = \beta$, we clearly have

$$A(y, z; d, a) = |\{y \le m < z : \mathbf{t}(m) = \mathtt{0} \text{ and } m \equiv a \bmod d\}|.$$

Our main theorem is the following result.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

THEOREM 2.1. *Let $\varepsilon > 0$. There exist $\eta > 0$ and C such that*

$$\sum_{1\le d\le D} \max_{\substack{y,z\ge 0\\ z-y\le x}} \max_{0\le a<d} \left| A(y,z;d,a) - \frac{z-y}{2d} \right| \le C x^{1-\eta}$$

for all $x \ge 1$ and $D = x^{1-\varepsilon}$.

Note that this theorem allows intervals $[y, z)$ for arbitrary $y \ge 0$, which is more general than our definition of a level of distribution. Noting that $1 - 2\mathbf{t}(n) = (-1)^{s(n)}$, we obtain the form of this theorem given in the introduction.

As a corollary we obtain an estimate for the least element m in an arithmetic progression such that $\mathbf{t}(m) = \mathtt{1}$. For most common differences d, we do not have to search for a long time until we encounter the first $\mathtt{1}$.

COROLLARY 2.2. *For $d \ge 1$ and $a \ge 0$ we define*

$$m(d,a) = \min\{n \in \mathbb{N} : \mathbf{t}(nd + a) = \mathtt{1}\}.$$

For each $\varepsilon > 0$ we have, as $D \to \infty$,

$$\left| \left\{ d < D : \max_{a\ge 0} m(d,a) \ge d^{\varepsilon} \right\} \right| = o(D).$$

We note that Dartyge and Tenenbaum considered (among many other things) the homogeneous problem concerning $a = 0$: they proved in particular [DT06, Théorème 2.5] that for *any* function $\xi(d)$ tending to ∞, we have $m(d, 0) \le \xi(d)$ for almost all d in the sense of asymptotic density. The added value of our corollary lies in the fact that the maximum is taken over all arithmetic progressions having a given common difference and a given number of terms. We also wish to note that Morgenbesser *et al.* [MSS11] proved in particular that $m(d, 0) \le d + 4$ for *all* nonnegative integers d.

Our second result concerns Piatetski-Shapiro subsequences of the Thue–Morse sequence.

THEOREM 2.3. *Let $1 < c < 2$. Then the sequence $n \mapsto \mathbf{t}(\lfloor n^c \rfloor)$ is simply normal. More precisely, there exist an exponent $\eta > 0$ and a constant C such that*

$$\left| \frac{1}{N} \left| \{0 \le n < N : \mathbf{t}(\lfloor n^c \rfloor) = \mathtt{0}\} \right| - \frac{1}{2} \right| \le C N^{-\eta}.$$

In order to prove this theorem, we use the general argument presented in § 4.2 of [MS17]. This argument uses linear approximation of $\lfloor n^c \rfloor$ by $\lfloor n\alpha + \beta \rfloor$ and thus reduces the problem to Beatty sequences. Therefore, Theorem 2.3 is a corollary of the following Beatty sequence version of a statement on the level of distribution.

THEOREM 2.4. *Let $0 < \theta_1 \le \theta_2 < 1$. There exist $\eta > 0$ and C such that*

$$\int_D^{2D} \max_{\substack{y,z\ge 0\\ z-y\le x}} \max_{\beta\ge 0} \left| A(y,z;\alpha,\beta) - \frac{z-y}{2\alpha} \right| d\alpha \le C x^{1-\eta}$$

for all x and D such that $x \ge 1$ and $x^{\theta_1} \le D \le x^{\theta_2}$.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

In order to derive Theorem 2.3 from this result, it is essential that we have the maximum over β inside the integral over α, since we need to approximate $\lfloor n^c \rfloor$ by inhomogeneous (shifted) Beatty sequences $\lfloor n\alpha + \beta \rfloor$.

Concerning Theorem 2.1, we can obtain a weakened version of this result, without the maximum over a, using Martin *et al.* [MMR14].

Remark. Martin *et al.* [MMR14, Proposition 3] proved an estimate of a sum of type II containing the following special case: let a_m and b_n be complex numbers satisfying $|a_m| \le 1$ and $|b_n| \le 1$. Assume that $x \ge 2$, $0 < \varepsilon \le 1/2$, $x^\varepsilon \le M, N \le x$ and $MN \le x$. Then

$$S_0 = \sum_{M < m \le 2M} \sum_{\substack{N < n \le 2N \\ mn \le x}} a_m b_n (-1)^{s(mn)} \le C x^{1-\eta}$$

for an absolute constant C and some $\eta > 0$ only depending on ε. By dyadic decomposition and using the trivial estimate for $n < x^\varepsilon$, we obtain

$$\sum_{M < m \le 2M} \left| \sum_{\substack{0 \le n \le 2N \\ mn \le x}} (-1)^{s(mn)} \right| \ll_\varepsilon x^{1-\eta} \log N + M x^\varepsilon$$

for M and N satisfying the same restrictions and with an implied constant that may depend on ε. Let x be given and assume that $x^\varepsilon \le M \le x^\theta$ for some $\theta \in (1/2, 1)$. Set $\varepsilon = 1 - \theta \le 1/2$ and $N = x/M$. Then $N \ge x^\varepsilon$ and the condition $mn \le x$ implies that $n \le 2N$. Using dyadic decomposition again, this time in the variable m, we obtain

$$\sum_{x^\varepsilon < m \le D} \left| \sum_{\substack{0 \le u \le x \\ u \equiv 0 \bmod m}} (-1)^{s(u)} \right| \ll_\varepsilon x^{1-\eta} \log^2 x + M x^\varepsilon \log x$$

for $D = x^\theta$. Finally, we use Fouvry and Mauduit [FM96b] in order to handle residue classes having small modulus m, that is, $m \le x^\varepsilon$. We note (as we did in [MS17]) that the error term in their estimate [FM96b, (1.6)] is in fact $x^{1-\eta}$ for some $\eta > 0$; this follows from Théorème 2 in the same paper [FM96b]. We obtain

$$\sum_{1 \le d \le D} \left| \sum_{\substack{0 \le u \le x \\ n \equiv 0 \bmod d}} (-1)^{s(u)} \right| \le C x^{1-\eta}$$

for $D = x^\theta$ and some $\eta > 0$ and C depending on θ. This is a weak version of a statement of the type 'the Thue–Morse sequence has level of distribution 1', where $\mathcal{Q}(d)$ has only one element. We note that we could also handle the maximum over $y \le x$, using the factor $e(\beta mn)$ that appears in [MMR14, Proposition 3]. The added value of our paper (compare also to the remark after Corollary 2.2) lies in the maximum over the residue classes modulo d.

Finally, we note the following open questions concerning Theorems 2.1 and 2.3.

(i) In Theorem 2.1, can we choose $D = x(\log x)^{-B}$ for some $B > 0$, using $x(\log x)^{-A}$ as error term?

(ii) Does Theorem 2.3 hold for $\lfloor x^2 (\log x)^{-C} \rfloor$ (and similar sequences, possibly with a worse error term) in place of $\lfloor x^c \rfloor$?

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Plan of the paper. In § 3 we state two results (Propositions 3.1 and 3.2) from which Theorems 2.1 and 2.4 follow; moreover, we prove an important Gowers uniformity norm estimate for the Thue–Morse sequence in Proposition 3.3. We also give an idea of the proof of Proposition 3.1. Using Proposition 3.1, the proof of Corollary 2.2 is very short and we present it in that section. In § 4 we state lemmas needed for proving the results from § 3. Section 5 is devoted to proving Propositions 3.1 and 3.2. Finally, in §§ 5.1 and 5.2, we prove Proposition 3.3 and a technical lemma appearing in the proof of Propositions 3.1 and 3.2.

3. Auxiliary results

It will be sufficient to prove the following two propositions in order to obtain our main theorems. To see this, we follow our earlier paper with Müllner [MS17, Section 4.1] and Fouvry and Mauduit [FM96b] for handling small d. In fact, as we noted before, their Théorème 2 holds with an improved error term. Moreover, the proof of this result also reveals that the result holds for arbitrarily shifted intervals $[y, z)$.

Proposition 3.1. *For real numbers $N, D \geq 1$ and ξ set*

$$S_0 = S_0(N, D, \xi) = \sum_{D \leq d < 2D} \max_{a \geq 0} \left| \sum_{0 \leq n < N} e\left(\frac{1}{2} s(nd + a)\right) e(n\xi) \right|. \tag{3.1}$$

Let $\rho_2 \geq \rho_1 > 0$. There exist an $\eta > 0$ and a constant C such that

$$\frac{S_0}{ND} \leq C N^{-\eta} \tag{3.2}$$

holds for all $\xi \in \mathbb{R}$ and all real numbers $N, D \geq 1$ satisfying $N^{\rho_1} \leq D \leq N^{\rho_2}$.

With the help of this proposition, it is not difficult to prove Corollary 2.2: we have $|\{d \in [D, 2D) : \max_{a \geq 0} m(d, a) \geq N\}| \leq CDN^{-\eta}$ for all N, D such that $N^{\rho_1} \leq D \leq N^{\rho_2}$ and some $C > 0$, $\eta > 0$. This is the case since we cannot have more than $CDN^{-\eta}$ many trivial sums in the expression S_0; this means that for each nontrivial summand we encounter at least one `1` for each a. It follows that $|\{d \in [D, 2D) : \max_{a \geq 0} m(d, a) \geq D^{\varepsilon}\}| \leq CD^{1-\eta'}$ for all $\varepsilon > 0$. By dyadic decomposition the statement of the corollary follows.

Proposition 3.2. *For real numbers $D, N \geq 1$ and ξ set*

$$S_0 = S_0(N, D, \xi) = \int_D^{2D} \max_{\beta \geq 0} \left| \sum_{0 \leq n < N} e\left(\frac{1}{2} s(\lfloor n\alpha + \beta \rfloor)\right) e(n\xi) \right| d\alpha. \tag{3.3}$$

Let $\rho_2 \geq \rho_1 > 0$. There exist $\eta > 0$ and a constant C such that

$$\frac{S_0}{ND} \leq C N^{-\eta} \tag{3.4}$$

holds for all real numbers $D, N \geq 1$ satisfying $N^{\rho_1} \leq D \leq N^{\rho_2}$ and for all $\xi \in \mathbb{R}$.

In the proof of these results, we will use the following essential estimate of a *Gowers uniformity norm* of the Thue–Morse sequence (see Konieczny [Kon19]).

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Proposition 3.3. *Let $k \geq 2$ be an integer. There exist some $\eta > 0$ and some C such that*

$$\frac{1}{2^{(k+1)\rho}} \sum_{\substack{0\leq n<2^\rho\\ 0\leq r_1,\ldots,r_k<2^\rho}} e\left(\frac{1}{2}\sum_{\varepsilon\in\{0,1\}^k} s_\rho(n+\varepsilon\cdot r)\right) \leq C2^{-\rho\eta}$$

for all $\rho \geq 0$, where $\varepsilon \cdot r = \sum_{1\leq i\leq k} \varepsilon_i r_i$.

Remark. Since the paper [Kon19] by Konieczny also handles the Rudin–Shapiro sequence, it is certainly possible to prove analogous theorems for this sequence instead of the Thue–Morse sequence.

We wish to give a rough idea of the proof of Proposition 3.1 (Proposition 3.2 being proved essentially in the same way).

Idea of the proof of Proposition 3.1. The key idea is to reduce the number of digits that have to be taken into account and thus to replace the sum-of-digits function s by its truncated version s_ρ. Here 2^ρ will be significantly smaller than N, so that (we simplify things a bit to convey the idea) we may replace the sum over $s(nd+a)$ by a full sum over the periodic function $s_\rho(n)$. This reducing of the digits is achieved by a refinement of the method used by Müllner and the author [MS17], which in turn builds on the ideas from the papers [MR09, MR10] by Mauduit and Rivat.

First, we apply Van der Corput's inequality and use a 'carry propagation lemma' in order to replace s by s_λ. In general, 2^λ will be much larger than N, so that we have to reduce λ further. The next step is to apply the generalized Van der Corput inequality repeatedly. With each application, we remove μ many digits. This is achieved by appealing to the Dirichlet approximation theorem, by which we can find a multiple of $\alpha = d/2^{j\mu}$ that is close to a multiple of 2^μ. This property can be used to discard the μ lowest digits.

By this repeated application the estimate is reduced to an estimate of a Gowers uniformity norm of the Thue–Morse sequence, and we use the method of proof of Konieczny [Kon19] in order to obtain this estimate. The application of Van der Corput's inequality in the context of digital problems is well established, beginning with the work of Mauduit and Rivat [MR09, MR10]. The combination with Gowers norms however is novel, and we think that this connection is a fruitful one: iterated application of Van der Corput's inequality leads to multiple correlations, which in a natural way lead to Gowers norms.

4. Lemmas

We have the following series of lemmas that can also be found in our earlier paper with Müllner [MS17]. The first lemma can be proved by elementary considerations.

Lemma 4.1. *Let $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$.*

$$\text{If } \|a\| < \varepsilon \text{ and } \|b\| \geq \varepsilon, \text{ then } \lfloor a+b \rfloor = \langle a \rangle + \lfloor b \rceil, \tag{4.1}$$

$$\|na\| \leq n\|a\|, \tag{4.2}$$

$$\text{If } \|a\| < \varepsilon \text{ and } 2n\varepsilon < 1, \text{ then } \langle na \rangle = n\langle a \rangle. \tag{4.3}$$

As an essential tool, we will use repeatedly the following generalized Van der Corput inequality [MR09, Lemme 17].

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

LEMMA 4.2. *Let I be a finite interval in $\mathbb{Z}$ containing N integers and let z_n be a complex number for $n \in I$. For all integers $K \geq 1$ and $R \geq 1$ we have*

$$\left|\sum_{n\in I} z_n\right|^2 \leq \frac{N+K(R-1)}{R} \sum_{0\leq |r|<R} \left(1-\frac{|r|}{R}\right) \sum_{\substack{n\in I\\ n+Kr\in I}} z_{n+Kr}\overline{z_n}. \tag{4.4}$$

Assume that α is a real number and N is a nonnegative integer. We define the *discrepancy* of the sequence $n\alpha$ modulo 1:

$$D_N(\alpha) = \sup_{\substack{0\leq x\leq 1\\ y\in\mathbb{R}}} \left|\frac{1}{N}\sum_{n<N} 1_{[0,x)+y+\mathbb{Z}}(n\alpha) - x\right|.$$

Applying this definition, using $x = 1/(KT)$, $y = t/(KT)$, and α/K instead of α, we obtain the following lemma.

LEMMA 4.3. *Let J be an interval in $\mathbb{R}$ containing N integers and let α and β be real numbers. Assume that t, T, ℓ and L are integers such that $0 \leq t < T$ and $0 \leq \ell < L$. Then*

$$\left|\left\{n \in J : \frac{t}{T} \leq \{n\alpha+\beta\} < \frac{t+1}{T}, \lfloor n\alpha+\beta\rfloor \equiv \ell \bmod L\right\}\right| = \frac{N}{LT} + \mathcal{O}\left(ND_N\left(\frac{\alpha}{L}\right)\right)$$

with an absolute implied constant.

In the estimation of our error terms, we will use the following mean discrepancy results (Lemma 3.4 in [MS17]).

LEMMA 4.4. *For integers $\mu \geq 0$ and $N \geq 1$ we have*

$$\sum_{0\leq d<2^\mu} D_N\left(\frac{d}{2^\mu}\right) \leq C_1 \frac{N+2^\mu}{N} (\log^+ N)^2.$$

Also, the estimate

$$\int_0^1 D_N(\alpha)\, d\alpha \leq C_2 \frac{(\log^+ N)^2}{N}$$

holds. The constants C_1 and C_2 in these estimates are absolute.

The following 'carry propagation lemma' will allow us to replace the sum-of-digits function s by its truncated version s_λ. Statements of this type were used by Mauduit and Rivat in their papers on the sum of digits of primes and squares [MR09, MR10].

LEMMA 4.5. *Let r, N, λ be nonnegative integers and $\alpha > 0, \beta \geq 0$ real numbers. Assume that I is an interval containing N integers. Then*

$$\begin{aligned}&|\{n \in I : s(\lfloor (n+r)\alpha+\beta\rfloor) - s(\lfloor n\alpha+\beta\rfloor) \neq s_\lambda(\lfloor (n+r)\alpha+\beta\rfloor) - s_\lambda(\lfloor n\alpha+\beta\rfloor)\}|\\ &\quad \leq r(N\alpha/2^\lambda + 2).\end{aligned}$$

Let $\mathcal{F}_n$ be the set of rational numbers p/q such that $1 \leq q \leq n$, the *Farey series of order n*. Each $a \in \mathcal{F}_n$ has two *neighbours* $a_L, a_R \in \mathcal{F}_n$ satisfying $a_L < a < a_R$ and $(a_L, a) \cap \mathcal{F}_n = (a, a_R) \cap \mathcal{F}_n = \emptyset$. We have the following elementary lemma concerning this set (see [HW54, Chapter 3]).

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Lemma 4.6. *Assume that a/b and c/d are reduced fractions such that $b, d > 0$ and $a/b < c/d$. Then $a/b < (a+c)/(b+d) < c/d$. If a/b and c/d are neighbours in the Farey series $\mathcal{F}_n$, then $bc - ad = 1$ and $b + d > n$; moreover,*

$$(a+c)/(b+d) - a/b < \frac{1}{bn} \quad \text{and} \quad c/d - (a+c)/(b+d) < \frac{1}{dn}.$$

Let $\alpha \in \mathbb{R}$ and Q a positive integer. We assign a fraction $p_Q(\alpha)/q_Q(\alpha)$ to α according to the Farey dissection of the reals: consider reduced fractions $a/b < c/d$ that are neighbours in the Farey series $\mathcal{F}_Q$ such that $a/b \le \alpha < c/d$. If $\alpha < (a+c)/(b+d)$, then set $p_Q(\alpha) = a$ and $q_Q(\alpha) = b$, otherwise set $p_Q(\alpha) = c$ and $q_Q(\alpha) = d$. Lemma 4.6 implies that

$$|q_Q(\alpha)\alpha - p_Q(\alpha)| < Q^{-1}. \tag{4.5}$$

We will call an interval of the form $\{\alpha \in \mathbb{R} : p_Q(\alpha) = p, q_Q(\alpha) = q\}$ a *Farey interval* around p/q.

5. Proof of Propositions 3.1 and 3.2

As in the proof of Proposition 2.5 in [MS17], for (3.2) and (3.4) to hold it is sufficient to prove that there exist $\eta > 0$ and a constant C such that

$$\frac{S_0(N, 2^\nu, \xi)}{N 2^\nu} \le C N^{-\eta}$$

for all real numbers ξ and for all positive integers N and ν such that there exists a real number $D \ge 1$ satisfying $N^{\rho_1} \le D \le N^{\rho_2}$ and $D < 2^\nu \le 2D$, where S_0 is defined according to (3.1) or (3.3).

In order to treat the two propositions to some extent in parallel, we will work with two measures $\boldsymbol{\mu}$: for Proposition 3.1, we take the measure defined by $\boldsymbol{\mu}(A) = |A \cap \mathbb{Z}|$, counting the number of integers inside a set, while for Proposition 3.2, $\boldsymbol{\mu}$ is the Lebesgue measure. We note that in this proof, implied constants in estimates depend only on the variable k, whose meaning will become clear later.

By Cauchy–Schwarz, followed by Van der Corput's inequality (4.4) (R_0 will be specified later), we obtain

$$\begin{aligned}|S_0(N, 2^\nu, \xi)|^2 \le 2^\nu \frac{N + R_0}{R_0} \int_{2^\nu}^{2^{\nu+1}} \sup_{\beta \ge 0} \sum_{0 \le |r_0| < R_0} \left(1 - \frac{|r_0|}{R_0}\right) e(r_0 \xi) \\ \times \sum_{\substack{0 \le n < N \\ 0 \le n + r_0 < N}} e\left(\frac{1}{2} s(\lfloor (n + r_0)\alpha + \beta \rfloor) - \frac{1}{2} s(\lfloor n\alpha + \beta \rfloor)\right) d\boldsymbol{\mu}(\alpha).\end{aligned}$$

We apply the carry propagation lemma (Lemma 4.5), treat the summand $r_0 = 0$ separately and omit the condition $0 \le n + r_0 < N$. Moreover, we consider r_0 and $-r_0$ synchronously. In this

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

way we obtain for all $\lambda \geq 0$

$$|S_0(N, 2^\nu, \xi)|^2 \ll (2^\nu N)^2 E_0 + \frac{2^\nu N}{R_0} \sum_{1 \leq r_0 < R_0} \times \int_{2^\nu}^{2^{\nu+1}} \sup_{\beta \geq 0} \left| \sum_{0 \leq n < N} e\left(\frac{1}{2} s_\lambda(\lfloor (n+r_0)\alpha + \beta \rfloor) - \frac{1}{2} s_\lambda(\lfloor n\alpha + \beta \rfloor)\right) \right| d\boldsymbol{\mu}(\alpha),$$

where

$$E_0 = \frac{1}{R_0} + \frac{R_0 2^\nu}{2^\lambda} + \frac{R_0}{N}.$$

We apply Cauchy–Schwarz on the sum over r_0 and the integral over α in order to prepare our expression for another application of Van der Corput's inequality. It follows that

$$|S_0(N, 2^\nu, \xi)|^4 \ll \frac{2^{3\nu} N^2}{R_0} \sum_{1 \leq r_0 < R_0} \int_{2^\nu}^{2^{\nu+1}} \sup_{\beta \geq 0} |S_1|^2 \, d\boldsymbol{\mu}(\alpha) + (2^\nu N)^4 E_0,$$

where

$$S_1 = \sum_{0 \leq n < N} e\left(\frac{1}{2} s_\lambda(\lfloor (n+r_0)\alpha + \beta \rfloor) - \frac{1}{2} s_\lambda(\lfloor n\alpha + \beta \rfloor)\right).$$

(Note that the error term is also squared, but if it is larger than or equal to 1, the estimate is trivial anyway. We will use this argument again in a moment.) We apply Van der Corput's inequality (4.4) with $R = R_1$ and $K = K_1$ to be chosen later:

$$|S_1|^2 \leq \frac{N + K_1(R_1 - 1)}{R_1} \sum_{0 \leq |r_1| < R_1} \left(1 - \frac{|r_1|}{R_1}\right) \times \sum_{\substack{0 \leq n < N \\ 0 \leq n + r_1 K_1 < N}} e\left(\frac{1}{2} \sum_{\varepsilon_0, \varepsilon_1 \in \{0,1\}} s_\lambda(\lfloor (n + \varepsilon_0 r_0 + \varepsilon_1 r_1 K_1)\alpha + \beta \rfloor)\right);$$

therefore, combining the summands for r_1 and $-r_1$ and omitting the condition $0 \leq n + r_1 K_1 < N$,

$$|S_0(N, 2^\nu, \xi)|^4 \ll \frac{2^{3\nu} N^3}{R_0 \, R_1} \sum_{\substack{1 \leq r_0 < R_0 \\ 0 \leq r_1 < R_1}} \int_{2^\nu}^{2^{\nu+1}} \sup_{\beta \geq 0} |S_2| \, d\boldsymbol{\mu}(\alpha) + (2^\nu N)^4 (E_0 + E_1),$$

where

$$S_2 = \sum_{0 \leq n < N} e\left(\frac{1}{2} \sum_{\varepsilon_0, \varepsilon_1 \in \{0,1\}} s_\lambda(\lfloor (n + \varepsilon_0 r_0 + \varepsilon_1 r_1 K_1)\alpha + \beta \rfloor)\right)$$

and

$$E_1 = \frac{R_1 K_1}{N}.$$

Cauchy–Schwarz over r_0, r_1 and α yields

$$|S_0(N, \nu, \xi)|^8 \ll \frac{2^{7\nu} N^6}{R_0 R_1} \sum_{\substack{1 \leq r_0 < R_0 \\ 0 \leq r_1 < R_1}} \int_{2^\nu}^{2^{\nu+1}} \sup_{\beta \geq 0} |S_2|^2 \, d\boldsymbol{\mu}(\alpha) + (2^\nu N)^8 (E_0 + E_1).$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

We apply Van der Corput's inequality with $R = R_2$ and $K = K_2$ to be chosen later:

$$\frac{|S_0(N,2^\nu,\xi)|^8}{(2^\nu N)^8} \ll (E_0+E_1+E_2) + \frac{1}{R_0R_1R_2 2^\nu N} \sum_{\substack{1\le r_0<R_0\\ 0\le r_1<R_1\\ 0\le r_2<R_2}} \int_{2^\nu}^{2^{\nu+1}} \sup_{\beta\ge 0} |S_3|\, d\boldsymbol{\mu}(\alpha),$$

where

$$S_3 = \sum_{0\le n<N} e\Biggl(\frac{1}{2} \sum_{\varepsilon_0,\varepsilon_1,\varepsilon_2\in\{0,1\}} s_\lambda(\lfloor n\alpha+\beta+\varepsilon_0 r_0\alpha+\varepsilon_1 r_1 K_1\alpha+\varepsilon_2 r_2 K_2\alpha\rfloor)\Biggr)$$

and $E_2 = R_2K_2/N$. Continuing in this manner and replacing the range of integration (we note that we are going to choose $\lambda > \nu$ later), we obtain

$$\begin{aligned}\left|\frac{S_0(N,2^\nu,\xi)}{2^\nu N}\right|^{2^{k+1}} &\ll (E_0+E_1+\cdots+E_k)\\ &\quad + \frac{1}{R_0R_1\cdots R_k 2^\nu N} \sum_{\substack{1\le r_0<R_0\\ 0\le r_i<R_i, 1\le i\le k}} \int_0^{2^\lambda} \sup_{\beta\ge 0}|S_4|\, d\boldsymbol{\mu}(\alpha),\end{aligned} \tag{5.1}$$

where

$$S_4 = \sum_{0\le n<N} e\Biggl(\frac{1}{2}\sum_{\varepsilon_0,\ldots,\varepsilon_k\in\{0,1\}} s_\lambda(\lfloor n\alpha+\beta+\varepsilon_0 r_0\alpha+\varepsilon_1 r_1K_1\alpha+\cdots+\varepsilon_k r_k K_k\alpha\rfloor)\Biggr)$$

and

$$\begin{aligned} E_0 &= \frac{1}{R_0} + \frac{R_0\, 2^\nu}{2^\lambda} + \frac{R_0}{N},\\ E_i &= \frac{R_i\, K_i}{N} \quad \text{for } 1\le i\le k.\end{aligned}$$

Now we choose the multiples $K_1,\ldots,K_k$ in such a way that the number of digits to be taken into account is reduced from λ to $\rho \coloneqq \lambda-(k+1)\mu$, where μ is chosen later. For this we use Farey series; see (4.5). Let

$$\begin{aligned} K_1 &= q_{2^{2\mu+2\sigma}}\left(\frac{\alpha}{2^{2\mu}}\right) q_{2^\sigma}\left(\frac{p_{2^{2\mu+2\sigma}}(\alpha/2^{2\mu})}{2^{(k-1)\mu}}\right);\\ K_i &= q_{2^{\mu+2\sigma}}\left(\frac{\alpha}{2^{(i+1)\mu}}\right) q_{2^\sigma}\left(\frac{p_{2^{\mu+2\sigma}}(\alpha/2^{(i+1)\mu})}{2^{(k-i)\mu}}\right) \quad \text{for } 2\le i<k;\\ K_k &= q_{2^{\mu+\sigma}}\left(\frac{\alpha}{2^{(k+1)\mu}}\right),\end{aligned}$$

where σ is chosen later. Moreover, we set

$$\begin{aligned} M_1 &= p_{2^{2\mu+2\sigma}}\left(\frac{\alpha}{2^{2\mu}}\right) q_{2^\sigma}\left(\frac{p_{2^{2\mu+2\sigma}}(\alpha/2^{2\mu})}{2^{(m-1)\mu}}\right);\\ M_i &= p_{2^{\mu+2\sigma}}\left(\frac{\alpha}{2^{(i+1)\mu}}\right) q_{2^\sigma}\left(\frac{p_{2^{\mu+2\sigma}}(\alpha/2^{(i+1)\mu})}{2^{(k-i)\mu}}\right) \quad \text{for } 2\le i<k;\end{aligned}$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

$$M_k = p_{2^{\mu+\sigma}}\left(\frac{\alpha}{2^{(k+1)\mu}}\right).$$

By Lemma 4.6, estimating the second factor in the definition of K_i and M_i by 2^σ, we have

$$\begin{aligned} &|K_1\alpha - 2^{2\mu}M_1| < 2^{-\sigma};\\ &\left|\frac{K_i\alpha}{2^{i\mu}} - 2^\mu M_i\right| < 2^{-\sigma} \quad \text{for } 2 \le i < k;\\ &\left|\frac{K_k\alpha}{2^{k\mu}} - 2^\mu M_k\right| < 2^{-\sigma}. \end{aligned} \tag{5.2}$$

We are going to use these inequalities in order to replace $r_iK_i\alpha$ in the sum S_4, starting with $r_1K_1\alpha$. We treat the case when α is an integer first: in this case, $K_1\alpha = 2^{2\mu}M_1$, and by the fact that the arguments of s_λ corresponding to $\varepsilon_1 = 0, 1$ differ by a multiple of $2^{2\mu}$, we may shift the argument by 2μ digits and thus reduce the number of digits to be taken into account from λ to $\lambda - 2\mu$:

$$\begin{aligned} S_4 &= \sum_{0\le n<N} e\biggl(\frac{1}{2}\sum_{\varepsilon_0,\ldots,\varepsilon_k\in\{0,1\}} s_{2\mu,\lambda}(\lfloor n\alpha+\beta\\ &\qquad + \varepsilon_0 r_0\alpha + \varepsilon_1 r_1 M_1 2^{2\mu} + \varepsilon_2 r_2 K_2\alpha + \cdots + \varepsilon_k r_k K_k\alpha\rfloor)\biggr)\\ &= \sum_{0\le n<N} e\biggl(\frac{1}{2}\sum_{\varepsilon_0,\ldots,\varepsilon_k\in\{0,1\}} s_{\lambda-2\mu}\biggl(\biggl\lfloor\frac{n\alpha+\beta}{2^{2\mu}} + \frac{\varepsilon_0 r_0\alpha}{2^{2\mu}} + \varepsilon_1 r_1 M_1\\ &\qquad + \frac{\varepsilon_2 r_2 K_2\alpha}{2^{2\mu}} + \cdots + \frac{\varepsilon_k r_k K_k\alpha}{2^{2\mu}}\biggr\rfloor\biggr)\biggr). \end{aligned}$$

In the case $\alpha \notin \mathbb{Z}$, we use the inequalities (5.2) and the argument that $n\alpha$-sequences are usually not close to an integer. This can be made precise as follows. Assume that

$$\|n\alpha + \beta'\| \ge R_1/2^\sigma, \tag{5.3}$$

where $\beta' = \beta + \varepsilon_0 r_0\alpha + \varepsilon_2 r_2 K_2\alpha + \cdots + \varepsilon_k r_k K_k\alpha$, and that $2R_1 < 2^\sigma$. Using the inequality (4.3) in Lemma 4.1 with $\varepsilon = 1/2^\sigma$, where $\sigma \ge 1$ is chosen later, and (4.5), we obtain

$$\langle r_1K_1\alpha\rangle = r_1\langle K_1\alpha\rangle = r_1 2^{2\mu}M_1.$$

Applying (4.1), setting $\varepsilon = R_1/2^\sigma$, we see that (5.3) together with (5.2) implies that

$$\lfloor n\alpha + r_1K_1\alpha + \beta'\rfloor = \lfloor n\alpha + r_1 2^{2\mu}M_1 + \beta'\rfloor.$$

The number of n where hypothesis (5.3) fails for some $\varepsilon_0, \varepsilon_2, \ldots, \varepsilon_k$ can be estimated by discrepancy estimates for $\{n\alpha\}$-sequences: for all positive integers N and $2R_1 < 2^\sigma$ we have

$$\begin{aligned} &|\{n \in [0, N-1] : \|n\alpha + \beta'\| \le R_1/2^\sigma\}|\\ &\quad = |\{n \in [0, N-1] : n\alpha + \beta' \in [-R_1/2^\sigma, R_1/2^\sigma] + \mathbb{Z}\}|\\ &\quad = |\{n \in [0, N-1] : n\alpha \in [0, 2R_1/2^\sigma] - \beta' - R_1/2^\sigma + \mathbb{Z}\}|\\ &\quad \le ND_N(\alpha) + 2R_1N/2^\sigma. \end{aligned}$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Therefore, the number of $n \in [0, N-1]$ such that there exist $\varepsilon_0, \varepsilon_2, \ldots, \varepsilon_k \in \{0,1\}$ with $\|n\alpha + \beta'\| \le R_1/2^\sigma$ is bounded by $2^k N(D_N(\alpha) + 2R_1/2^\sigma)$, which is $\ll N(D_N(\alpha) + 2R_1/2^\sigma)$ by our convention that implied constants may depend on k.

We replace $K_1\alpha$ by $2^{2\mu} M_1$ and subsequently shift the digits by 2μ and obtain

$$S_4 = \sum_{0 \le n < N} e\biggl(\frac{1}{2} \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} s_{\lambda - 2\mu}\biggl(\biggl\lfloor \frac{n\alpha + \beta}{2^{2\mu}} + \frac{\varepsilon_0 r_0 \alpha}{2^{2\mu}} + \varepsilon_1 r_1 M_1 + \frac{\varepsilon_2 r_2 K_2 \alpha}{2^{2\mu}} + \cdots + \frac{\varepsilon_k r_k K_k \alpha}{2^{2\mu}} \biggr\rfloor\biggr)\biggr) + \mathcal{O}(N D_N(\alpha) + N R_1/2^\sigma).$$

Repeating this argument for all $i \in \{2, \ldots, k\}$ we obtain

$$S_4 = N\mathcal{O}\biggl(\widetilde{D}_N(\alpha) + D_N\biggl(\frac{\alpha}{2^{2\mu}}\biggr) + \cdots + D_N\biggl(\frac{\alpha}{2^{k\mu}}\biggr) + \frac{R_1 + \cdots + R_k}{2^\sigma}\biggr) + \sum_{0 \le n < N} e\biggl(\frac{1}{2} \sum_{\varepsilon_1, \ldots, \varepsilon_k \in \{0,1\}} s_{\lambda - (k+1)\mu}\biggl(\biggl\lfloor \frac{n\alpha + \beta}{2^{(k+1)\mu}} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} + \sum_{1 \le i \le k} \frac{\varepsilon_i r_i M_i}{2^{(k-i)\mu}} \biggr\rfloor\biggr)\biggr),$$

where $\widetilde{D}_N(\alpha) = D_N(\alpha)$ if $\alpha \notin \mathbb{Z}$ and $\widetilde{D}_N(\alpha) = 0$ otherwise.

Now the second factor in the definition of K_i comes into play. We use the definition of M_i together with the approximation property (4.5), and apply the discrepancy estimate for $\{n\alpha\}$-sequences again, to obtain

$$S_4 = N\mathcal{O}\biggl(\widetilde{D}_N(\alpha) + D_N\biggl(\frac{\alpha}{2^{2\mu}}\biggr) + \cdots + D_N\biggl(\frac{\alpha}{2^{(k+1)\mu}}\biggr) + \frac{R_1 + \cdots + R_k}{2^\sigma}\biggr) + S_5, \tag{5.4}$$

where

$$S_5 = \sum_{0 \le n < N} e\biggl(\frac{1}{2} \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} s_{\lambda - (k+1)\mu}\biggl(\biggl\lfloor \frac{n\alpha + \beta}{2^{(k+1)\mu}} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \biggr\rfloor + \sum_{1 \le i \le k} \varepsilon_i r_i \mathfrak{p}_i\biggr)\biggr)$$

and

$$\begin{aligned} \mathfrak{p}_1 &= p_{2^\sigma}\biggl(\frac{p_{2^{2\mu+2\sigma}}(\alpha/2^{2\mu})}{2^{(k-1)\mu}}\biggr); \\ \mathfrak{p}_i &= p_{2^\sigma}\biggl(\frac{p_{2^{\mu+2\sigma}}(\alpha/2^{(i+1)\mu})}{2^{(k-i)\mu}}\biggr) \quad \text{for } 2 \le i < k; \\ \mathfrak{p}_k &= p_{2^{\mu+\sigma}}\biggl(\frac{\alpha}{2^{(k+1)\mu}}\biggr). \end{aligned} \tag{5.5}$$

Our next goal is to remove the Beatty sequence occurring in S_5, and also to remove the integers $\mathfrak{p}_i$. The resulting expression can be handled by the Gowers norm estimate given in Proposition 3.3, which will finish the proof.

We start by splitting the Beatty sequence into two summands. Let t, T be integers such that $0 \le t < T$ and define

$$S_6 = \sum_{\substack{0 \le n < N \\ t/T \le \{(n\alpha+\beta)/2^{(k+1)\mu}\} < (t+1)/T}} e\biggl(\frac{1}{2} \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} s_{\lambda - (k+1)\mu}\biggl(\biggl\lfloor \frac{n\alpha + \beta + \varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \biggr\rfloor + \sum_{1 \le i \le k} \varepsilon_i r_i \mathfrak{p}_i\biggr)\biggr).$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

We define

$$G = \left\{1 \leq t < T : \left[\frac{t}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}}, \frac{t+1}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}}\right) \cap \mathbb{Z} = \emptyset\right\}.$$

Clearly, we have $|G| \geq T-2$, since we have to exclude at most one t. For $t \in \{0, \ldots, T-1\} \setminus G$ we estimate S_6 trivially, using Lemma 4.3: we obtain

$$S_6 \ll \frac{N}{T} + N D_N\left(\frac{\alpha}{2^{(k+1)\mu}}\right). \tag{5.6}$$

Assume that $t \in G$ and that $t/T \leq \{(n\alpha+\beta)/2^{(k+1)\mu}\} < (t+1)/T$. Then

$$\left\lfloor \frac{n\alpha+\beta}{2^{(k+1)\mu}} \right\rfloor + \frac{t}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \leq \frac{n\alpha+\beta+\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} < \left\lfloor \frac{n\alpha+\beta}{2^{(k+1)\mu}} \right\rfloor + \frac{t+1}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}}$$

and the assumption $t \in G$ gives

$$\left\lfloor \frac{n\alpha+\beta+\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \right\rfloor = \left\lfloor \frac{n\alpha+\beta}{2^{(k+1)\mu}} \right\rfloor + \left\lfloor \frac{t}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \right\rfloor$$

for $\varepsilon_0 \in \{0,1\}$. From these observations we obtain for $t \in G$:

$$S_6 = \sum_{0 \leq m < 2^\rho} \sum_{\substack{0 \leq n < N \\ t/T \leq \{(n\alpha+\beta)/2^{(k+1)\mu}\} < (t+1)/T \\ \lfloor (n\alpha+\beta)/2^{(k+1)\mu} \rfloor \equiv m \bmod 2^\rho}} \times e\left(\frac{1}{2} \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} s_\rho\left(m + \left\lfloor \frac{t}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \right\rfloor + \sum_{1 \leq i \leq k} \varepsilon_i r_i \mathfrak{p}_i\right)\right).$$

Note that the Beatty sequence $\lfloor (n\alpha+\beta)/2^{(k+1)\mu} \rfloor$ does not occur in the summand any more. We may therefore remove the second summation by estimating the number of times the three conditions under the summation sign are satisfied. At this point we want to stress the fact that N is going to be significantly larger than $2^\rho = 2^{\lambda-(k+1)\mu}$. Using Lemma 4.3 and the usually very small discrepancy of $n\alpha$-sequences, this fact will enable us to remove the summation over n, while introducing only a negligible error term for most α. This is the point in the proof where the successive 'cutting away' of binary digits with the help of Farey series pays off.

By Lemma 4.3, applied with $L = 2^\rho$, and noting that $\lambda = (k+1)\mu + \rho$, we obtain for $t \in G$

$$S_6 = \frac{N}{2^\rho T} S_7 + \mathcal{O}\left(2^\rho N D_N\left(\frac{\alpha}{2^\lambda}\right)\right), \tag{5.7}$$

where

$$S_7 = \sum_{0 \leq m < 2^\rho} e\left(\frac{1}{2} \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} s_\rho\left(m + \left\lfloor \frac{t}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \right\rfloor + \sum_{1 \leq i \leq k} \varepsilon_i r_i \mathfrak{p}_i\right)\right).$$

We note the important fact that this expression is independent of β. This will allow us to remove the maximum over β inside the integral over α and thus prove the strong statement on the level of distribution.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

We wish to simplify this expression in such a way that Proposition 3.3 is applicable. To this end, we use the summation over r_i and the integral over α. We define

$$S_8 = \int_0^{2^\lambda} \sum_{0\le r_1,\ldots,r_k<2^\rho} |S_7|\, d\boldsymbol{\mu}(\alpha),$$

which is an expression that will appear when we expand the original sum S_0.

We are going to apply the argument that for most $\alpha < 2^\lambda$ (with respect to $\boldsymbol{\mu}$) the 2-adic valuation of $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ is small. For these α, the term $r_i\mathfrak{p}_i \bmod 2^\rho$ attains each $m \in \{0, \ldots, 2^\rho - 1\}$ not too often, as r_i varies. We may therefore replace $r_i\mathfrak{p}_1$ by r_i and thus obtain full sums over r_i: at this point, we set

$$R_i = 2^\rho \quad \text{for } 1 \le i \le k.$$

In order to make this argument work, we are going to utilize the following technical result, the proof of which we give in § 5.2.

Lemma 5.1. *Let $\mu, \lambda, \sigma, \gamma, k$ be nonnegative integers such that $k \ge 2$ and*

$$\begin{aligned} &\lambda \ge (k+1)\mu, && \gamma \le \lambda - (k+1)\mu,\\ &\mu \ge 4\sigma, && \sigma \ge \gamma \ge 1. \end{aligned} \tag{5.8}$$

Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ be defined by (5.5) and set

$$A = \{\alpha \in \{0, \ldots, 2^\lambda - 1\} : 2^{3\gamma} \mid \mathfrak{p}_i \textit{ for some } i = 1, \ldots, k\}.$$

Then

$$|A| = \mathcal{O}(2^{\lambda-\gamma}).$$

Analogously, if

$$A = \{\alpha \in [0, 2^\lambda] : 2^{3\gamma} \mid \mathfrak{p}_i \textit{ for some } i = 1, \ldots, k\},$$

then

$$\boldsymbol{\lambda}(A) = \mathcal{O}(2^{\lambda-\gamma}),$$

where $\boldsymbol{\lambda}$ is the Lebesgue measure. The implied constants only depend on m (and are independent of μ, λ, σ and γ).

Let A be defined as in this lemma. We choose $R_i = 2^\rho$ for $1 \le i \le k$.

Assume that $\alpha \notin A$. Then, by an elementary argument, $r_i\mathfrak{p}_i \bmod 2^\rho$ attains each value not more than $2^{3\gamma}$ times, as r_i runs through $\{0, \ldots, 2^\rho - 1\}$. The contribution for $\alpha \in A$ will be estimated trivially by the lemma. We obtain

$$S_8 \le 2^{3\gamma k} \int_0^{2^\lambda} \sum_{0\le r_1,\ldots,r_k<2^\rho} |S_9| d\boldsymbol{\mu}(\alpha) + \mathcal{O}(2^{\lambda+(k+1)\rho-\gamma}),$$

where

$$S_9 = \sum_{0\le n<2^\rho} e\left(\frac{1}{2} \sum_{\varepsilon_0,\ldots,\varepsilon_k\in\{0,1\}} s_\rho\left(n + \left\lfloor \frac{t}{T} + \frac{\varepsilon_0 r_0 \alpha}{2^{(k+1)\mu}} \right\rfloor + \sum_{1\le i\le k} \varepsilon_i r_i\right)\right).$$

The next step is removing the remaining floor function, using the integral over α. In the continuous case, the expression $\lfloor t/T + r_0K_0\alpha/2^{(k+1)\mu}\rfloor \bmod 2^\rho$ runs through $\{0, \ldots, 2^\rho - 1\}$ in a

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

completely uniform manner. That is, for $r_0 \neq 0$ and $0 \le m < 2^\rho$ we have

$$\boldsymbol{\lambda}(\{\alpha \in [0, 2^\lambda] : \lfloor t/T + r_0\alpha/2^{(k+1)\mu} \rfloor \equiv m \bmod 2^\rho\}) = 2^{\lambda-\rho},$$

where $\boldsymbol{\lambda}$ is the Lebesgue measure. We consider the discrete case. Assume that $r_0 \le 2^{(k+1)\mu}$ (we will choose R_0 very small at the end of the proof, so that this will be satisfied). Then the set of $\alpha \in \{0, \ldots, 2^\lambda - 1\}$ such that $\lfloor t/T + r_0\alpha/2^{(k+1)\mu} \rfloor \equiv m \bmod 2^\rho$ decomposes into at most $r_0 + 1$ intervals (note that $\lambda = (k+1)\mu + \rho$), each having $\le 2^{(k+1)\mu}/r_0 + 1$ elements. In total we have $\ll 2^{\lambda-\rho}$ elements, where the implied constant is absolute. It follows that

$$S_8 \ll 2^{\lambda+(k+1)\rho-\gamma} + 2^{\lambda-\rho+3\gamma k} \sum_{0 \le r_0, \ldots, r_k < 2^\rho} |S_{10}(r_0, \ldots, r_k)|,$$

where

$$S_{10}(r_0, \ldots, r_k) = \sum_{0 \le n < 2^\rho} e\biggl(\frac{1}{2} \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} s_\rho\biggl(n + \sum_{0 \le i \le k} \varepsilon_i r_i\biggr)\biggr).$$

As a final step in the procedure of reducing the main theorems to Proposition 3.3, we are going to remove the absolute value around S_{10}. For brevity, we set

$$g(n) = \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} s_\rho\biggl(n + \sum_{0 \le i \le k} \varepsilon_i r_i\biggr).$$

By the 2^ρ-periodicity of g we have

$$\begin{aligned}
&\sum_{0 \le r_0, \ldots, r_k < 2^\rho} |S_{10}(r_0, \ldots, r_k)|^2 \\
&\quad = \sum_{0 \le r_0, \ldots, r_k < 2^\rho} \sum_{0 \le n_1, n_2 < 2^\rho} e\biggl(\frac{1}{2} g(n_1) + \frac{1}{2} g(n_2)\biggr) \\
&\quad = \sum_{0 \le r_0, \ldots, r_k < 2^\rho} \sum_{0 \le n_1 < 2^\rho} \sum_{0 \le r_{k+1} < 2^\rho} e\biggl(\frac{1}{2} g(n_1) + \frac{1}{2} g(n_1 + r_{k+1})\biggr) \\
&\quad = \sum_{0 \le r_0, \ldots, r_{k+1} < 2^\rho} \sum_{0 \le n_1 < 2^\rho} e\biggl(\frac{1}{2} g(n_1) + \frac{1}{2} g(n_1 + r_{k+1})\biggr) \\
&\quad = \sum_{0 \le r_0, \ldots, r_{k+1} < 2^\rho} \sum_{0 \le n_1 < 2^\rho} e\biggl(\frac{1}{2} \sum_{\varepsilon_0, \ldots, \varepsilon_k \in \{0,1\}} \sum_{\varepsilon_{k+1} \in \{0,1\}} s_\rho(n_1 + \varepsilon \cdot r + \varepsilon_{k+1} r_{k+1})\biggr) \\
&\quad = \sum_{0 \le r_0, \ldots, r_{k+1} < 2^\rho} S_{10}(r_0, \ldots, r_{k+1}).
\end{aligned}$$

We have therefore removed the absolute value around S_{10} for the price of an additional variable r_{k+1}; see also [GT10, Section 4] for this type of argument. This means that we have reduced our main theorems to Proposition 3.3.

By this proposition and Cauchy–Schwarz we obtain

$$S_8 \ll 2^{\lambda+(k+1)\rho}(2^{-\gamma} + 2^{3\gamma k - \eta\rho}) \tag{5.9}$$

for some $\eta > 0$.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

It remains to collect the error terms and to choose values for the free variables. Using (5.7) and (5.6), we obtain

$$S_5 \ll \sum_{t \notin G} \left(\frac{N}{T} + N D_N\left(\frac{\alpha}{2^{(k+1)\mu}}\right) \right) + \sum_{t \in G} \left(\frac{N}{2^\rho T} S_7 + 2^\rho N D_N\left(\frac{\alpha}{2^\lambda}\right) \right)$$
$$\ll \frac{N}{2^\rho T} \sum_{t \in G} S_7 + \frac{N}{T} + N D_N\left(\frac{\alpha}{2^{(k+1)\mu}}\right) + 2^\rho N T D_N\left(\frac{\alpha}{2^\lambda}\right)$$

and by (5.4) and (5.1) we obtain

$$\begin{aligned}
&\left| \frac{S_0(N,\nu,\xi)}{2^\nu N} \right|^{2^{k+1}} \\
&\ll \mathcal{O}\left(\frac{1}{R_0} + \frac{R_0 2^\nu}{2^\lambda} + \frac{R_0}{N} + \frac{R_1 K_1}{N} + \cdots + \frac{R_k K_k}{N} \right) \\
&\quad + \frac{1}{2^\nu N} \int_0^{2^\lambda} N \mathcal{O}\left(\widetilde{D}_N(\alpha) + D_N\left(\frac{\alpha}{2^{2\mu}}\right) + \cdots + D_N\left(\frac{\alpha}{2^{(k+1)\mu}}\right) + \frac{R_1 + \cdots + R_k}{2^\sigma} \right) d\boldsymbol{\mu}(\alpha) \\
&\quad + \frac{1}{2^\nu N} \int_0^{2^\lambda} \mathcal{O}\left(\frac{N}{T} + N D_N\left(\frac{\alpha}{2^{(k+1)\mu}}\right) + 2^\rho T N D_N\left(\frac{\alpha}{2^\lambda}\right) \right) d\boldsymbol{\mu}(\alpha) \\
&\quad + \frac{1}{R_0 \cdots R_k 2^\nu N} \frac{N}{2^\rho T} \sum_{t \in G} \sum_{1 \le r_0 < R_0} \int_0^{2^\lambda} \sum_{0 \le r_1, \ldots, r_k < 2^\rho} |S_7| \, d\boldsymbol{\mu}(\alpha). \qquad (5.10)
\end{aligned}$$

We employ the mean discrepancy estimates from Lemma 4.4. Assume that $\delta \le \lambda$. In the continuous case we have

$$\frac{1}{2^\nu} \int_0^{2^\lambda} D_N\left(\frac{\alpha}{2^\delta}\right) d\alpha \ll 2^{\lambda-\nu-\delta} \int_0^{2^\delta} D_N\left(\frac{\alpha}{2^\delta}\right) d\alpha \ll 2^{\lambda-\nu} \frac{(\log^+ N)^2}{N},$$

while the discrete case gives

$$\frac{1}{2^\nu} \sum_{0 \le d < 2^\lambda} D_N\left(\frac{d}{2^\delta}\right) \ll 2^{\lambda-\delta-\nu} \frac{N + 2^\delta}{N} (\log^+ N)^2 = 2^{\lambda-\nu} (\log^+ N)^2 \left(\frac{1}{N} + \frac{1}{2^\delta} \right).$$

In total, noting that $\lambda \ge (k+1)\mu$, the discrepancy terms can be estimated by

$$\ll 2^{\lambda-\nu} (\log^+ N)^2 2^\rho T \left(\frac{1}{N} + \frac{1}{2^{2\mu}} \right).$$

By (5.9), the last summand in (5.10) can be estimated by

$$\ll 2^{\lambda-\nu} (2^{-\gamma} + 2^{3\gamma k - \eta\rho}).$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Moreover, using the facts that $R_1 = \cdots = R_k = 2^\rho$ and $K_i \le 2^{2\mu+3\sigma}$ for $1 \le i \le k$, we obtain

$$\left|\frac{S_0(N,\nu,\xi)}{2^\nu N}\right|^{2^{k+1}} \ll \frac{1}{R_0} + \frac{R_0 2^\nu}{2^\lambda} + \frac{R_0}{N} + \frac{2^{\rho+2\mu+3\sigma}}{N} + 2^{\lambda-\nu}(\log^+ N)^2 2^\rho T\left(\frac{1}{N} + \frac{1}{2^{2\mu}}\right) + 2^{\rho-\sigma+\lambda-\nu} + \frac{1}{T} + 2^{\lambda-\nu}(2^{-\gamma} + 2^{3\gamma k-\eta\rho}) \tag{5.11}$$

with some implied constant only depending on k. Collecting also the requirements on the variables we assumed in the course of our calculation, we see that this estimate is valid as long as

$$\begin{array}{ll} R_0,\ T \ge 1,\ k \ge 2,\ \gamma,\nu,\lambda,\rho,\mu \ge 0, & R_1 = \cdots = R_k = 2^\rho, \\ \lambda > \nu, & \rho = \lambda - (k+1)\mu, \\ \gamma \le \rho < \sigma - 1, & \mu \ge 4\sigma, \\ R_0 \le 2^{(k+1)\mu}. & \end{array} \tag{5.12}$$

It remains to choose the variables within these constraints. Choose the integer $j \ge 1$ in such a way that $N^{j-1} \le 2^\nu < N^j$ and set $k = 3j - 1$. Clearly, $k \ge 2$. We define

$$\mu = \left\lfloor \frac{\nu}{k+1+1/8} \right\rfloor, \quad \sigma = \lfloor \mu/4 \rfloor, \quad \widetilde{\rho} = \nu - (k+1)\mu.$$

We obtain the inequalities $N \ge 2^{3\mu}$, $\mu \ge 4\sigma$, $\widetilde{\rho} \ge 0$. Moreover, for large ν we obtain $\widetilde{\rho} \sim \mu/8$.

Choose $\gamma = \lfloor \widetilde{\rho}\eta/(6k) \rfloor$ and $R_0 = \lfloor 2^{\gamma/4} \rfloor$. Then the last summand in (5.11) is $\ll 2^{\lambda-\nu}(2^{-\gamma} + 2^{-\widetilde{\rho}\eta/2}) \ll 2^{\lambda-\nu-\gamma}$. Finally, set $\lambda = \nu + \lfloor \gamma/2 \rfloor$, $T = 2^\gamma$ and $\rho = \lambda - (k+1)\mu$. It follows that $\rho = \widetilde{\rho} + \lfloor \gamma/2 \rfloor \sim (\mu/8)(1 + \eta/(12k)) \le \mu/8 + \mu/192$. Using these definitions, it is not hard to see that, for large N and ν, the requirements (5.12) are met.

Using the statements $N^{\rho_1} \le D \le N^{\rho_2}$ and $D < 2^\nu \le 2D$ we can easily estimate (5.11) term by term and conclude that $S_0(N,\nu,\xi)/(2^\nu N) \le CN^{-\eta'}$ for some $\eta' > 0$ and some constant C. This finishes the proof of Propositions 3.1 and 3.2 and therefore of our main theorems. It remains to prove our auxiliary results.

5.1 Proof of Proposition 3.3

We utilize ideas from the paper [Kon19] by Konieczny. In that paper, he uses the Gowers norm on intervals in $\mathbb{Z}$, while we are concerned with the cyclic group $\mathbb{Z}/2^\rho\mathbb{Z}$. The proof of Proposition 3.3 is analogous to Konieczny's proof. In fact, it is possible to relate the two notions of Gowers norms to each other and therefore avoid going into the details of the proof in [Kon19] (Konieczny, private communication; we also thank the anonymous referee for pointing out this possibility). In this paper, however, we chose to follow the proof from [Kon19], as the argument is interesting and not unreasonably long.

Set

$$A_\rho(\mathbf{a}) = \frac{1}{2^{(k+1)\rho}} \sum_{\substack{0 \le n < 2^\rho \\ 0 \le r_1,\ldots,r_k < 2^\rho}} e\left(\frac{1}{2} \sum_{\varepsilon \in \{0,1\}^k} s_\rho(n + \varepsilon \cdot r + \mathbf{a}_\varepsilon)\right).$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Then, in analogy to equation (16) of [Kon19], we get after a similar calculation (using $k \geq 2$) that

$$A_{\rho+1}(\mathbf{a}) = \frac{(-1)^{|\mathbf{a}|}}{2^{k+1}} \sum_{e_0,\ldots,e_k \in \{0,1\}} A_\rho(\delta(\mathbf{a}, e)), \tag{5.13}$$

where $|\mathbf{a}| = \sum_{\varepsilon \in \{0,1\}^k} \mathbf{a}_\varepsilon$ and

$$\delta(\mathbf{a}, e)_\varepsilon = \left\lfloor \frac{\mathbf{a}_\varepsilon + e_0 + \sum_{1 \leq i \leq k} \varepsilon_i e_i}{2} \right\rfloor.$$

We define a directed graph with weighted edges according to (5.13). The set of vertices is given by the set of families $\mathbf{a} \in \mathbb{Z}^{\{0,1\}^k}$. There is an edge from $\mathbf{a}$ to $\mathbf{b}$ if and only if there is an $e = (e_0, \ldots, e_k) \in \{0,1\}^{k+1}$ such that $\delta(\mathbf{a}, e) = \mathbf{b}$ and this edge has the weight

$$w(\mathbf{a}, \mathbf{b}) = \frac{(-1)^{|\mathbf{a}|}}{2^{k+1}} |\{e \in \{0,1\}^{k+1} : \delta(\mathbf{a}, e) = \mathbf{b}\}|.$$

Note that

$$\sum_{\mathbf{b} \in \mathbb{Z}^{\{0,1\}^k}} |w(\mathbf{a}, \mathbf{b})| = 1, \tag{5.14}$$

which we will need later. We are interested in the subgraph (V, E, w) induced by the set of vertices reachable from $\mathbf{0}$. This graph is finite: we have

$$\max_{\varepsilon \in \{0,1\}^k} |\delta(\mathbf{a}, e)_\varepsilon| \leq \frac{1}{2}\Big(\max_{\varepsilon \in \{0,1\}^k} |\mathbf{a}_\varepsilon| + k + 1\Big)$$

and, by induction, it follows that $\max_{\varepsilon \in \{0,1\}^k} |\mathbf{a}_\varepsilon| < k + 1$ for all $\mathbf{a} \in V$, which implies the finiteness of V.

This subgraph is strongly connected. We prove this by showing that $\mathbf{0}$ is reachable from each $\mathbf{a} \in V$. This follows immediately by considering the path given by the edges $(\mathbf{a}^{(0)}, \mathbf{a}^{(1)})$, $(\mathbf{a}^{(1)}, \mathbf{a}^{(2)})$,$\ldots$,$(\mathbf{a}^{(j)}, \mathbf{a}^{(j+1)})$ defined by $\mathbf{a}^{(0)} = \mathbf{a}$ and $\mathbf{a}^{(i+1)} = \delta(\mathbf{a}^{(i)}, (0, \ldots, 0))$. It is clear from the definition of δ that such a path reaches $\mathbf{0}$ if j is large enough.

We wish to apply (5.13) recursively. We therefore define, for two vertices $\mathbf{a}, \mathbf{b} \in V$ and a positive integer j, the weight $w_j(\mathbf{a}, \mathbf{b})$ as the sum of all weights of paths of length j from $\mathbf{a}$ to $\mathbf{b}$. (Here the weight of a path is the product of the weights of the edges.)

In order to prove Proposition 3.3, it is sufficient to prove that there is a j such that

$$\sum_{\mathbf{b} \in V} |w_j(\mathbf{a}, \mathbf{b})| < 1$$

for all $\mathbf{a} \in V$. In order to prove this, it is sufficient, by the strong connectedness of the graph and (5.14), to prove that there are two paths of the same length from $\mathbf{0}$ to $\mathbf{0}$ such that their respective weights have different sign. One of these paths is the trivial one, choosing $e_0 = \cdots = e_j = 0$ in each step. This path has positive weight.

For the second path, we follow Konieczny [Kon19, proof of Proposition 2.3]. As in that paper, we define $\mathbf{a}^{(0)} = \mathbf{a}^{(j+1)} = \mathbf{0}$ and, for $1 \leq i \leq j$,

$$\mathbf{a}^{(i)}_\varepsilon = \begin{cases} 1, & \text{if } \varepsilon_1 = \cdots = \varepsilon_i = 1; \\ 0, & \text{otherwise.} \end{cases}$$

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Assuming for a moment that there is an edge from $\mathbf{a}^{(i)}$ to $\mathbf{a}^{(i+1)}$ for all $i \in \{0, \ldots, j\}$, it is easy to see that each edge $(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)})$ has positive weight for $0 \le i < j$, while $(\mathbf{a}^{(j)}, \mathbf{a}^{(j+1)})$ has negative weight. Proving that these vertices indeed define a path is contained completely in the argument given in [Kon19]. This finishes the proof of Lemma 3.3.

5.2 Proof of Lemma 5.1

We choose an integer $\gamma > 0$ and bound the size of the set of $\alpha < 2^\lambda$ such that $2^{3\gamma} \mid \mathfrak{p}_i$ for some $i \in \{1, \ldots, k\}$. We will need the following two lemmas.

Lemma 5.2. *Let $\boldsymbol{\lambda}$ be the Lebesgue measure. Assume that $K \ge 1$ and $\gamma \ge 0$ are integers. Then*

$$\boldsymbol{\lambda}(\{x \in [0,1] : 2^\gamma \mid q_K(x)\}) \ll \frac{1}{2^\gamma} + \frac{1}{K}.$$

The constant in this estimate is absolute.

Proof. We have to sum up the lengths of the Farey intervals around p/q such that $2^\gamma \mid q$. By Lemma 4.6, each such fraction contributes at most $2/(Kq)$. By summing over $p \in \{1, \ldots, q\}$, this gives a contribution $2/K$ for each multiple q of 2^γ and we obtain a total contribution

$$\ll \sum_{\substack{1 \le q \le K \\ 2^\gamma \mid q}} \frac{1}{K} \le \frac{1}{2^\gamma} + \frac{1}{K}. \qquad \square$$

Lemma 5.3. *Let $x_0, \ldots, x_{M-1} \in [0,1]$ and $\delta > 0$. Assume that $\|x_i - x_j\| \ge \delta$ for $i \ne j$. Then*

$$|\{n \in \{0, \ldots, M-1\} : 2^\gamma \mid q_K(x_i)\}| \ll \frac{K^2}{2^\gamma} + \frac{1}{\delta}\left(\frac{1}{2^\gamma} + \frac{1}{K}\right).$$

The implied constant is absolute.

Proof. In each Farey interval around p/q such that q is divisible by 2^γ there are at most $2/(Kq\delta) + 1$ many points x_i. By summing over p and q, we can bound the number of points in such intervals by

$$\ll \sum_{\substack{1 \le q \le K \\ 2^\gamma \mid q}} \sum_{1 \le p \le q} \left(\frac{1}{qK\delta} + 1\right) = \sum_{\substack{1 \le q \le K \\ 2^\gamma \mid q}} \left(\frac{1}{K\delta} + q\right) = (K2^{-\gamma} + 1)\frac{1}{K\delta} + \sum_{\substack{1 \le q \le K \\ 2^\gamma \mid q}} q$$

$$\le \frac{1}{2^\gamma \delta} + \frac{1}{K\delta} + 2^\gamma \sum_{1 \le q' \le \lfloor K2^{-\gamma} \rfloor} q' \ll \frac{K^2}{2^\gamma} + \frac{1}{2^\gamma \delta} + \frac{1}{K\delta}. \qquad \square$$

We proceed to the proof of Lemma 5.1. Consider $\mathfrak{p}_1$ and the case 'α discrete'. In this case, we have $p_{2^{2\mu+2\sigma}}(\alpha/2^{2\mu}) = \alpha$. Assume therefore that $\alpha = \alpha_0 + 2^{(k-1)\mu}\alpha_1$, where $\alpha_0 \in \{0, \ldots, 2^{(k-1)\mu} - 1\}$ and $\alpha_1 \in \{0, \ldots, 2^{\lambda-(k-1)\mu} - 1\}$.

Then

$$\mathfrak{p}_1 = p_{2^\sigma}(\alpha/2^{(k-1)\mu}) = p_{2^\sigma}(\alpha_0/2^{(k-1)\mu}) + q_{2^\sigma}(\alpha_0/2^{(k-1)\mu})\alpha_1.$$

By Lemma 5.3, using also (5.8), it follows that the number of $\alpha_0 \in \{0, \ldots, 2^{(k-1)\mu} - 1\}$ such that $2^\gamma \nmid q_{2^\sigma}(\alpha_0/2^{(k-1)\mu})$ is $2^{(k-1)\mu}(1 - \mathcal{O}(2^{-\gamma}))$. For each such α_0, we let α_1 run through $\{0, \ldots, 2^{\lambda-(k-1)\mu} - 1\}$. Then two occurrences α_1, α_1' such that $2^{2\gamma} \mid \mathfrak{p}_1$ are separated by at

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

least 2^γ steps; it follows that the number of such α_1 is bounded by $2^{\lambda-(k-1)\mu-\gamma}$. Putting these errors together, we see that the number of $\alpha \in \{0,\ldots,2^\lambda-1\}$ such that $2^{2\gamma} \nmid \mathfrak{p}_1$ is given by $2^{(k-1)\mu}(1-\mathcal{O}(2^{-\gamma}))2^{\lambda-(k-1)\mu}(1-\mathcal{O}(2^{-\gamma})) = 2^\lambda(1-\mathcal{O}(2^{-\gamma}))$.

Next, we consider the continuous case. We write $\alpha = \alpha_0 + 2^{2\mu}\alpha_1 + 2^{(k+1)\mu}\alpha_2$, where $\alpha_0 \in [0, 2^{2\mu})$ is real and $\alpha_1 < 2^{(k-1)\mu}$ and $\alpha_2 < 2^{\lambda-(k+1)\mu}$ are nonnegative integers. Set $p = p_{2^{2\mu+2\sigma}}(\alpha_0/2^{2\mu})$ and $q = q_{2^{2\mu+2\sigma}}(\alpha_0/2^{2\mu})$. Then

$$\frac{p_{2^{2\mu+2\sigma}}(\alpha/2^{2\mu})}{2^{(k-1)\mu}} = \frac{p+(\alpha_1+2^{(k-1)\mu}\alpha_2)q}{2^{(k-1)\mu}} = \frac{p+\alpha_1 q}{2^{(k-1)\mu}} + \alpha_2 q.$$

By the approximation property (4.5) (note that $\sigma \geq 1$) we have

$$\begin{aligned}\mathfrak{p}_1 &= \left\langle \left(\frac{p+\alpha_1 q}{2^{(k-1)\mu}} + \alpha_2 q\right) q_{2^\sigma}\left(\frac{p+\alpha_1 q}{2^{(k-1)\mu}}\right)\right\rangle \\ &= \left\langle \frac{p+\alpha_1 q}{2^{(k-1)\mu}} q_{2^\sigma}\left(\frac{p+\alpha_1 q}{2^{(k-1)\mu}}\right)\right\rangle + \alpha_2 q\, q_{2^\sigma}\left(\frac{p+\alpha_1 q}{2^{(k-1)\mu}}\right)\end{aligned}$$

and we note that the first summand does not depend on α_2.

As α_0 runs through $[0, 2^{2\mu}]$, we have by Lemma 5.2 that $2^\gamma \nmid q$ in a set of measure $2^{2\mu}(1-\mathcal{O}(2^{-\gamma}+2^{-2\mu-2\sigma}))$. By (5.8), this is $2^{2\mu}(1-\mathcal{O}(2^{-\gamma}))$. Assume that α_0 is such that $2^\gamma \nmid q$ and set $\gamma' = \nu_2(q) < \gamma$. Next, we let α_1 run. We choose $x_j = \{(p+jq)/2^{(k-1)\mu}\}$ for $0 \leq j < 2^{(k-1)\mu-\gamma'}$ and we note that these points satisfy $\|x_i - x_j\| \geq 1/2^{(k-1)\mu-\gamma'}$ for $i \neq j$. By Lemma 5.3 it follows that

$$\left\{\alpha_1 \in \{0,\ldots,2^{(k-1)\mu-\gamma'}-1\} : 2^\gamma \mid q_{2^\sigma}\left(\frac{p+\alpha_1 q}{2^{(k-1)\mu}}\right)\right\} \ll \frac{2^{2\sigma}}{2^\gamma} + 2^{(k-1)\mu-\gamma'}\left(\frac{1}{2^\gamma}+\frac{1}{2^\sigma}\right).$$

By (5.8), this is $\ll 2^{(k-1)\mu-\gamma'-\gamma}$. Performing this also for the other intervals of length $2^{(k-1)\mu-\gamma'}$, we obtain

$$\left\{\alpha_1 \in \{0,\ldots,2^{(k-1)\mu}-1\} : 2^\gamma \mid q_{2^\sigma}\left(\frac{p+\alpha_1 q}{2^{(k-1)\mu}}\right)\right\} \ll 2^{(k-1)\mu-\gamma}.$$

Finally, α_2 runs through $\{0,\ldots,2^{\lambda-(k+1)\mu}-1\}$ and we consider $\mathfrak{p}_1$. For given good α_1 and α_0 (such that $2^\gamma \nmid q$ and $2^\gamma \nmid q_{2^\sigma}((p+\alpha_1 q)/2^{(k-1)\mu})$), $\mathfrak{p}_1$ is an arithmetic progression in α_2 whose common difference is not divisible by $2^{2\gamma}$. Similarly to the discrete case, it follows that $\mathfrak{p}_1$ is divisible by $2^{3\gamma}$ for at most $2^{\lambda-(k+1)\mu-\gamma}$ many α_2. It follows that there is a set of measure

$$2^{2\mu}(1-\mathcal{O}(2^{-\gamma}))2^{(k-1)\mu}(1-\mathcal{O}(2^{-\gamma}))2^{\lambda-(k+1)\mu}(1-\mathcal{O}(2^{-\gamma})) = 2^\lambda(1-\mathcal{O}(2^{-\gamma}))$$

of $\alpha < 2^\lambda$ such that $2^{3\gamma} \nmid \mathfrak{p}_1$.

The cases $2 \leq i \leq k$ do not require any new ideas; we only give a sketch of a proof. Let $2 \leq i < k$. We treat the discrete and continuous cases in parallel. We write $\alpha = \alpha_0 + 2^{(i+1)\mu}\alpha_1 + 2^{(k+1)\mu}\alpha_2$, where $\alpha_0 < 2^{(i+1)\mu}$, and $\alpha_1 < 2^{(k-i)\mu}$ and $\alpha_2 < 2^{\lambda-(k+1)\mu}$ are nonnegative integers. Set $p = p_{2^{\mu+2\sigma}}(\alpha_0/2^{(i+1)\mu})$ and $q = q_{2^{\mu+2\sigma}}(\alpha_0/2^{(i+1)\mu})$. Then

$$\mathfrak{p}_i = \left\langle \frac{p+\alpha_1 q}{2^{(k-i)\mu}} q_{2^\sigma}\left(\frac{p+\alpha_1 q}{2^{(k-i)\mu}}\right)\right\rangle + \alpha_2 q q_{2^\sigma}\left(\frac{p+\alpha_1 q}{2^{(k-i)\mu}}\right),$$

as before. By Lemmas 5.2 and 5.3, we have $2^\gamma \nmid q$ for α_0 in a set of measure $2^{(i+1)\mu}(1-\mathcal{O}(2^{-\gamma}))$, where we used $2\mu + 4\sigma \leq (i+1)\mu$ in the discrete case. (We note that this last inequality is the

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

reason for defining $\mathfrak{p}_1$ separately, using $2^{2\mu}$ instead of 2^{μ}.) The remaining steps are as before and this case is finished.

Finally, in the case $i = k$, we write $\alpha = \alpha_0 + 2^{(k+1)\mu}\alpha_1$, where $\alpha_0 < (k+1)\mu$ and $\alpha_1 \in \{0, \dots, 2^{\lambda-(k+1)\mu} - 1\}$. Then

$$\mathfrak{p}_k = p_{2^{\mu+\sigma}}(\alpha_0/2^{(k+1)\mu}) + q_{2^{\mu+\sigma}}(\alpha_0/2^{(k+1)\mu})\alpha_1.$$

By Lemmas 5.2 and 5.3 and (5.8), we have $2^{\gamma} \mid q_{2^{\mu+\sigma}}(\alpha_0/2^{(k+1)\mu})$ for α_0 in a set of measure $\mathcal{O}(2^{(k+1)\mu-\gamma})$ and the statement follows as before.

In total, we have a set of measure $2^{\lambda}(1 - \mathcal{O}(2^{-\gamma}))$ of $\alpha < 2^{\lambda}$ such that $2^{3\gamma} \nmid \mathfrak{p}_i$ for all i.

Acknowledgements

The author wishes to thank Thomas Stoll for helpful discussions during his stay in Nancy, France, where the work on this project began. Moreover, the author wishes to thank Michael Drmota and Clemens Müllner for several fruitful discussions on the topic, and the anonymous referees for numerous suggestions. Finally, the author is indebted to Etienne Fouvry, Jakub Konieczny, Christian Mauduit and Gwendal Collet for valuable advice.

References

AS99 J.-P. Allouche and J. Shallit, *The ubiquitous Prouhet–Thue–Morse sequence*, in *Sequences and their applications (Singapore, 1998)*, Springer Series in Discrete Mathematics and Theoretical Computer Science (Springer, London, 1999), 1–16.

AS03 J.-P. Allouche and J. Shallit, *Automatic sequences: theory, applications, generalizations* (Cambridge University Press, Cambridge, 2003).

BFI86 E. Bombieri, J. B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. **156** (1986), 203–251.

BFI87 E. Bombieri, J. B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli. II*, Math. Ann. **277** (1987), 361–393.

BFI89 E. Bombieri, J. B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli. III*, J. Amer. Math. Soc. **2** (1989), 215–224.

DT06 C. Dartyge and G. Tenenbaum, *Congruences de sommes de chiffres de valeurs polynomiales*, Bull. Lond. Math. Soc. **38** (2006), 61–69.

DDM12 J.-M. Deshouillers, M. Drmota and J. F. Morgenbesser, *Subsequences of automatic sequences indexed by* $\lfloor n^c \rfloor$ *and correlations*, J. Number Theory **132** (2012), 1837–1866.

DMR19 M. Drmota, C. Mauduit and J. Rivat, *Normality along squares*, J. Eur. Math. Soc. **21** (2019), 507–548.

DR05 M. Drmota and J. Rivat, *The sum-of-digits function of squares*, J. Lond. Math. Soc. (2) **72** (2005), 273–292.

EH70 P. D. T. A. Elliott and H. Halberstam, *A conjecture in prime number theory*, in *Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69)* (Academic Press, London, 1970), 59–72.

Fou82 E. Fouvry, *Répartition des suites dans les progressions arithmétiques*, Acta Arith. **41** (1982), 359–382.

Fou84 E. Fouvry, *Autour du théorème de Bombieri–Vinogradov*, Acta Math. **152** (1984), 219–244.

FI80 E. Fouvry and H. Iwaniec, *On a theorem of Bombieri–Vinogradov type*, Mathematika **27** (1980), 135–152.

FM96a E. Fouvry and C. Mauduit, *Méthodes de crible et fonctions sommes des chiffres*, Acta Arith. **77** (1996), 339–351.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

FM96b E. Fouvry and C. Mauduit, *Sommes des chiffres et nombres presque premiers*, Math. Ann. **305** (1996), 571–599.

FI85 J. B. Friedlander and H. Iwaniec, *Incomplete Kloosterman sums and a divisor problem*, Ann. of Math. (2) **121** (1985), 319–350. With an appendix by Bryan J. Birch and Enrico Bombieri.

FI10 J. Friedlander and H. Iwaniec, *Opera de cribro* (American Mathematical Society, Providence, RI, 2010).

Gel68 A. O. Gel'fond, *Sur les nombres qui ont des propriétés additives et multiplicatives données*, Acta Arith. **13** (1967–1968), 259–265.

GPY09 D. A. Goldston, J. Pintz and C. Y. Yıldırım, *Primes in tuples. I*, Ann. of Math. (2) **170** (2009), 819–862.

Gow98 W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **8** (1998), 529–551.

Gow01 W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

GT08 B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), 481–547.

GT10 B. Green and T. Tao, *An arithmetic regularity lemma, an associated counting lemma, and applications*, in *An irregular mind*, Bolyai Society Mathematical Studies, vol. 21 (János Bolyai Mathematical Society, Budapest, 2010), 261–334.

HW54 G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, third edition (Clarendon Press, Oxford, 1954).

Kon14 A. Kontorovich, *Levels of distribution and the affine sieve*, Ann. Fac. Sci. Toulouse Math. (6) **23** (2014), 933–966.

Kon19 J. Konieczny, *Gowers norms for the Thue–Morse and Rudin–Shapiro sequences*, Ann. Inst. Fourier (Grenoble) **69** (2019), 1897–1913.

MMR14 B. Martin, C. Mauduit and J. Rivat, *Théoréme des nombres premiers pour les fonctions digitales*, Acta Arith. **165** (2014), 11–45.

Mau01 C. Mauduit, *Multiplicative properties of the Thue–Morse sequence*, Period. Math. Hungar. **43** (2001), 137–153.

MR95 C. Mauduit and J. Rivat, *Répartition des fonctions q-multiplicatives dans la suite* $([n^c])_{n\in\mathbf{N}}$, $c>1$, Acta Arith. **71** (1995), 171–179.

MR05 C. Mauduit and J. Rivat, *Propriétés q-multiplicatives de la suite* $\lfloor n^c\rfloor$, $c>1$, Acta Arith. **118** (2005), 187–203.

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

MR10 C. Mauduit and J. Rivat, *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Ann. of Math. (2) **171** (2010), 1591–1646.

May15 J. Maynard, *Small gaps between primes*, Ann. of Math. (2) **181** (2015), 383–413.

Mor11 J. F. Morgenbesser, *The sum of digits of* $\lfloor n^c\rfloor$, Acta Arith. **148** (2011), 367–393.

MSS11 J. F. Morgenbesser, J. Shallit and T. Stoll, *Thue–Morse at multiples of an integer*, J. Number Theory **131** (2011), 1498–1512.

Mos07 Y. Moshe, *On the subword complexity of Thue–Morse polynomial extractions*, Theoret. Comput. Sci. **389** (2007), 318–329.

MS17 C. Müllner and L. Spiegelhofer, *Normality of the Thue–Morse sequence along Piatetski-Shapiro sequences, II*, Israel J. Math. **220** (2017), 691–738.

Pia53 I. I. Piatetski-Shapiro, *On the distribution of prime numbers in sequences of the form* $[f(n)]$, Mat. Sb. **33** (1953), 559–566.

RS01 J. Rivat and P. Sargos, *Nombres premiers de la forme* $\lfloor n^c\rfloor$, Canad. J. Math. **53** (2001), 414–433.

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press

Spi14 L. Spiegelhofer, *Piatetski-Shapiro sequences via Beatty sequences*, Acta Arith. **166** (2014), 201–229.

Spi15 L. Spiegelhofer, *Normality of the Thue–Morse sequence along Piatetski-Shapiro sequences*, Q. J. Math. **66** (2015), 1127–1138.

Tao12 T. Tao, *Higher order Fourier analysis*, Graduate Studies in Mathematics, vol. 142 (American Mathematical Society, Providence, RI, 2012).

Zha14 Y. Zhang, *Bounded gaps between primes*, Ann. of Math. (2) **179** (2014), 1121–1174.

Lukas Spiegelhofer lukas.spiegelhofer@unileoben.ac.at

Institute of Discrete Mathematics and Geometry, Vienna University of Technology, Wiedner Hauptstrasse 8–10, 1040 Vienna, Austria

and

Department of Mathematics and Information Technology, Montanuniversität Leoben, Franz-Josef-Straße 18, 8700 Leoben, Austria

https://doi.org/10.1112/S0010437X20007563 Published online by Cambridge University Press