JOURNAL de Théorie des Nombres de BORDEAUX

anciennement Séminaire de Théorie des Nombres de Bordeaux

Terence TAO et Joni TERÄVÄINEN

Odd order cases of the logarithmically averaged Chowla conjecture

Tome 30, n° 3 (2018), p. 997-1015.

<http://jtnb.cedram.org/item?id=JTNB_2018__30_3_997_0>

Journal de Théorie des Nombres
de Bordeaux **30** (2018), 997–1015

Odd order cases of the logarithmically averaged Chowla conjecture

par Terence TAO et Joni TERÄVÄINEN

Résumé. Une conjecture bien connue de Chowla affirme que les corrélations des translatés de la fonction $\lambda(n)$ de Liouville sont asymptotiquement nulles. Dans un article récent, les auteurs ont démontré un résultat partiel en direction de la conjecture d'Elliott logarithmiquement pondérée concernant les corrélations des fonctions multiplicatives, qui à son tour implique tous les cas de la conjecture de Chowla avec un nombre impair de translatés. Dans cet article, nous donnons une nouvelle démonstration de ce dernier résultat sur la conjecture de Chowla. En particulier, celle-ci évite l'usage de la théorie ergodique qui joue un rôle crucial dans notre démonstration précédente.

Abstract. A famous conjecture of Chowla states that the Liouville function $\lambda(n)$ has negligible correlations with its shifts. Recently, the authors established a weak form of the logarithmically averaged Elliott conjecture on correlations of multiplicative functions, which in turn implied all the odd order cases of the logarithmically averaged Chowla conjecture. In this note, we give a new proof of the odd order cases of the logarithmically averaged Chowla conjecture. In particular, this proof avoids all mention of ergodic theory, which had an important role in the previous proof.

1. Introduction

Let $\lambda(n)$ be the Liouville function, defined as $\lambda(n) := (-1)^{\Omega(n)}$, with $\Omega(n)$ being the number of prime factors of the integer n counting multiplicity. The distribution of $\lambda(n)$ has been extensively studied. For instance, the statement

$$\frac{1}{x}\sum_{n\leq x}\lambda(an+b) = o_{x\to\infty}(1)$$

for any fixed $a \in \mathbb{N}$, $b \in \mathbb{Z}$ is equivalent to the prime number theorem in arithmetic progressions by an elementary argument. It was conjectured by Chowla [2] that we have the significantly more general correlation estimate

$$\frac{1}{x}\sum_{n\leq x}\lambda(a_1n+b_1)\cdots\lambda(a_kn+b_k) = o_{x\to\infty}(1) \tag{1.1}$$

Manuscrit reçu le 5 octobre 2017, accepté le 10 avril 2018.
2010 *Mathematics Subject Classification.* 11N37.
Mots-clefs. Liouville function, Chowla's conjecture, Gowers uniformity norms.

for any $k \geq 1$, $a_1, \ldots, a_k, b_1, \ldots, b_k \in \mathbb{N}$ satisfying the non-degeneracy condition $a_i b_j - a_j b_i \neq 0$ for $1 \leq i < j \leq k$. The non-degeneracy condition may be omitted when k is odd, since a degenerate pair $\lambda(a_i n + b_i)\lambda(a_j n + b_j)$ with $a_i b_j - a_j b_i = 0$ is constant in n and can therefore be deleted. One can of course extend this conjecture to the case where the $b_1, \ldots, b_k$ are integers rather than natural numbers (after defining λ arbitrarily on negative numbers), but this leads to an equivalent conjecture after applying a translation in the n variable.

Chowla's conjecture (1.1) can be thought of as a simpler analogue of the famous Hardy–Littlewood prime k-tuple conjecture [13], [10, Section 1], which predicts an asymptotic for the correlations of the von Mangoldt function $\Lambda(n)$. Any rigorous implication between (1.1) and the Hardy–Littlewood k-tuples conjecture, however, would require good savings of the type $O((\log x)^{-A})$ for the error term $o_{x\to\infty}(1)$ in (1.1) and a large regime of uniformity in the parameters $a_1, \ldots, a_k$, $b_1, \ldots, b_k$; none of the currently known partial progress on Chowla's conjecture for $k > 1$ fulfills these additional requirements. Nevertheless, Chowla's conjecture is subject to the well-known *parity problem* of sieve theory, which also obstructs sieve theoretic approaches to the Hardy–Littlewood prime k-tuple conjecture. The parity problem states the fact, first observed by Selberg (see [8, Chapter 16]), that classical combinatorial sieves are unable to distinguish numbers with an odd and even number of prime factors from each other.

One can also view Chowla's conjecture as a special case of Elliott's conjecture on correlations of multiplicative functions (see [22, Section 1] for a modern version of this conjecture, avoiding a technical counterexample to the original conjecture in [3]).

In [19], Matomäki, Radziwiłł and the first author showed that Chowla's conjecture holds on average over the shifts $b_1, \ldots, b_k$, and this was generalised by Frantzikinakis [5] to averages over independent polynomials. Nevertheless, not much is known in the case of individual shifts, unless one considers the logarithmically averaged[1] version of the conjecture, which states that

$$\frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(a_1 n + b_1) \cdots \lambda(a_k n + b_k)}{n} = o_{x\to\infty}(1), \tag{1.2}$$

provided again that $a_i b_j - a_j b_i \neq 0$ for $1 \leq i < j \leq k$. These logarithmically averaged correlations are certainly easier, since (1.1) implies (1.2) by partial summation. For the logarithmically averaged variant (1.2) of Chowla's conjecture, it was shown by the first author [20] that (1.2) is $o_{x\to\infty}(1)$ for $k = 2$, and we recently showed in [22] that the same conclusion holds for

[1] If $1 \leq \omega(x) \leq x$ is any function tending to infinity, one could equally well consider (1.2) with a sum over $\frac{x}{\omega(x)} \leq n \leq x$, with the $\log x$ normalisation replaced by $\log \omega(x)$. In fact, this is what is done in [20], [22].

all odd k. Both of these works actually handle more general correlations of bounded multiplicative functions, with [20] having the same assumptions as in Elliott's conjecture, and [22] having a non-pretentious assumption for the product of the multiplicative functions (see [22, Corollary 1.4] for a precise statement). In addition, it was recently shown by Frantzikinakis and Host [6, Theorem 1.4] that if one replaces the weight $\frac{1}{n}$ in (1.2) with $\frac{e^{2\pi i\alpha n}}{n}$ for any irrational α, then the analogue of (1.2) holds for all k. When it comes to conditional results, Frantzikinakis [4] showed that the logarithmically averaged Chowla conjecture would follow from ergodicity of the measure preserving system associated with the Liouville function.

The proof in [22] of the odd order cases of the logarithmically averaged Chowla conjecture relies on deep results of Leibman [17] and Le [16] on ergodic theory, and is not much simpler than the proof of the structural theorem for correlations of general bounded multiplicative functions in that paper. Here we give a different proof of the odd order cases of Chowla's conjecture, which avoids all use of ergodic theory, although it now requires the Gowers uniformity of the von Mangoldt function, established by Green, the first author and Ziegler [10], [11], [12]. The proof we give here is also shorter than the earlier proof, given the mentioned Gowers uniformity result. More precisely, we will prove the following.

Theorem 1.1 (Odd order cases of the logarithmic Chowla conjecture)**.** *Let $k \geq 1$ be an odd natural number, and let $a_1, \dots, a_k, b_1, \dots, b_k$ be natural numbers. Then we have*

$$\frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(a_1 n + b_1) \cdots \lambda(a_k n + b_k)}{n} = o_{x\to\infty}(1).$$

Remark 1.2. *As remarked previously, as we are dealing with an odd number of shifts of the Liouville function, there is no need to impose any non-degeneracy assumptions on the coefficients $a_1, \dots, a_k, b_1, \dots, b_k$.*

Remark 1.3. *Using the same proof as for Theorem 1.1, one could establish an analogous statement for the Möbius function $\mu(\cdot)$, namely that*

$$\frac{1}{\log x} \sum_{n \leq x} \frac{\mu(a_1 n + b_1)^{c_1} \cdots \mu(a_k n + b_k)^{c_k}}{n} = o_{x\to\infty}(1) \tag{1.3}$$

whenever $c_j \geq 1$ are fixed integers with $c_1 + \cdots + c_k$ odd and a_j, b_j are as above (see also [22, Corollary 1.6]*).*[2]

[2] The only small modification in the proof of (1.3) compared to that of Theorem 1.1 is in the approximate functional equation (Theorem 3.1). The approximate functional equation holds in the same form for correlations of the Möbius function, but in its proof the multiplicativity relation $\lambda(pn) = -\lambda(n)$ is to be replaced with $\mu(pn)^c = (-1)^c \mu(n)^c + O(1_{p|n})$. The contribution of $O(1_{p|n})$ is negligible by the triangle inequality and the fact that p will be moderately large.

Remark 1.4. *From the proof of Theorem 1.1, we see that for the three-point case $k = 3$ of Theorem 1.1, we only need U^3-uniformity of the von Mangoldt function, which was established in* [9] *and is simpler than the general U^k-uniformity result. In contrast, in* [22] *the $k = 3$ case was no easier than the general case.*

It was shown by the first author in [21] that the logarithmically averaged Chowla conjecture (1.2) for all k is equivalent to two difficult conjectures, namely the logarithmically averaged Sarnak conjecture [21, Conjecture 1.5] and the (logarithmic) local Gowers uniformity of the Liouville function [21, Conjecture 1.6]. We manage to avoid these problems, since we will only be dealing with odd values of k. Indeed, it is natural that the even order cases of Chowla's conjecture are harder than the odd order ones, since one can use the Kátai–Bourgain–Sarnak–Ziegler orthogonality criterion [15], [1] to show that the even order cases imply the odd order ones (see [22, Remark 1.7]). Another indication that the even order cases are more challenging is Elliott's result [3] that for odd k the $\limsup$ of the absolute value of (1.2) is strictly less than 1; in the even order cases this has not been shown in general. We also remark that the proof of Theorem 1.1 does not require the Matomäki–Radziwiłł theorem [18], in contrast to the $k = 2$ result in [20] which relied crucially on this theorem.

Acknowledgments. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1266164.

JT was supported by UTUGS Graduate School and project number 293876 of the Academy of Finland.

We thank the anonymous referees for careful reading of the paper and valuable comments. Part of this paper was written while the authors were in residence at MSRI in spring 2017, which is supported by NSF grant DMS-1440140. We thank Kaisa Matomäki for helpful discussions and encouragement and Maksym Radziwiłł for suggesting the use of semiprimes in the entropy decrement argument.

2. Notation

We use standard notation for arithmetic functions throughout this paper. In particular, $\lambda(n)$ is the Liouville function, $\mu(n)$ is the Möbius function, $\Lambda(n)$ is the von Mangoldt function, and $\varphi(n)$ is the Euler totient function. Various letters, such as m, n, d, a_j, b_j, are reserved for integer variables. We use (n, m) to denote the greatest common divisor of n and m. The variable p in turn will always be a prime; in particular, summations such as $\sum_{p \in A} f(p)$ will always be understood to restricted to primes. We will use the standard Landau asymptotic notations $O(\cdot)$, $o(\cdot)$, with $o_{\eta\to 0}(1)$

for instance signifying a quantity that tends to 0 as $\eta \to 0$; we also use the Vinogradov notation $X \ll Y$ for $X = O(Y)$.

For a proposition $P(n)$ depending on n, we denote by $1_{P(n)}$ the function that takes value 1 if $P(n)$ is true and 0 if it is false. We also use the expectation notations

$$\mathbb{E}_{n\in A} f(n) := \frac{\sum_{n\in A} f(n)}{\sum_{n\in A} 1}$$

and

$$\mathbb{E}^{\log}_{n\in A} f(n) := \frac{\sum_{n\in A} \frac{f(n)}{n}}{\sum_{n\in A} \frac{1}{n}}$$

whenever A is a finite non-empty set and $f : A \to \mathbb{C}$ is a function. If we replace the symbol n by p, it is understood that all sums involved are over primes, thus for instance

$$\mathbb{E}^{\log}_{p\in A} f(p) := \frac{\sum_{p\in A} \frac{f(p)}{p}}{\sum_{p\in A} \frac{1}{p}}.$$

Strictly speaking, this average may be undefined if A contains no primes, but in practice we will always be in a regime in which A contains plenty of primes.

3. The two key subtheorems

Let k be a natural number, and let $a_1, \dots, a_k, b_1, \dots, b_k$ be natural numbers. All implied constants in asymptotic notation (and in assertions such as "X is sufficiently large depending on Y" are henceforth allowed to depend on these quantities. For any natural number a and any $x \geq 1$, define the quantity

$$f_x(a) := \mathbb{E}^{\log}_{n\leq x} \lambda(a_1 n + ab_1) \cdots \lambda(a_k n + ab_k). \tag{3.1}$$

To prove Theorem 1.1, it will suffice to show that

$$f_x(1) \ll \varepsilon \tag{3.2}$$

whenever $\varepsilon > 0$, k is odd, and x is sufficiently large depending on ε (and, by the preceding convention, on $k, a_1, \dots, a_k, b_1, \dots, b_k$).

To obtain (3.2), we will rely crucially on the following approximate functional equation for f_x, which informally asserts that $f_x(ap) \approx (-1)^k f_x(a)$ for "most" a and p:

Theorem 3.1 (Approximate functional equation)**.** *Let $k, a_1, \dots, a_k$, $b_1, \dots, b_k$ be natural numbers. For any $0 < \varepsilon < 1$, $x > 1$, and any natural number a, one has*

$$\mathbb{E}_{2^m < p \leq 2^{m+1}} |f_x(ap) - (-1)^k f_x(a)| \ll \varepsilon \tag{3.3}$$

for all natural numbers $m \leq \log\log x$ outside of an exceptional set $\mathcal{M}$ with

$$\sum_{m\in\mathcal{M}} \frac{1}{m} \ll a\varepsilon^{-3}, \tag{3.4}$$

where the quantity $f_x(a)$ is defined in (3.1).

Results similar to these appear in [6, Theorem 3.6], [22, Theorem 3.6]. As in these references, we will prove Theorem 3.1 in Section 4 via the entropy decrement argument introduced in [20]; we will use the modification of that argument in [22] to obtain the relatively strong bound (3.4).

From (3.3) we have

$$\mathbb{E}^{\log}_{2^m<p\leq 2^{m+1}} |f_x(ap) - (-1)^k f_x(a)| \ll \varepsilon$$

(since $1/p$ is comparable to $1/2^m$ in the range $2^m < p \leq 2^{m+1}$), and hence from the triangle inequality we have

$$f_x(a) = (-1)^k \mathbb{E}^{\log}_{2^m<p\leq 2^{m+1}} f_x(ap) + O(\varepsilon)$$

for all m with $2^m \leq (\log x)^{1/2}$ outside of the exceptional set $\mathcal{M}$. The fact that the average on the right-hand side is over primes will be inconvenient for our argument. To overcome this, we will establish the following comparison.

Theorem 3.2 (Comparison)**.** *Let $k, a_1, \dots, a_k, b_1, \dots, b_k$ be natural numbers. Let $0 < \varepsilon < 1$, and let*

$$1 < w < H_- < H_+ < x$$

be parameters with w be sufficiently large depending on ε; H_- sufficiently large depending on w, ε; H_+ sufficiently large depending on H_-, w, ε; and x sufficiently large depending on H_+, H_-, w, ε. Set $W := \prod_{p\leq w} p$. Then, for any natural number $a \leq H_+$ and any m with $H_- \leq 2^m \leq H_+$, one has

$$\mathbb{E}^{\log}_{2^m<p\leq 2^{m+1}} f_x(ap) = \mathbb{E}^{\log}_{2^m<n\leq 2^{m+1}:(n,W)=1} f_x(an) + O(\varepsilon)$$

where the quantity $f_x(a)$ is defined in (3.1).

We will prove this assertion in Section 5. Our main tool will be the theory of the Gowers uniformity norms, and in particular the Gowers uniformity of the W-tricked von Mangoldt function proven in [10], [11], [12]. In contrast to Theorem 3.1, the bounds in Theorem 3.2 (particularly with regards to what "sufficiently large" means) are qualitative rather than quantitative; this is primarily due to the qualitative nature of the bounds currently available for the Gowers uniformity of the W-tricked von Mangoldt function. A key technical point in the above theorem is that the parameter a is permitted to be large compared to the parameter w (or W); this will be important in the argument below.

In the remainder of this section we show how Theorem 3.1 and Theorem 3.2 yield (3.2) when k is odd and x is sufficiently large depending on ε.

Fix $0 < \varepsilon < 1/2$. We will need parameters

$$\frac{1}{\varepsilon} < w < H_1 < H_2 < H_3 < H_4 < x \tag{3.5}$$

with w sufficiently large depending on ε, each H_i for $i = 1, 2, 3, 4$ sufficiently large depending on w, ε and $H_1, \dots, H_{i-1}$, and x sufficiently large depending on $H_4, H_3, H_2, H_1, w, \varepsilon$.

From Theorem 3.1 and the hypothesis that k is odd, one has

$$f_x(1) = -\mathbb{E}^{\log}_{2^m < p_1 \le 2^{m+1}} f_x(p_1) + O(\varepsilon)$$

for all m in the range $H_1 \le 2^m \le H_2$, outside of an exceptional set $\mathcal{M}_1$ with

$$\sum_{m \in \mathcal{M}_1} \frac{1}{m} \ll \varepsilon^{-3}.$$

For m in this exceptional set, we of course have

$$f_x(1) = -\mathbb{E}^{\log}_{2^m < p_1 \le 2^{m+1}} f_x(p_1) + O(1).$$

Averaging over all such m and using the prime number theorem, we conclude (given the hypotheses on the parameters (3.5)) that

$$f_x(1) = -\mathbb{E}^{\log}_{H_1 < p_1 \le H_2} f_x(p_1) + O(\varepsilon). \tag{3.6}$$

A similar application of Theorem 3.1 yields

$$f_x(1) = -\mathbb{E}^{\log}_{H_3 < p \le H_4} f_x(p) + O(\varepsilon). \tag{3.7}$$

Also, applying Theorem 3.1 with a replaced by p_1, we have

$$f_x(p_1) = -\mathbb{E}^{\log}_{H_3 < p_2 \le H_4} f_x(p_1 p_2) + O(\varepsilon)$$

for all primes p_1 with $H_1 < p_1 \le H_2$; inserting this into (3.6), we obtain

$$f_x(1) = +\mathbb{E}^{\log}_{H_1 < p_1 \le H_2} \mathbb{E}^{\log}_{H_3 < p_2 \le H_4} f_x(p_1 p_2) + O(\varepsilon). \tag{3.8}$$

Crucially, the sign in (3.8) is the opposite of the sign in (3.7). To conclude the proof of (3.2) from (3.7), (3.8), it will suffice to show that the average (3.7) involving primes p and the average (3.8) involving semiprimes $p_1 p_2$ are comparable in the sense that

$$\mathbb{E}^{\log}_{H_3 < p \le H_4} f_x(p) = \mathbb{E}^{\log}_{H_1 < p_1 \le H_2} \mathbb{E}^{\log}_{H_3 < p_2 \le H_4} f_x(p_1 p_2) + O(\varepsilon). \tag{3.9}$$

To do this, we use Theorem 3.2 several times. Firstly, from this theorem we see that

$$\mathbb{E}^{\log}_{2^m < p \le 2^{m+1}} f_x(p) = \mathbb{E}^{\log}_{2^m < n \le 2^{m+1} : (n,W)=1} f_x(n) + O(\varepsilon)$$

whenever $H_3 \leq 2^m \leq H_4$; averaging over m (and noting that the error terms that arise can be easily absorbed into the $O(\varepsilon)$ error) we conclude that

$$\mathbb{E}^{\log}_{H_3<p\leq H_4} f_x(p) = \mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} f_x(n) + O(\varepsilon).$$

Similarly, we have

$$\mathbb{E}^{\log}_{H_3<p_2\leq H_4} f_x(p_1p_2) = \mathbb{E}^{\log}_{H_3<n_2\leq H_4:(n_2,W)=1} f_x(p_1n_2) + O(\varepsilon)$$

whenever $H_1 < p_1 \leq H_2$ (note that this is despite p_1 being large compared with w or W). Thus it will suffice to show that

$$\begin{aligned} \mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} & f_x(n) \\ &= \mathbb{E}^{\log}_{H_1<p_1\leq H_2} \mathbb{E}^{\log}_{H_3<n_2\leq H_4:(n_2,W)=1} f_x(p_1n_2) + O(\varepsilon). \end{aligned} \tag{3.10}$$

By making the change of variables $n = p_1n_2$, and noting that n is coprime to W if and only if n_2 is, we can write

$$\mathbb{E}^{\log}_{H_3<n_2\leq H_4:(n_2,W)=1} f_x(p_1n_2) = \mathbb{E}^{\log}_{p_1H_3<n\leq p_1H_4:(n,W)=1} f_x(n)p_1 1_{p_1|n} + O(\varepsilon),$$

and one can modify the range $p_1H_3 < n \leq p_1H_4$ to $H_3 < n \leq H_4$ incurring a further error of $O(\varepsilon)$. We may thus rearrange (3.10) as

$$\mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} f_x(n)(g(n)-1) = O(\varepsilon)$$

where g is the weight

$$g(n) := \mathbb{E}^{\log}_{H_1<p_1\leq H_2} p_1 1_{p_1|n}.$$

By the Cauchy–Schwarz inequality and the boundedness of f_x, it thus suffices to establish the bound

$$\mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} (g(n)-1)^2 \ll \varepsilon^2$$

which will follow in turn from the bounds

$$\mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} g(n) = 1 + O(\varepsilon^2) \tag{3.11}$$

and

$$\mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} g(n)^2 = 1 + O(\varepsilon^2). \tag{3.12}$$

The left-hand side of (3.11) can be rewritten as

$$\mathbb{E}^{\log}_{H_1<p_1\leq H_2} p_1 \mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} 1_{p_1|n}$$

and the claim (3.11) follows since one can easily compute that

$$\mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} 1_{p_1|n} = \frac{1+O(\varepsilon^2)}{p_1}.$$

Similarly, the left-hand side of (3.12) can be rewritten as

$$\mathbb{E}^{\log}_{H_1<p_1\leq H_2} \mathbb{E}^{\log}_{H_1<p_1'\leq H_2} p_1p_1' \mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} 1_{p_1,p_1'|n}$$

and the claim (3.12) follows since $\mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} 1_{p_1,p_1'|n}$ is equal to $\frac{1+O(\varepsilon^2)}{p_1p_1'}$ when $p_1 \neq p_1'$, and can be bounded crudely by $O(1/p_1)$ when $p_1 = p_1'$. This concludes the proof of Theorem 1.1, except for the proofs of Theorem 3.1 and Theorem 3.2 which will be accomplished in the next two sections respectively.

4. Using the entropy decrement argument

We now prove Theorem 3.1. Let $k, a_1, \dots, a_k, b_1, \dots, b_k, \varepsilon, a, x$ be as in that theorem. We may assume that

$$x \geq \exp\exp\exp(a\varepsilon^{-3}) \tag{4.1}$$

since otherwise the claim is trivial by setting $\mathcal{M}$ to consist of all $m \leq \log\log x$. We may also restrict attention to proving (3.3) for m satisfying

$$\exp(a\varepsilon^{-3}) \leq m \leq \frac{1}{100}\log\log x \tag{4.2}$$

since all the m between $\frac{1}{100}\log\log x$ and $\log\log x$, or less than $\exp(a\varepsilon^{-3})$, can be placed in the exceptional set $\mathcal{M}$ without significantly affecting (3.4). Finally, we can assume that $\varepsilon \leq 1/2$, since for $1/2 < \varepsilon \leq 1$ the bound (3.3) holds from the triangle inequality.

For any prime p, one has the identity

$$\lambda(n) = -\lambda(pn)$$

for any natural number n, and hence

$$\lambda(a_1n + ab_1)\cdots\lambda(a_kn + ab_k) = (-1)^k\lambda(a_1pn + apb_1)\cdots\lambda(a_kn + apb_k).$$

From (3.1) we thus have

$$f_x(a) = (-1)^k\mathbb{E}^{\log}_{n\leq x}\lambda(a_1pn + apb_1)\cdots\lambda(a_kpn + apb_k).$$

If $p \leq \log x$, then (using (4.1)) we have $\sum_{x<n\leq px}\frac{1}{n} \ll \varepsilon\sum_{n\leq x}\frac{1}{n}$, and hence[3] that

$$\mathbb{E}^{\log}_{n\leq px} g(n) = \mathbb{E}^{\log}_{n\leq x} g(n) + O(\varepsilon)$$

whenever $g : \mathbb{N} \to \mathbb{C}$ is bounded in magnitude by 1. Thus we have

$$f_x(a) = (-1)^k\mathbb{E}^{\log}_{n\leq px}\lambda(a_1pn + apb_1)\cdots\lambda(a_kpn + apb_k) + O(\varepsilon)$$

for all $p \leq \log x$. Making the change of variables $n' := pn$, we conclude that

$$f_x(a) = (-1)^k\mathbb{E}^{\log}_{n'\leq x}\lambda(a_1n' + apb_1)\cdots\lambda(a_kn' + apb_k)p1_{p|n'} + O(\varepsilon).$$

---

[3] Here it is essential that we are using logarithmic averaging; the argument breaks down completely at this point if one uses ordinary averaging.

Replacing n' with n, and comparing with (3.1) with a replaced by ap, we conclude that

$$\begin{aligned} &f_x(a) - (-1)^k f_x(ap) \\ &\qquad = (-1)^k \mathbb{E}^{\log}_{n \le x} \lambda(a_1 n + apb_1) \cdots \lambda(a_k n + apb_k)(p 1_{p|n} - 1) + O(\varepsilon). \end{aligned}$$

The contribution of those n with $n \le x^\varepsilon$ is $O(\varepsilon)$, so we have

$$\begin{aligned} &f_x(a) - (-1)^k f_x(ap) \\ &\qquad = (-1)^k \mathbb{E}^{\log}_{x^\varepsilon < n \le x} \lambda(a_1 n + apb_1) \cdots \lambda(a_k n + apb_k)(p 1_{p|n} - 1) + O(\varepsilon) \end{aligned}$$

for all $p \le \log x$. If we set $c_p \in \{-1, 0, +1\}$ to be the signum of $\mathbb{E}^{\log}_{n \le x} \lambda(a_1 n + apb_1) \cdots \lambda(a_k n + apb_k)(p 1_{p|n} - 1)$, it will thus suffice to show that

$$\begin{aligned} \mathbb{E}_{2^m < p \le 2^{m+1}} c_p \mathbb{E}^{\log}_{x^\varepsilon < n \le x} \lambda(a_1 n + apb_1) \cdots \lambda(a_k n + apb_k)(p 1_{p|n} - 1) \\ = O(\varepsilon) \end{aligned} \tag{4.3}$$

for all m obeying (4.2), outside of an exceptional set $\mathcal{M}$ obeying (3.4).

Let m obey (4.2). If j is a natural number less than or equal to 2^m (and hence of size $O(\log^{1/10} x)$), one easily computes the total variation bound

$$\sum_{x^\varepsilon < n \le x^\varepsilon + j} \frac{1}{n} + \sum_{x^\varepsilon + j < n \le x + j} \left| \frac{1}{n} - \frac{1}{n+j} \right| \ll \frac{\log^{1/10} x}{x^\varepsilon}$$

and thus

$$\mathbb{E}^{\log}_{x^\varepsilon \le n \le x} g(n) = \mathbb{E}^{\log}_{x^\varepsilon \le n \le x} g(n+j) + O\left(\frac{\log^{1/10} x}{x^\varepsilon \log x} \right)$$

for any function $g : \mathbb{N} \to \mathbb{C}$ bounded in magnitude by 1. By (4.1), the error term is certainly of size $O(\varepsilon)$. In particular, the left-hand side of (4.3) can be written as

$$\begin{aligned} &\mathbb{E}_{2^m \le p \le 2^{m+1}} c_p \mathbb{E}^{\log}_{x^\varepsilon < n \le x} \lambda(a_1 n + a_1 j + apb_1) \\ &\qquad\qquad \cdots \lambda(a_k n + a_k j + apb_k)(p 1_{p|n+j} - 1) + O(\varepsilon) \end{aligned}$$

for any $1 \le j \le 2^m$. Averaging in j and rearranging, we can thus write the left-hand side of (4.3) in probabilistic language[4] as

$$\mathbf{E}\mathbf{Z}_m + O(\varepsilon),$$

where $\mathbf{E}$ denotes expectation, $\mathbf{Z}_m$ is the random variable

$$\begin{aligned} \mathbf{Z}_m := \mathbb{E}_{2^m < p \le 2^{m+1}} \mathbb{E}_{j \le 2^m} c_p \lambda(a_1 \mathbf{n} + a_1 j + apb_1) \\ \cdots \lambda(a_k \mathbf{n} + a_k j + apb_k)(p 1_{p|\mathbf{n}+j} - 1), \end{aligned}$$

[4] We will use boldface symbols such as $\mathbf{n}, \mathbf{X}_m, \mathbf{Y}_m, \mathbf{Z}_m$ to denote random variables, with non-boldface symbols such as X_m being used to denote deterministic variables instead.

and $\mathbf{n}$ is a random natural number in the interval $(x^\varepsilon, x]$ drawn using the logarithmic distribution

$$\mathbf{P}(\mathbf{n} = n) = \frac{1/n}{\sum_{x^\varepsilon < n' \leq x} \frac{1}{n'}}$$

for all $x^\varepsilon < n \leq x$.

We now "factor" the random variable $\mathbf{Z}_m$ into a function of two other random variables $\mathbf{X}_m, \mathbf{Y}_m$, defined as follows. Let $B := \max_i b_i$ and

$$C := \sum_{i=1}^{k} (2aB + 1)a_i,$$

and let $\mathbf{X}_m \in \{-1, +1\}^{C2^m}$ and $\mathbf{Y}_m \in \prod_{2^m < p \leq 2^{m+1}} \mathbb{Z}/p\mathbb{Z}$ be the random variables

$$\mathbf{X}_m := (\lambda(a_i \mathbf{n} + r))_{1 \leq i \leq k; 1 \leq r \leq (2aB+1)a_i 2^m}$$

and

$$\mathbf{Y}_m := (\mathbf{n} \bmod p)_{2^m < p \leq 2^{m+1}}.$$

Then we may write $\mathbf{Z}_m = F_m(\mathbf{X}_m, \mathbf{Y}_m)$, where $F_m : \{-1, +1\}^{C2^m} \times \prod_{2^m < p \leq 2^{m+1}} \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ is the function defined by

$$\begin{aligned} &F_m((b_{i,r})_{1 \leq i \leq k; 1 \leq r \leq (2aB+1)a_i 2^m}, (n_p)_{2^m < p \leq 2^{m+1}}) \\ &\qquad := \mathbb{E}_{2^m < p \leq 2^{m+1}} \mathbb{E}_{j \leq 2^m} c_p b_{1, a_1 j + apb_1} \cdots b_{k, a_k j + apb_k} (p 1_{p | n_p + j} - 1). \end{aligned}$$

for all $b_{i,r} \in \{-1, +1\}$ and $n_p \in \mathbb{Z}/p\mathbb{Z}$. It will now suffice to show that

$$\mathbf{E} F_m(\mathbf{X}_m, \mathbf{Y}_m) = O(\varepsilon)$$

for all m obeying (4.2), outside of an exceptional set $\mathcal{M}$ obeying (3.4).

At this point we recall some information-theoretic concepts:

Definition 4.1 (Entropy and conditional expectation)**.** *Let* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *be random variables taking finitely many values. Then we have the entropy*

$$\mathbf{H}(\mathbf{X}) := \sum_x \mathbf{P}(\mathbf{X} = x) \log \frac{1}{\mathbf{P}(\mathbf{X} = x)}$$

where the sum is over all x *for which* $\mathbf{P}(\mathbf{X} = x) \neq 0$*. Similarly we have the conditional entropy*

$$\mathbf{H}(\mathbf{X}|E) := \sum_x \mathbf{P}(\mathbf{X} = x|E) \log \frac{1}{\mathbf{P}(\mathbf{X} = x|E)}$$

for any event E *of positive probability, and*

$$\mathbf{H}(\mathbf{X}|\mathbf{Y}) := \sum_y \mathbf{P}(\mathbf{Y} = y) \mathbf{H}(\mathbf{X}|\mathbf{Y} = y).$$

Finally, we define the mutual information

$$\mathbf{I}(\mathbf{X} : \mathbf{Y}) = \mathbf{H}(\mathbf{X}) - \mathbf{H}(\mathbf{X}|\mathbf{Y}) = \mathbf{H}(\mathbf{Y}) - \mathbf{H}(\mathbf{Y}|\mathbf{X}),$$

and similarly define the conditional mutual information

$$\mathbf{I}(\mathbf{X} : \mathbf{Y}|\mathbf{Z}) = \mathbf{H}(\mathbf{X}|\mathbf{Z}) - \mathbf{H}(\mathbf{X}|\mathbf{Y}, \mathbf{Z}) = \mathbf{H}(\mathbf{Y}|\mathbf{Z}) - \mathbf{H}(\mathbf{Y}|\mathbf{X}, \mathbf{Z}).$$

For each m obeying (4.2), let $\mathbf{Y}_{<m}$ be the random variable $\mathbf{Y}_{<m} := (\mathbf{Y}_{m'})_{m'<m}$. We can control the expectation $\mathbf{E}F_m(\mathbf{X}_m, \mathbf{Y}_m)$ by the conditional mutual information $\mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m})$ as follows:

Proposition 4.2. *Suppose m obeys* (4.2) *and is such that*

$$\mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m}) \leq \varepsilon^3 \frac{2^m}{m}. \tag{4.4}$$

Then one has

$$\mathbf{E}F_m(\mathbf{X}_m, \mathbf{Y}_m) \ll \varepsilon.$$

Proof. We argue as in [22], which are in turn a modification of the arguments in [20]. Let $\mathbf{U}_m$ be drawn uniformly at random from $\prod_{2^m<p\leq 2^{m+1}} \mathbb{Z}/p\mathbb{Z}$. We first show that for any sign pattern $X_m \in \{-1, +1\}^{C2^m}$, one has

$$\mathbf{P}(|F_m(X_m, \mathbf{U}_m)| \geq \varepsilon) \ll \exp(-c\varepsilon^2 2^m/m) \tag{4.5}$$

for an absolute constant $c > 0$. If we write $\mathbf{U}_m = (\mathbf{n}_p)_{2^m<p\leq 2^{m+1}}$, then the $\mathbf{n}_p$ are jointly independent in p and uniformly distributed on $\mathbb{Z}/p\mathbb{Z}$. If $X_m = (b_{i,r})_{1\leq i\leq k; 1\leq r\leq (2aB+1)a_i 2^m}$, then one can write

$$F_m(X_m, \mathbf{U}_m) = \mathbb{E}_{2^m<p\leq 2^{m+1}} \mathbf{W}_p$$

where $\mathbf{W}_p$ is the random variable

$$\mathbf{W}_p := \mathbb{E}_{j\leq 2^m} c_p b_{1,a_1 j+apb_1} \cdots b_{k,a_k j+apb_k} (p 1_{p|\mathbf{n}_p+j} - 1).$$

Observe that the $\mathbf{W}_p$ are jointly independent, bounded in magnitude by $O(1)$, and have mean zero. The claim (4.5) now follows from Hoeffding's inequality [14].

Applying the Pinsker-type inequality from [22, Lemma 3.4] (see also [20, Lemma 3.3]), we conclude that

$$\mathbf{P}(|F_m(X_m, \mathbf{Y})| \geq \varepsilon) \ll \frac{m}{\varepsilon^2 2^m}(\mathbf{H}(\mathbf{U}_m) - \mathbf{H}(\mathbf{Y}) + 1)$$

for any random variable $\mathbf{Y}$ taking values in $\prod_{2^m<p\leq 2^{m+1}}(\mathbb{Z}/p\mathbb{Z})$; in particular, applying this to the probability measure $\mathbf{P}'(E) := \mathbf{P}(E|\mathbf{X}_m = X_m, \mathbf{Y}_{<m} = Y_{<m})$, we have

$$\begin{aligned}\mathbf{P}(|F_m(\mathbf{X}_m, \mathbf{Y}_m)| \geq \varepsilon|\mathbf{X}_m = X_m, \mathbf{Y}_{<m} = Y_{<m}) \\ \ll \frac{m}{\varepsilon^2 2^m}(\mathbf{H}(\mathbf{U}_m) - \mathbf{H}(\mathbf{Y}_m|\mathbf{X}_m = X_m, \mathbf{Y}_{<m} = Y_{<m}) + 1).\end{aligned}$$

Averaging over $X_m, Y_{<m}$, we conclude that

$$\mathbf{P}(|F_m(\mathbf{X}_m, \mathbf{Y}_m)| \geq \varepsilon) \ll \frac{m}{\varepsilon^2 2^m}(\mathbf{H}(\mathbf{U}_m) - \mathbf{H}(\mathbf{Y}_m|\mathbf{X}_m, \mathbf{Y}_{<m}) + 1),$$

and hence (since F_m is bounded by $O(1)$, and m is large compared to $1/\varepsilon$)

$$\mathbf{E}|F_m(\mathbf{X}_m, \mathbf{Y}_m)| \ll \frac{m}{\varepsilon^2 2^m}(\mathbf{H}(\mathbf{U}_m) - \mathbf{H}(\mathbf{Y}_m|\mathbf{X}_m, \mathbf{Y}_{<m})) + \varepsilon.$$

We can write

$$\mathbf{H}(\mathbf{Y}_m|\mathbf{X}_m, \mathbf{Y}_{<m}) = \mathbf{H}(\mathbf{Y}_m|\mathbf{Y}_{<m}) - \mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m})$$

and hence by (4.4) we have

$$\mathbf{E}|F_m(\mathbf{X}_m, \mathbf{Y}_m)| \ll \frac{m}{\varepsilon^2 2^m}(\mathbf{H}(\mathbf{U}_m) - \mathbf{H}(\mathbf{Y}_m|\mathbf{Y}_{<m})) + \varepsilon. \tag{4.6}$$

Uniformly for $1 \le b \le q \le x^\varepsilon$, we have the simple estimate

$$\sum_{\substack{x^\varepsilon \le n \le x \\ n \equiv b \pmod q}} \frac{1}{n} = \left(\frac{1}{q} + O\left(\frac{q}{x^\varepsilon}\right)\right) \sum_{x^\varepsilon \le n \le x} \frac{1}{n},$$

so from the Chinese remainder theorem (and the prime number theorem), we see that the random variable $\mathbf{Y}_m$, after conditioning to any event of the form $\mathbf{Y}_{<m} = Y_{<m}$, is almost uniformly distributed in the sense that

$$\mathbf{P}(\mathbf{Y}_m = Y_m|\mathbf{Y}_{<m} = Y_{<m}) = \frac{1}{\prod_{2^m < p \le 2^{m+1}} p} + O\left(\frac{\exp(O(2^m))}{x^\varepsilon}\right). \tag{4.7}$$

We have for any distinct $x, y \in (0, 1]$ the elementary inequality[5]

$$\left|x \log \frac{1}{x} - y \log \frac{1}{y}\right| \le C|x - y| \log \frac{2}{|x - y|} \le 2C|x - y|^{\frac{1}{2}}$$

for some constant $C > 0$, so if $\mathbf{X}$ and $\mathbf{X}'$ are any random variables having the same finite range $\mathcal{X}$, then we can compare their entropies by

$$|\mathbf{H}(\mathbf{X}) - \mathbf{H}(\mathbf{X}')| \le 2C \cdot \max_{x \in \mathcal{X}} |\mathbf{P}(\mathbf{X} = x) - \mathbf{P}(\mathbf{X}' = x)|^{\frac{1}{2}} \cdot |\mathcal{X}|. \tag{4.8}$$

From this and (4.7) we compute that

$$\mathbf{H}(\mathbf{U}_m) - \mathbf{H}(\mathbf{Y}_m|\mathbf{Y}_{<m}) \ll \frac{\exp(O(2^m))}{x^{\varepsilon/2}}.$$

Inserting this into (4.6) and using (4.1), (4.2) we conclude that

$$\mathbf{E}|F_m(\mathbf{X}_m, \mathbf{Y}_m)| \ll \varepsilon$$

as required. □

Theorem 3.1 now follows from the preceding proposition and the following estimate.

Proposition 4.3 (Entropy decrement argument)**.** *One has*

$$\sum_{\exp(a\varepsilon^{-3}) \le m \le \frac{1}{100} \log\log x} \frac{1}{2^m} \mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m}) \ll a.$$

[5]This inequality follows from the mean value theorem applied to $x \mapsto x \log \frac{1}{x}$.

Proof. For any m obeying (4.2), consider the quantity

$$\mathbf{H}(\mathbf{X}_{m+1}|\mathbf{Y}_{<m+1}).$$

We can view $\mathbf{X}_{m+1}$ as a pair $(\mathbf{X}_m, \mathbf{X}'_m)$, where

$$\mathbf{X}'_m := (\lambda(a_i\mathbf{n}' + r))_{1\leq i\leq k; 1\leq r\leq (2aB+1)a_i 2^m}$$

and $\mathbf{n}' := \mathbf{n} + (2aB+1)2^m$. By the Shannon entropy inequalities, we thus have

$$\mathbf{H}(\mathbf{X}_{m+1}|\mathbf{Y}_{<m+1}) \leq \mathbf{H}(\mathbf{X}_m|\mathbf{Y}_{<m+1}) + \mathbf{H}(\mathbf{X}'_m|\mathbf{Y}_{<m+1}).$$

If we write

$$\mathbf{Y}'_{<m+1} := (\mathbf{n}' \bmod p)_{p\leq 2^{m+1}}$$

then $\mathbf{Y}_{<m+1}$ and $\mathbf{Y}'_{<m+1}$ define the same σ-algebra (each random variable is a deterministic function of the other), and so we have

$$\mathbf{H}(\mathbf{X}_{m+1}|\mathbf{Y}_{<m+1}) \leq \mathbf{H}(\mathbf{X}_m|\mathbf{Y}_{<m+1}) + \mathbf{H}(\mathbf{X}'_m|\mathbf{Y}'_{<m+1}).$$

The total variation distance between $\mathbf{n}$ and $\mathbf{n}'$ can be computed to be $O(\exp(O(2^m))/x^\varepsilon)$. Since $\mathbf{Y}_{<m+1}$ takes on $O(\exp(O(2^m)))$ values, we see from (4.8) that

$$\mathbf{H}(\mathbf{Y}'_{<m+1}) = \mathbf{H}(\mathbf{Y}_{<m+1}) + O(\exp(O(2^m))/x^{\varepsilon/2}).$$

Similarly, since the random variables $(\mathbf{X}_m, \mathbf{Y}_{<m+1})$ and $(\mathbf{X}'_m, \mathbf{Y}'_{<m+1})$ take on $O(\exp(O(2^m)))$ values and are deterministic functions of $\mathbf{n}$ and $\mathbf{n}'$, respectively, by (4.8) we again have

$$\mathbf{H}(\mathbf{X}'_m, \mathbf{Y}'_{<m+1}) = \mathbf{H}(\mathbf{X}_m, \mathbf{Y}_{<m+1}) + O(\exp(O(2^m))/x^{\varepsilon/2}),$$

and hence on subtracting

$$\mathbf{H}(\mathbf{X}'_m|\mathbf{Y}'_{<m+1}) = \mathbf{H}(\mathbf{X}_m|\mathbf{Y}_{<m+1}) + O(\exp(O(2^m))/x^{\varepsilon/2}).$$

Thus we have

$$\mathbf{H}(\mathbf{X}_{m+1}|\mathbf{Y}_{<m+1}) \leq 2\mathbf{H}(\mathbf{X}_m|\mathbf{Y}_{<m+1}) + O(\exp(O(2^m))/x^{\varepsilon/2}).$$

But we can write $\mathbf{Y}_{<m+1}$ as a pair $(\mathbf{Y}_{<m}, \mathbf{Y}_m)$, to conclude that

$$\mathbf{H}(\mathbf{X}_m|\mathbf{Y}_{<m+1}) = \mathbf{H}(\mathbf{X}_m|\mathbf{Y}_{<m}) - \mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m}).$$

Inserting this identity and rearranging, we conclude that

$$\begin{aligned}&\frac{1}{2^m}\mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m})\\ &\quad\leq \frac{1}{2^m}\mathbf{H}(\mathbf{X}_m|\mathbf{Y}_{<m}) - \frac{1}{2^{m+1}}\mathbf{H}(\mathbf{X}_{m+1}|\mathbf{Y}_{<m+1}) + O(\exp(O(2^m))/x^{\varepsilon/2})\end{aligned}$$

and thus on summing the telescoping series

$$\sum_{m\leq \frac{1}{100}\log\log x} \frac{1}{2^m}\mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m}) \ll \mathbf{H}(\mathbf{X}_1) + 1$$

(say). Since $\mathbf{X}_1$ takes at most $\exp(O(a))$ values, we have $\mathbf{H}(\mathbf{X}_1) = O(a)$, and the claim follows. $\square$

5. Using the Gowers norms

We now prove Theorem 3.2. As stated previously, we will rely heavily on the theory of the Gowers norms, which we now recall.

Definition 5.1 (Gowers norms)**.** *Given integers $k \geq 1$ and $N \geq 1$ and a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, we define the Gowers norms $U^k(\mathbb{Z}/N\mathbb{Z})$ by*

$$\|f\|_{U^k(\mathbb{Z}/N\mathbb{Z})} := \left(\mathbb{E}_{n\in\mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{h_1,\ldots,h_k\in\mathbb{Z}/N\mathbb{Z}} \prod_{\omega\in\{0,1\}^k} \mathcal{C}^{|\omega|} f(n+\omega\cdot\mathbf{h})\right)^{2^{-k}},$$

where $\mathcal{C}$ is the complex conjugation operator, $|\omega|$ is the number of ones in $\omega \in \{0,1\}^k$, $\mathbf{h} = (h_1,\ldots,h_k)$, and $\cdot$ denotes the inner product of two vectors. One easily sees that $\|f\|_{U^k(\mathbb{Z}/N\mathbb{Z})}$ is a well-defined nonnegative quantity. We can then define the Gowers $U^k[N]$-norm of a function $f : \{1,\ldots,N\} \to \mathbb{C}$ defined on a finite interval by

$$\|f\|_{U^k[N]} := \frac{\left\|f\cdot 1_{[1,N]}\right\|_{U^k(\mathbb{Z}_{N'})}}{\left\|1_{[1,N]}\right\|_{U^k(\mathbb{Z}_{N'})}}$$

where $N' = 3N$, say (one easily sees that the definition is independent of the choice of $N' > 2N$) and $f \cdot 1_{[1,N]}$ is to be interpreted as a function of period N', and hence as a function on $\mathbb{Z}'_N$.

For the basic properties of Gowers norms, see [23, Chapter 11]. The main general fact we will need about these norms is the following.

Lemma 5.2 (A generalised von Neumann theorem)**.** *For $k \in \mathbb{N}$, let θ, $\phi_1,\ldots,\phi_k : \mathbb{Z} \to \mathbb{C}$ be functions with $|\phi_j| \leq 1$. Also let $a_j, b_j, r_j \in \mathbb{Z}$ for $1 \leq j \leq k$, and $W \in \mathbb{N}$ with $W \leq N^{0.1}$. Then*

$$\left|\mathbb{E}_{d\leq\frac{N}{W}}\mathbb{E}_{n\leq N}\theta(d)\phi_1(a_1n+Wb_1d+r_1)\cdots\phi_k(a_kn+Wb_kd+r_k)\right| \leq C\|\theta\|_{U^k[\frac{N}{W}]} + o_{N\to\infty}(1)$$

for some constant $C > 0$ depending only on k and the numbers $a_1,\ldots,a_k$, $b_1,\ldots,b_k$, but independent of W and $r_1,\ldots,r_k$.

Without the W-aspect, this is standard; see for instance [7, Lemma 2]. However, the uniformity of the bounds in W (and $r_1,\ldots,r_k$) will be crucial in our arguments.

Proof. We shall adapt the proof of [21, Proposition 3.3]. By splitting the variable n into residue classes $\pmod W$ and setting $N' := \frac{N}{W}$ it suffices to show that

$$\left|\mathbb{E}_{d\le N'}\mathbb{E}_{n\le N'}\theta(d)\phi_1(W(a_1n+b_1d)+r_1')\cdots\phi_k(W(a_kn+b_kd)+r_k')\right| \\ \le C\|\theta\|_{U^k[N']}+o_{N'\to\infty}(1)$$

for all integers $r_1',\dots,r_k'$. To simplify notation, we will call N' just N. By considering the functions $\widetilde{\phi}_j(n) := \phi_j(Wn+r_j')$, we see that it suffices to prove for all functions $|\phi_j|\le 1$ that

$$\text{(5.1)}\quad |\mathbb{E}_{d\le N}\mathbb{E}_{n\le N}\theta(d)\phi_1(a_1n+b_1d)\cdots\phi_k(a_kn+b_kd)| \\ \le C\|\theta\|_{U^k[N]}+o_{N\to\infty}(1).$$

Since the statement of (5.1) involves the values of the functions θ and ϕ_i only on $(-HN,HN)$, where $H=\max_{i\le k}(|a_i|+|b_i|)+1$, we may assume that the functions θ and ϕ_i are $2HN$-periodic, and hence they can be interpreted as functions on $\mathbb{Z}_{2HN}$. We are then reduced to showing that

$$\left|\mathbb{E}_{d\in\mathbb{Z}_{2HN}}\mathbb{E}_{n\in\mathbb{Z}_{2HN}}\theta(d)1_{[0,N]}(d)\prod_{i=1}^k\phi_i(a_in+db_i)1_{[0,N]}(n)\right| \\ \le C'\|\theta\|_{U^k[N]}+o_{N\to\infty}(1)$$

for some constant C', since one can then set $C:=(2H)^2C'$. By approximating $1_{[0,N]}(n)$ with a Lipschitz function, and then further with a finite Fourier series as in [10, Appendix C], and redefining the functions ϕ_j, we may eliminate the factor $1_{[0,N]}(n)$. Then, making a change of variables $d=d_1+\dots+d_k$, $n=n'-d_1b_1-\dots-d_kb_k$, we are left with showing that

$$\text{(5.2)}\quad \left|\mathbb{E}_{d_1,\dots,d_k\in\mathbb{Z}_{2HN}}\theta'(d_1+\dots+d_k)\prod_{i=1}^k\phi_i\left(a_in'+\sum_{\ell=1}^k d_\ell(b_i-b_\ell)\right)\right| \\ \le C''\|\theta\|_{U^k[N]}+o_{N\to\infty}(1),$$

for all $n'\in\mathbb{Z}_{2HN}$, where $\theta'(d):=\theta(d)1_{[0,N]}(d)$. By the Gowers–Cauchy–Schwarz inequality (see e.g., [10, (B.7)]), we have

$$\left|\mathbb{E}_{d_1,\dots,d_k\in\mathbb{Z}_{2HN}}\theta'(d_1+\dots+d_k)\prod_{i=1}^k\phi_i'(L_i(d_1,\dots,d_k))\right|\le\|\theta'\|_{U^k(\mathbb{Z}_{2HN})}$$

for any functions θ' and ϕ_i' bounded by 1 in modulus and any linear forms $L_i:\mathbb{Z}_{2HN}^k\to\mathbb{Z}_{2HN}$, with L_i independent of the ith coordinate. Applying this to the left-hand side of (5.2), where each term involving ϕ_i is independent of the variable d_i, we see that

$$\left|\mathbb{E}_{d_1,\dots,d_k\in\mathbb{Z}_{2HN}}\theta'(d_1+\cdots+d_k)\prod_{i=1}^{k}\phi_i\left(a_in'+\sum_{\ell=1}^{k}d_\ell(b_i-b_\ell)\right)\right| \leq \|\theta'\|_{U^k(\mathbb{Z}_{2HN})}.$$

Then, by noting that

$$\|\theta(n)1_{[0,N]}(n)\|_{U^k(\mathbb{Z}_{2HN})} = \|\theta\|_{U^k[N]}\cdot\|1_{[0,N]}\|_{U^k(\mathbb{Z}_{2HN})} \leq \|\theta\|_{U^k[N]},$$

the lemma follows. □

Next, we need control on the Gowers norms for the primes.

Lemma 5.3 (Gowers uniformity of the primes)**.** *Let $k\in\mathbb{N}$, and let $w\in\mathbb{N}$ be a large parameter. Further, let $W=\prod_{p\leq w}p$, and let $b\in[1,W]$ be coprime to W. Then for any N large enough in terms of w, the W-tricked von Mangoldt function*

$$\Lambda_{b,W}(n) := \frac{\varphi(W)}{W}\Lambda(Wn+b) \tag{5.3}$$

enjoys the Gowers uniformity bound

$$\|\Lambda_{b,W}-1\|_{U^{k+1}[N]} = o_{w\to\infty}(1).$$

Proof. This was proven in [10], subject to conjectures that were later verified in [11], [12]. □

We now prove Theorem 3.2. Let $k, a_1,\dots,a_k, b_1,\dots,b_k,\varepsilon, w, H_-, H_+, x, W, a, m$ be as in that theorem. Because $\Lambda(p)=\log(2^m)+O(1)$ when p is a prime with $2^m<p\leq 2^{m+1}$, and Λ is non-zero for only $O(2^{2m/3})$ (say) other integers in the interval $(2^m,2^{m+1}]$, we have

$$\mathbb{E}^{\log}_{2^m<p\leq 2^{m+1}}f_x(ap) = \mathbb{E}^{\log}_{2^m<d\leq 2^{m+1}}f_x(ad)\Lambda(d)+O(\varepsilon),$$

since m is assumed to be sufficiently large depending on ε. The contribution to the right-hand side of those d that share a common factor with W is negligible (as $\Lambda(d)$ will then vanish unless n is a power of a prime less than or equal to w), thus

$$\mathbb{E}^{\log}_{2^m<p\leq 2^{m+1}}f_x(ap) = \frac{W}{\phi(W)}\mathbb{E}^{\log}_{2^m<d\leq 2^{m+1}:(d,W)=1}f_x(ad)\Lambda(d)+O(\varepsilon).$$

It therefore suffices to show that

$$\mathbb{E}^{\log}_{2^m<d\leq 2^{m+1}:(d,W)=1}f_x(ad)\left(\frac{W}{\phi(W)}\Lambda(d)-1\right) \ll \varepsilon.$$

Partitioning into residue classes modulo W and using (5.3), it suffices to show that

$$\mathbb{E}^{\log}_{2^m/W<d\leq 2^{m+1}/W} f_x(a(Wd+b))(\Lambda_{b,W}(d)-1) \ll \varepsilon$$

whenever $1 \leq b \leq W$ is coprime to W.

Fix b. By summation by parts, it will suffice to show that

$$\mathbb{E}_{d\leq H} f_x(a(Wd+b))(\Lambda_{b,W}(d)-1) \ll \varepsilon$$

whenever $2^m/W \leq H \leq 2^{m+1}/W$. From (3.1), and replacing the average $n \leq x$ with the average $x^\varepsilon < n \leq x$, we have

$$\begin{aligned} &f_x(a(Wd+b)) \\ &\quad = \mathbb{E}^{\log}_{x^\varepsilon<n\leq x} \lambda(a_1 n + Wab_1 d + abb_1)\dots\lambda(a_k n + Wab_k d + abb_k) + O(\varepsilon), \end{aligned}$$

so it suffices to show that

$$\begin{aligned} \mathbb{E}_{d\leq H}\mathbb{E}^{\log}_{x^\varepsilon<n\leq x}(\Lambda_{b,W}(d)-1)\lambda(a_1 n + Wab_1 d + abb_1) \\ \dots\lambda(a_k n + Wab_k d + abb_k) \ll \varepsilon. \end{aligned} \tag{5.4}$$

The quantity x (or x^ε) is large compared with aHW. Thus we can shift n by any quantity $1 \leq n' \leq aHW$ without affecting the above average by more than $O(\varepsilon)$. Performing this shift and then averaging in n', the left-hand side of (5.4) may be written as

$$\begin{aligned} \mathbb{E}^{\log}_{x^\varepsilon<n\leq x}\mathbb{E}_{d\leq H}\mathbb{E}_{n'\leq aHW}(\Lambda_{b,W}(d)-1)[\lambda(a_1 n' + Wab_1 d + a_1 n + abb_1) \\ \dots\lambda(a_k n' + Wab_k d + a_1 n + abb_k)] + O(\varepsilon). \end{aligned}$$

Applying Lemma 5.2 with N replaced by aHW, W replaced by aW, n replaced by n', and the r_j replaced by $a_j n + abb_j$ for $1, \dots, k$, we can bound this as

$$O(\|\Lambda_{b,W}-1\|_{U^k[H]}) + o_{H\to\infty}(1) + O(\varepsilon),$$

but by Lemma 5.3 this is $O(\varepsilon)$ as required.

References

[1] J. Bourgain, P. Sarnak & T. Ziegler, "Disjointness of Moebius from horocycle flows", in *From Fourier analysis and number theory to Radon transforms and geometry*, Developments in Mathematics, vol. 28, Springer, 2013, p. 67-83.

[2] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*, Mathematics and its Applications, vol. 4, Gordon and Breach Science Publishers, 1965.

[3] P. D. T. A. Elliott, "On the correlation of multiplicative functions", *Notas Soc. Mat. Chile* **11** (1992), no. 1, p. 1-11.

[4] N. Frantzikinakis, "Ergodicity of the Liouville system implies the Chowla conjecture", *Discrete Anal.* **2017** (2017), Art. ID 19, 41 pages.

[5] ———, "An averaged Chowla and Elliott conjecture along independent polynomials", *Int. Math. Res. Not.* **2018** (2018), no. 12, p. 3721-3743.

[6] N. Frantzikinakis & B. Host, "The logarithmic Sarnak conjecture for ergodic weights", *Ann. Math.* **187** (2018), no. 3, p. 869-931.

[7] N. Frantzikinakis, B. Host & B. Kra, "Multiple recurrence and convergence for sequences related to the prime numbers", *J. Reine Angew. Math.* **611** (2007), p. 131-144.
[8] J. Friedlander & H. Iwaniec, *Opera de cribro*, Colloquium Publications, vol. 57, American Mathematical Society, 2010.
[9] B. Green & T. Tao, "An inverse theorem for the Gowers U^3-norm, with applications", *Proc. Edinb. Math. Soc.* **51** (2008), no. 1, p. 73-153.
[10] ———, "Linear equations in primes", *Ann. Math.* **171** (2010), p. 1753-1850.
[11] ———, "The Möbius function is strongly orthogonal to nilsequences", *Ann. Math.* **175** (2012), no. 2, p. 541-566.
[12] B. Green, T. Tao & T. Ziegler, "An inverse theorem for the Gowers $U^{s+1}[N]$-norm", *Ann. Math.* **176** (2012), no. 2, p. 1231-1372.
[13] G. H. Hardy & J. E. Littlewood, "Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes", *Acta Math.* **44** (1923), no. 1, p. 1-70.
[14] W. Hoeffding, "Probability inequalities for sums of bounded random variables", *J. Am. Stat. Assoc.* **58** (1963), p. 13-30.
[15] I. Kátai, "A remark on a theorem of H. Daboussi", *Acta Math. Hung.* **47** (1986), p. 223-225.
[16] A. N. Le, "Nilsequences and multiple correlations along subsequences", `https://arxiv.org/abs/1708.01361`, 2017.
[17] A. Leibman, "Nilsequences, null-sequences, and multiple correlation sequences", *Ergodic Theory Dyn. Syst.* **35** (2015), no. 1, p. 176-191, corrected version available at `people.math.osu.edu/leibman.1/preprints/msqx.pdf`.
[18] K. Matomäki & M. Radziwiłł, "Multiplicative functions in short intervals", *Ann. Math.* **183** (2016), no. 3, p. 1015-1056.
[19] K. Matomäki, M. Radziwiłł & T. Tao, "An averaged form of Chowla's conjecture", *Algebra & Number Theory* **9** (2015), p. 2167-2196.
[20] T. Tao, "The logarithmically averaged Chowla and Elliott conjectures for two-point correlations", *Forum Math. Pi* **4** (2016), Art. ID e8, 36 pages.
[21] ———, "Equivalence of the logarithmically averaged Chowla and Sarnak conjectures", in *Number theory—Diophantine problems, uniform distribution and applications*, Springer, 2017, p. 391-421.
[22] T. Tao & J. Teräväinen, "The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures", to appear in *Duke Math. J.*, `https://arxiv.org/abs/1708.02610`, 2017.
[23] T. Tao & V. H. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, 2006.

Terence Tao
Department of Mathematics, UCLA
405 Hilgard Ave
Los Angeles CA 90095, USA
E-mail: `tao@math.ucla.edu`

Joni Teräväinen
Department of Mathematics and Statistics
University of Turku
20014 Turku, Finland
E-mail: `joni.p.teravainen@utu.fi`