Turun yliopisto
University of Turku

TOPICS IN MULTIPLICATIVE NUMBER THEORY

Joni Teräväinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
Sarja – ser. AI osa – tom. 589 | Astronomica – Chemica – Physica – Mathematica | Turku 2018

Turun yliopisto
University of Turku

TOPICS IN MULTIPLICATIVE NUMBER THEORY

Joni Teräväinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
Sarja - ser. A I osa - tom. 589 | Astronomica - Chemica - Physica - Mathematica | Turku 2018

University of Turku

Faculty of Science and Engineering
Department of Mathematics and Statistics

Supervised by

Docent Kaisa Matomäki
Department of Mathematics and Statistics
University of Turku
Turku, Finland

Reviewed by

Assistant Professor Adam Harper
Mathematics Institute
University of Warwick
Coventry, England, UK

Associate Professor Dimitris Koukoulopoulos
Department of Mathematics and Statistics
Université de Montréal
Montreal, Canada

Opponent

Professor Ben Green
Mathematical Institute
University of Oxford
Oxford, England, UK

ISBN 978-951-29-7339-2 (PRINT)
ISBN 978-951-29-7340-8 (PDF)
ISSN 0082-7002 (PRINT)
ISSN 2343-3175 (PDF)
Painosalama Oy – Turku, Finland 2018

Abstract

This thesis is comprised of four articles in multiplicative number theory, a subfield of analytic number theory that studies questions related to prime numbers and multiplicative functions. A central principle in multiplicative number theory is that multiplicative structures, such as the primes or the values of a multiplicative function, should not correlate with additive structures of various types. The results in this thesis can be interpreted as instances of this principle.

In the first article, we consider the problem of finding almost primes in almost all short intervals, which is a natural approximation to the problem of finding primes in short intervals. We show that almost all intervals of nearly optimal length contain a product of exactly three primes. For products of exactly two primes, we improve a result of Harman. The proofs are based on careful analysis of Dirichlet polynomials related to almost primes.

The second article is about the Goldbach problem for a sparse subset of the primes. Vinogradov famously showed that any large odd number is the sum of three primes, so it is natural to study the same problem with the summands coming from a subset of the primes. Improving a result of Matomäki, we show that a special set of primes, consisting of primes representable as one plus the sum of two squares, satisfies the ternary Goldbach problem. We also establish a number of other additive results for this same set of primes. The proofs use sieve methods and transference principles for additive equations in primes.

We also study the Möbius function and its autocorrelations. A famous conjecture of Chowla asserts that products of shifts of the Möbius function should have mean zero. In the third article, together with T. Tao we settle a logarithmic version of this conjecture in all the cases involving an odd number of shifts. This complements Tao's earlier result that the two-point Chowla conjecture holds with logarithmic weights.

Lastly, in the fourth article, we study binary correlations of multiplicative functions with logarithmic weights. We prove an asymptotic formula for these correlations for a wide class of multiplicative functions, extending an earlier result of Tao. We then derive a number of applications regarding the largest prime factors of consecutive integers, including a logarithmic version of a conjecture of Erdős and Turán. Moreover, we prove a new estimate for character sums over reducible quadratic polynomials.

Tiivistelmä

Tämä väitöskirja koostuu neljästä artikkelista multiplikatiivisessa lukuteoriassa, joka on alkulukuja ja multiplikatiivisia funktioita tutkiva analyyttisen lukuteorian haara. Keskeinen periaate multiplikatiivisessa lukuteoriassa on, että multiplikatiivisten objektien (kuten alkulukujen tai multiplikatiivisten funktioiden arvojen) ei pitäisi korreloida additiivisten objektien kanssa. Tämän väitöskirjan tulokset voidaankin tulkita kyseisen periaatteen ilmentyminä.

Ensimmäisessä artikkelissa tarkastelemme melkein alkulukujen löytämistä melkein kaikilta lyhyiltä väleiltä; tämä on luonnollinen approksimaatio alkulukujen löytämiselle lyhyiltä väleiltä. Osoitamme, että melkein kaikki välit, joiden pituus on lähes optimaalisen lyhyt, sisältävät tasan kolmen alkuluvun tulon. Tasan kahden alkuluvun tulojen tapauksessa parannamme Harmanin tulosta. Todistukset perustuvat melkein alkulukuihin liitettyjen Dirichlet'n polynomien tarkkaan analysointiin.

Toinen artikkeli koskee Goldbach-ongelmaa eräälle harvalle osajoukolle alkulukuja. Vinogradov osoitti kuuluisassa työssään, että jokainen riittävän suuri pariton luku on kolmen alkuluvun summa, joten on luonnollista tarkastella vastaavaa ongelmaa alkulukujen osajoukoille. Parantaen Matomäen tulosta osoitamme, että vastaus ternääriseen Goldbach-oneglmaan on positiivinen niiden alkulukujen joukolle, jotka voidaan esittää ykkösen ja kahden neliöluvun summana. Osoitamme myös useita muita additiivisia tuloksia samalle alkulukujen osajoukolle. Todistukset käyttävät seulamenetelmiä sekä ns. traansferenssiperiaatteita additiivisille yhtälöille alkulukujen joukossa.

Tutkimme myös Möbiuksen funktiota ja sen autokorrelaatioita. Chowlan kuuluisa konjektuuri väittää, että Möbiuksen funktioiden translaatioiden tuloilla pitäisi olla keskiarvo nolla. Kolmannessa artikkelissa yhdessä T. Taon kanssa ratkaisemme logaritmisen version tästä konjektuurista kaikissa tapauksissa, joissa translaatioiden määrä on pariton. Tämä täydentää Taon aikaisempaa tulosta, jonka mukaan kahden pisteen Chowlan konjektuuri pätee logaritmisilla painoilla.

Lopuksi neljännessä artikkelissa tutkimme multiplikatiivisten funktioiden binäärisiä korrelaatioita logaritmisilla painoilla. Todistamme asymptoottisen kaavan näille korrelaatioille, joka pätee laajalle luokalle multiplikatiivisia funktioita ja parantaa Taon aikaisempaa tulosta. Johdamme sitten useita sovelluksia koskien peräkkäisten lukujen suurimpia alkutekijöitä – mukaan lukien logaritmisen version eräästä Erdősin ja Turánin konjektuurista. Lisäksi todistamme uuden arvion karakterisummille yli jaollisen toisen asteen polynomin arvojen.

Acknowledgments

I am sincerely grateful to my advisor, Prof. Kaisa Matomäki, for all the excellent advice and guidance she has given me throughout my PhD studies. Our regular meetings, the funding she has offered for my studies and travels, and her constant valuable feedback on my works have greatly progressed my career. It has been an honor learning the latest techniques in analytic number theory from one of the masters in the field.

I am grateful to the reviewers of this thesis, Prof. Adam Harper and Prof. Dimitris Koukoulopoulos, for spending a lot of their valuable time to carefully read my dissertation and for giving detailed comments on it. Equally, I am grateful to Prof. Ben Green for agreeing to act as the opponent at my doctoral defense.

I thank Prof. Terry Tao for all the fruitful collaboration projects we have had so far, for answering so many of my questions on additive combinatorics, and for hosting my visit to UCLA in April 2018. Working with him has been a great pleasure and has made me more mature as a mathematician.

I would like to thank the sources of funding of my research, namely the Vilho, Yrjö and Kalle Väisälä Fund of the Finnish Academy of Science and Letters, the UTUGS Graduate School of the University of Turku, and the Academy of Finland. The support from these grants has made my work and participation in various conferences possible.

I express my thanks to my colleagues at the universities of Oxford, KTH and Nancy for inviting me for research visits and for making my stays very enjoyable and scientifically productive. I extend my thanks to my number theory colleagues at Åbo Akademi for organizing a stimulating joint seminar. I also thank Alexander Mangerel for careful reading of my thesis.

All my colleagues at the Department of Mathematics and Statistics at the University of Turku deserve thanks for making the work environment pleasant, productive, and relaxed. Specifically, I thank Matti Vuorinen for all the help and friendship he has offered me, and for the lovely dinners I have had at Matti and Sinikka's home. In addition, I thank my fellow number theory students, Jori Merikoski and Juho Salmensuu, for our many interesting number-theoretic lunch discussions.

Special thanks go to all my friends for making my time as a graduate student so fun and enjoyable. Especially, I thank Jesse Jääsaari for the many fascinating discussions we have had about mathematics and other topics, and for great company during our many conference trips together. Likewise, I thank Otte Heinävaara for the countless hours we have spent solving and devising mathematical problems, and for the long, adventurous cycling trips we have taken through Finland.

I warmly thank my family for always supporting me and believing in me. Especially, I thank my parents, grandparents and siblings, whose help I can always count on and who have always encouraged me to pursue my dreams. Lastly, and most

importantly, I thank my fiancée Parisa for unconditional love and support, and for bringing joy to the long days I spent working on this thesis.

Turku, June 2018
Joni Teräväinen

List of original publications

This thesis contains the following four publications.

[I] J. Teräväinen: Almost primes in almost all short intervals. *Math. Proc. Cambridge Philos. Soc.*, 161(2):247–281, 2016. DOI: 10.1017/S0305004116000232

[II] J. Teräväinen: The Goldbach problem for primes that are sums of two squares plus one. *Mathematika*, 64(1):20–70, 2018. DOI: 10.1112/S0025579317000341

[III] T. Tao and J. Teräväinen: Odd order cases of the logarithmically averaged Chowla conjecture. To appear in *J. Théor. Nombres Bordeaux*. arXiv: 1710.02112 [Math.NT]

[IV] J. Teräväinen: On binary correlations of multiplicative functions. *Forum Math. Sigma*, 6:e10, 41, 2018. DOI: 10.1017/fms.2018.10

Table of Contents

Summary

1. Notations and conventions

We collect here the notations used in the summary sections 2–6. The Articles [I]–[IV] have their own notation sections.

1.1. Sets

- $\mathbb{N}$ – the set of positive integers $\{1, 2, 3, \ldots\}$.
- $\mathbb{Z}$ – the set of all integers.
- $\mathbb{Z}_N$ – the ring of integers mod N.
- $\mathbb{P}$ – the set of prime numbers.
- Squarefree integers – integers $n \geq 1$ such that n is not divisible by p^2 for primes p.
- P_k – the set of integers with at most k prime factors, counting multiplicities.
- E_k – the set of integers with exactly k prime factors, counting multiplicities.
- $\mathbb{D}$ – the unit disk $\{z \in \mathbb{C} : \ |z| \leq 1\}$.
- $1_S(n)$ – the indicator function of a set S, equaling 1 if $n \in S$ and 0 otherwise.

1.2. Letters

- d, k, ℓ, m, n – positive integers.
- $p, p_1, p_2, \ldots$ – prime numbers.
- ε – an arbitrarily small positive constant.
- W – the product of primes in $[1, w]$ for some large w.

1.3. Arithmetic functions

- $\varphi(n)$ – the Euler function, giving the number of integers $1 \leq j \leq n$ coprime to n.
- $\Lambda(n)$ – the von Mangoldt function, which equals $\log p$ if $n = p^k$ for some prime p and some $k \geq 1$, and equals 0 if no such p exists.
- $\Omega(n)$ – number of prime factors of n, counted with multiplicities.
- $\lambda(n)$ – the Liouville function, given by $\lambda(n) := (-1)^{\Omega(n)}$.
- $\mu(n)$ – the Möbius function, given by $\mu(n) := \lambda(n) 1_{n \text{ squarefree}}$.
- $P^+(n)$ – the largest prime factor of n, with $P^+(1) := 1$.
- $\pi(x)$ – the number of prime numbers in $[1, x]$.
- Multiplicative function – a function $g : \mathbb{N} \to \mathbb{C}$ satisfying $g(mn) = g(m)g(n)$ whenever $m, n \in \mathbb{N}$ are coprime.
- $\lambda_d^{+,\mathrm{LIN}}$, $\lambda_d^{-,\mathrm{LIN}}$ – the upper and lower bound linear sieve weights. Given a level D and a sifting parameter z, they are equal to 1 for $d = 1$ and are equal to the Möbius function $\mu(d)$ for $d \geq 2$ belonging to the sets

  $$\mathcal{D}^{+,\mathrm{LIN}} := \{p_1 \cdots p_r \leq D : \ z > p_k > p_{k+1}, \ p_1 \cdots p_{2k-2} p_{2k-1}^3 \leq D \ \forall k \geq 1\},$$

  $$\mathcal{D}^{-,\mathrm{LIN}} := \{p_1 \cdots p_r \leq D : \ z > p_k > p_{k+1}, \ p_1 \cdots p_{2k-1} p_{2k}^3 \leq D \ \forall k \geq 1\}.$$

  For other values of d, they are equal to 0.
- $\lambda_d^{+,\mathrm{SEM}}$, $\lambda_d^{-,\mathrm{SEM}}$ – the upper and lower bound semilinear sieve weights. Given a level D and a sifting parameter z, they are equal to 1 for $d = 1$ and are equal to the Möbius function $\mu(d)$ for $d \geq 2$ belonging to the sets

$\mathcal{D}^{+,\mathrm{SEM}} := \{p_1 \cdots p_r \le D : \ z > p_k > p_{k+1}, \ p_1 \cdots p_{2k-2} p_{2k-1}^2 \le D \ \forall k \ge 1\}$,
$\mathcal{D}^{-,\mathrm{SEM}} := \{p_1 \cdots p_r \le D : \ z > p_k > p_{k+1}, \ p_1 \cdots p_{2k-1} p_{2k}^2 \le D \ \forall \ k \ge 1\}$.

For other values of d, they are equal to 0.

1.4. **Analysis**

- $f(x) = o(g(x))$ – we have $\lim_{x\to\infty} f(x)/g(x) = 0$.
- $f(x) \ll g(x)$ – we have, for some constant C, $|f(x)| \le C|g(x)|$ for all large enough x.
- $f(x) \asymp g(x)$ – we have $f(x) \ll g(x)$ and $g(x) \ll f(x)$.
- $o_{\varepsilon\to 0}(1)$ – an unspecified function $f(\varepsilon)$ tending to 0 as $\varepsilon \to 0$.
- Almost all – for a proposition $P(n)$, we say that it holds for almost all $n \in \mathbb{N}$ if $|\{n \le X : \ P(n) \text{ fails}\}| = o(X)$. Analogously, we say that $P(n)$ holds for almost all even n if $|\{n \le X : \ n \equiv 0 \pmod 2, \ P(n) \text{ fails}\}| = o(X)$.
- $e(\alpha)$ – the additive character $e^{2\pi i\alpha}$.
- $\|x\|$ – the distance from $x \in \mathbb{R}$ to the nearest integer.
- $\sum_{p\in I}$ – summation over primes in I, whenever I is an interval.
- $\widehat{f}(\xi)$ – the discrete Fourier transform of $f : \mathbb{Z}_N \to \mathbb{C}$, given by

$$\widehat{f}(\xi) := \frac{1}{N} \sum_{n\in\mathbb{Z}_N} f(n) e\left(-\frac{\xi n}{N}\right).$$

1.5. **Probability theory**

- $\mathbf{P}(A)$ – the logarithmic probability of $A \subset [1, x]$, given by

$$\mathbf{P}(A) := \frac{\sum_{\substack{n\le x \\ n\in A}} \frac{1}{n}}{\sum_{n\le x} \frac{1}{n}}.$$

- $\mathbf{H}(\mathbf{X})$ – the entropy of a random variable $\mathbf{X}$ having a finite range $\mathcal{X}$. This is defined by

$$\mathbf{H}(\mathbf{X}) := \sum_{x\in\mathcal{X}} \mathbf{P}(\mathbf{X} = x) \log \frac{1}{\mathbf{P}(\mathbf{X} = x)}.$$

- $\mathbf{H}(\mathbf{X}, \mathbf{Y})$ – the joint entropy of two random variables $\mathbf{X}$ and $\mathbf{Y}$ with finite ranges $\mathcal{X}$ and $\mathcal{Y}$, respectively. This equals the entropy of the random variable $(\mathbf{X}, \mathbf{Y})$ that takes values in $\mathcal{X} \times \mathcal{Y}$.
- $\mathbf{H}(\mathbf{X}|\mathbf{Y})$ – the conditional entropy of $\mathbf{X}$ given $\mathbf{Y}$;

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

- $\mathbf{I}(\mathbf{X}, \mathbf{Y})$ – the mutual information between two random variables $\mathbf{X}$ and $\mathbf{Y}$;

$$\mathbf{I}(\mathbf{X}, \mathbf{Y}) = \mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y}) - \mathbf{H}(\mathbf{X}, \mathbf{Y}).$$

1.6. **Miscellaneous**

- (m,n) – the greatest common divisor of m and n.
- $\mathbb{D}(f,g;x)$ – the pretentious distance between f and g, defined in formula (6.1).
- $\mathcal{P}(z)$ – the product of all the primes in $[1,z)$.
- $\|a\|_{U^k(\mathbb{Z}_N)}$ the U^k Gowers norm of $a:\mathbb{Z}_N\to\mathbb{C}$, defined recursively as

$$\|a\|_{U^1(\mathbb{Z}_N)}:=\left|\frac{1}{N}\sum_{n\in\mathbb{Z}_N}f(n)\right|,\quad \|a\|_{U^{k+1}(\mathbb{Z}_N)}:=\left(\frac{1}{N}\sum_{t\in\mathbb{Z}_N}\|a\cdot\overline{a_t}\|_{U^k(\mathbb{Z}_N)}^{2^k}\right)^{1/2^{k+1}},$$

  where $a_t(n):=a(n+t)$.
- $\|a\|_{U^k[N]}$ – the U^k Gowers norm of $a:[1,N]\to\mathbb{C}$, defined by

$$\|a\|_{U^k[N]}:=\frac{\|1_{[1,N]}\cdot a\|_{U^k(\mathbb{Z}_{2N+1})}}{\|1_{[1,N]}\|_{U^k(\mathbb{Z}_{2N+1})}}.$$

2. Introduction

Multiplicative number theory is an area of analytic number theory where one studies the distributional properties of multiplicative functions, prime numbers, and other sets possessing multiplicative structure. A fundamental principle in this area is that *multiplicative structures* (such as the primes, or the values of a multiplicative function) should behave independently of *additive structures* (such as intervals, additive equations, or arithmetic progressions). This gives rise to a number of conjectures, many of which are still open.

One instance of this principle is that the primes are expected to be distributed somewhat uniformly on very short intervals, such as $[x, x + (\log x)^c]$ with $c > 1$ and $x \in \mathbb{N}$ large. Under the Riemann hypothesis, Selberg proved that for $c = 2 + \varepsilon$ this holds, at least for *almost all* such intervals. A famous conjecture of Cramér [8] asserts that for $c = 2 + \varepsilon$ there should be a prime on $[x, x + (\log x)^c]$ for large x, even without any exceptional intervals, but this is not known even conditional on the Riemann hypothesis. In Article [I], we show, improving the work of Harman [35], that almost all intervals $[x, x + (\log x)^{3.51}]$ contain a product of exactly two primes, and almost all intervals $[x, x + (\log x)^{1+\varepsilon}]$ contain a product of exactly three primes for any $\varepsilon > 0$, the latter result being nearly optimal. Numbers that are the product of exactly two or three primes can be thought of as approximations to the primes, and they have a rigid multiplicative structure in particular. It turns out that these *almost primes* (discussed in detail in Section 3) have more flexibility than the primes, and this is what enables us to prove much stronger results about them.

Another conjecture that combines multiplicative and additive structures is the binary Goldbach conjecture, dating from 1742 and stating that every even $n \geq 4$ is the sum of two primes. This remains an important open problem. On the other hand, the ternary version of the problem, to the effect that every odd number $n \geq 7$ is the sum of three primes, was proved by Vinogradov [108] in 1937 for all sufficiently large integers, and by Helfgott [43] in 2013 in the remaining cases. The essence of many problems of Goldbach-type is showing that the primes do not correlate with "additive sets" (such as the so-called Bohr sets, defined in Section 4.2). The binary Goldbach conjecture has been known since the 1930s to be true for almost all even n, so a natural question to examine is whether the conjecture remains true in almost all cases when one only uses summands coming from a specific subset of the primes. Improving a result of Matomäki [70], we show in Article [II] that this question has a positive answer for primes represented by the polynomial $x^2 + y^2 + 1$; this subset of the primes has also been studied in several other contexts [49], [115], [69] and is an example of a sparse subset of the primes (that is, it has relative density 0 within the primes). Continuing with the theme of additive problems in the primes, we show that the primes of the form $x^2 + y^2 + 1$ also contain infinitely many three-term arithmetic progressions, and that the numbers αp, where α is a fixed irrational and

p runs through such primes, are "well-distributed" modulo 1.

Turning our attention from primes to multiplicative functions, we also study the Liouville function, $\lambda(n)$, whose value is determined by the parity of the number of prime factors of n. Chowla [7] posed in the 1960s the famous conjecture that the Liouville function should behave independently along strings of consecutive integers, taking any sequence of $+1$s and -1s with equal probability. More precisely, Chowla's conjecture can be written as the statement that

$$\frac{1}{x}\sum_{n\leq x}\lambda(n+h_1)\cdots\lambda(n+h_k)=o(1)$$

for any fixed $k\geq 1$ and distinct $h_1,\dots,h_k\in\mathbb{N}$. Thus, for example, the probability that the Liouville function takes value $+1$ at both n and $n+1$ should be $\frac{1}{4}$, the product of the individual probabilities of the events $\lambda(n)=1$ and $\lambda(n+1)=1$ (which have probability $\frac{1}{2}$). During the last few years, there has been a lot of research activity surrounding Chowla's conjecture, and several approximations to the conjecture have been proved (see Section 5 for descriptions of them). In particular, Tao [98] showed in 2015 that the two-point case of Chowla's conjecture holds with logarithmic weights, in the sense that

$$\frac{1}{\log x}\sum_{n\leq x}\frac{\lambda(n+h_1)\lambda(n+h_2)}{n}=o(1)$$

for any distinct $h_1,h_2\in\mathbb{N}$. In Article [III], jointly with Tao, we consider the higher order cases and show that for odd values of k the k-point Chowla conjecture holds with logarithmic weights. Our proof uses combinatorial tools, such as the theory of Gowers norms, and is independent of and simpler than our earlier proof of the same result in [100].

The Liouville function is an archetypal example of a multiplicative function, so it is natural to believe that also shifts of more general multiplicative functions are independent of each other under suitable assumptions. This was made precise by Elliott [11] in the 1990s; he conjectured that one has the discorrelation estimate

$$\frac{1}{x}\sum_{n\leq x}g_1(n+h_1)\cdots g_k(n+h_k)=o(1),$$

whenever $g_1,\dots,g_k$ are multiplicative functions that take values in the unit disc, $h_1,\dots,h_k\in\mathbb{N}$ are distinct shifts, and one of the functions g_j is *non-pretentious* in a suitable sense (we elaborate on this in Section 6). In 2015, Tao [98] proved that Elliott's conjecture holds for $k=2$ with logarithmic weights, in the sense that

$$\frac{1}{\log x}\sum_{n\leq x}\frac{g_1(n+h_1)g_2(n+h_2)}{n}=o(1)$$

under the same assumptions. In Article [IV], we show that Tao's result on the two-point logarithmic Elliott conjecture can be extended to a wider class of real-valued multiplicative functions (with a main term in the asymptotic). This wider class turns out to contain many functions of interest, such as indicator functions related to *smooth numbers* (see Section 6 for details). Making use of this, we prove a logarithmic version of a conjecture of Erdős and Turán [94] on the largest prime factors of n and $n+1$. We also show that certain other sets constructed from multiplicative functions behave independently at n and $n+1$, as one would expect from the heuristic discussed above.

The structure of this thesis is as follows. In Sections 3, 4, 5 and 6, we introduce the topics of the articles [I], [II], [III] and [IV], respectively, and give a wealth of references to the literature on these and related questions. This is followed by the original publications in the same order. The preprint versions of these publications can also be found on the arXiv.org preprint server.

3. Almost primes in very short intervals

In article [I], we study the problem of finding almost primes in almost all short intervals. Since almost primes (defined subsequently) are an approximation to prime numbers, we begin with an overview of conjectures and results on primes in short intervals.

3.1. **Heuristics and conjectures for primes in short intervals**

The prime number theorem, a cornerstone in classical analytic number theory, states that the number of primes $\pi(x)$ up to x satisfies the asymptotic relation

$$\pi(x) = (1 + o(1))\frac{x}{\log x}.$$

Interpreted probabilistically, this means that an integer $n \leq x$ chosen uniformly at random is prime with probability $(1 + o(1))/(\log x)$. Based on this, H. Cramér [8] introduced in the 1930s the heuristic model that the indicator function $1_{\mathbb{P}}(n)$ of primes should behave for $n \leq x$ like a random variable $\mathbf{X}_n \in \{0, 1\}$ that equals 1 with probability $1/\log x$. Moreover, he made the strong assumption that the $\mathbf{X}_n$ are jointly independent of each other; this property of course does not hold for the primes as such (since both n and $n + 1$ cannot be prime for $n \geq 3$), but it serves as a good approximation in various problems[1]. Cramér then deduced from basic probability theory that if the $\mathbf{X}_n$ are as above, then the sum $\sum_{x-\lambda\log x\leq k<x} \mathbf{X}_k$ is Bernoulli distributed with mean λ, and further that the Bernoulli distribution is very closely approximated by the Poisson distribution with the same mean λ (in the regime where $\lambda > 0$ is fixed and $x \to \infty$). Thus, if the model of the primes as the random variables $\mathbf{X}_n$ is adequate, the primes follow the Poisson distribution in short intervals, in the sense that

$$\frac{1}{x}|\{n \leq x :\ \pi(n + \lambda\log x) - \pi(n) = k\}| = (1 + o(1))e^{-\lambda}\frac{\lambda^k}{k!} \tag{3.1}$$

for any fixed $\lambda > 0$ and $k \in \mathbb{N}$. There is strong evidence in support of (3.1), as Gallagher [20] showed that it would follow from a certain uniform version of the widely believed Hardy–Littlewood prime tuples conjecture (for the non-uniform version, see Subsection 5.1). From (3.1) one can deduce many further (yet unproved) properties of the primes in short intervals; in particular, letting λ grow slowly with x and taking $k = 0$, (3.1) would imply that, for any function $\psi(x) \to \infty$ as $x \to \infty$, we have

$$\pi(x + \psi(x)\log x) - \pi(x) \geq 1 \tag{3.2}$$

for almost all $x \in \mathbb{N}$. By a more careful analysis of the tails of the Poisson distribution, one could similarly infer the stronger statement

$$\pi(x + \psi(x)\log x) - \pi(x) = (1 + O(\varepsilon))\psi(x) \tag{3.3}$$

[1]There are more elaborate versions of Cramér's model that take into account local obstructions; see [23].

for almost all $x \in \mathbb{N}$, for any for any fixed $\varepsilon > 0$ and any given function $\psi(x) \to \infty$ with $\psi(x) \leq x$ (and with the implied constant in the $O(\cdot)$ notation being absolute). Moreover, under the assumption that (3.1) holds uniformly for $\lambda \leq C \log x$ with $C \geq 1$, the right-hand side of (3.1) is $\ll 1/x$, so one ends up with the following conjecture[2].

3.1. **Conjecture** (Cramér's conjecture)**.** There exists a constant $C > 0$ such that the interval $[x, x + C(\log x)^2]$ contains a prime for all large enough x.

It seems that Cramér's conjecture is out of reach even under the Riemann hypothesis. Nevertheless, Selberg [91] showed that the Riemann hypothesis implies a version of Cramér's conjecture for *almost all* x.

3.2. **Theorem** (Selberg)**.** *Assume the Riemann hypothesis. Then, for any function $\psi(x)$ tending to infinity as $x \to \infty$, almost all intervals $[x, x + \psi(x)(\log x)^2]$ contain a prime.*

When it comes to the existence of primes in all short intervals, the best statement known under the Riemann hypothesis is that $[x, x + x^{1/2} \log x]$ contains a prime for all large x; see [90]. This remains very far from intervals of polylogarithmic length. Let us mention in passing that Cramér's model also gives probabilistic evidence for the Riemann hypothesis; namely, if one redefines the random variables $\mathbf{X}_n$ slightly to take the value 1 with probability $1/\log n$, then one can use basic properties of random walks to show that for any fixed $\varepsilon > 0$ we have

$$\sum_{n \leq x} \mathbf{X}_n - \int_2^x \frac{dt}{\log t} \ll x^{1/2+\varepsilon}$$

with probability 1, and the corresponding statement for $1_{\mathbb{P}}(n)$ in place of $\mathbf{X}_n$ is well-known to be equivalent to the Riemann hypothesis.

The above indicates that results (whether conditional or unconditional) one can prove about primes in almost all intervals tend to be considerably stronger than what can be proved about primes in all short intervals. One fact that complicates the study of primes in all short intervals is that there are actually some short intervals where the primes notably deviate from their typical behavior. Namely, Maier [67] showed in 1985 in a seminal work that, given any $C > 0$, there is a constant $\eta(C) > 0$ such that for an infinite sequence of $x \in \mathbb{N}$ we have

$$\pi(x + (\log x)^C) - \pi(x) > (1 + \eta(C))(\log x)^{C-1},$$

and an analogous statement holds with the inequality reversed and $1 - \eta(C)$ in place of $1 + \eta(C)$. This is however not in contradiction with the Cramér model, as that model only predicts how the primes should behave on typical intervals, instead of

[2]The above heuristic in fact suggests that $C = 1$ in Conjecture 3.1. There is however some reason to doubt this choice of C, since a more refined version of the Cramér heuristic due to Granville [23], which takes into account the local distribution of primes, predicts that $C \geq 2e^{-\gamma} = 1.12\ldots$. It is nevertheless generally believed that there is a constant C such that Conjecture 3.1 is true.

all intervals (at least if λ is fixed in (3.1)). In particular, the asymptotic (3.3) is believed to be true on almost all intervals, and in [I] we prove some analogues of (3.2) for the counting function of almost primes, with $\psi(x)$ a very slowly growing function, such as $\psi(x) = (\log x)^{\varepsilon}$.

3.2. **Primes in all short intervals**

The problem of detecting primes in short intervals has attracted wide interest in analytic number theory over several decades; see for instance [38], [116], [48, Chapter 12] for treatises on this topic. There are still many open conjectures in this topic, including the Cramér conjecture (Conjecture 3.1) mentioned above. A much more approachable problem than Conjecture 3.1 is that of finding a real number $\theta \in (0,1)$ as small as possible such that every interval $[x, x+x^{\theta}]$ with x large enough contains a prime number. It is expected that any $\theta > 0$ is admissible, as would follow from Conjecture 3.1. The first result in this direction is Hoheisel's result [46] from 1930, with $\theta = 1 - \delta$ for some small $\delta > 0$ (he had $\delta = \frac{1}{33000}$). The exponent θ was improved several times during the following decades by various authors, by using results on the theory of the Riemann zeta function and in particular zero density estimates for it. In 1972, Huxley [47] proved that any $\theta > \frac{7}{12}$ is admissible, and this was slightly improved to $\theta = \frac{7}{12}$ by Heath-Brown [40]. All of the above mentioned results in fact provide an asymptotic of the form

$$\pi(x + x^{\theta}) - \pi(x) = (1 + o(1))\frac{x^{\theta}}{\log x}, \tag{3.4}$$

where $\pi(x)$ is the number of primes up to x; furthermore, when it comes to asymptotics of the type (3.4), the result $\theta = \frac{7}{12}$ is still the best one known.

Subsequent authors have considered the problem of obtaining lower bounds of the correct order of magnitude for the number of primes in an interval, meaning estimates of the form

$$\pi(x + x^{\theta}) - \pi(x) \gg \frac{x^{\theta}}{\log x}. \tag{3.5}$$

To achieve such bounds, one can utilize sieve methods in addition to zero density estimates for the Riemann zeta function to obtain stronger results than for the problem (3.4). Such improvements were achieved for instance in [51], [42], [86], [53], [2], and the best result to date is that of Baker, Harman and Pintz [3], who reached $\theta = 0.525$.

The exponent $\theta = \frac{1}{2}$ certainly appears to be the limit of all known methods; as we mentioned, even under the Riemann hypothesis it is only known that (3.4) is true for all $\theta > \frac{1}{2}$. For the same conclusion $\theta > \frac{1}{2}$, it would suffice to assume the density

hypothesis[3], which is implied by the Riemann hypothesis. In conclusion, conjectures related to the Riemann zeta function do not seem to enable getting information on primes in all intervals of polylogarithmic length (unlike Conjecture 3.1).

3.3. **Primes in almost all short intervals**

Since known results on the problem of primes in all short intervals remain far from what is being conjectured, it is worthwhile to consider the problem of primes in *almost all* short intervals. Naturally, we say that almost all intervals $[x, x + y(x)]$ contain a prime if

$$|\{n \le X : \ [n, n + y(n)] \cap \mathbb{P} = \emptyset\}| = o(X). \tag{3.6}$$

Several authors have obtained much better results for (3.6) than for the problem of finding primes in all short intervals. Regarding asymptotics for primes in almost all short intervals, the best result is Huxley's [47] with $y(x) = x^\theta$ and $\theta > \frac{1}{6}$ in (3.6). When one gives up asymptotics, one can again obtain much better results, as was done in [36], [112], [54], and most recently by Jia in [55] with $\theta > \frac{1}{20}$.

The natural barrier for the known methods is $\theta > 0$; in particular, one is still far from reaching unconditionally intervals that are as short as in Conjecture 3.1. In analogy with the case of all short intervals, $y(x) = x^\theta$ for all $\theta > 0$ in (3.6) would follow from the density hypothesis (and thus also from the Riemann hypothesis), but has not been attained without resorting to such conjectures.

Nevertheless, if one assumes the full strength of the Riemann hypothesis, then Selberg's result (Theorem 3.2) nearly establishes Conjecture 3.1 in almost all cases. Since Gallagher [20] proved that the Poisson distribution property (3.1) of the primes holds under a uniform version of the Hardy–Littlewood prime tuples conjecture, by the discussion of Subsection 3.1 even the optimally short intervals $[x, x+\psi(x)\log x]$ contain a prime almost always under the uniform Hardy–Littlewood conjecture. Heath-Brown [39] showed that the same result can be obtained by assuming the Riemann hypothesis and a suitable uniform version of the pair correlation conjecture for the zeroes of the Riemann zeta function. Needless to say, proving any of these hypotheses seems to be out of reach for all known methods.

If we seek unconditional results in almost all short intervals that are of similar length as in Theorem 3.2, we must relax the notion of primes somewhat. This leads to the study of almost primes in short intervals.

---

[3]This hypothesis states that if $\sigma \in [\frac{1}{2}, 1]$ and $N(\sigma, T)$ is the number of zeros of the Riemann zeta function in the rectangle $[\sigma, 1] \times [-T, T]$ of the complex plane, then $N(\sigma, T) \ll T^{2(1-\sigma)+\varepsilon}$ for any fixed $\varepsilon > 0$.

3.4. **Almost primes in almost all short intervals**

We define two classes of almost primes, the E_k *numbers*

$$E_k = \{n \in \mathbb{N} : \ \Omega(n) = k\}$$

and the P_k *numbers*

$$P_k = \{n \in \mathbb{N} : \ \Omega(n) \leq k\},$$

where $\Omega(n)$ is the number of prime factors of n, counted with multiplicities. Then we trivially have the relations $\mathbb{P} \subset P_k$, $E_k \subset P_k$, and $\mathbb{P} = E_1 = P_1 \setminus \{1\}$. Many of the questions of interest for the set $\mathbb{P}$ of primes have also been investigated for these sets of almost primes, often with significantly better unconditional results. For works that study analogues of classical questions on the primes for the E_k numbers see [21], [35], and for P_k numbers see [5], [81].

There are several reasons why the sets E_k and P_k can be viewed as good approximations[4] to the set $\mathbb{P}$. An obvious reason is of course that the E_k and P_k numbers have only a bounded number of prime factors. In addition, if we denote by $\pi_k(x)$ and $\pi_k^*(x)$ the counting functions of E_k and P_k numbers up to x, respectively, then it is a classical result of Landau (see [102, Section II.6.1]) that we have

$$\pi_k(x) = (1 + o(1))\pi_k^*(x) = (1 + o(1))\frac{x}{\log x} \cdot \frac{(\log\log x)^{k-1}}{(k-1)!} \tag{3.7}$$

for fixed k, so the sets E_k and P_k have nearly the same density $1/(\log x)$ on $[1, x]$ as $\mathbb{P}$ has. In article [I] and in many earlier works, one actually considers numbers $p_1 \cdots p_k \leq x$ with the constraint $P_i \leq p_i \leq P_i^c$ for $i \leq k-1$ for some suitably chosen $P_i \leq x$ and $c > 1$, and it is not difficult to show that such numbers have cardinality $\asymp_c \frac{x}{\log x}$ up to x, just like the primes.

Another reason for the abundance of results on P_k numbers in analytic number theory is that they are exactly the kind of numbers detected by sieve methods. Indeed, sieve methods typically produce numbers $n \leq x$ with no prime factors $p \leq x^c$ for some $c < \frac{1}{2}$, which then means that $n \in P_{\lceil 1/c \rceil - 1}$. Here we see however an important contrast between the E_k and P_k numbers, namely that the E_k numbers (just like the primes) cannot be produced using only classical combinatorial sieves. Indeed, the notorious parity problem in sieve theory, first discovered by Selberg (and discussed for example in [19, Chapter 16]), states that classical combinatorial sieves cannot distinguish numbers with an odd and even number of prime factors from each other. As E_k numbers have exactly k prime factors, they cannot be distinguished from E_{k+1} numbers in such a manner (and so in particular, primes and P_2 numbers cannot be distinguished). Due to this, many results are significantly weaker for E_k numbers than for P_k numbers, and the E_k numbers are a much closer

[4]Of course, the sets $\mathbb{P}$ and E_k are disjoint for $k > 1$, but so are for instance $\mathbb{P}$ and $\{2p : p \in \mathbb{P}\}$, yet they have essentially the same distributional properties.

approximation to the primes, since the sets $\mathbb{P}$ and E_k are both subject to the parity problem. In problems involving E_k numbers in short intervals one therefore wants to make use of the theory of Dirichlet polynomials (and in particular, the theory of the Riemann zeta function), and the advantage compared to primes is that E_k numbers offer more variables to work with in these Dirichlet polynomial bounds, a substantial benefit in proving various estimates.

Concerning P_k numbers, there are very satisfactory short interval results. Notably, Friedlander and Iwaniec [19, Chapters 6 and 11] proved that, for any function $\psi(x)$ tending to infinity with x, almost all intervals $[x, x+\psi(x)\log x]$ contain a P_4 number. They also hinted how to obtain the same result for P_3 numbers. Mikawa proved in turn that almost all intervals $[x, x+(\log x)^{5+\varepsilon}]$ contain a P_2 number. As both proofs are based on classical sieve methods, they are not applicable to the corresponding question for E_k numbers.

It is nevertheless the case that considerably stronger short interval results have been obtained for the E_k numbers than for the primes. Motohashi [83] proved that, for any $\varepsilon > 0$, almost all intervals $[x, x+x^{\varepsilon}]$ contain an E_2 number[5]. Soon after that, Wolke [114] improved this to almost all intervals $[x, x+(\log x)^c]$ for some large constant c (he had $c = 5\cdot 10^6$). This was the first result for E_2 numbers that involved intervals of merely polylogarithmic length, as in Conjecture 3.1 and Theorem 3.2. Harman [35] then gave a reasonable value of c, namely $c = 7+\varepsilon$ for any $\varepsilon > 0$. In Article [I], we improve the exponent $7+\varepsilon$ for E_2 numbers to 3.51 and obtain a nearly optimal result for E_3 numbers.

3.3. **Theorem** (Article [I]). *(a) Almost all intervals* $[x, x+(\log x)^{1+\varepsilon}]$ *contain an* E_3 *number, for any fixed* $\varepsilon > 0$.
(b) Almost all intervals $[x, x+(\log x)^{3.51}]$ *contain an* E_2 *number.*

The result for E_3 numbers is close to optimal, since by (3.7) there exists a positive proportion of intervals $[x, x+(\log x)(\log\log x)^{-2}]$ with no E_3 numbers in them. When it comes to E_2 numbers, significantly improving the exponent 3.51 appears difficult, since even under the density hypothesis the method used in [I] would only improve the exponent to $3+\varepsilon$ (for this, see [I, Remark 10]).

As a matter of fact, we prove the following quantitative version of Theorem 3.3, where the prime factors of the E_3 and E_2 numbers that we detect are of specific sizes.

3.4. **Theorem** (Article [I]). *Let* $\varepsilon > 0$ *be small but fixed. Let* $X \geq 1$ *be large enough. Define the parameters* $P_1 = (\log\log X)^{6+10\sqrt{\varepsilon}}$, $P_2 = (\log X)^{\varepsilon^{-2}}$ *and* $P_1' = (\log X)^{2.51}$.

[5] In the works [83], [114], [35], the numbers under consideration are E_2 numbers, although the wording "P_2 numbers" is used there for lack of better terminology.

Then, for $P_1 \log X \le h \le X$ we have

$$\frac{1}{X}\int_X^{2X}\left|\frac{1}{h}\sum_{\substack{x\le p_1p_2p_3\le x+h\\ p_1\in[P_1,P_1^{1+\varepsilon}]\\ p_2\in[P_2,P_2^{1+\varepsilon}]}}1-\frac{1}{X}\sum_{\substack{X\le p_1p_2p_3\le 2X\\ p_1\in[P_1,P_1^{1+\varepsilon}]\\ p_2\in[P_2,P_2^{1+\varepsilon}]}}1\right|^2 dx = o\left(\frac{1}{(\log X)^2}\right), \tag{3.8}$$

and for $P_1' \log X \le h' \le X$

$$\frac{1}{h'}\sum_{\substack{x\le p_1p_2\le x+h'\\ p_1\in[P_1',(P_1')^{1+\varepsilon}]}}1 \ge \frac{\delta}{X}\sum_{\substack{X\le p_1p_2\le 2X\\ p_1\in[P_1',(P_1')^{1+\varepsilon}]}}1 \tag{3.9}$$

for some small absolute constant $\delta > 0$ and almost all $x \in [X, 2X]$.

By the prime number theorem and a simple application of Chebyshev's inequality, one can show that Theorem 3.4 indeed implies Theorem 3.3. We remark that in [I] we also find E_k numbers on intervals whose lengths approach $\log x$ as k grows. More precisely, almost all intervals $[x, x+(\log x)(\log_{k-1} x)^{C_k}]$ contain an E_k number for some constant $C_k > 0$. Subsequently, Goudout [22] considered E_k numbers in almost all short intervals $[x, x + h_k(x)]$ uniformly in the k aspect. He gave optimal results for $k \asymp \log\log x$ and nearly optimal results for $5 \le k \le \log\log x$.

3.5. **Proof ideas for products of three primes**

As in many previous works on primes and almost primes in short intervals, we reduce proving (3.8), and hence Theorem 3.3(a), to the study of Dirichlet polynomials.

More precisely, we use Perron's formula and a Parseval-type inequality (which utilizes the mean square present in (3.8)) to essentially reduce (3.8) to the corresponding bound for Dirichlet polynomials:

$$\int_{X^{0.01}}^{X/h}|F(1+it)|^2\,dt = o\left(\frac{1}{(\log X)^2}\right), \quad \text{where } F(s) := \sum_{\substack{X\le p_1p_2p_3\le 2X\\ p_1\in[P_1,P_1^{1+\varepsilon}]\\ p_2\in[P_2,P_2^{1+\varepsilon}]}}(p_1p_2p_3)^{-s}; \tag{3.10}$$

strictly speaking, we also need to consider the integral over other intervals than $[X^{0.01}, X/h]$, but this turns out to be the most difficult regime. See [I, Lemma 1, formula (4)] for a more precise version of (3.10). Reducing a problem about short intervals to Dirichlet polynomials is advantageous, because the sum $F(s)$ now runs over a long interval and we can make use of various pointwise, mean value and large values estimates for Dirichlet polynomials to estimate the mean square of $F(s)$.

To effectively estimate these Dirichlet polynomials, we incorporate the method that Matomäki and Radziwiłł [74] developed in 2015 for analyzing multiplicative functions in very short intervals to the setting of almost primes in almost all short intervals. Matomäki and Radziwiłł proved, as a special case of their breakthrough

on multiplicative functions, that the Möbius function $\mu(n)$ has mean $o(1)$ on almost all intervals $[x, x+\psi(x)]$, for any $\psi(x)$ tending to infinity with x. This result suggests that their method in [74] might imply something about primes or almost primes in almost all short intervals, as well. However, for the case of primes the method is not amenable, as a vital element of the proof is a combinatorial factorization identity available for multiplicative functions (the Ramaré identity, [74, Formula (9)]). For the primes there certainly is no analogous identity[6]. The indicator function of those E_k numbers that we will consider, on the other hand, does have a useful factorization, owing to the constraints for their prime factors in Theorem 3.4. Using this, (3.10) roughly speaking takes the factorized form

(3.11)
$$\int_{X^{0.01}}^{X/h} |P_1(1+it)P_2(1+it)P_3(1+it)|^2\,dt = o\left(\frac{1}{(\log X)^2}\right), \quad P_j(s) := \sum_{P_j \le p \le P_j^{1+\varepsilon}} p^{-s}$$

and $P_3 := X/P_1P_2$ and the sum $P_3(s)$ is over a dyadic interval. Above we have separated the contribution of each of the variables p_i and can estimate the polynomials corresponding to different variables in different ways.

An estimate of the shape (3.11) is our goal in the proof of Theorem 3.3(a), but a number of aspects of the Matomäki–Radziwiłł method require modifications when working with E_k numbers; in particular, one needs to obtain logarithmic savings in places where $o(1)$ savings would suffice for multiplicative functions (for instance, in [I, Lemma 4]). This is due to the fact that the E_k numbers are a sparse set, of density roughly $1/(\log x)$ up to x. Additionally, there is a part of the proof ([I, Proposition 3]), where we need a product of three Dirichlet polynomials of "significant length", in order to apply a $L^2 - L^\infty$ bound to the mean square of their product (if we had only one or two polynomials, we could not afford to apply a pointwise bound to one of them; this is reminiscent of the differences in difficulty between binary and ternary problems in applications of the circle method; see [107, Chapter 3]). We do obtain three Dirichlet polynomials when dealing with E_3 numbers, but two of them are of minuscule length[7] (reflecting the fact that we want to minimize the length of the intervals on which we detect E_3 numbers). We go around this issue by applying Heath-Brown's identity [52, Chapter 13] to decompose one of the "long" Dirichlet polynomials into a product of either two "zeta sums" or three "prime-factored polynomials" (for these concepts, see [I, Section 1.2, Section 2.5]). We then employ a result of Watt [111] (which generalizes the fourth moment bound of the Riemann zeta function) to deal with mean squares of the resulting zeta sums,

[6]It is the case that the indicator function of the primes can be "factorized" into a Dirichlet convolution, by means of Vaughan's or Heath-Brown's identities [52, Chapter 13], but these factorizations are not nearly flexible enough.

[7]The lengths of the Dirichlet polynomials involved will be roughly $(\log\log x)^6$, $(\log x)^{\varepsilon^{-2}}$ and $x(\log x)^{-\varepsilon^{-2}}$, the first two of which are too short for pointwise bounds.

whereas the mean square of the product of three prime-factored polynomials can be dealt with the mentioned $L^2 - L^\infty$ approach. These are the main ingredients we use for the E_3 part of Theorem 3.3.

3.6. **Proof ideas for products of two primes**

For obtaining good results about E_2 numbers, we make use of all the above-mentioned ideas, as well as some additional ones. One could readily apply the strategy we used for E_3 numbers to obtain the exponent $5+\varepsilon$ for E_2 numbers (see [I, Section 4.1]; in the case of the exponent $5+\varepsilon$, we would even get an asymptotic formula on almost all short intervals for the number of E_2 numbers with prime factors in certain ranges, as in the E_3 case). The fact that we have only two variables to work with in the case of E_2 numbers appears to make improving the exponent hard. However, we can apply the principle of Harman's sieve to gain more flexibility. Firstly, we can increase the number of variables by applying the Buchstab identity, a number-theoretic form of the inclusion-exclusion identity. This identity allows us to decompose

$$S_h(x) := \sum_{\substack{x\le p_1p_2\le x+h\\ P_1'\le p_1\le (P_1')^{1+\varepsilon}}} 1,$$

for any choice of $1 \le w < \sqrt{x}$, as

$$\begin{aligned} S_h(x) &= \sum_{\substack{x\le p_1n\le x+h\\ P_1'\le p_1\le (P_1')^{1+\varepsilon}\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 - \sum_{\substack{x\le p_1q_1n\le x+h\\ P_1'\le p_1\le (P_1')^{1+\varepsilon}\\ w\le q_1<\sqrt{x}\\ (n,\mathcal{P}(q_1))=1\\ n>1}} 1 \\ &= \sum_{\substack{x\le p_1n\le x+h\\ P_1'\le p_1\le (P_1')^{1+\varepsilon}\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 - \sum_{\substack{x\le p_1q_1n\le x+h\\ P_1'\le p_1\le (P_1')^{1+\varepsilon}\\ w\le q_1<\sqrt{x}\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 + \sum_{\substack{x\le p_1q_1q_2n\le x+h\\ P_1'\le p_1\le (P_1')^{1+\varepsilon}\\ w\le q_2<q_1<\sqrt{x}\\ (n,\mathcal{P}(q_2))=1\\ n>1}} 1 \\ &:= \Sigma_1(h) - \Sigma_2(h) + \Sigma_3(h). \end{aligned}$$

We take here $w = X^{\eta(X)}$ for a suitable function $\eta(X)$ tending to 0. (In particular, w is small enough for the fundamental lemma of sieve theory [19, Chapter 6] to be applicable). As we will see later, the first two sums $\Sigma_1(h), \Sigma_2(h)$ are asymptotically equal to their dyadic counterparts $\frac{h}{X}\Sigma_1(X)$ and $\frac{h}{X}\Sigma_2(X)$, respectively, for almost all $x \in [X, 2X]$. For the sum $\Sigma_3(h)$, however, we are not able to prove an asymptotic unless we impose some additional conditions on the sizes of the variables q_i. Let $\Sigma_3'(h)$ be the part of $\Sigma_3(h)$ that we can evaluate asymptotically (to be asymptotic to the normalized dyadic version of the same sum), and let $\Sigma_3''(h) \ge 0$ be the rest (the part that we can evaluate is expressed precisely in [I, Subsection 6.3] as the

sums $\Sigma_3^{(1)}(h)$ and $\Sigma_3^{(2)}(h)$). Then, for almost all $x \in [X, 2X]$, we get

$$\begin{aligned}\frac{1}{h}S_h(x) &= \frac{1}{h}(\Sigma_1(h) - \Sigma_2(h) + \Sigma_3(h)) \\ &= \frac{1}{X}\Sigma_1(X) - \frac{1}{X}\Sigma_2(X) + \frac{1}{h}\Sigma_3'(h) + \frac{1}{h}\Sigma_3''(h) + o\left(\frac{1}{\log X}\right) \\ &\geq \frac{1}{X}(\Sigma_1(X) - \Sigma_2(X) + \Sigma_3'(X)) + o\left(\frac{1}{\log X}\right) \\ &= \frac{1}{X}S_X(X) - \frac{1}{X}\Sigma_3''(X) + o\left(\frac{1}{\log X}\right).\end{aligned}$$

Hence, proving (3.9) has been reduced to establishing the mentioned asymptotics for $\Sigma_1(h), \Sigma_2(h)$ and $\Sigma_3'(h)$, and additionally to showing that $\Sigma_3''(X) \leq (1 - 2\delta)S_X(X)$ for some fixed $\delta > 0$. Clearly, the part $\Sigma_3'(h)$ of $\Sigma_3(h)$ that we can evaluate must be large enough for this upper bound to hold. It turns out that if we look for E_2 numbers on intervals of length $[x, x + (\log x)^c]$, then we should take $P_1' = (\log x)^{c-1}$ in the definitions of $S_h(x)$ and $\Sigma_i(h)$, and the smaller P_1' is, the harder $\Sigma_3(h)$ is to estimate. We can give by the prime number theorem an asymptotic for $\Sigma_3''(X)$ (since the sum is over a dyadic interval) in terms of multidimensional "Buchstab integrals" [38, Chapter 3], [I, Section 6.3.3], and we compute that if $c = 3.51$ above, the sum $\Sigma_3''(X)$ is indeed smaller than the main term $S_X(X)$. We are thus left with showing asymptotics for $\Sigma_1(h), \Sigma_2(h), \Sigma_3'(h)$

The proofs of the asymptotics of the sums $\Sigma_1(h), \Sigma_2(h)$ and $\Sigma_3'(h)$ follow the same strategy as in the E_3 case, but make use of some additional inputs. We reduce the problem to the setting of Dirichlet polynomials, so that the aim is to prove that (3.10) holds for the Dirichlet polynomials $F(s)$ that correspond to $\Sigma_1(h), \Sigma_2(h)$ and $\Sigma_3'(h)$. By applying a simple sieve to $\Sigma_1(h)$ and $\Sigma_2(h)$, they become *type I sums* (meaning a sum having a long, unrestricted integer variable), and therefore we can employ Watt's mean value theorem as in the E_3 case to handle them.

The sum $\Sigma_3'(h)$, in turn, is a *type II sum* (it has several variables of substantial length, but these variables come with weights), and is more difficult to estimate. However, we have restricted the sizes of the variables in a suitable manner in this sum, making asymptotic evaluation possible. We utilize the ideas from the E_3 case together with the theory of exponent pairs [I, Section 5.1] and better large values theorems for Dirichlet polynomials [I, Lemma 7] to obtain the bound (3.10) for the Dirichlet polynomial corresponding to $\Sigma_3'(h)$, and this then implies that $\Sigma_3'(h)$ has the desired asymptotic.

We have now outlined the main strategy for proving Theorem 3.3; the details can be found in [I].

4. The Goldbach problem for primes of a special form

4.1. **The Goldbach conjectures**

The Goldbach conjectures, proposed by Goldbach in a letter to Euler in 1742, are some of the most influential and well-known problems in analytic number theory. The ternary Goldbach conjecture asserts that every odd integer $n \geq 7$ is the sum of three primes. The binary Goldbach conjecture in turn claims that every even integer $n \geq 4$ can be written as the sum of two primes; this is still unsolved. The binary Goldbach conjecture is evidently stronger than the ternary one, since if $n = p_1 + p_2$ is a sum of two primes, then $n+3 = p_1 + p_2 + 3$ is a sum of three primes.

The ternary conjecture was settled in all but finitely many cases by Vinogradov [108] in 1937 in a work that redefined the Hardy–Littlewood circle method.

4.1. **Theorem** (Vinogradov)**.** *Every large enough odd integer n can be written as $n = p_1 + p_2 + p_3$ with $p_1, p_2, p_3 \in \mathbb{P}$.*

For a modern proof of Theorem 4.1, see [107, Chapter 3]. It took until 2013 before Theorem 4.1 was extended to all $n \geq 7$; this was achieved by Helfgott [43], by introducing new ideas both on the analytic and numerical sides. Although the binary analogue of Vinogradov's result has resisted all attempts to a full resolution, shortly after Vinogradov's proof it was shown independently by Chudakov, van der Corput and Estermann that we have the following approximation (see [107, Chapter 3]).

4.2. **Theorem** (Almost all cases of binary Goldbach)**.** *Almost all even integers n can be expressed as $n = p_1 + p_2$, where $p_1, p_2 \in \mathbb{P}$.*

Here and in what follows, by "almost all" we mean that the number of exceptional even integers $n \leq N$ is $o(N)$.

Given that one has such an approximation to the binary Goldbach conjecture, one may contemplate a number of refinements, such as strengthening the bound for the number of exceptions

$$E(X) := |\{n \leq X : \ n \equiv 0 \pmod 2, \ n \text{ not a sum of two primes}\}|.$$

This question was considered most notably by Montgomery and Vaughan [82], who obtained $E(X) \ll X^{1-\delta}$ for some fixed $\delta > 0$. The bound was improved by Chen and Pan [6], Li [64] and Lu [66], among others, the last of whom holds the record $\delta = 0.121$. In a somewhat different direction, one can try to minimize $\theta > 0$ such that every interval $[X, X + X^\theta]$ with X large contains a sum of two primes. This was investigated in [84], [85], [63], among others, and in Jia's work [56], where the best known result $\theta = \frac{7}{108} + \varepsilon$ was obtained.

In Article [II], we had a different generalization of Theorems 4.1 and 4.2 in mind, namely a version of the problem where only a specific subset of the primes are allowed as summands.

4.2. **The Goldbach problem for subsets of the primes**

When looking for a problem which is more challenging than the ternary Goldbach problem, but (hopefully) more manageable than the binary Goldbach problem, the following problem naturally arises.

4.3. **Problem** (Ternary Goldbach for subsets of the primes)**.** *Let $\mathcal{P} \subset \mathbb{P}$ be a given interesting subset of the primes. Is it the case that all large enough odd n can be represented as $n = p_1 + p_2 + p_3$ with $p_1, p_2, p_3 \in \mathcal{P}$?*

Whether Problem 4.3 has a positive or negative answer crucially depends on the distribution of the set $\mathcal{P}$ in arithmetic progressions and more general *Bohr sets*. These are sets of the form

$$\bigcup_{i \leq m} \{n \in \mathbb{N} : \|\alpha_i n\| \leq \eta_i\} \tag{4.1}$$

with $\alpha_i \in \mathbb{R}$, $\eta_i \in (0,1)$. If we take $\alpha_i \in \mathbb{Q}$, we see that arithmetic progressions are a special case of Bohr sets. Now, if for example $\mathcal{P} = \{p \in \mathbb{P} : \ p \equiv 1 \pmod 3\}$, then only integers of the form $n \equiv 0 \pmod 3$ can be represented as a sum of three primes from $\mathcal{P}$. Similarly, if $\mathcal{P} = \{p \in \mathbb{P} : \ \|\sqrt{2}p\| < \frac{1}{10}\}$, then every n representable as a sum of three primes from $\mathcal{P}$ satisfies $\|\sqrt{2}n\| < \frac{3}{10}$, a property that fails for a positive proportion of odd n. In light of these examples where the answer to Problem 4.3 is negative, we would like the set $\mathcal{P}$ studied in Problem 4.3 to contain a fair proportion of elements from each Bohr set of the form (4.1).

It is only in recent years that interesting subcases of Problem 4.3 have been solved. In 2014, Shao [92] showed that Problem 4.3 has an affirmative answer for any subset $\mathcal{P} \subset \mathbb{P}$ of relative lower density[8] greater than $\frac{5}{8}$. Perhaps surprisingly, this is optimal when taking only the density into consideration: the subset

$$\mathcal{P} := \{p \in \mathbb{P} : \ p \equiv 1, 2, 4, 7, 13 \pmod{15}\}$$

has density $\frac{5}{8}$ and, by simple modular arithmetic, sums of three of its elements are never $\equiv 14 \pmod{15}$. Matomäki and Shao [77] considered Problem 4.3 for significantly sparser but specific subsets of the primes. Their subsets of interest are the *Chen primes* and the *bounded gap primes*[9]. Chen primes are primes p such that $p + 2$ has at most two prime factors; the infinitude of such primes was proved by Chen [5] in 1973. The bounded gap primes are primes p such that the interval $[p, p + C]$ contains at least two primes for some large, fixed C, and their infinitude was proved in the celebrated work of Zhang [117] in 2013 and in a more general form by Maynard [80] and Tao (unpublished) in 2014.

---

[8]We define the relative lower density of $B \subset A$ with respect to A as $\liminf_{N\to\infty} \frac{|B \cap [1,N]|}{|A \cap [1,N]|}$.

[9]The latter set does not have a standardized name.

Problem 4.3 should be compared to the problem of locating three-term arithmetic progressions in the set $\mathcal{P}$, which also involves studying a linear equation in the primes.

4.4. **Problem** (Three-term arithmetic progressions in subsets of the primes)**.** *Let $\mathcal{P} \subset \mathbb{P}$ be a given interesting subset of the primes. Is it the case that $\mathcal{P}$ contains infinitely many solutions to $p_1 + p_3 = 2p_2$ with $p_1, p_2, p_3 \in \mathcal{P}$ distinct?*

As with Problem 4.3, a number of special cases of Problem 4.4 have been solved. Concerning this, Green [25] proved Roth's theorem for the primes[10], stating that any subset of the primes of positive relative upper density[11] contains infinitely many non-constant three-term arithmetic progressions. This was famously generalized to k-term arithmetic progressions by Green and Tao [27]. In another work, Green and Tao [26] showed that the Chen primes satisfy Roth's theorem.

The approach that Green and Green–Tao developed for this type of problems is called a *transference principle*, as it allows one to transfer information (such as Roth's theorem) from dense subsets of the integers to sparse ones (such as the primes) under suitable conditions. Intuitively speaking, the principle says that if $A \subset [1, N]$ is a set with $|A| = \delta N$ and $\delta = \delta(N) > 0$, then A contains many three-term arithmetic progressions, provided that the normalized indicator $\delta^{-1}1_A(n)$ has a *pseudorandom majorant* (that is, a majorizing function $\nu(n)$ that has mean $\asymp 1$ and has small Fourier coefficients) and that $\delta^{-1}1_A(n)$ is "Fourier bounded" (that is, its Fourier transform has small L^r norm for $r > 2$). This version of the transference principle is specific to the translation-invariant linear equation $x + z = 2y$, and does as such not apply to the setting of Problem 4.3. Indeed, the set $\{p \in \mathbb{P} : \ \|\sqrt{2}p\| < \frac{1}{10}\}$ is an example of a set that contains an abundance of arithmetic progressions (since it is a positive relative density set of the primes) but, as mentioned earlier, has no solutions to $n = p_1 + p_2 + p_3$ for many odd n. Therefore, to deal with Problem 4.3, one needs a different version of the transference principle, which takes into account the distribution of $\mathcal{P}$ in Bohr sets (an example of which is the fractional part set above); we shall discuss this later in this section.

4.3. **Statements of results**

In Article [II], we study Problem 4.3 for the specific subset

$$\mathscr{P} := \{p \in \mathbb{P} : \ p = x^2 + y^2 + 1, x, y \in \mathbb{Z}\},$$

consisting of primes representable as values of the polynomial $x^2 + y^2 + 1$. There are a number of reasons why the set $\mathscr{P}$ is interesting. Firstly, it is perhaps the simplest non-trivial example of a sparse subset of the primes consisting of the values of a multivariate polynomial. When it comes to single-variable polynomials,

[10]The theorem is named so, since Roth [88] proved that positive upper density subsets of the integers contain infinitely many non-trivial three-term arithmetic progressions.

[11]The relative upper density of a set $A \subset B$ with respect to $B \subset \mathbb{N}$ is $\limsup_{N\to\infty} \frac{|B \cap [1,N]|}{|A \cap [1,N]|}$.

only degree one polynomials have been proved to produce infinitely many primes (by Dirichlet's theorem, under a coprimality condition), so one should turn to multivariate polynomials for interesting unconditional results. Concerning irreducible binary quadratic forms $ax^2+bxy+cy^2$, which are some of the simplest multivariate polynomials, the Chebotarev density theorem can be used to characterize when such a form represents infinitely many primes. When the form does represent infinitely many primes, Chebotarev's theorem implies that the relative density of such primes is $\frac{1}{2}$. Therefore, binary quadratic forms do not produce sparse subsets of the primes.

We mention that there are also some interesting higher degree multivariate polynomials that are known to represent infinitely many primes. Friedlander and Iwaniec [18] showed that the polynomial x^2+y^4 takes infinitely many prime values, Heath-Brown [41] showed that x^3+2y^3 has the same property, and Maynard [79] showed this property for an infinite class of more general polynomials called norm forms. The sets of primes corresponding to these polynomials have cardinalities $\ll X^{1-\delta}$ up to X for some $\delta>0$, so they are certainly sparse subsets of the primes. Since primes represented by these polynomials have not been studied in arithmetic progressions to large moduli, the Goldbach problem appears formidable for them.

The set $\mathscr{P}$ of primes represented by x^2+y^2+1 is also a sparse subset of the primes; an application of Selberg's sieve provides the bound $|\mathscr{P}\cap[1,N]|\ll N(\log N)^{-\frac{3}{2}}$ (to see this, note that if $m\in\mathscr{P}\cap[\frac{N}{2},N]$, then $(m,\prod_{p\le z}p)=(\frac{m-1}{k^2},\prod_{p\le z,p\equiv 3 \pmod 4}p)=1$ for $z=N^{0.01}$ and for some $k\in\mathbb{N}$). It is known that $\mathscr{P}$ is infinite, a result first shown by Linnik [65] in 1960, using his dispersion method. Later, a sieve-theoretic proof of this was given by Iwaniec [50], making use of the linear and semilinear sieves. Iwaniec's proof also established the matching lower bound $|\mathscr{P}\cap[1,N]|\gg N(\log N)^{-\frac{3}{2}}$. Subsequently, various properties of the set $\mathscr{P}$ have been investigated; in particular, it has been studied over short intervals [115], [69], and variants of the Goldbach problem have been studied for this set. In 2008, Matomäki [70] showed that almost all even $n\not\equiv 2 \pmod 6$ can be expressed as $n=p+q$ with $p\in\mathscr{P}$ and $q\in\mathbb{P}$ a generic prime. Next, Tolev [104] gave an asymptotic formula for the representations of such n as $n=p+q$ with $p\in\mathscr{P}$, $q\in\mathbb{P}$, again for almost all n. In another work [105], he considered the corresponding ternary Goldbach problem and showed that every large enough odd n can be written as $n=p+q+r$ with $p,q\in\mathscr{P}$ and $r\in\mathbb{P}$. We strengthen these results by solving Problem 4.3 for the set $\mathscr{P}$.

4.5. **Theorem** (Article [II]). *Every large enough odd n can be represented as $n=p_1+p_2+p_3$ with $p_1,p_2,p_3\in\mathscr{P}$.*

We also improve Matomäki's result [70] by settling almost all cases of the binary Goldbach problem for $\mathscr{P}$.

4.6. **Theorem** (Article [II]). *Almost all even integers $n\not\equiv 5,8 \pmod 9$ can be represented as $n=p_1+p_2$ with $p_1,p_2\in\mathscr{P}$.*

Note that the condition $n \not\equiv 5, 8 \pmod 9$ is necessary, since in the complementary case the fact that $p \equiv 1, 2, 5$ or $8 \pmod 9$ for primes $p = x^2 + y^2 + 1 \neq 3$ shows that p_1 or p_2 equals 3 in the representation $n = p_1 + p_2$, in which case we could only represent $\ll N(\log N)^{-1}$ integers up to N.

We also investigate Problem 4.4 in Article [II], again for the specific set $\mathscr{P}$. We are able to resolve this problem and, more generally, to prove Roth's theorem for $\mathscr{P}$.

4.7. **Theorem** (Article [II]). *The set $\mathscr{P}$ contains infinitely many non-trivial three-term arithmetic progressions. More generally, any subset of*

$$\mathscr{P}^* := \{p \in \mathbb{P} : \ p = x^2 + y^2 + 1, \ x, y \text{ coprime}\}$$

having positive relative upper density contains infinitely many non-trivial three-term arithmetic progressions.

We remark that subsequently Sun and Pan [95] generalized this result by proving that the set $\mathscr{P}$ contains arbitrarily long arithmetic progressions.

One more topic considered in Article [II] is the distribution of irrational multiples of primes belonging to the subset $\mathscr{P}$. For the whole set of primes, such results take the form:

(4.2)
For all $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, $\beta \in \mathbb{R}$ we have $\|\alpha p + \beta\| < p^{-\theta}$ for infinitely many $p \in \mathbb{P}$,

where θ is a constant whose value we are attempting to maximize. The first result in this direction was that of Vinogradov [109] with $\theta = \frac{1}{5} - \varepsilon$. This was improved several times, notably by Vaughan [106] to $\theta = \frac{1}{4} - \varepsilon$, by Harman [37] to $\theta = \frac{3}{10}$, and by Jia [57] to $\theta = \frac{9}{28}$. In the special case $\beta = 0$, the record is Matomäki's result [72] with $\theta = \frac{1}{3} - \varepsilon$.

The problem (4.2) has also been studied for Chen primes in [71], [93], Piatetski-Shapiro primes in [32], and Gaussian primes in [1]. Here we obtain the first result concerning (4.2) for the subset $\mathscr{P}$ of the primes.

4.8. **Theorem** (Article [II]). *Let $\varepsilon > 0$. For any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and $\beta \in \mathbb{R}$, we have*

$$\|\alpha p + \beta\| < p^{-\frac{1}{80} + \varepsilon}$$

for infinitely many $p \in \mathscr{P}$.

This establishes that the elements of $\mathscr{P}$ are somewhat uniformly distributed in Bohr sets. We remark that the exponent $\frac{1}{80}$ in Theorem 4.8 could be improved by a more careful analysis in [II, Sections 8-9]; we however confined ourselves to showing that one can get some positive, explicit exponent.

4.4. **Method of proof**

Our proofs of Theorems 4.5 and 4.6 do not apply the classical circle method, but rather a transference principle for ternary equations. Let us first describe why the traditional circle method approach is not applicable to the set $\mathscr{P}$.

The starting point of Vinogradov's proof of Theorem 4.1, and many subsequent developments of the circle method, is the reduction of the problem to analyzing exponential sums via the identity

$$|\{(p_1,p_2,p_3)\in\mathbb{P}^3:\ N=p_1+p_2+p_3\}| = \int_0^1 S(\alpha)^3 e(-N\alpha)\,d\alpha,\ \ S(\alpha):=\sum_{p\le N} e(\alpha p).$$

This identity is seen to hold by expanding out $S(\alpha)^3$ and applying the orthogonality identity

$$1_{n=0} = \int_0^1 e(n\alpha)\,d\alpha.$$

One then examines separately the *major arc* case $\alpha\in\mathfrak{M}$, where α is close to a rational number with small denominator, and the opposite *minor arc* case $\alpha\in\mathfrak{m}:=[0,1]\setminus\mathfrak{M}$. For $\alpha\in\mathfrak{M}$, the sum $S(\alpha)$ is often "large", and one can evaluate it asymptotically by the Siegel–Walfisz theorem. For $\alpha\in\mathfrak{m}$, in turn, one expects $S(\alpha)$ to be "small", and this can be proved with the help of Vaughan's identity, which transforms sums over primes to bilinear sums. We refer to [107, Chapter 3] for details of the method.

If we applied the same strategy to Theorem 4.5, we would run into trouble, since we do not have a good understanding of

$$S_{\mathscr{P}}(\alpha):=\sum_{\substack{p\le N\\ p\in\mathscr{P}}} e(\alpha p),$$

neither in the major arc nor in the minor arc case. In the major arcs, the problem is that we only have upper and lower bounds for $|\mathscr{P}\cap[1,N]|$ that are off by a constant factor, and hence we have no asymptotic even for $S_{\mathscr{P}}(0)$. In the minor arc case, the difficulty is that no analogue of Vaughan's identity is known for the indicator function $1_{\mathscr{P}}(n)$. For these reasons, the classical circle method is not the right line of attack for Theorem 4.5. We mention though that if one studies the ternary Goldbach problem with *two* of the three prime variables coming from the subset $\mathscr{P}$, then the circle method coupled with sieve methods is applicable; see [70], [105].

The proofs of Theorems 4.5 and 4.6 are instead based on a transference type principle of Matomäki and Shao [77, Theorem 2.3]. Roughly speaking, the principle says that if N is large and a set $A\subset[1,N]$ with $|A|=\delta N$ obeys, for some fixed $\delta_0>0$ (and small enough $\eta=\eta(\delta_0)$) the properties

(i) (well-distribution in Bohr sets) $|A \cap (B-t)| \geq \delta_0 |A||B|/N$ for $t \in [N/4, N/2]$ and all Bohr sets[12] B with $|B| \geq \eta N$;
(ii) (Fourier boundedness) $\sum_{\xi \in \mathbb{Z}_N} |\delta^{-1} \widehat{1_A}(\xi)|^{\frac{5}{2}} \leq \delta_0^{-1}$;
(iii) (Non-sparseness) $|A \cap [0.1N, 0.4N]| \geq \delta_0 \cdot \delta N$,

then there exist $a_1, a_2, a_3 \in A$ such that $N = a_1 + a_2 + a_3$. The actual formulation of the transference-type principle is somewhat more involved; we refer to [77, Theorem 2.3] for the details. The principle can also be generalized to work for almost all cases of binary problems; see [II, Proposition 2.1] for this. We also note that Matomäki, Maynard and Shao [73] developed a different transference-type principle for Goldbach-type problems; this version allowed them to improve the exponent for the Goldbach problem with almost equal variables [73, Theorem 1.1].

We will apply the transference principle essentially to

$$A := \{n \leq N : \ Wn + b \in \mathscr{P}\} \tag{4.3}$$

with

$$\delta \asymp \left(\frac{W}{\varphi(W)}\right)^{\frac{3}{2}} (\log N)^{-\frac{3}{2}},$$

where $(b, W) = 1$ and $W = \prod_{p \leq w} p$ for some large, fixed w. This "W-trick" of restricting to primes in a residue class is necessary to guarantee the well-distribution of A in arithmetic progressions (which are a special case of Bohr sets).

Intuitively, condition (i) of the transference principle guarantees that A contains a fair proportion of each Bohr set (which, as we indicated in Subsection 4.2, is necessary); condition (ii) is related to the existence of a pseudorandom majorant[13]; and condition (iii) says that A is not too concentrated on certain subintervals.

The main task in the proofs of Theorems 4.5 and 4.6 is then verifying the conditions (i)–(iii) of the transference principle for the specific set A given by (4.3). Condition (iii) is the simplest to check and follows with minor modifications from Iwaniec's proof of the infinitude of $\mathscr{P}$. Condition (ii), the Fourier boundedness condition, is closely related to the restriction theory of the primes, a topic studied by Green [25] and Green–Tao [26]. To obtain (ii), we roughly speaking want to construct a function $\beta : \mathbb{N} \to \mathbb{R}_{\geq 0}$ that enjoys the majorization property $\delta^{-1} 1_A(n) \leq C\beta(n)$, has mean value $\asymp 1$ (so that β is essentially a probability measure on $[1, N]$), and has a Fourier expansion that is of "low enough complexity". In other words, β is a pseudorandom majorant for 1_A in a suitable sense. Then, under these conditions,

[12] We defined Bohr sets in formula (4.1).
[13] In [II, Section 4], we show that the existence of a suitable pseudorandom majorant implies (ii), and then we construct such a majorant.

[26, Proposition 4.2] implies a *restriction estimate*

$$\left(\sum_{\xi\in\mathbb{Z}_N}\left|\frac{1}{N}\sum_{n\le N}a_n\beta(n)e\left(-\frac{\xi n}{N}\right)\right|^r\right)^{1/r}\le C_r\left(\frac{1}{N}\sum_{n\le N}|a_n|^2\beta(n)\right)^{1/2}$$

for all complex numbers a_n and fixed $r>2$. Taking

$$a_n=\begin{cases}\frac{\delta^{-1}1_A(n)}{\beta(n)}, & \beta(n)\neq 0\\ 0, & \beta(n)=0,\end{cases}$$

(so that $|a_n|\ll 1$) we deduce, in particular, that

$$\sum_{\xi\in\mathbb{Z}_N}|\delta^{-1}\widehat{1_A}(\xi)|^{\frac{5}{2}}\ll\left(\frac{1}{N}\sum_{n\le N}\beta(n)\right)^{5/4}\ll 1.$$

Naturally, we still need to prove that such a pseudorandom majorant β exists. It turns out that the Selberg upper bound sieve does the job, and to see this we closely follow a paper of Ramaré and Ruzsa [87].

The majority of the proofs of Theorems 4.5 and 4.6 is then devoted to proving condition (i), well-distribution in Bohr sets. In other words, we wish to analyze sums of the form

$$\sum_{\substack{n\le N\\ n\in\mathscr{P}}}1_B(n),$$

where B is a Bohr set (or a smoothed version thereof). By applying a weighted form of the linear and semilinear sieves (as developed in [II, Section 6], following Iwaniec's work in [49]), we reduce the problem to showing that the count of primes in Bohr sets has a good enough level of distribution. More precisely, we want to find levels of distribution $\rho_1,\rho_2\in(0,1)$ as large as possible, such that the following holds. For a set $\mathcal{L}\subset\mathbb{N}$ of "bilinear type" (in the sense that it consists of integers having a certain type of factorization), we have

$$\sum_{d\le N^{\rho_1}}\lambda_d^{+,\mathrm{LIN}}\sum_{\substack{\ell\le N^{0.9}\\ \ell\in\mathcal{L}}}\left(\sum_{\substack{N\le n\le 2N\\ n=\ell p+1\\ n\equiv 0 \pmod d}}1_B(n)-\frac{1}{\varphi(d)}\sum_{N\le n\le 2N}\frac{1_B(n)}{\ell\log\frac{n}{\ell}}\right)\ll\frac{N}{(\log N)^{100}},\tag{4.4}$$

and

$$\sum_{d\le N^{\rho_2}}\lambda_d^{-,\mathrm{SEM}}\left(\sum_{\substack{N\le p\le 2N\\ p\equiv 1 \pmod d}}1_B(p)-\frac{1}{\varphi(d)}\sum_{N\le p\le 2N}1_B(p)\right)\ll\frac{N}{(\log N)^{100}},\tag{4.5}$$

where B is a Bohr set (or a smoothed version of it), $\lambda_d^{+,\mathrm{LIN}}$ are the upper bound linear sieve weights with level $D_1=N^{\rho_1}$ and sifting parameter $z_1=N^{1/5}$ and $\lambda_d^{-,\mathrm{SEM}}$

are the lower bound semilinear sieve weights with level $D_2 = N^{\rho_2}$ and sifting parameter $z_2 = N^{1/3-\varepsilon}$ (see Section 1 for a precise definition of sieve weights and [II, Hypothesis 6.4] for the exact, slightly more complicated statements of interest).

By expanding $1_B(p)$ as a finite Fourier series (and a small error), we then need to bound the Bombieri–Vinogradov type averages (4.4) and (4.5) with an additive character $e(\alpha p)$ in place of $1_B(p)$. In the case of the linear sieve weights (which have the well-factorability property defined in [19, Chapter 12]), we manage to obtain the good value $\rho_1 = \frac{1}{2} - \varepsilon$ for the level of distribution by following [71]. The semilinear sieve weights, however, are not well-factorable, and the level of distribution $\rho_2 = \frac{1}{3} - \varepsilon$ obtained for general weights in [103, Lemma 1] is not good enough for our purposes. We therefore prove a combinatorial factorization of semilinear sieve weights [II, Lemma 9.2] by following the principle of Harman's sieve [38, Chapter 3], and this enables us to show that the weights $\lambda_d^{-,\mathrm{SEM}}$ have "enough flexibility" in their factorizations as Dirichlet convolutions. This amount of flexibility allows us to achieve the better level of distribution $\rho_2 = \frac{3}{7} - \varepsilon$, which is good enough for our needs. The details of the proof can be found in [II].

5. On the logarithmic Chowla conjecture

5.1. **Chowla's conjecture**

The Liouville function, a fundamentally important function in multiplicative number theory, is defined as $\lambda(n) := (-1)^{\Omega(n)}$, where $\Omega(n)$ is the number of prime factors of the integer n counted with multiplicities. This function is closely related to the more well-known Möbius function, given by $\mu(n) := (-1)^{\Omega(n)} \cdot 1_{n \text{ squarefree}}$; in particular, they are both multiplicative functions having value -1 at the primes. The distribution of the Liouville function (or equally well of the Möbius function) appears highly random (like a series of coin flips) and in particular, consecutive values of the Liouville function should be asymptotically independent of each other. This was formalized by Chowla [7] in 1965 as the following assertion.

5.1. **Conjecture** (Chowla's conjecture)**.** For any $k \geq 1$ and any distinct shifts $h_1, \dots, h_k \in \mathbb{N}$, we have

$$\frac{1}{x} \sum_{n \leq x} \lambda(n + h_1) \cdots \lambda(n + h_k) = o(1). \tag{5.1}$$

The conjecture can be interpreted as stating that shifted products of the Liouville function have mean 0. Alternatively, the conjecture can be stated in the following equivalent form from which it is clearer that it is a statement about the independence of simultaneous values of the Liouville function.

5.2. **Conjecture** (Chowla's conjecture, sign pattern formulation)**.** For any $k \geq 1$, any signs $\varepsilon_1, \dots, \varepsilon_k \in \{-1, +1\}$, and any distinct shifts $h_1, \dots, h_k \in \mathbb{N}$, we have

$$\lim_{x \to \infty} \frac{1}{x} |\{n \leq x : \ \lambda(n + h_1) = \varepsilon_1, \dots, \lambda(n + h_k) = \varepsilon_k\}| = 2^{-k}.$$

To see that Conjectures 5.1 and 5.2 are indeed equivalent, one can simply substitute $\lambda(n + h_i) = \varepsilon_i(2 \cdot 1_{\lambda(n+h_i)=\varepsilon_i} - 1)$ into (5.1) and expand the product.

We remark that Conjecture 5.1 could be generalized to the assertion that

$$\frac{1}{x} \sum_{n \leq x} \lambda(a_1 n + h_1) \cdots \lambda(a_k n + h_k) = o(1), \tag{5.2}$$

whenever the non-degeneracy condition $a_i h_j \neq a_j h_i$ for $1 \leq i < j \leq k$ is fulfilled. One could also formulate Conjecture 5.1 with the Möbius function in place of the Liouville function; one can show by elementary sieve theory that such a conjecture would still follow from (5.2). Conjecture 5.2, however, takes a more complicated form for the Möbius function, as for example the events $\mu(n) = 1$, $\mu(n + 1) = 1$, $\mu(n+2) = 1$ and $\mu(n+3) = 1$ are not independent (at most three of them can hold simultaneously, since one of the values is 0).

Conjectures 5.1 and 5.2 resemble the famous Hardy-Littlewood prime tuples conjecture [34], and they can be thought of as simpler analogues of it.

5.3. **Conjecture** (Hardy-Littlewood prime tuples conjecture)**.** Let $h_1, \dots, h_k \in \mathbb{N}$ be distinct integers. Then the von Mangoldt function $\Lambda(n)$ has the correlation asymptotics

$$\frac{1}{x}\sum_{n\le x}\Lambda(n+h_1)\cdots\Lambda(n+h_k) = \mathfrak{S}(h_1,\dots,h_k) + o(1),$$

with $\mathfrak{S}(h_1,\dots,h_k)$ an effectively computable constant satisfying $\mathfrak{S}(h_1,\dots,h_k) > 0$ if and only if the polynomial $(n+h_1)\cdots(n+h_k)$ has no fixed prime divisor.

A connection between the Chowla and Hardy-Littlewood conjectures is hinted by the identity

$$\Lambda(n) = \sum_{d\mid n}\mu(d)\log\frac{n}{d},$$

which binds together the Möbius and von Mangoldt functions. Nevertheless, one would need a strong error term of the form $O((\log x)^{-A})$ on the right-hand side of (5.2) to be able to have a rigorous implication from Conjecture 5.1 to Conjecture 5.3. None of the current progress on Chowla's conjecture (for $k \ge 2$) has produced such good error terms.

In its original form, Chowla's conjecture is open for all $k \ge 2$. The simplest $k = 1$ case

$$\frac{1}{x}\sum_{n\le x}\lambda(n) = o(1)$$

can be shown to be equivalent to the prime number theorem and, more generally,

$$\frac{1}{x}\sum_{n\le x}\lambda(an+h) = o(1)$$

is equivalent to the prime number theorem in arithmetic progressions.

Despite this lack of progress on the original conjecture, starting from 2015 there has been major progress on different variants of Chowla's conjecture. Matomäki and Radziwiłł [74] proved, while showing cancellation in very short averages of multiplicative functions, that, for any $h \in \mathbb{N}$,

$$\limsup_{x\to\infty}\left|\frac{1}{x}\sum_{n\le x}\lambda(n)\lambda(n+h)\right| \le 1-\delta(h)$$

for some $\delta(h) > 0$. This was the first nontrivial progress towards the two-point Chowla conjecture (for the odd order cases, the analogous result was proved by Elliott [11]). Soon after this result, Matomäki, Radziwiłł and Tao [75] showed that

Chowla's conjecture (as well as the more general Elliott's conjecture, discussed in Chapter 6) holds on average over the shifts, in the sense that

$$\frac{1}{H(x)^k} \sum_{h_1,\dots,h_k \leq H(x)} \left| \frac{1}{x} \sum_{n \leq x} \lambda(n+h_1) \cdots \lambda(n+h_k) \right| = o(1) \tag{5.3}$$

for any $H(x) \leq x$ tending to infinity with x. If one could take $H(x)$ bounded, one would of course obtain Chowla's conjecture. The result (5.3) was generalized by Frantzikinakis [15] to averages where the shifts are given by independent multivariate polynomials.

Another interesting approximation to Chowla's conjecture is obtained by adding weights to the conjecture. The logarithmic weights $\frac{1}{n}$ are a fruitful choice, since they have the property that

$$\frac{\sum_{x/2 \leq n \leq x} \frac{1}{n}}{\sum_{n \leq x} \frac{1}{n}} = o(1),$$

or in other words that the measure of the interval $[\frac{x}{2}, x]$ is small. We will see the usefulness of this in Subsection 5.3

In this direction, Tao [98] made a breakthrough by settling the two-point case of Chowla's conjecture with logarithmic weights.

5.4. **Theorem** (Tao)**.** *For any distinct $h_1, h_2 \in \mathbb{N}$, we have*

$$\frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(n+h_1)\lambda(n+h_2)}{n} = o(1). \tag{5.4}$$

In fact, Tao proved an analogous approximation to Elliott's conjecture (see Section 6) from which (5.4) follows as a special case.

In light of the result (5.4), it is natural to study in detail the logarithmic variant of Chowla's conjecture.

5.5. **Conjecture** (The logarithmic Chowla conjecture)**.** For any $k \geq 1$ and any distinct shifts $h_1, \dots, h_k \in \mathbb{N}$, we have

$$\frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(n+h_1) \cdots \lambda(n+h_k)}{n} = o(1).$$

Tao's result [98] is the $k = 2$ case of this. In [100] and [III], Tao and the author settled Conjecture 5.5 for all odd k. Therefore, the cases $k = 1, 2, 3, 5, 7, 9 \dots$ of the conjecture are now known, whereas the even cases $k \geq 4$ remain open.

5.6. **Theorem** (Tao-T., [100], [III]). *Let $k \geq 1$ be odd and $a_1, \dots, a_k, h_1, \dots, h_k \in \mathbb{N}$. Then we have*

$$\frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(a_1 n + h_1) \cdots \lambda(a_k n + h_k)}{n} = o(1). \tag{5.5}$$

There is no need to assume a non-degeneracy condition on a_i and h_i here, since such a condition makes a difference only for even k.

Our first proof of Theorem 5.6 in [100] utilized deep results of Leibman [62] and Le [61] from ergodic theory, as well as the theory of nilsequences. On the other hand, the proof in [100] gave a general structural theorem for correlations of multiplicative functions, of which Theorem 5.6 is a special case.

The second proof, which we present in [III], proceeds along rather different lines, since after applying the so-called *entropy decrement argument* from [98, Section 3], we do not employ ergodic theory machinery, but use combinatorial results instead. The proof via this method turns out to be both shorter and simpler; in the proof given in [100], the case of the Liouville function was not significantly easier than the case of arbitrary multiplicative functions. Using this combinatorial proof, it would in addition be possible to obtain quantitative error bounds for the right hand side of (5.5); however, these error terms were not analyzed in [III], due to the fact that the error bounds would be very weak[14].

Since Chowla's conjecture can be stated as a claim about the sign patterns of the Liouville function, it is natural that Theorem 5.6 also implies something about sign patterns. We showed in [100] that Theorem 5.6, together with the two-point result and some additional considerations, gives the following.

5.7. **Theorem** (Tao-T., [100], [III]). *Let $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 \in \{-1, +1\}^4$. Then we have*

$$\lim_{x \to \infty} \frac{1}{x} |\{n \leq x : \ \lambda(n+1) = \varepsilon_1, \lambda(n+2) = \varepsilon_2, \lambda(n+3) = \varepsilon_3\}| = \frac{1}{8}$$

and

$$\liminf_{x \to \infty} \frac{1}{x} |\{n \leq x : \ \lambda(n+1) = \varepsilon_1, \lambda(n+2) = \varepsilon_2, \lambda(n+3) = \varepsilon_3, \lambda(n+4) = \varepsilon_4\}| > 0.$$

We also proved the analogous results for the Möbius function[15]. These improve the result of Matomäki, Radziwiłł and Tao [76] on sign patterns of length 3, as well as Tao's result on sign patterns in [98, Corollary 1.7].

[14] For the $k = 2$ case of (5.5), Tao's method [98] gives an error term of the form $O((\log\log\log x)^{-c})$ for some $c > 0$. For $k \geq 3$, we expect even worse error terms.

[15] In the case of $\mu(n)$, we of course need to exclude from the four-point result those sign patterns $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) \in \{-1, 0, +1\}^4$ which cannot occur for trivial reasons, and in the three point result the density of the set is some function of $\varepsilon_i \in \{-1, 0, +1\}$ instead of $\frac{1}{8}$.

5.2. **Connections to other conjectures**

Chowla's conjecture can be viewed as one instance of the *Möbius randomness law* from [52, p. 338], a heuristic stating that the Möbius function (or the Liouville function) should behave randomly. Another manifestation of this heuristic is a conjecture of Sarnak [89], which states that $\mu(n)$ does not correlate with any bounded sequence of "low complexity". The complexity of a sequence $a : \mathbb{N} \to \mathbb{C}$ is measured in terms of its *topological entropy*, which is the infimum of all $\sigma > 0$ such that sets of the form $\{a(n), a(n+1), \dots, a(n+k-1)\} \subset \mathbb{C}^k$ can be covered with $\leq \exp((\sigma + o(1))k)$ balls of any fixed radius. With this definition, Sarnak's conjecture takes the form below.

5.8. **Conjecture** (Sarnak)**.** Let $a : \mathbb{N} \to \mathbb{C}$ be a bounded sequence of topological entropy 0. Then we have

$$\frac{1}{x}\sum_{n\leq x}\mu(n)a(n) = o(1).$$

Sarnak's conjecture has been extensively studied in the ergodic theory literature, and many important special cases have been verified; see [14] for a survey. In the ergodic theory literature, one usually assumes in Conjecture 5.8 the (equivalent) condition for the sequence a that it can be written as $a(n) = F(T^n X)$ for (X, T) a *topological dynamical system* of zero topological entropy and $F : X \to \mathbb{C}$ continuous. Here we will not work with the dynamical systems definition, and instead refer to [89] for its details.

It was already observed by Sarnak that his conjecture would follow from Chowla's conjecture. In [99], Tao strengthened the connection between the two conjectures by showing that their logarithmic forms are equivalent (that is, Conjecture 5.5 is equivalent to Conjecture 5.8 with $(1/(\log x))\sum_{n\leq x}\mu(n)a(n)/n$ in place of $(1/x)\sum_{n\leq x}\mu(n)a(n)$). He also showed that both of these conjectures are equivalent to the yet unproved "logarithmic local Gowers uniformity of the Liouville function", which can be thought of as a short exponential sum estimate for the Liouville function and contains as the simplest case the Matomäki-Radziwiłł theorem [74]. Further works that lie at the intersection of the Sarnak and Chowla conjectures include [10] and [17]. In the latter, Frantzikinakis and Host verify many new cases of the logarithmic Sarnak conjecture, and as a byproduct obtain also a "minor arc" Chowla-type result

$$\frac{1}{\log x}\sum_{n\leq x}\lambda(n+h_1)\cdots\lambda(n+h_k)\frac{e(\alpha n)}{n} = o(1)$$

for any $k \geq 1$ and any fixed irrational α. In Article [III], however, we do not make progress on Sarnak's conjecture, since it is the even order cases of Chowla's conjecture that are needed in the proof that Chowla's conjecture implies Sarnak's. It seems therefore that the even order cases lie deeper, and indeed in [100, Remark

1.7] we observed that the $2k$-point case of the logarithmic Chowla conjecture (with dilations as in (5.2)) implies the $(k+1)$-point case.

5.3. **Proof ideas**

The proof of the odd order cases of the logarithmic Chowla conjecture in [III] starts with the averaging over small primes and entropy decrement arguments devised in [98] to deal with the two-point case of the conjecture. The averaging argument enables us to replace a correlation average over n with a double average over n and p for p belonging to some small scale, thus offering more flexibility. More precisely, if we define

$$f_x(a) := \frac{1}{\log x} \sum_{n \le x} \frac{\lambda(n+a) \cdots \lambda(n+ak)}{n}$$

(assuming for simplicity that $a_1 = \cdots = a_k = 1$ and $h_j = j$ in Theorem 5.6), then the multiplicativity property $\lambda(pn) = -\lambda(n)$ for all primes p allows us to write

$$f_x(1) = -\frac{1}{\log x} \sum_{p \le n' \le px} \frac{\lambda(n'+p) \cdots \lambda(n'+pk)}{n'} p 1_{p|n'} + o(1) \tag{5.6}$$

for any prime p and for odd k (for even k we would have a $+$ sign). Taking averages over p, we get the identity

$$f_x(1) = -\frac{m}{2^m} \sum_{2^m \le p < 2^{m+1}} \frac{1}{\log x} \sum_{p \le n' \le px} \frac{\lambda(n'+p) \cdots \lambda(n'+pk)}{n'} p 1_{p|n'} + O(\varepsilon)$$

for $\varepsilon^{-1} \le m \le \log\log x$. Since logarithmic averages are slowly varying, we can replace the average over $p \le n' \le px$ with an average over $n \le x$ (this is the benefit of logarithmic averaging). Thus we have

$$f_x(1) = -\frac{m}{2^m} \sum_{2^m \le p < 2^{m+1}} \frac{1}{\log x} \sum_{n \le x} \frac{\lambda(n+p) \cdots \lambda(n+pk)}{n} p 1_{p|n} + O(\varepsilon). \tag{5.7}$$

for $\varepsilon^{-1} \le m \le \log\log x$.

We wish to replace the factor $p1_{p|n'}$ with its average value $1 + O(\varepsilon)$ in order to get a bilinear sum over n' and p, for which there are many tools available. This is enabled by Tao's entropy decrement argument [98] (with refinements in [96], [100], [III]), which draws ideas from probability and information theory to show that this replacement can be done for "almost all" scales m in (5.7).

We elaborate on this part of the argument. Firstly, by using the approximate translation invariance of averages, (5.7) becomes

$$f_x(1) = -\frac{1}{\log x} \sum_{n \le x} \frac{m}{2^{2m}} \sum_{2^m \le p < 2^{m+1}} \sum_{j \le 2^m} \frac{\lambda(n+j+p) \cdots \lambda(n+j+pk)}{n} p 1_{p|n+j} + O(\varepsilon) \tag{5.8}$$

which is a more convenient form to work with. The task is then to show that, for most choices of the scale m, the sign pattern $\mathbf{X}_m(n) := (\lambda(n), \lambda(n+1), \dots, \lambda(n + 2^{m+2} - 1))$ and the divisibility conditions $\mathbf{Y}_m(n) := (n \mod p)_{2^m \leq p < 2^{m+1}}$ behave essentially independently (with respect to the logarithmic probability on $[1, x]$).

In the language of information theory, we thus want to show that if $\mathbf{X}_m$ and $\mathbf{Y}_m$ are interpreted as random variables, the *entropy*[16] $\mathbf{H}(\mathbf{Y}_m)$ is essentially the same as the *conditional entropy* $\mathbf{H}(\mathbf{Y}_m|\mathbf{X}_m)$ (the two entropies are equal if $\mathbf{Y}_m$ and $\mathbf{X}_m$ are independent, so a small difference between them amounts to near independence). In other words, we want the *mutual information*

$$\mathbf{I}(\mathbf{X}_m, \mathbf{Y}_m) := \mathbf{H}(\mathbf{Y}_m) - \mathbf{H}(\mathbf{Y}_m|\mathbf{X}_m) \tag{5.9}$$

to be small for most m; more precisely, it should be of size $\varepsilon^{10} \cdot 2^m/m$, whereas the trivial upper bound is $\leq \mathbf{H}(\mathbf{Y}_m) \ll 2^m$. As mentioned above, mutual information reflects how close two random variables are to being independent (in particular, the information is maximal when one of the two random variables is a deterministic function of the other).

By applying inequalities from information theory, and an insightful pigeonholing argument, Tao showed in [98] that one can indeed bound (5.9) by $\leq \varepsilon^{10} \cdot 2^m/m$, not for all scales m, but for infinitely many m. In [III, Section 3], we need a refinement of this, to the effect that if $\mathcal{M}(x, \varepsilon)$ is the set of scales $m \leq \log\log x$ for which (5.9) is $> \varepsilon^{10} 2^m/m$, then

$$\sum_{m \in \mathcal{M}(x,\varepsilon)} \frac{1}{m} \ll \varepsilon^{-20},$$

say. In paricular, the set of suitable scales has logarithmic density 1. We refer to [96], [III, Proposition 4.3] for the details[17].

After applying the entropy decrement argument, we know that we can replace in (5.7) the factor $p1_{p|n}$ with $1 + O(\varepsilon)$ for all $m \leq \log\log x$ outside a set whose sum of reciprocals over $[1, \log\log x]$ is $\ll \varepsilon^{-20}$. In particular, we can average logarithmically over different scales m to reach

$$f_x(1) = -\frac{1}{\log_2 H_2 - \log_2 H_1} \sum_{H_1 \leq p \leq H_2} \frac{1}{p} \frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(n+p) \cdots \lambda(n+kp)}{n} + O(\varepsilon) \tag{5.10}$$

for $H_j = H_j(x)$ tending to infinity slowly enough and $H_1(x)$ growing slowly enough in terms of $H_2(x)$. Here $\log_2 x$ is the second iterate of $\log x$.

[16] For the definitions of entropy and other related notions from information theory, see Section 1.

[17] For technical reasons, those works deal with a more general notion of information, namely *conditional mutual information*.

We can apply the same argument again to the right-hand side of (5.10), finding

$$f_x(1) = +\frac{1}{\log_2 H_2 - \log_2 H_1} \sum_{H_1 \leq p_1 \leq H_2} \frac{1}{p_1} \frac{1}{\log_2 H_4 - \log_2 H_3} \sum_{H_3 \leq p_2 \leq H_4} \frac{1}{p_2} \frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(n+p_1p_2)\cdots\lambda(n+kp_1p_2)}{n} + O(\varepsilon), \tag{5.11}$$

where $H_1 < H_2 < H_3 < H_4$ and $H_j(x)$ grows slowly enough in terms of $H_{j+1}(x)$, and $H_4(x)$ tends to infinity slowly. Crucially, we have a + sign in (5.11) and a − sign in (5.10); this allows us to break the symmetry of the correlations.

We can easily replace the averages over primes with averages over the integers weighted by the von Mangoldt function $\Lambda(d)$, so (5.10) (with H_1 and H_2 replaced with H_3 and H_4) and (5.11) take the forms

$$f_x(1) = -\frac{1}{\log_2 H_4 - \log_2 H_3} \sum_{H_3 \leq d \leq H_4} \frac{\Lambda(d)}{d \log d} \frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(n+d)\cdots\lambda(n+kd)}{n} + O(\varepsilon) \tag{5.12}$$

and

$$f_x(1) = +\frac{1}{\log_2 H_2 - \log_2 H_1} \sum_{H_1 \leq d_1 \leq H_2} \frac{\Lambda(d_1)}{d_1 \log d_1} \frac{1}{\log_2 H_4 - \log_2 H_3} \sum_{H_3 \leq d_2 \leq H_4} \frac{\Lambda(d_2)}{d_2 \log d_2} \frac{1}{\log x} \sum_{n \leq x} \frac{\lambda(n+d_1d_2)\cdots\lambda(n+kd_1d_2)}{n} + O(\varepsilon), \tag{5.13}$$

respectively.

We now encounter multilinear averages of the form

$$\frac{1}{N^2} \sum_{d \leq N} \sum_{n \leq N} \theta(d) f_1(n+d) \cdots f_k(n+kd), \tag{5.14}$$

where $f_1, \ldots, f_k : \mathbb{N} \to \mathbb{C}$ are some functions with $|f_i| \leq 1$ and $\theta : \mathbb{N} \to \mathbb{C}$ is some other function (in this case a normalized version of $\Lambda(d)$). The expression (5.14) thus counts patterns of the form $(d, n+d, \ldots, n+kd)$ with weights. Such averages have been widely studied both in the additive combinatorics and the ergodic theory literature (see for instance [101, Chapter 11]), and by a version of the so-called generalized von Neumann theorem [III, Lemma 5.2], it turns out that one has the

bound[18]

$$\left| \frac{1}{N^2} \sum_{d \leq N} \sum_{n \leq N} \theta(d) f_1(n+d) \cdots f_k(n+kd) \right| \leq C_k \|\theta\|_{U^k[N]} + o(1),$$

where $\|\theta\|_{U^k[N]}$ is the U^k *Gowers norm* of θ on $[1, N]$ (see [101, Chapter 11]). Thus, analyzing (5.12) and (5.13) has been reduced to understanding the Gowers norm of $\Lambda(Wn + b) - 1$, where $W = \prod_{p \leq w} p$, $(b, W) = 1$, and w is a large constant.

It is known that the W-tricked von Mangoldt function has negligible Gowers norms; this was proved by Green, Tao and Ziegler in a series of breakthroughs [28],[29],[30], [31]. Therefore, we can remove the von Mangoldt function weight both in the average (5.12) and the average (5.13), after splitting the sums into residue classes (mod W). This leads, after some considerations, to

(5.15)

$$\begin{aligned} f_x(1) &= \frac{W}{\varphi(W)} \frac{1}{\log_2 H_4 - \log_2 H_3} \sum_{\substack{H_3 \leq d \leq H_4 \\ (d,W)=1}} \frac{1}{d \log d \log x} \sum_{n \leq x} \frac{\lambda(n+d) \cdots \lambda(n+dk)}{n} + O(\varepsilon) \\ &= -f_x(1) + O(\varepsilon). \end{aligned}$$

Importantly, $f_x(1)$ appears with different signs in (5.15), so $f_x(1) = O(\varepsilon)$, after which we can send $\varepsilon \to 0$. This concludes the sketch of the proof; for the full proof, see [III].

[18] As is shown for example in [101, Chapter 11], the weighted arithmetic progression patterns $(n+d, \ldots, n+kd)$ are controlled by the U^{k-1} Gowers norm, but the pattern $(d, n+d, \ldots, n+kd)$ in (5.14) has "complexity" one higher, and should thus be controlled by the U^k Gowers norm.

6. Binary correlations of multiplicative functions and applications

Article [IV] is concerned with binary correlations of multiplicative functions with logarithmic averaging. Before stating the results of that article, we review what is known and conjectured on correlations of multiplicative functions.

6.1. **Correlations of multiplicative functions**

A function $g : \mathbb{N} \to \mathbb{C}$ is called multiplicative if it satisfies $g(mn) = g(m)g(n)$ whenever $m, n \in \mathbb{N}$ are coprime. In what follows, we will restrict attention to 1-bounded multiplicative functions, that is, multiplicative functions taking values in the unit disk $\mathbb{D} := \{z \in \mathbb{C} : |z| \leq 1\}$, since much less is known about the behavior of unbounded multiplicative functions.

A fundamental notion in multiplicative number theory is the *pretentious distance* $\mathbb{D}(f, g; x)$ between two multiplicative functions $f, g : \mathbb{N} \to \mathbb{D}$, introduced by Granville and Soundararajan [24]. This quantity is defined as

$$\mathbb{D}(f, g; x) := \left(\sum_{p \leq x} \frac{1 - \mathrm{Re}(f(p)\overline{g(p)})}{p} \right)^{1/2}, \tag{6.1}$$

and it is a pseudometric[19] and, heuristically, if f and g "behave similarly" (when it comes to their mean values or correlations), then the distance between them is "small".

The Dirichlet characters $\chi(n)$ and the Archimedean characters n^{it} are important classes of 1-bounded multiplicative functions, and although their complexity is relatively low in the sense that $\chi(n)$ is periodic and n^{it} is slowly varying, one usually wants to exclude these functions when studying mean values or correlations of multiplicative functions, as these two classes of functions exhibit different behavior from other functions in this context. One thus classifies 1-bounded multiplicative functions as either

(i) pretentious, in the sense that $\mathbb{D}(g, \chi(n)n^{it}; \infty) < \infty$ for some Dirichlet character χ and some $t \in \mathbb{R}$,

or

(ii) non-pretentious, in the sense that $\mathbb{D}(g, \chi(n)n^{it}; \infty) = \infty$ for all Dirichlet characters χ and all $t \in \mathbb{R}$.

By the zero-free region for the Dirichlet L-functions, the Liouville function $\lambda(n)$ from Section 5 is non-pretentious, whereas any multiplicative function $f : \mathbb{N} \to \mathbb{D}$ with $f(p) \neq 1$ for only finitely many primes p is an example of a pretentious function (one can take $\chi \equiv 1$, $t = 0$ in (i)).

[19] This means that it satisfies the axioms of a metric, excluding the property that $\mathrm{d}(x, y) = 0 \Rightarrow x = y$.

The mean values

$$\frac{1}{x}\sum_{n\le x} g(n) \tag{6.2}$$

of multiplicative functions are connected to many topics of interest in multiplicative number theory, including the prime number theorem and its generalizations, sieve methods, and probabilistic number theory. The asymptotics of these mean values are described by a theorem of Halász [33] from the 1960s (generalizing a theorem of Wirsing [113] from the real-valued case), and the result demonstrates the need for distinguishing pretentious and non-pretentious functions from each other.

6.1. **Theorem** (Halász)**.** *Let $g : \mathbb{N} \to \mathbb{D}$ be a 1-bounded multiplicative function. Then*
(i) If there exists $t \in \mathbb{R}$ such that $\mathbb{D}(g, n^{it}; \infty) < \infty$, we have

$$\frac{1}{x}\sum_{n\le x} g(n) = (1+o(1))\frac{x^{it}}{1+it}\prod_p \left(1-\frac{1}{p}\right)\left(1+\frac{g(p)}{p^{1+it}}+\frac{g(p^2)}{p^{2(1+it)}}+\cdots\right).$$

(ii)If no such t exists, we have

$$\frac{1}{x}\sum_{n\le x} g(n) = o(1).$$

For a proof of the theorem, see [102, Section III.4]. Among other things, Theorem 6.1 implies that if $g : \mathbb{N} \to [-1, 1]$ is real-valued, then the mean value of g always exists (that is, (6.2) converges as $x \to \infty$).

We wish to understand the much more general *correlation averages* of bounded multiplicative functions $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$, defined as

$$\frac{1}{x}\sum_{n\le x} g_1(n+h_1)\cdots g_k(n+h_k), \tag{6.3}$$

where $h_1, \dots, h_k \in \mathbb{N}$ are fixed, distinct integers. These correlations have a number of applications; most notably, in the case of the Liouville function showing that the correlations are small reduce to the celebrated Chowla conjecture, discussed in Section 5 and in particular, gives information on the sign patterns of the Liouville function, studied in [76], [98], [100], [17]. In a very different and surprising direction, Tao [97] used his breakthrough on two-point correlations to settle the Erdős discrepancy problem [13] in combinatorics. There are further applications to discrepancy of multiplicative functions in [58], rigidity theorems for multiplicative functions in [60], and to distribution laws of additive functions in [12]. In article [IV], we give further applications, discussed in Subsection 6.2.

The central conjecture pertaining to (6.3) is that of Elliott [11], [12] from the 1990s. His conjecture states that, in the case where at least one of $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$ is

non-pretentious, distinct shifts of the functions g_j should behave independently of each other.

6.2. **Conjecture** (Elliott)**.** Let $k \geq 1$ and let $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions and $h_1, \dots, h_k \in \mathbb{N}$ distinct shifts. Then we have

$$\frac{1}{x} \sum_{n \leq x} g_1(n + h_1) \cdots g_k(n + h_k) = o(1) \tag{6.4}$$

unless for all $1 \leq j \leq k$ there exists a Dirichlet character χ_j such that

$$\liminf_{x \to \infty} \inf_{|t| \leq x} \mathbb{D}(g_j, \chi_j(n) n^{it}; x) < \infty.$$

The formulation above takes into account the observation in [75, Appendix B] that the original conjecture in [11], [12] has to be slightly modified in the complex-valued case. As in the case of Halász's theorem (Theorem 6.1), the property (6.4) often fails in the pretentious case; for example

$$\frac{1}{x} \sum_{n \leq x} n^{it} (n+1)^{-it} = 1 + o(1) \quad \text{and} \quad \frac{1}{x} \sum_{n \leq x} \chi_3(n) \chi_3(n+1) = -\frac{1}{3} + o(1),$$

where χ_3 is the real non-principal Dirichlet character modulo 3. On the other hand, a theorem of Klurman [58, Theorem 1.3] gives a formula for (6.3) in the case where $g_1, \dots, g_k$ are fixed pretentious functions.

In the form presented above, Conjecture 6.2 is open for all $k \geq 2$, whereas the $k = 1$ case follows from Theorem 6.1. However, several variants of (6.4) have been established in the last few years. In particular, Matomäki, Radziwiłł and Tao [75] showed that Elliott's conjecture holds on average over the shifts $h_1, \dots, h_k$. Tao [98] made another breakthrough by proving the binary case $k = 2$ of Elliott's conjecture with logarithmic averaging.

6.3. **Theorem** (Tao)**.** *Let $g_1, g_2 : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions and $h_1 \neq h_2$ natural numbers. Then we have*

$$\frac{1}{\log x} \sum_{n \leq x} \frac{g_1(n + h_1) g_2(n + h_2)}{n} = o(1)$$

unless for both $j \in \{1, 2\}$ there exists a Dirichlet character χ_j such that

$$\liminf_{x \to \infty} \inf_{|t| \leq x} \mathbb{D}(g_j, \chi_j(n) n^{it}; x) < \infty.$$

For many purposes, this logarithmic averaging is acceptable; see [97], [58], [60] for some applications. In [100] we generalized Theorem 6.3 to the higher order cases $k \geq 3$, under an additional non-pretentiousness assumption on the product of the functions involved.

6.4. **Theorem** (Tao-T., [100]). *Let $k \geq 1$ and let $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions and $h_1, \dots, h_k \in \mathbb{N}$ natural numbers. Then we have*

$$\frac{1}{\log x} \sum_{n \leq x} \frac{g_1(n+h_1) \cdots g_k(n+h_k)}{n} = o(1)$$

unless there exists a Dirichlet character χ for which the product $g_1 \cdots g_k$ weakly pretends *to be χ, in the sense that $\mathbb{D}(g_1 \cdots g_k, \chi; x)^2 = o(\log \log x)$.*

We applied this result to settle the odd order cases of the logarithmically averaged Chowla conjecture; see Section 5.

Let us also mention a different line of study to Elliott's conjecture, namely two-dimensional variants of it. The two-dimensional version of Elliott's conjecture states that

$$\frac{1}{x^2} \sum_{d \leq x} \sum_{n \leq x} g_1(n+dh_1) \cdots g_k(n+dh_k) = o(1),$$

given the assumptions of Conjecture 6.2. This was proved by Frantzikinakis and Host in [16], and further works on two-dimensional correlations include those of Matthiesen [78] and Klurman–Mangerel [59].

6.2. **The result and its applications**

In Article [IV], we generalize Tao's result on the binary logarithmic Elliott conjecture, but in a different direction than in [100], where higher order correlations were considered. Namely, we show that for a large class of real-valued multiplicative functions $g_1, g_2 : \mathbb{N} \to [-1, 1]$ we can give an asymptotic formula for their correlation (and typically the asymptotic formula has a nonzero main term). The class of functions we consider is defined as follows.

6.5. **Definition** (Uniformity assumption). Let $x \geq 1$, $1 \leq Q \leq x$ and $\eta > 0$. For a function $g : \mathbb{N} \to \mathbb{D}$, denote $g \in \mathcal{U}(x, Q, \eta)$ if we have the estimate

$$\left| \frac{1}{x} \sum_{\substack{x \leq n \leq 2x \\ n \equiv a \pmod q}} g(n) - \frac{1}{qx} \sum_{x \leq n \leq 2x} g(n) \right| \leq \frac{\eta}{q} \quad \text{for all} \quad 1 \leq a \leq q \leq Q.$$

From Halász's theorem we see (as was observed in [IV, Remark 1.3]) that if $g : \mathbb{N} \to \mathbb{D}$ is non-pretentious in the sense that $\inf_{|t| \leq x} \mathbb{D}(g, \chi(n) n^{it}; x) \geq \varepsilon^{-10}$ for all Dirichlet characters χ of modulus $\leq \varepsilon^{-10}$ (and with $\varepsilon > 0$ small), then $g \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$ for $x \geq x_0(\varepsilon)$. This means that the class of functions in Definition 6.5 is larger than the class of real-valued functions considered in Conjecture 6.2 or in [98]. Very importantly, Definition 6.5 allows the multiplicative function g to depend on the summation length x, as will be the case in our applications. One can for example show that if $\alpha \in (0, 1)$ is given, then the indicator of smooth numbers[20] $g(n) := 1_{n \text{ is } x^\alpha\text{−smooth}}$ is a multiplicative function satisfying $g \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$ for $x \geq x_0(\varepsilon, \alpha)$, although g

[20] We say that n is y–smooth (also called y-friable) if n has no prime factor larger than y.

pretends to be 1 on the interval $[1, x]$.

The main result in [IV] then states that if $g_1, g_2 : \mathbb{N} \to [-1, 1]$ are two multiplicative functions, possibly depending on x, and g_1 is uniformly distributed at scale x in the sense of Definition 6.5, then the shifts of g_1 and g_2 are independent of each other.

6.6. **Theorem** (Article [IV]). *Let a small real number $\varepsilon > 0$, a fixed integer shift $h \neq 0$, and a function $\omega : \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 1}$ with $1 \leq \omega(X) \leq \log(3X)$ and $\omega(X) \xrightarrow{X\to\infty} \infty$ be given. Let $x \geq x_0(\varepsilon, h, \omega)$. Then, for any multiplicative functions $g_1, g_2 : \mathbb{N} \to [-1, 1]$ with $g_1 \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$, we have*

(6.5)
$$\frac{1}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \leq n \leq x} \frac{g_1(n) g_2(n+h)}{n} = \left(\frac{1}{x} \sum_{x \leq n \leq 2x} g_1(n)\right) \left(\frac{1}{x} \sum_{x \leq n \leq 2x} g_2(n)\right) + o_{\varepsilon\to 0}(1).$$

Here $o_{\varepsilon\to 0}(1)$ denotes some function that tends uniformly to 0 as $\varepsilon \to 0$. Note also that even if $h < 0$ in Theorem 6.6, $g_2(n + h)$ is still well-defined, as the function $x_0(\cdot)$ above can be chosen to be large enough, so that $\frac{x}{\omega(x)} > h$ for $x \geq x_0(\varepsilon, h, \omega)$.

As mentioned, Theorem 6.6 contains the real-valued case of Tao's result [98] and shows that g_1 and g_2 are discorrelated in the sense that the correlation of g_1 and g_2 is the product of their mean values. In the complex-valued case, Theorem 6.6 does not hold as such, as is seen by taking g_1 and g_2 to be suitable Archimedean characters (such as $g_1(n) = g_2(n) = n^{it}$ with $t \neq 0$). It would nevertheless be possible to generalize it to the case where g_1 and g_2 take values in roots of unity of a fixed order.

Theorem 6.6 could also be generalized to the case of functions that are uniformly distributed only in coprime residue classes, instead of all residue classes as in Definition 6.5. However, this would significantly complicate the main term on the right-hand side of (6.5) and make it dependent on the shift h (as is seen by considering the simple example $g_1(n) = g_2(n) = 1_{n\equiv 1 \pmod 2}$). Therefore, we do not pursue this generalization.

The utility of Theorem 6.6 lies in its uniformity over the choice of the functions g_1, g_2. For example, the theorem can be applied to the interesting cases
(i) $g(n) = 1_{n \text{ is } x^\alpha\text{−smooth}}$
and
(ii) $g(n) = \chi_Q(n)$ where χ_Q is a real non-principal character $\pmod Q$ with $Q = Q(x) \leq x^{4-\varepsilon}$ cube-free[21] (so Q can be very large in terms of x).
In the case of (i), the result of [98] is clearly not applicable, and also in case (ii), for

[21] We say that n is cube-free if $p^3 \nmid n$ for all primes p.

all we know, it could be that[22] $\mathbb{D}(\chi_Q, 1; x) \ll 1$, in which case [98] does not apply. The range $Q \leq x^{4-\varepsilon}$ in (ii) is the same as in a celebrated estimate of Burgess [4], a special case of which implies that

$$\sum_{n \leq x} \chi_Q(n) = o(x), \tag{6.6}$$

uniformly for $Q \leq x^{4-\varepsilon}$ cube-free. We note that (6.6) is not enough to exclude the (unlikely) scenario that $\chi_Q(p) = 1$ for all $p \leq x^{\varepsilon}$.

We mentioned earlier the result of Klurman [58, Theorem 1.3] that gives an asymptotic formula for the correlations (6.3) in the case where all the functions $g_1, \dots, g_k$ are fixed and pretentious. Nevertheless, this asymptotic cannot be applied to (i) or (ii), since both of these two functions depend on x in a very essential way, whereas in [58] it is necessary that the functions are (almost) independent of x (in fact, the asymptotic formula in [58, Theorem 1.3] does not predict the correct asymptotic for the autocorrelations of the functions in (i) or (ii)).

As the examples (i) and (ii) indicate, Theorem 6.6 should yield new results on consecutive smooth (friable) numbers and quadratic residues. We confirm this in [IV]. Define the function $P^+(n)$ that outputs the largest prime factor of $n \in \mathbb{N}$, with the convention that $P^+(1) = 1$. Then n is y–smooth if and only if $P^+(n) \leq y$. The distribution of smooth numbers is well-understood (see [45] for a survey), but much more elusive is the simultaneous distribution of two or more consecutive smooth numbers. Related to this, Erdős and Turán [94] posed the following problem.

6.7. **Conjecture.** The asymptotic density[23] of the set

$$\{n \in \mathbb{N} : \ P^+(n) < P^+(n+1)\}$$

exists and equals $\frac{1}{2}$.

By applying Theorem 6.6 to the indicator function of x^{α}–smooth numbers for various α, and doing some additional deductions, we were able to prove a logarithmic variant of this conjecture.

6.8. **Theorem** (Article [IV])**.** *The logarithmic density*[24] *of the set*

$$\{n \in \mathbb{N} : \ P^+(n) < P^+(n+1)\}$$

exists and equals $\frac{1}{2}$.

---

[22]It is a well-known conjecture, due to Vinogradov, that, for any $\varepsilon > 0$ and any $Q \geq Q_0(\varepsilon)$, there is a quadratic nonresidue modulo Q on $[1, Q^{\varepsilon}]$. But this is open, and if it fails, then χ_Q pretends to be 1 on $[1, Q]$.

[23]We define the asymptotic density of $A \subset \mathbb{N}$ as $\lim_{x\to\infty} \frac{1}{x} \sum_{n\leq x, n\in A} 1$, whenever the limit exists. The upper and lower asymptotic densities are defined analogously with lim sup and lim inf.

[24]We define the logarithmic density of $A \subset \mathbb{N}$ as $\lim_{x\to\infty} \frac{1}{\log x} \sum_{n\leq x, n\in A} \frac{1}{n}$, whenever the limit exists. The upper and lower logarithmic densities are defined analogously with lim sup and lim inf.

We mention in passing that we also proved in [IV] some generalizations of Theorem 6.8, including a logarithmic version of a conjecture of Erdős and Pomerance ([IV, Theorem 1.12]).

We can also say something about *asymptotic* densities of sets related to two consecutive smooth numbers. In this case, we are not able to determine the precise value of the density, but we can at least show that the lower density is positive.

6.9. **Theorem** (Article [IV]). *Let $a, b, c, d \in (0,1)$ be real numbers with $a < b$ and $c < d$. Then the set*

$$\{n \in \mathbb{N} : \ n^a < P^+(n) < n^b, \ n^c < P^+(n+1) < n^d\}$$

has positive asymptotic lower density.

This theorem implies a result of Hildebrand [44], which is the special case $(a, b) = (c, d)$ (Hildebrand also considers more general "stable sets", in addition to sets of smooth numbers). Our theorem also reproves a recent result of Wang [110, Théorème 2] on the truncated largest prime factor $P_y^+(n) := \max\{p \le y : \ p \mid n\}$ at two consecutive integers. This result states that, if $a \in (0,1)$ is fixed, then $P_{x^a}^+(n) < P_{x^a}^+(n+1)$ for a positive lower density of integers $n \le x$.

Another source for applications of Theorem 6.6 is the collection of real non-principal Dirichlet characters whose modulus $Q(x)$ grows moderately fast in terms of x. A fundamental result of Burgess [4] from 1963 says, among other things, that if χ_Q is a non-principal Dirichlet character of cube-free[25] modulus $Q = Q(x)$, and $\varepsilon > 0$ is fixed, then

$$\sum_{n \le x} \chi_Q(n) = o(x), \tag{6.7}$$

uniformly for $Q \le x^{4-\varepsilon}$. The range of Q here is still the best one known up to the ε in the exponent.

By employing the Burgess bound (6.7), we can show that if χ_Q is as above with $Q \le x^{4-\varepsilon}$ cube-free, then the uniformity assumption $\chi_Q \in \mathcal{U}(x, \eta^{-1}, \eta)$ holds for $x \ge x_0(\eta, \varepsilon)$; see [IV, Section 4]. Therefore, Theorem 6.6 implies a result on the sums of χ_Q along reducible quadratics $n(n+h)$.

6.10. **Theorem** (Article [IV]). *Let a small number $\varepsilon > 0$, a fixed integer $h \neq 0$, and a function $1 \le \omega(X) \le \log(3X)$ tending to infinity be given. For $x \ge x_0(\varepsilon, h, \omega)$, let $Q = Q(x) \le x^{4-\varepsilon}$ be a cube-free natural number with $Q(x) \xrightarrow{x \to \infty} \infty$. Then, the real primitive Dirichlet character χ_Q modulo Q satisfies*

$$\frac{1}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \le n \le x} \frac{\chi_Q(n(n+h))}{n} = o(1). \tag{6.8}$$

[25] We say that Q is cube-free if it is not divisible by the cube of any prime.

Moreover, if Q is as before and QNR denotes quadratic nonresidue[26]*, we have*

$$\frac{1}{\log x} \sum_{\substack{n\le x \\ n,\, n+1 \text{ QNR } (\mathrm{mod}\ Q)}} \frac{1}{n} = \frac{1}{4}\prod_{p\mid Q}\left(1-\frac{2}{p}\right) + o(1) \tag{6.9}$$

and

$$\frac{1}{x}|\{n\le x:\ n \text{ and } n+1 \text{ QNR } \ (\mathrm{mod}\ Q)\}| \gg \prod_{p\mid Q}\left(1-\frac{2}{p}\right). \tag{6.10}$$

We remark that the well-known Weil bound [52, Theorem 11.23] for character sums would give, for prime values of Q, the estimate (6.8) only in the smaller range $Q = o(\frac{x^2}{\log x})$.

Lastly, we employ Theorem 6.6 to study the number of large prime factors of consecutive integers. For $y \ge 1$, define the truncated count of prime factors as $\omega_{>y}(n) := |\{p > y :\ p \mid n\}|$. It is natural to conjecture that the numbers of large prime factors (say $> n^{\varepsilon}$) of n and $n+1$ are independent. Choosing in Theorem 6.6 multiplicative functions of the form $z^{\omega_{>x^a}(n)}$ with $z \in [-1,1]$, and using a generating function argument, we show that this independence property indeed holds, at least in the logarithmic sense.

6.11. **Theorem** (Article [IV])**.** *Let $a, b \in (0,1)$ be real numbers and $0 \le k < \frac{1}{a}$, $0 \le \ell < \frac{1}{b}$ integers. Then, if $\delta(\cdot)$ stands for logarithmic density, we have*

$$\begin{aligned}&\delta(\{n\in\mathbb{N}:\ \omega_{>n^a}(n)=k,\ \omega_{>n^b}(n+1)=\ell\})\\ &=\delta(\{n\in\mathbb{N}:\ \omega_{>n^a}(n)=k\})\cdot\delta(\{n\in\mathbb{N}:\ \omega_{>n^b}(n)=\ell\}).\end{aligned}$$

Moreover, the set $\{n\in\mathbb{N}:\ \omega_{>n^a}(n)=k,\ \omega_{>n^b}(n+1)=\ell\}$ has positive asymptotic lower density.

Theorem 6.11 in a sense complements the result of Daboussi–Sárközy [9] and Mangerel [68], which states that if $\omega_{<y}(n) = |\{p < y :\ p \mid n\}|$ is the count of the *small* prime factors of n, then we have the independence of small primes property

$$\frac{1}{x}\sum_{n\le x}(-1)^{\omega_{<x^{\varepsilon}}(n)}(-1)^{\omega_{<x^{\varepsilon}}(n+1)} = o_{\varepsilon\to 0}(1). \tag{6.11}$$

In comparison, Theorem 6.11 implies among other things the independence of large primes property

$$\frac{1}{\log x}\sum_{n\le x}\frac{(-1)^{\omega_{>x^{\varepsilon}}(n)}(-1)^{\omega_{>x^{\varepsilon}}(n+1)}}{n} = o_{\varepsilon\to 0}(1). \tag{6.12}$$

The methods used to prove (6.11) and (6.12) are however completely different, the proof of (6.11) being based on sieve theory.

[26] We say that n is a quadratic nonresidue $(\mathrm{mod}\ Q)$ if $\chi_Q(n) = -1$.

6.3. **Proof sketch for the main result**

The proof of Theorem 6.6 makes use of the ideas Tao [98] developed for his proof of Theorem 6.3; these are an averaging over small primes argument and the entropy decrement argument, also discussed in Section 5.

The averaging over small primes works as follows. Suppose for simplicity that g_1, g_2 are completely multiplicative and take only values ± 1. Then, for any prime $p \le \log \omega(x)$, we have

(6.13)
$$\begin{aligned}\frac{1}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \le n \le x} \frac{g_1(n) g_2(n+h)}{n} &= \frac{g_1 g_2(p)}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \le n \le x} \frac{g_1(pn) g_2(pn+ph)}{n} \\ &= \frac{g_1 g_2(p)}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \le n' \le x} \frac{g_1(n') g_2(n'+ph)}{n'} p 1_{p|n'} + O(\varepsilon)\end{aligned}$$

where we wrote $n' = pn$ and used the fact that the average is a logarithmic one. We can then sum (6.13) over p to conclude that

(6.14)
$$\begin{aligned}&\frac{1}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \le n \le x} \frac{g_1(n) g_2(n+h)}{n} \\ &= \frac{m}{2^m} \sum_{2^m \le p < 2^{m+1}} g_1(p) g_2(p) \frac{1}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \le n' \le x} \frac{g_1(n') g_2(n'+ph)}{n'} p 1_{p|n'} + O(\varepsilon),\end{aligned}$$

where $\varepsilon^{-1} \le m \le \log\log \omega(x)$. By the entropy decrement argument, developed by Tao in [98] and based on inequalities from information theory, we can replace $p 1_{p|n'}$ with its average value $1 + O(\varepsilon)$ for some suitable, large $m = m(\varepsilon)$. The advantage gained is that now (6.14) becomes a bilinear average

$$\frac{m}{2^m} \sum_{2^m \le p < 2^{m+1}} g_1(p) g_2(p) \frac{1}{\log \omega(x)} \sum_{\frac{x}{\omega(x)} \le n \le x} \frac{g_1(n) g_2(n+ph)}{n} + o(1),$$

where n and p have been decoupled. This enables us to apply the circle method. In the same spirit as in [98], the circle method gives the anticipated asymptotic for this sum, provided that we prove the short exponential sum bound[27]

$$\sup_{\alpha \in \mathbb{R}} \frac{1}{x} \int_x^{2x} \left| \frac{1}{H} \sum_{y \le n \le y+H} (g_1(n) - \delta_1) e(n\alpha) \right| dy = o_{\varepsilon \to 0}(1), \tag{6.15}$$

where δ_1 is the mean value of g_1 on $[x, 2x]$ and $H \asymp 2^{(1+O(\varepsilon))m}$ with $m = m(\varepsilon)$ large. This estimate deviates from what was used in [98], since there (6.15) was used in the non-pretentious case covered by a result of Matomäki, Radziwiłł and Tao [75,

[27] In reality, we need to consider the integral of the exponential sum over more general intervals $[y, 2y]$ with $\frac{x}{\omega(x)} \le y \le x$.

Theorem 1.7]. The case where g_1 is uniformly distributed in the sense of Definition 6.5 is not addressed in [75], but can be dealt with using the tools employed there.

The proof of (6.15) naturally splits into the *major arc* case, where α is close to a rational number with small denominator, and the opposite *minor arc* case. In the minor arc case, we can ignore the constant term δ_1 in (6.15) and then follow the argument in [75, Section 3], as that is based solely on the multiplicativity and boundedness of g_1.

In the major arc case, in contrast, we plainly need to use the uniform distribution property of g_1, as the result fails for example for Dirichlet characters, which are not equidistributed. If α is on a major arc, then $e(n\alpha)$ is essentially periodic, so we may make it essentially constant by splitting n into residue classes. Then we end up with the need to prove that

$$\frac{1}{x}\int_x^{2x}\left|\frac{1}{H}\sum_{\substack{y\le n\le y+H\\ n\equiv b \pmod q}} g_1(n) - \frac{1}{qH}\sum_{y\le n\le y+H} g_1(n)\right| dy = \frac{o_{\varepsilon\to 0}(1)}{q} \tag{6.16}$$

uniformly for $1 \le b \le q \le \varepsilon^{-1}$. Here we used the fact that δ_1 is the mean of g_1 also in arithmetic progressions of modulus $\le \varepsilon^{-1}$.

The estimate (6.16) follows for $q = 1$ from the Matomäki–Radziwiłł theorem [74] (since g_1 is real-valued), and it turns out that for $q > 1$, by expanding $1_{n\equiv b \pmod q}$ in terms of characters, we can use the complex-valued case of that theorem from [75, Appendix A] together with some pretentious distance estimates. This then leads to the desired conclusion (6.16). For the proof in its entirety, we refer to [IV].

References

[1] S. Baier. A note on Diophantine approximation with Gaussian primes. *ArXiv e-prints*, September 2016.

[2] R. C. Baker and G. Harman. The difference between consecutive primes. *Proc. London Math. Soc. (3)*, 72(2):261–280, 1996.

[3] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes. II. *Proc. London Math. Soc. (3)*, 83(3):532–562, 2001.

[4] D. A. Burgess. On character sums and L-series. II. *Proc. London Math. Soc. (3)*, 13:524–536, 1963.

[5] J. R. Chen. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica*, 16:157–176, 1973.

[6] J. R. Chen and C. D. Pan. The exceptional set of Goldbach numbers. I. *Sci. Sinica*, 23(4):416–430, 1980.

[7] S. Chowla. *The Riemann hypothesis and Hilbert's tenth problem*. Mathematics and Its Applications, Vol. 4. Gordon and Breach Science Publishers, New York-London-Paris, 1965.

[8] H. Cramér. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arith.*, 2:23–46, 1936.

[9] H. Daboussi and A. Sárközy. On the correlation of the truncated Liouville function. *Acta Arith.*, 108(1):61–76, 2003.

[10] H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, and T. de la Rue. The Chowla and the Sarnak conjectures from ergodic theory point of view. *Discrete Contin. Dyn. Syst.*, 37(6):2899–2944, 2017.

[11] P. D. T. A. Elliott. On the correlation of multiplicative functions. *Notas Soc. Mat. Chile*, 11(1):1–11, 1992.

[12] P. D. T. A. Elliott. On the correlation of multiplicative and the sum of additive arithmetic functions. *Mem. Amer. Math. Soc.*, 112(538):viii+88, 1994.

[13] P. Erdős. Some unconventional problems in number theory. In *Journées Arithmétiques de Luminy (Colloq. Internat. CNRS, Centre Univ. Luminy, Luminy, 1978)*, volume 61 of *Astérisque*, pages 73–82. Soc. Math. France, Paris, 1979.

[14] S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk. Sarnak's Conjecture – what's new. *ArXiv e-prints*, October 2017.

[15] N. Frantzikinakis. An averaged Chowla and Elliott conjecture along independent polynomials. *Int. Math. Res. Not. IMRN*, 2018(12):3721–3743, 2018.

[16] N. Frantzikinakis and B. Host. Asymptotics for multilinear averages of multiplicative functions. *Math. Proc. Cambridge Philos. Soc.*, 161(1):87–101, 2016.

[17] N. Frantzikinakis and B. Host. The logarithmic Sarnak conjecture for ergodic weights. *Ann. of Math. (2)*, 187(3):869–931, 2018.

[18] J. Friedlander and H. Iwaniec. The polynomial $X^2 + Y^4$ captures its primes. *Ann. of Math. (2)*, 148(3):945–1040, 1998.

[19] J. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society,

Providence, RI, 2010.

[20] P. X. Gallagher. On the distribution of primes in short intervals. *Mathematika*, 23(1):4–9, 1976.

[21] D. A. Goldston, S. W. Graham, J. Pintz, and C. Y. Yıldırım. Small gaps between almost primes, the parity problem, and some conjectures of Erdős on consecutive integers. *Int. Math. Res. Not. IMRN*, 2011(7):1439–1450, 2011.

[22] É. Goudout. Lois locales de la fonction ω dans presque tous les petits intervalles. *Proc. Lond. Math. Soc. (3)*, 115(3):599–637, 2017.

[23] A. Granville. Harald Cramér and the distribution of prime numbers. *Scand. Actuar. J.*, 1995(1):12–28, 1995. Harald Cramér Symposium (Stockholm, 1993).

[24] A. Granville and K. Soundararajan. Large character sums: pretentious characters and the Pólya-Vinogradov theorem. *J. Amer. Math. Soc.*, 20(2):357–384, 2007.

[25] B. Green. Roth's theorem in the primes. *Ann. of Math. (2)*, 161(3):1609–1636, 2005.

[26] B. Green and T. Tao. Restriction theory of the Selberg sieve, with applications. *J. Théor. Nombres Bordeaux*, 18(1):147–182, 2006.

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

[28] B. Green and T. Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

[29] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)*, 175(2):541–566, 2012.

[30] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012.

[31] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2)*, 176(2):1231–1372, 2012.

[32] V. Z. Guo. Piatetski-Shapiro primes in a Beatty sequence. *J. Number Theory*, 156:317–330, 2015.

[33] G. Halász. Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen. *Acta Math. Acad. Sci. Hungar.*, 19:365–403, 1968.

[34] G. H. Hardy and J. E. Littlewood. Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes. *Acta Math.*, 44(1):1–70, 1923.

[35] G. Harman. Almost-primes in short intervals. *Math. Ann.*, 258(1):107–112, 1981/82.

[36] G. Harman. Primes in short intervals. *Math. Z.*, 180(3):335–348, 1982.

[37] G. Harman. On the distribution of αp modulo one. *J. London Math. Soc. (2)*, 27(1):9–18, 1983.

[38] G. Harman. *Prime-detecting sieves*, volume 33 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2007.

[39] D. R. Heath-Brown. Gaps between primes, and the pair correlation of zeros of the zeta function. *Acta Arith.*, 41(1):85–99, 1982.
[40] D. R. Heath-Brown. The number of primes in a short interval. *J. Reine Angew. Math.*, 389:22–63, 1988.
[41] D. R. Heath-Brown. Primes represented by $x^3 + 2y^3$. *Acta Math.*, 186(1):1–84, 2001.
[42] D. R. Heath-Brown and H. Iwaniec. On the difference between consecutive primes. *Invent. Math.*, 55(1):49–69, 1979.
[43] H. A. Helfgott. The ternary Goldbach problem. *ArXiv e-prints*, January 2015.
[44] A. Hildebrand. On a conjecture of Balog. *Proc. Amer. Math. Soc.*, 95(4):517–523, 1985.
[45] A. Hildebrand and G. Tenenbaum. Integers without large prime factors. *J. Théor. Nombres Bordeaux*, 5(2):411–484, 1993.
[46] G. Hoheisel. Primzahlprobleme in der Analysis. *Sitz. Preuss. Akad. Wiss.*, 33:580–588, 1930.
[47] M. N. Huxley. On the difference between consecutive primes. *Invent. Math.*, 15:164–170, 1972.
[48] A. Ivić. *The Riemann zeta-function: theory and applications*. Dover Publications, Inc., Mineola, NY, 2003.
[49] H. Iwaniec. Primes of the type $\phi(x, y) + A$ where ϕ is a quadratic form. *Acta Arith.*, 21:203–234, 1972.
[50] H. Iwaniec. The half dimensional sieve. *Acta Arith.*, 29(1):69–95, 1976.
[51] H. Iwaniec and M. Jutila. Primes in short intervals. *Ark. Mat.*, 17(1):167–176, 1979.
[52] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.
[53] H Iwaniec and J. Pintz. Primes in short intervals. *Monatsh. Math.*, 98(2):115–143, 1984.
[54] C. H. Jia. Difference between consecutive primes. *Sci. China Ser. A*, 38(10):1163–1186, 1995.
[55] C. H. Jia. Almost all short intervals containing prime numbers. *Acta Arith.*, 76(1):21–84, 1996.
[56] C. H. Jia. On the exceptional set of Goldbach numbers in a short interval. *Acta Arith.*, 77(3):207–287, 1996.
[57] C. H. Jia. On the distribution of αp modulo one. II. *Sci. China Ser. A*, 43(7):703–721, 2000.
[58] O. Klurman. Correlations of multiplicative functions and applications. *Compos. Math.*, 153(8):1622–1657, 2017.
[59] O. Klurman and A. P. Mangerel. Effective Asymptotic Formulae for Multilinear Averages of Multiplicative Functions. *ArXiv e-prints*, August 2017.
[60] O. Klurman and A. P. Mangerel. Rigidity Theorems for Multiplicative Functions. *ArXiv e-prints*, July 2017.

[61] A. N. Le. Nilsequences and Multiple Correlations along Subsequences. *ArXiv e-prints*, August 2017.
[62] A. Leibman. Nilsequences, null-sequences, and multiple correlation sequences. *Ergodic Theory Dynam. Systems*, 35(1):176–191, 2015.
[63] H. Z. Li. Goldbach numbers in short intervals. *Sci. China Ser. A*, 38(6):641–652, 1995.
[64] H. Z. Li. The exceptional set of Goldbach numbers. II. *Acta Arith.*, 92(1):71–88, 2000.
[65] Ju. V. Linnik. An asymptotic formula in an additive problem of Hardy-Littlewood. *Izv. Akad. Nauk SSSR Ser. Mat.*, 24:629–706, 1960.
[66] W. C. Lu. Exceptional set of Goldbach number. *J. Number Theory*, 130(10):2359–2392, 2010.
[67] H. Maier. Primes in short intervals. *Michigan Math. J.*, 32(2):221–225, 1985.
[68] A. P. Mangerel. On the Bivariate Erdős-Kac Theorem and Correlations of the Möbius Function. *ArXiv e-prints*, December 2016.
[69] K. Matomäki. Prime numbers of the form $p = m^2 + n^2 + 1$ in short intervals. *Acta Arith.*, 128(2):193–200, 2007.
[70] K. Matomäki. The binary Goldbach problem with one prime of the form $p = k^2 + l^2 + 1$. *J. Number Theory*, 128(5):1195–1210, 2008.
[71] K. Matomäki. A Bombieri-Vinogradov type exponential sum result with applications. *J. Number Theory*, 129(9):2214–2225, 2009.
[72] K. Matomäki. The distribution of αp modulo one. *Math. Proc. Cambridge Philos. Soc.*, 147(2):267–283, 2009.
[73] K. Matomäki, J. Maynard, and X. Shao. Vinogradov's theorem with almost equal summands. *Proc. Lond. Math. Soc. (3)*, 115(2):323–347, 2017.
[74] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math. (2)*, 183(3):1015–1056, 2016.
[75] K. Matomäki, M. Radziwiłł, and T. Tao. An averaged form of Chowla's conjecture. *Algebra Number Theory*, 9(9):2167–2196, 2015.
[76] K. Matomäki, M. Radziwiłł, and T. Tao. Sign patterns of the Liouville and Möbius functions. *Forum Math. Sigma*, 4:e14, 44, 2016.
[77] K. Matomäki and X. Shao. Vinogradov's three primes theorem with almost twin primes. *Compos. Math.*, 153(6):1220–1256, 2017.
[78] L. Matthiesen. Linear correlations of multiplicative functions. *ArXiv e-prints*, June 2016.
[79] J. Maynard. Primes represented by incomplete norm forms. *ArXiv e-prints*, July 2015.
[80] J. Maynard. Small gaps between primes. *Ann. of Math. (2)*, 181(1):383–413, 2015.
[81] H. Mikawa. Almost-primes in arithmetic progressions and short intervals. *Tsukuba J. Math.*, 13(2):387–401, 1989.
[82] H. L. Montgomery and R. C. Vaughan. The exceptional set in Goldbach's problem. *Acta Arith.*, 27:353–370, 1975. Collection of articles in memory of

Juriĭ Vladimirovič Linnik.

[83] Y. Motohashi. A note on almost-primes in short intervals. *Proc. Japan Acad. Ser. A Math. Sci.*, 55(6):225–226, 1979.

[84] A. Perelli and J. Pintz. On the exceptional set for the $2k$-twin primes problem. *Compositio Math.*, 82(3):355–372, 1992.

[85] A. Perelli and J. Pintz. On the exceptional set for Goldbach's problem in short intervals. *J. London Math. Soc. (2)*, 47(1):41–49, 1993.

[86] J. Pintz. On primes in short intervals. II. *Studia Sci. Math. Hungar.*, 19(1):89–96, 1984.

[87] O. Ramaré and I. Z. Ruzsa. Additive properties of dense subsets of sifted sequences. *J. Théor. Nombres Bordeaux*, 13(2):559–581, 2001.

[88] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[89] P. Sarnak. Mobius randomness and dynamics. *Not. S. Afr. Math. Soc.*, 43(2):89–97, 2012.

[90] L. Schoenfeld. Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. II. *Math. Comp.*, 30(134):337–360, 1976.

[91] A. Selberg. On the normal density of primes in small intervals, and the difference between consecutive primes. *Arch. Math. Naturvid.*, 47(6):87–105, 1943.

[92] X. Shao. A density version of the Vinogradov three primes theorem. *Duke Math. J.*, 163(3):489–512, 2014.

[93] S.-Y. Shi. On the distribution of αp modulo one for primes p of a special form. *Osaka J. Math.*, 49(4):993–1004, 2012.

[94] V. T. Sós. Turbulent years: Erdős in his correspondence with Turán from 1934 to 1940. In *Paul Erdős and his mathematics, I (Budapest, 1999)*, volume 11 of *Bolyai Soc. Math. Stud.*, pages 85–146. János Bolyai Math. Soc., Budapest, 2002.

[95] Y.-C. Sun and H. Pan. The Green-Tao theorem for primes of the form $x^2 + y^2 + 1$. *ArXiv e-prints*, August 2017.

[96] T. Tao. Special cases of Shannon entropy. `https://terrytao.wordpress.com/2017/03/01/special-cases-of-shannon-entropy/`.

[97] T. Tao. The Erdős discrepancy problem. *Discrete Anal.*, pages Paper No. 1, 29, 2016.

[98] T. Tao. The logarithmically averaged Chowla and Elliott conjectures for two-point correlations. *Forum Math. Pi*, 4:e8, 36, 2016.

[99] T. Tao. Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. In *Number theory—Diophantine problems, uniform distribution and applications*. Springer, Cham, 2017.

[100] T. Tao and J. Teräväinen. The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures. *ArXiv e-prints*, August 2017.

[101] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[102] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, third edition, 2015. Translated from the 2008 French edition by Patrick D. F. Ion.

[103] T. L. Todorova and D. I. Tolev. On the distribution of αp modulo one for primes p of a special form. *Math. Slovaca*, 60(6):771–786, 2010.

[104] D. I. Tolev. The binary Goldbach problem with arithmetic weights attached to one of the variables. *Acta Arith.*, 142(2):169–178, 2010.

[105] D. I. Tolev. The ternary Goldbach problem with arithmetic weights attached to two of the variables. *J. Number Theory*, 130(2):439–457, 2010.

[106] R. C. Vaughan. On the distribution of αp modulo 1. *Mathematika*, 24(2):135–141, 1977.

[107] R. C. Vaughan. *The Hardy-Littlewood method*, volume 125 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, second edition, 1997.

[108] I. M. Vinogradov. A new method in analytic number theory (Russian). *Tr. Mat. Inst. Steklova*, 10:5–122, 1937.

[109] I. M. Vinogradov. *The Method of Trigonometric Sums in the Theory of Numbers*. Translated from the Russian, revised and annotated by K. F. Roth and A. Davenport. Wiley-Interscience, 1954.

[110] Z. Wang. Sur les plus grands facteurs premiers d'entiers consécutifs. *Mathematika*, 64(2):343–379, 2018.

[111] N. Watt. Kloosterman sums and a mean value for Dirichlet polynomials. *J. Number Theory*, 53(1):179–210, 1995.

[112] N. Watt. Short intervals almost all containing primes. *Acta Arith.*, 72(2):131–167, 1995.

[113] E. Wirsing. Das asymptotische Verhalten von Summen über multiplikative Funktionen. II. *Acta Math. Acad. Sci. Hungar.*, 18:411–467, 1967.

[114] D. Wolke. Fast-Primzahlen in kurzen Intervallen. *Math. Ann.*, 244(3):233–242, 1979.

[115] J. Wu. Primes of the form $p = 1 + m^2 + n^2$ in short intervals. *Proc. Amer. Math. Soc.*, 126(1):1–8, 1998.

[116] C. Y. Yıldırım. A survey of results on primes in short intervals. In *Number theory and its applications (Ankara, 1996)*, volume 204 of *Lecture Notes in Pure and Appl. Math.*, pages 307–343. Dekker, New York, 1999.

[117] Y. Zhang. Bounded gaps between primes. *Ann. of Math. (2)*, 179(3):1121–1174, 2014.

Original publications

Publication I

J. Teräväinen: Almost primes in almost all short intervals. *Math. Proc. Cambridge Philos. Soc.*, 161(2):247–281, 2016. DOI: 10.1017/S0305004116000232

Math. Proc. Camb. Phil. Soc. (2016), **161**, 247–281
doi:10.1017/S0305004116000232
First published online 13 April 2016

Almost primes in almost all short intervals

BY JONI TERÄVÄINEN

Department of Mathematics, University of Turku, 20014 *Turku, Finland.*
e-mail: joni.p.teravainen@utu.fi

(*Received* 16 *November* 2015; *revised* 07 *March* 2016)

Abstract

Let E_k be the set of positive integers having exactly k prime factors. We show that almost all intervals $[x, x + \log^{1+\varepsilon} x]$ contain E_3 numbers, and almost all intervals $[x, x + \log^{3.51} x]$ contain E_2 numbers. By this we mean that there are only $o(X)$ integers $1 \leqslant x \leqslant X$ for which the mentioned intervals do not contain such numbers. The result for E_3 numbers is optimal up to the ε in the exponent. The theorem on E_2 numbers improves a result of Harman, which had the exponent $7+\varepsilon$ in place of 3.51. We also consider general E_k numbers, and find them on intervals whose lengths approach $\log x$ as $k \to \infty$.

1. *Introduction*

When studying E_k numbers (products of exactly k primes), it is natural to ask, how short intervals include such numbers almost always. Since Wolke's work [**21**], the essential question has been minimising the number c such that almost all intervals $[x, x + \log^c x]$ contain an E_k number, meaning that all but $o(X)$ such intervals with integer $x \in [1, X]$ contain such a number. Wolke showed in 1979 that the value $c = 5 \cdot 10^6$ is admissible for E_2 numbers. This was improved to $c = 7 + \varepsilon$ for E_2 numbers by Harman [**9**] in 1982. Wolke's and Harman's methods are based on reducing the problem to estimates for sums over the zeros of the Riemann zeta function, and on the fact that the density hypothesis is known to hold in a non-trivial strip (namely Jutila's [**14**] region $\sigma \geqslant 11/14$ in Harman's argument[1]). To the author's knowledge, Harman's exponent for E_2 numbers was the best one known also for E_k numbers with $k \geqslant 3$.

If one considers P_k numbers, which are products of no more than k primes, one can obtain improvements. Mikawa [**16**] showed in 1989 that for any function $\psi(x)$ tending to infinity, the interval $[x, x + \psi(x) \log^5 x]$ contains a P_2 number almost always. Furthermore, Friedlander and Iwaniec [**4**, chapters 6 and 11] proved that for any such function $\psi(x)$ the interval $[x, x+\psi(x) \log x]$ contains a P_4 number almost always. They also hint how to prove the same result for P_3 numbers. There is however a crucial difference between E_k and P_k numbers, since the E_k numbers are subject to the famous parity problem, and hence cannot be dealt with using only classical combinatorial sieves, which are the basis of the arguments on P_k numbers. Therefore, the E_k numbers are also a much closer analogue of primes than the P_k numbers.

[1] In fact, introducing into Harman's argument the widest known density hypothesis region $\sigma \geqslant 25/32$, due to Bourgain [**2**], would give $c = 6.86$.

One would naturally expect almost all intervals $[x, x+\psi(x)\log x]$ to have also prime numbers in them, and this would follow from the heuristic that the proportion of x for which $[x, x+\lambda\log x]$ contains exactly m primes for fixed m and $\lambda > 0$ should be given by the Poisson distribution $(\lambda^m/m!)e^{-\lambda}$. Such results are however far beyond the current knowledge, as the shortest intervals, almost all of which are known to contain primes, are $[x, x+x^{\frac{1}{20}+\varepsilon}]$ by a result of Jia [**13**]. However, the results of Goldston–Pintz–Yıldırım [**6**],[**7**] on short gaps between primes tell that for any $\lambda > 0$ there is a positive proportion of integers $x \leqslant X$ for which $[x, x+\lambda\log x]$ contains a prime, but it is not known whether this proportion approaches 1 as λ increases. A recent result of Freiberg [**3**], in turn, gives exactly m primes on an interval $[x, x+\lambda\log x]$ for at least $X^{1-o(1)}$ integers $x \leqslant X$. Concerning conditional results, Gallagher [**5**] showed that the Poisson distribution of primes in short intervals would follow from a certain uniform form of the Hardy–Littlewood prime k-tuple conjecture. Under the Riemann hypothesis, it was shown by Selberg [**18**] in 1943 that almost all intervals $[x, x+\psi(x)\log^2 x]$ contain primes. For E_2 numbers, under the density hypothesis, Harman's argument from [**9**] would give the exponent $c = 3+\varepsilon$.

In this paper, we establish the exponent $c = 1+\varepsilon$ for E_3 numbers and the exponent $c = 3.51$ for E_2 numbers. Our results for E_2, E_3 and E_k numbers are stated as follows.

THEOREM 1. *Almost all intervals* $[x, x+(\log\log x)^{6+\varepsilon}\log x]$ *contain a product of exactly three distinct primes.*

THEOREM 2. *For any integer* $k \geqslant 4$, *there exists* $C_k > 0$ *such that almost all intervals* $[x, x+(\log_{k-1} x)^{C_k}\log x]$ *contain a product of exactly* k *distinct primes. Here* $\log_\ell$ *is the* ℓ*th iterated logarithm.*

THEOREM 3. *Almost all intervals* $[x, x+\log^{3.51} x]$ *with* $x \leqslant X$ *contain a product of exactly two distinct primes.*

Theorems 1 and 2 are direct consequences of the following theorem.

THEOREM 4. *Let* X *be large enough,* $k \geqslant 3$ *a fixed integer, and* $\varepsilon > 0$ *small enough but fixed. Define the numbers* $P_1, \ldots, P_{k-1}$ *by setting* $P_{k-1} = (\log X)^{\varepsilon^{-2}}$, $P_{k-2} = (\log\log X)^{6+10\sqrt{\varepsilon}}$ *and* $P_j = (\log P_{j+1})^{\varepsilon^{-1}}$ *for* $1 \leqslant j \leqslant k-3$. *For* $P_1\log X \leqslant h \leqslant X$, *we have*

$$\left|\frac{1}{h}\sum_{\substack{x\leqslant p_1\cdots p_k\leqslant x+h\\ P_i\leqslant p_i\leqslant P_i^{1+\varepsilon},\, i\leqslant k-1}} 1 - \frac{1}{X}\sum_{\substack{X\leqslant p_1\cdots p_k\leqslant 2X\\ P_i\leqslant p_i\leqslant P_i^{1+\varepsilon},\, i\leqslant k-1}} 1\right| \ll \frac{1}{(\log X)(\log_k X)} \tag{1·1}$$

for almost all $x \leqslant X$.

In the theorem above, the average over the dyadic interval is $\gg 1/\log X$ by the prime number theorem, so Theorems 1 and 2 indeed follow from Theorem 4. Similarly, Theorem 3 is a direct consequence of the following.

THEOREM 5. *Let* X *be large enough,* $P_1 = \log^a X$ *with* $a = 2.51$, $\varepsilon > 0$ *fixed, and* $P_1\log X \leqslant h \leqslant X$. *We have*

$$\frac{1}{h}\sum_{\substack{x\leqslant p_1p_2\leqslant x+h\\ P_1\leqslant p_1<P_1^{1+\varepsilon}}} 1 \gg \frac{1}{X}\sum_{\substack{X\leqslant p_1p_2\leqslant 2X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}} 1 \tag{1·2}$$

for almost all $x \leqslant X$.

Remark 1. Since $h \geqslant P_1 \log X$, we have the dependence $c = a + 1$ between the exponent a in Theorem 5 and the smallest exponent c for which we can show that the interval $[x, x + \log^c x]$ contains an E_2 number almost always.

Remark 2. Note that Theorems 4 and 5 tell us that there are $\gg h/\log X$ E_k numbers in almost all intervals $[x, x + h]$, where h and k are as in one of the theorems. However, we are not quite able to find E_k numbers on intervals $[x, x + \psi(x) \log x]$ with ψ tending to infinity arbitrarily slowly, unlike in the result of Friedlander and Iwaniec on P_k numbers. In addition, our bound for the number of exceptional values is at best $\ll x/\log^{\varepsilon} x$ and often weaker, while the methods used in [**10**], [**13**] and [**20**] for primes in almost all short intervals have a tendency to give the bound $\ll x/\log^A x$ for any $A > 0$, when they work. The limit of our method for E_2 numbers is the exponent $3+\varepsilon$, as will be seen later, so proving for example unconditionally the analogue of Selberg's result for E_2 numbers would require some further ideas.

To prove our results, we adapt the ideas of the paper [**15**] of Matomäki and Radziwiłł on multiplicative functions in short intervals to considering almost primes in short intervals. In that paper, a groundbreaking result is that for any multiplicative function, with values in $[-1, 1]$, its average over $[x, x + h]$ is almost always asymptotically equal to its dyadic average over $[x, 2x]$, with $h = h(x) \leqslant x$ any function tending to infinity. The error terms obtained there for general multiplicative functions are not quite good enough for our purposes. Nevertheless, using similar techniques, and replacing the multiplicative function with the indicator function of the numbers $p_1 \cdots p_k$, with p_i primes from carefully chosen intervals, allows us to find E_k numbers on intervals $[x, x + h]$, with $h/\log x$ growing very slowly. In this setting, we can apply various mean, large and pointwise value results for Dirichlet polynomials, some of which work specifically with primes or the zeta function, but not with general multiplicative functions (such as Watt's theorem on the twisted moment of the Riemann zeta function, a large values theorem from [**15**] for Dirichlet polynomials supported on primes, and Vinogradov's zero-free region). In many places in the argument, we cannot afford to lose even factors of $\log^{\varepsilon} x$, so we need to factorise Dirichlet polynomials in a manner that is nearly nearly lossless, and use an improved form of the mean value theorem for Dirichlet polynomials. To deal with some of the arising Dirichlet polynomials, we also need some sieve methods, similar to those that have been successfully applied to finding primes in short intervals for example in [**10**], [**13**] and [**20**]. In the case of E_2 numbers, in addition to these methods, we benefit from the theory of exponent pairs and Jutila's large values theorem.

The structure of the proofs of Theorems 4 and 5 is the following. We will first present the lemmas necessary for proving Theorem 4, and hence Theorems 1 and 2. Besides employing these lemmas to prove Theorem 4, we notice that they are already sufficient for finding products of exactly two primes in almost all intervals $[x, x + \log^{5+\varepsilon} x]$, which is as good as Mikawa's result for P_2 numbers up to ε in the exponent (one could also get c slightly below 5 using exponent pairs, which are just one of the additional ideas required for Theorem 5). The rest of the paper is then concerned with reducing the exponent $5+\varepsilon$ to 3.51 for products of two primes, and this requires some further ingredients, as well as all the lemmas that were needed for products of three or more primes.

1·1. *Notation*

The symbols p, q, p_i and q_i are reserved for primes, and d, k, ℓ, m and n are always positive integers. We often use the same capital letter for a Dirichlet polynomial and its length. We call *zeta sums* partial sums of $\zeta(s)$ or $\zeta'(s)$ of the form $\sum_{n\sim N} n^{-s}$ or $\sum_{n\sim N}(\log n)n^{-s}$.

The function $\nu(\cdot)$ counts the number of distinct prime divisors of a number, $\mu(\cdot)$ is the Möbius function, $\Lambda(\cdot)$ is the von Mangoldt function, and $d_r(m)$ is the number of solutions to $a_1 \cdots a_r = m$ in positive integers. The function $\omega(\cdot)$ is Buchstab's function (see Harman's book [**10**, chapter 1]), defined as $\omega(u) = 1/u$ for $1 \leqslant u \leqslant 2$ and via the differential equation $(d/du)(u\omega(u)) = \omega(u-1)$ for $u > 2$, imposing the requirement that ω be continuous on $[1, \infty)$. We make the convention that $\omega(u) = 0$ for $u < 1$. In addition, let $\mathcal{P}(z) = \prod_{p<z} p$, and let $S(A, \mathbb{P}, z)$ count the numbers in A coprime to $\mathcal{P}(z)$.

The quantity $\varepsilon > 0$ is always small enough but fixed. The symbols $C_1, C_2, \ldots$ denote unspecified, positive, absolute constants. By writing $n \sim X$ in a summation, we mean $X \leqslant n < 2X$. The expression 1_S is the indicator function of the set S, so that $1_S(n) = 1$ if $n \in S$ and $1_S(n) = 0$ otherwise. We use the usual Landau and Vinogradov asymptotic notation $o(\cdot)$, $O(\cdot)$ and $\ll, \gg$. The notation $X \asymp Y$ is shorthand for $X \ll Y \ll X$.

2. *Preliminary lemmas*

2·1. *Reduction to mean values of Dirichlet polynomials*

We present several lemmas that are required for proving both Theorems 4 and 5. Later on, we give some additional lemmas that are needed only for proving Theorem 5.

The plan of the proofs of Theorems 4 and 5, and hence of Theorems 1, 2 and 3, is to transform the problem of comparing almost primes in short and long intervals to finding cancellation in the mean square of the corresponding Dirichlet polynomial. The polynomial can be factorised after it is divided into short intervals, and different methods can be applied to different factors. This approach is utilised in many earlier works on primes and almost primes in short intervals; see e.g. [**10**], [**15**]. We then apply several mean, large and pointwise value theorems, which are presented in Subsection 2·3, to find the desired cancellation in the Dirichlet polynomial.

The following Parseval-type lemma allows us to reduce the problem of finding almost primes in short intervals to finding cancellation in a Dirichlet polynomial.

LEMMA 1. *Let*

$$S_h(x) = \frac{1}{h} \sum_{x \leqslant n \leqslant x+h} a_n,$$

where a_n are complex numbers, and let $2 \leqslant h_1 \leqslant h_2 \leqslant \frac{X}{T_0^3}$ with $T_0 \geqslant 1$. Also let $F(s) = \sum_{n\sim X} a_n/n^s$. Then

$$\frac{1}{X}\int_X^{2X} \left|\frac{1}{h_1}S_{h_1}(x) - \frac{1}{h_2}S_{h_2}(x)\right|^2 dx \ll \frac{1}{T_0} + \int_{T_0}^{\frac{X}{h_1}} |F(1+it)|^2 dt + \max_{T \geqslant \frac{X}{h_1}} \frac{X}{Th_1}\int_T^{2T} |F(1+it)|^2 dt. \tag{2·3}$$

Proof. This is lemma 14 in the paper [**15**] (except that we do not specify the value of T_0). A related bound can be found for example in [**10**, chapter 9].

We choose $T_0 = X^{0.01}$, and $h_2 = X/T_0^3$ in Lemma 1, and the average function $S_h(x)$ is given by the short average in (1·1) or (1·2). Now, defining

$$F(s) = \sum_{\substack{p_1\cdots p_k\sim X\\ P_i\leqslant p_i\leqslant P_i^{1+\varepsilon}, i\leqslant k-1}} (p_1\cdots p_k)^{-s},$$

where P_i are as in Theorem 4 or 5, proving Theorems 4 and 5 is reduced to showing that

$$\int_{T_0}^{T} |F(1+it)|^2 dt = o\left(\left(\frac{Th}{X}+1\right)\cdot\frac{1}{(\log^2 X)(\log_k X)^2}\right), \tag{2·4}$$

for $T_0 = X^{0.01}$ and $h \geqslant P_1 \log X$. Indeed, substituting this to Lemma 1 shows that

$$\frac{1}{X}\int_X^{2X}\left|\frac{1}{h}S_h(x) - \frac{1}{h_2}S_{h_2}(x)\right|^2 dx = o\left(\frac{1}{(\log^2 X)(\log_k X)^2}\right),$$

where $h_2 = X/T_0^3$. It actually suffices to prove (2·4) for $T \leqslant X$, since otherwise the mean value theorem (Lemma 3) gives a good enough bound for the last term in (2·3).

Note that for $T \leqslant X$ the trivial bound for the integral in (2·4), coming from the mean value theorem, is $\ll (\log X)^{-1}$. Thus our task is to save slightly more than one additional logarithm in this integral (for $T \leqslant X/h$, at least).

Once the required estimates for Dirichlet polynomials have been established, we can apply the prime number theorem in short intervals with Vinogradov's error term (see [**12**, chapter 10]) to see that

$$\frac{1}{h_2}S_{h_2}(x) - \frac{1}{X}S_X(X) \ll \exp(-(\log X)^{\frac{3}{5}-\varepsilon}),$$

for $h_2 = x^{0.97}$, $x \sim X$, and hence deduce Theorems 4 and 5 (and consequently 1, 2 and 3). For example, we compute

$$\begin{aligned}\frac{1}{h_2}\sum_{\substack{x\leqslant p_1p_2p_3\leqslant x+h_2\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ P_2\leqslant p_2\leqslant P_2^{1+\varepsilon}}} 1 &= \frac{1}{h_2}\sum_{\substack{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ P_2\leqslant p_2\leqslant P_2^{1+\varepsilon}}}\left(\pi\left(\frac{x+h_2}{p_1p_2}\right)-\pi\left(\frac{x}{p_1p_2}\right)\right)\\ &= \frac{1}{h_2}\sum_{\substack{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ P_2\leqslant p_2\leqslant P_2^{1+\varepsilon}}}\frac{h_2}{p_1p_2\log\frac{x}{p_1p_2}}\\ &\quad + O\left(\exp(-(\log x)^{\frac{3}{5}-\frac{\varepsilon}{2}})\right)\\ &= \sum_{\substack{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ P_2\leqslant p_2\leqslant P_2^{1+\varepsilon}}}\frac{1}{p_1p_2\log\frac{X}{p_1p_2}} + O(\exp(-(\log X)^{\frac{3}{5}-\varepsilon})),\end{aligned}$$

and the same asymptotics hold for the dyadic sum. Sometimes we end up comparing the sums $(1/h_2)S_{h_2}(x)$ and $(1/x)S_2(x)$ with a_n not quite equal to the coefficients of $F(s)$, but equal to the indicator function of the numbers p_1p_2n with p_1 and p_2 from the intervals $[P_1, P_1^{1+\varepsilon}]$ and $[P_2, P_2^{1+\varepsilon}]$, respectively, and n having no prime factors smaller than p_2. There may also be a simple cross-conditions on p_1 and p_2, but comparing the sums still causes no difficulty.

Thus, in the rest of the paper we can concentrate on bounding Dirichlet polynomials. Although there is a close analogy in the formulations of Theorems 4 and 5, estimating the polynomial arising from the latter is more difficult, and will require several additional ideas.

2·2. *Factorisations for Dirichlet polynomials*

In bounding Dirichlet polynomials, factorisations play an important role. We encounter situations where the only cross-condition on the variables in the polynomial is that their product belongs to a certain range, so the variables can be separated by diving them into short ranges and estimating the mean values of the resulting polynomials. The factorisation is provided by the following lemma, which also takes into account the improved mean value theorem (Lemma 4).

LEMMA 2. *Let $\mathcal{S} \subset [-T, T]$ be measurable and*

$$F(s) = \sum_{\substack{mn \sim X \\ M \leqslant m \leqslant M'}} \frac{a_m b_n}{(mn)^s}$$

for some $M' > M \geqslant 2$ and for some complex numbers a_m, b_n. Let $H \geqslant 1$ be such that $H \log M$ and $H \log M'$ are integers. Denote

$$A_{v,H}(s) = \sum_{e^{\frac{v}{H}} \leqslant m < e^{\frac{v+1}{H}}} \frac{a_m}{m^s}, \quad B_{v,H}(s) = \sum_{n \sim X e^{-\frac{v}{H}}} \frac{b_n}{n^s}.$$

Then

$$\begin{aligned} \int_{\mathcal{S}} |F(1+it)|^2 dt \ll |I|^2 \int_{\mathcal{S}} |A_{v_0,H}(1+it) B_{v_0,H}(1+it)|^2 dt \\ + T \sum_{\substack{n \in [Xe^{-\frac{1}{H}}, Xe^{\frac{1}{H}}] \text{ or} \\ n \in [2X, 2Xe^{\frac{1}{H}}]}} |c_n|^2 + T \sum_{1 \leqslant h \leqslant \frac{2X}{T}} \sum_{\substack{m-n=h \\ m,n \in [Xe^{-\frac{1}{H}}, Xe^{\frac{1}{H}}] \text{ or} \\ m,n \in [2X, 2Xe^{\frac{1}{H}}]}} |c_m||c_n|, \end{aligned}$$

with

$$c_n = \frac{1}{n} \sum_{\substack{n = k\ell \\ M \leqslant k \leqslant M'}} |a_k b_\ell|,$$

$I = [H \log M, H \log M')$ and $v_0 \in I$ a suitable integer.

Remark 3. In applications we have $M' \geqslant 2M$, so the conditions that $H \log M$ and $H \log M'$ be integers can be ignored, since we can always afford to vary H and M' by the necessary amount.

Remark 4. When proving Theorem 4, we cannot afford to lose any powers of logarithm in some factorisations, and indeed the second term in the lemma crucially has the factor T instead of the factor X occurring in the mean value theorem, and in the first term we will lose a factor of size $\ll H^2 \log^2(M'/M)$, which in practice is minuscule.

Proof. This resembles [**15**, lemma 12] by Matomäki and Radziwiłł (where, in addition to factorisation in short intervals, a Ramaré-type identity is used). We split $F(s)$ into short

intervals, obtaining

$$F(s) = \sum_{v \in I \cap \mathbb{Z}} \sum_{e^{\frac{v}{H}} \leqslant m < e^{\frac{v+1}{H}}} \frac{a_m}{m^s} \sum_{\substack{Xe^{-\frac{v+1}{H}} \leqslant n < 2Xe^{-\frac{v}{H}} \\ mn \sim X}} \frac{b_n}{n^s}.$$

Observe that $Xe^{-\frac{v+1}{H}} \leqslant n < Xe^{-\frac{v}{H}}$ can hold above only for $mn \in [Xe^{-\frac{1}{H}}, Xe^{\frac{1}{H}}]$. Furthermore, we always have $mn \in [Xe^{-\frac{1}{H}}, 2Xe^{\frac{1}{H}}]$. This allows us to write

$$F(s) = \sum_{v \in I \cap \mathbb{Z}} A_{v,H}(s) B_{v,H}(s) + \sum_{\substack{k \in [Xe^{-\frac{1}{H}}, Xe^{\frac{1}{H}}] \text{ or} \\ k \in [2X, 2Xe^{\frac{1}{H}}]}} \frac{d_k}{k^s} \tag{2·5}$$

with

$$|d_k| \leqslant \sum_{k=mn} |a_m b_n|.$$

Now the claim of the lemma follows by taking mean squares on both sides of (2·5) on the line $\Re(s) = 1$, applying the improved mean value theorem (Lemma 4), and taking the maximum in the sum over I.

2·3. *Bounds for Dirichlet polynomials*

We need several mean, large and pointwise value results on Dirichlet polynomials. The following lemma is one of the basic tools.

LEMMA 3 (Mean value theorem for Dirichlet polynomials). *Let* $N \geqslant 1$ *and* $F(s) = \sum_{n \sim N} \frac{a_n}{n^s}$, *where* a_n *are any complex numbers. Then*

$$\int_{-T}^{T} |F(it)|^2 dt \ll (N + T) \sum_{n \sim N} |a_n|^2.$$

Proof. See for example Iwaniec and Kowalski's book [**12**, chapter 9].

If the coefficients a_n are supported on the primes or almost primes and are of size $\asymp 1/n$, the sum $\sum_{n \sim N} |a_n|^2$ is essentially $1/N \log N$. However, in some places in the proofs of Theorems 1, 2 and 3, it is vital to save one more logarithm in such a situation. This is enabled by an improved mean value theorem.

LEMMA 4 (Improved mean value theorem). *Let* N *and* $F(s)$ *be as above. We have*

$$\int_{-T}^{T} |F(it)|^2 dt \ll T \sum_{n \sim N} |a_n|^2 + T \sum_{1 \leqslant h \leqslant \frac{N}{T}} \sum_{\substack{m-n=h \\ m,n \sim N}} |a_m||a_n|. \tag{2·6}$$

Remark 5. The number of solutions to $m - n = h$, with m and n primes and $m, n \sim N$, is $\ll (N^2/\log^2 N \cdot h/\varphi(h))$ (with φ Euler's totient function), which follows easily from Brun's sieve, for example. If $T \leqslant N/h$, $h \geqslant \log N$ and a_n is supported on the primes, the first sum in (2·6) turns out not to be problematic, so we indeed save essentially one additional logarithm with this lemma. We remark that if we have polynomials of length $N \leqslant T$, Lemma 4 reduces to the basic mean value theorem.

Proof. This follows from [**12**, chapter 7, lemma 7·1], taking $Y = 10T$ there.

We also put into use a discrete mean value theorem, which is particularly useful when we take the mean square over a rather small set of points.

LEMMA 5 (Halász–Montgomery inequality). *Let N and $F(s)$ be as before. Let $\mathcal{T} \subset [-T, T]$ be* well-spaced*, meaning that $t, u \in \mathcal{T}$ and $t \neq u$ imply $|t - u| \geqslant 1$. Then*

$$\sum_{t \in \mathcal{T}} |F(it)|^2 \ll (N + |\mathcal{T}|T^{\frac{1}{2}})(\log T) \sum_{n \sim N} |a_n|^2.$$

Proof. For a proof, see Iwaniec and Kowalski's book [**12**, chapter 9].

In addition to mean value theorems, we need some large values theorems. We come across some very short Dirichlet polynomials, say of length $\ll T^{o(1)}$, and we make use of the fact that the coefficients of these polynomials are supported on the primes.

LEMMA 6. *Let $P \geqslant 1$, $V > 0$ and*

$$F(s) = \sum_{p \sim P} \frac{a_p}{p^s}$$

with $|a_p| \leqslant 1$. Let $\mathcal{T} \subset [-T, T]$ be a well-spaced set of points such that $|F(1 + it)| \geqslant V$ for each $t \in \mathcal{T}$. Then we have

$$|\mathcal{T}| \ll T^{2\frac{\log V^{-1}}{\log P}} V^{-2} \exp\left((1 + o(1))\frac{\log T}{\log P} \log\log T\right).$$

Remark 6. We may also apply this lemma to polynomials not supported on primes, provided that $P \gg X^{\varepsilon}$ for some $\varepsilon > 0$. In this case, the lemma is essentially the mean value theorem applied to a suitable moment of the polynomial.

Proof. This is [**15**, lemma 8]. There a factor of 2 occurs instead of $1 + o(1)$ in the last exponential, but the exact same proof works with the factor $1 + o(1)$.

For proving Theorem 5, we also need a large values theorem designed for long polynomials. The reason for presenting it along with the lemmas for Theorem 4 is that combining it with the other lemmas already gives the exponent $c = 5 + \varepsilon$ for E_2 numbers. The large values result is a theorem of Jutila that improves on the better known Huxley's large values theorem.

LEMMA 7 (Jutila's large values theorem). *Let $F(s) = \sum_{n \sim N} a_n/n^s$ with $|a_n| \leqslant d_r(n)$ for some fixed r. Let $\mathcal{T} \subset [-T, T]$ be a well-spaced set such that $|F(1 + it)| \geqslant V$ for $t \in \mathcal{T}$, and let k be any positive integer. We have*

$$|\mathcal{T}| \ll \left(V^{-2} + \frac{T}{N^2} V^{-6+\frac{2}{k}} + V^{-8k} \frac{T}{N^{2k}}\right) (NT)^{o(1)}.$$

Proof. The proof can be found in Jutila's paper [**14**]. We apply formula (1.4) there to $F(s)^{\ell}$, and have $G = \sum_{n \sim N} |a_n|^2/n^2 \ll (NT)^{o(1)} N^{-1}$ in the notation of that paper.

In some cases in the proof of Theorem 4, there will be polynomials supported on primes or almost primes for which the best we can do is apply a pointwise bound. These bounds follow in the end from Vinogradov's zero-free region.

LEMMA 8. *Let*

$$P(s) = \sum_{n_1 \cdots n_k \sim N} g_1(n_1) \cdots g_k(n_k)(n_1 \cdots n_k)^{-s},$$

where $k \geqslant 1$ is a fixed integer and each g_i is either the Möbius function, the characteristic function of the primes, the identity function, or the logarithm function. We have

$$|P(1+it)| \ll \exp\left(-(\log N)^{\frac{1}{10}}\right)$$

when $\exp((\log N)^{\frac{1}{3}}) \leqslant |t| \leqslant N^{A \log\log N}$ *for any fixed* $A > 0$.

Proof. For $k = 1$, the claim follows directly from Perron's formula and Vinogradov's zero-free region, so let $k \geqslant 2$. We may assume that $n_1, \ldots, n_k$ belong to some dyadic intervals $I_1, \ldots, I_k$ such that $I_k = [a, b]$ with $a \gg N^{\frac{1}{k}}$, $b \ll N$. Now

$$\sum_{n_1 \in I_1, \ldots, n_{k-1} \in I_{k-1}} g(n_1) \cdots g(n_{k-1})(n_1 \cdots n_{k-1})^{-1-it} \sum_{\substack{n_k \in I_k \\ n_k \sim \frac{N}{n_1 \cdots n_{k-1}}}} g(n_k) n_k^{-1-it}$$

$$\ll (\log N)^{O(1)} \sum_{n_1 \in I_1, \ldots, n_{k-1} \in I_{k-1}} (n_1 \cdots n_{k-1})^{-1} \cdot \exp\left(-\frac{\log N^{\frac{1}{k}}}{(\log t)^{\frac{2}{3}+\varepsilon}}\right)$$

$$\ll \exp\left(-(\log N)^{\frac{1}{10}}\right),$$

as wanted.

2·4. *Moments of Dirichlet polynomials*

We need Watt's result on the twisted fourth moment of zeta sums (see Subsection 1·1 for the definition of zeta sums). This bound comes into play when we estimate the mean square of a product of Dirichlet polynomials where one of the polynomials is a long zeta sum.

LEMMA 9 (Watt). *Let $T \geqslant T_0 \geqslant T^{\varepsilon}$, $T^{1+o(1)} \gg M$, $N \geqslant 1$. Define the Dirichlet polynomials $N(s) = \sum_{n \sim N} n^{-s}$ or $N(s) = \sum_{n \sim N} (\log n) n^{-s}$ and $M(s) = \sum_{m \sim M} a_m / m^s$ with a_m any complex numbers. We have*

$$\int_{T_0}^{T} |N(1+it)|^4 |M(1+it)|^2 dt \ll \left(\frac{T}{MN^2}(1 + M^2 T^{-\frac{1}{2}}) + \frac{1}{T_0^3}\right) T^{o(1)} \max_{m \sim M} |a_m|^2.$$

Proof. An easy partial summation argument shows that we may assume $N(s) = \sum_{n \sim N} n^{-s}$. The lemma will be reduced to Watt's original twisted moment result [**19**], where $N(s)$ is replaced with $\zeta(s)$. It is well–known that $|N(1+it)| \ll 1/t$ for $N \geqslant t \geqslant 1$ (see [**12**, chapter 8]), so

$$\begin{aligned}\int_{T_0}^{N} |N(1+it)|^4 |M(1+it)|^2 dt &\ll \max_{m \sim M} |a_m|^2 \int_{T_0}^{T} \frac{1}{t^4} dt \cdot T^{o(1)} \\ &\ll \frac{T^{o(1)}}{T_0^3} \max_{m \sim M} |a_m|^2.\end{aligned}$$

Now it suffices to consider the integrals over dyadic intervals $[U, 2U]$ with $N \leqslant U \leqslant T$. These are bounded as in [**1**, lemma 2] (using Watt's result and simple considerations), since translating the results there from the line $\Re(s) = 1/2$ to the line $\Re(s) = 1$ is an easy matter (and the bound in [**1**] should be multiplied by $\max_{m \sim M} |a_m|^2$, as we do not assume $|a_m| \leqslant 1$).

2·5. *Sieve estimates*

There are occasions in the proofs of Theorems 4 and 5 where our Dirichlet polynomials are too long, and we need a device for splitting them into shorter ones. This is enabled by Heath–Brown's identity and the decomposition resulting from it, which tells that either our Dirichlet polynomial can be replaced with a product of many polynomials, which is desirable, or it can be replaced with products of zeta sums, in which case we can make use of Watt's theorem.

Definition 1. A Dirichlet polynomial $M(s) = \sum_{m\sim M} a_m/m^s$ with $|a_n| \ll d_r(n)$ for fixed r is called prime-factored if, for each $A > 0$, we have $|M(1+it)| \ll_A (\log M)^{-A}$ for $\exp((\log M)^{\frac{1}{3}}) \leqslant t \leqslant M^{A\log\log M}$.

LEMMA 10 (Heath–Brown's decomposition). *Let an integer $k \geqslant 1$ and a real number $\delta > 0$ be fixed, and let $T \geqslant 2$. Define $P(s) = \sum_{P\leqslant p<P'} p^{-s}$ with $P \gg T^{\delta}$, $P' \in [P + P/\log T, 2P]$. There exist Dirichlet polynomials $G_1(s), \dots, G_L(s)$ and a constant $C > 0$ such that*

$$|P(1+it)| \ll (\log^C X)(|G_1(1+it)| + \cdots + |G_L(1+it)|) \quad \textit{for all} \quad t \in [-T, T],$$

with $L \leqslant \log^C X$, each $G_j(s)$ being of the form

$$G_j(s) = \prod_{i\leqslant J_j} M_i(s), \quad J_j \leqslant 2k,$$

with $M_i(s)$ prime-factored Dirichlet polynomials (which depend on j), whose lengths satisfy $M_1 \cdots M_J = X^{1+o(1)}$, $M_i \gg \exp(\log P/\log\log P)$. Additionally, each $M_i(s)$ with $M_i > X^{\frac{1}{k}}$ is a zeta sum.

Proof. For a similar bound, see Harman's book [**10**, chapter 7]. It suffices to prove an analogous result for the polynomial $\sum_{P\leqslant n<P'} \Lambda(n)n^{-s}$ and use summation by parts. We take $f(n) = n^{-1-it}1_{[P,P']}(n)$ in the general Heath–Brown identity [**11**] for $\sum_{n\leqslant N} f(n)\Lambda(n)$, splitting each resulting variables into dyadic intervals, and separating the variables with Perron's formula. The summation condition in Heath–Brown's identity guarantees that of the arising polynomials only the zeta sums can have length $> X^{\frac{1}{k}}$. If there are any polynomials of length $\ll \exp(\log P/\log\log P)$, these can simply be estimated trivially. The fact that the remaining polynomials of length $\gg \exp(\log P/\log\log P)$ are prime-factored follows from the fact that they have as their coefficients one of the sequences (1), $(\log n)$ and $(\mu(n))$, so that Lemma 8 gives a pointwise saving of $\ll_A (\log P)^{-A}$.

There is one more lemma that we need on the coefficients of Dirichlet polynomials arising from almost primes. We need to bound the following quantities that are related to the quantities occurring in the improved mean value theorem for Dirichlet polynomials.

Definition 2. For any sequence (a_n) of complex numbers, set $X_1 = \exp(\log X/(\log\log X)^4)$ and

$$S_1(X, (a_n)) = \max_{\substack{\frac{X}{X_1}\leqslant Y\leqslant 4X \\ 1\leqslant H\leqslant \log^{10} X}} H \sum_{Y\leqslant n\leqslant Y+\frac{Y}{H}} \frac{|a_n|^2}{n},$$

$$S_2(X, (a_n)) = \max_{\substack{\frac{X}{X_1}\leqslant Y\leqslant 4X \\ 1\leqslant H\leqslant \log^{10} X}} H \sum_{1\leqslant h\leqslant \frac{X}{T}} \sum_{Y\leqslant n\leqslant Y+\frac{Y}{H}} \frac{|a_n||a_{n+h}|}{n}.$$

We get bounds of size essentially $1/\log X$ and $X/T\log^2 X$ for $S_1(X,(a_n))$ and $S_2(X,(a_n))$, respectively, under the assumptions of the next lemma.

LEMMA 11. *Let $Z_r \geqslant \cdots \geqslant Z_1 \geqslant 1$ for a fixed r with $Z_r \geqslant \exp(\log X/(\log\log X)^3)$, $Z_r \leqslant z \leqslant 4X$, and*

$$\mathcal{Q} = \left\{n \leqslant 4X : n = p_1\cdots p_r m,\ \ p_i \in [Z_i, Z_i^2],\ \ (m, \mathcal{P}(z)) = 1)\right\}.$$

Let $|a_n| \leqslant 1_{\mathcal{Q}}(n)$, and let $S_1(X,(a_n))$ and $S_2(X,(a_n))$ be as defined above. Then

$$S_1(X,(a_n)) \ll \frac{1}{\log z} \quad \text{and} \quad S_2(X,(a_n)) \ll \frac{1}{\log^2 z}\cdot\frac{X}{T}.$$

Remark 7. Notice that we could also take as the set $\mathcal{Q}$ the set

$$\mathcal{Q}' = \left\{n \leqslant 4X : n = p_1\cdots p_r m,\ \ p_i \in [Z_i, Z_i^2],\ \ (m, \mathcal{P}(p_r)) = 1\right\}$$

or the set

$$\mathcal{Q}'' = \left\{n \leqslant 4X : n = p_1\cdots p_r,\ \ p_i \in [Z_i, Z_i^2]\right\}.$$

Indeed, the sizes of $\mathcal{Q}'$ and $\mathcal{Q}''$ can be bounded by sizes of sets of the form given in the lemma (with the parameter $z = Z_r$ or $z = X^{\frac{1}{r-1}}$). This observation will be used subsequently.

Proof. Let $S(A,\mathbb{P},z)$ count the numbers in A having no prime factors below z, and let Π be the product of all primes in $\bigcup_{i=1}^r [Z_i, Z_i^2] \cap [1,z]$. Brun's sieve yields

$$\begin{aligned}
S_1(X,(a_n)) &\ll \max_{\substack{\frac{X}{X_1}\leqslant Y\leqslant 4X\\ 1\leqslant H\leqslant \log^{10} X}} \frac{H}{Y}\cdot\left|\left[Y, Y+\frac{Y}{H}\right]\cap\mathcal{Q}\right| \\
&\ll \max_{\substack{\frac{X}{X_1}\leqslant Y\leqslant 4X\\ 1\leqslant H\leqslant \log^{10} X}} \frac{H}{Y}\cdot\left|\left\{n\in\left[Y, Y+\frac{Y}{H}\right] : \left(n, \frac{\mathcal{P}(z)}{\Pi}\right) = 1\right\}\right| \\
&\ll \max_{\substack{\frac{X}{X_1}\leqslant Y\leqslant 4X\\ 1\leqslant H\leqslant \log^{10} X}} \frac{H}{Y}\cdot\left(\frac{Y}{H\log z} + z^{\frac{1}{2}}\right) \\
&\ll \frac{1}{\log z},
\end{aligned}$$

since $z^{\frac{1}{2}} \leqslant (4X)^{\frac{1}{2}} \ll \frac{Y}{H\log^2 z}$.

Furthermore, Brun's sieve also yields

$$\begin{aligned}
S_2(X,(a_n)) &\ll \max_{\substack{X_1\leqslant Y\leqslant 4X\\ 1\leqslant H\leqslant \log^{10} X}} \frac{H}{Y}\sum_{1\leqslant h\leqslant \frac{X}{T}}\left|\left\{n\in\left[Y, Y+\frac{Y}{H}\right] : \left(n(n+h), \frac{\mathcal{P}(z)}{\Pi}\right) = 1\right\}\right| \\
&\ll \max_{\substack{X_1\leqslant Y\leqslant 4X\\ 1\leqslant H\leqslant \log^{10} X}} \frac{H}{Y}\cdot\sum_{1\leqslant h\leqslant \frac{X}{T}}\frac{h}{\varphi(h)}\left(\frac{Y}{H\log^2 z} + z^{\frac{1}{2}}\right) \\
&\ll \frac{1}{\log^2 z}\cdot\frac{X}{T},
\end{aligned}$$

by the elementary bound $\sum_{m\leqslant M} m/\varphi(m) \ll M$. This proves the statement.

3. *Mean squares of Dirichlet polynomials*

With all the necessary lemmas available, we are ready to present the propositions that quickly lead to Theorem 4 and are also necessary in proving Theorem 5.

PROPOSITION 1. *Let $X \geqslant 1, T \geqslant T_0 = X^{0.01}, 0 \leqslant \alpha_1 \leqslant 1$ and $1 \leqslant P \ll X^{o(1)}$, where P is a function of X. Define*

$$K(s) = \sum_{n \sim \frac{X}{P}} \frac{a_n}{n^s} \quad \text{and} \quad P(s) = \sum_{p \sim P} \frac{b_p}{p^s},$$

where a_n and b_p are arbitrary complex numbers. Denoting

$$\mathcal{T}_1 = \{t \in [T_0, T] : |P(1+it)| \leqslant P^{-\alpha_1}\}$$

we have

$$\int_{\mathcal{T}_1} |K(1+it)P(1+it)|^2 dt \ll \frac{T}{X} \cdot P^{1-2\alpha_1} \left(S_1\left(\frac{X}{P}, (a_n)\right) + S_2\left(\frac{X}{P}, (a_n)\right)\right).$$

Proof. The improved mean value theorem (Lemma 4) and definition of $\mathcal{T}_1$ give

$$\begin{aligned}
\int_{\mathcal{T}_1} |K(1+it)P(1+it)|^2 dt &\ll P^{-2\alpha_1} \int_{\mathcal{T}_1} |K(1+it)|^2 dt \\
&\ll P^{-2\alpha_1} \left(T \sum_{k \sim \frac{X}{P}} |a_k|^2 + T \sum_{1 \leqslant h \leqslant \frac{X}{PT}} \sum_{\substack{k,k' \sim \frac{X}{P} \\ k-k'=h}} |a_k||a_{k'}| \right) \\
&\ll P^{-2\alpha_1} \left(\frac{TP}{X} S_1\left(\frac{X}{P}, (a_n)\right) + \frac{TP}{X} S_2\left(\frac{X}{P}, (a_n)\right)\right) \\
&= \frac{T}{X} \cdot P^{1-2\alpha_1} \left(S_1\left(\frac{X}{P}, (a_n)\right) + S_2\left(\frac{X}{P}, (a_n)\right)\right),
\end{aligned}$$

which was the claim.

PROPOSITION 2. *Let $X \geqslant 1, T \geqslant T_0 = X^{0.01}$ and $1 \leqslant P \ll X^{o(1)}$. Also let $0 \leqslant \alpha_1, \alpha_2 \leqslant 1$ and let the Dirichlet polynomials $K(s)$ and $M(s)$ with $K = X/M \gg X^{\varepsilon}$ be*

$$K(s) = \sum_{n \sim K} \frac{a_n}{n^s} \quad \text{and} \quad M(s) = \sum_{m \sim M} \frac{c_m}{m^s},$$

where $|c_m| \leqslant d_r(m)$ for fixed r, and $|a_n| = 1_S(n)$ for some set S whose elements have at most r prime factors from $[P, 2P]$ and have no prime factors in $[1, X^{0.01}] \setminus \bigcup_{i=1}^{r} [Z_i, Z_i^2]$ for some $Z_i \geqslant 1$. Write

$$P(s) = \sum_{p \sim P} \frac{b_p}{p^s} \quad \text{with} \quad |b_p| \leqslant 1 \quad \text{and}$$

$$\mathcal{T} = \{t \in [T_0, T] : \ |P(1+it)| \geqslant P^{-\alpha_1} \text{ and } |M(1+it)| \leqslant M^{-\alpha_2}\}.$$

We have

$$\int_{\mathcal{T}} |K(1+it)M(1+it)|^2 dt \ll M^{-2\alpha_2} P^{(2+10\varepsilon)\alpha_1 \ell} \cdot (\ell!)^{1+o(1)} \cdot \left(\frac{T}{X} \cdot \frac{1}{\log X} + \frac{1}{\log^2 X} \right),$$

where $\ell = \lceil \log \frac{X}{K} / \log P \rceil$.

Remark 8. For products of three primes, our variables are picked so that the bound given by this proposition saves X^{ε} over the trivial bound. However, for products of $k \geqslant 4$ primes, our savings are much more modest, and the factor $(T/X) \cdot (1/\log X) + 1/\log^2 X$ becomes necessary.

Proof. This result is inspired by [**15**, lemma 13]. Using the fact that $|M(1+it)|^2 \leqslant M^{-2\alpha_2}(P^{\alpha_1}|P(1+it)|)^{2\ell}$ for $t \in \mathcal{T}$ and splitting polynomials into shorter ones, we have

$$\begin{aligned}\int_{\mathcal{T}} |K(1+it)M(1+it)|^2 dt &\ll M^{-2\alpha_2} P^{2\alpha_1 \ell} \int_{\mathcal{T}} |K(1+it)P(1+it)^{\ell}|^2 dt \\ &\ll M^{-2\alpha_2} P^{2\alpha_1 \ell} \ell^2 \int_{\mathcal{T}} |A(1+it)|^2 dt, \end{aligned} \tag{3·7}$$

where

$$A(s) = \sum_{n \sim Y} \frac{A_n}{n^s}$$

for some $K P^{\ell} \leqslant Y \leqslant 2K(2P)^{\ell}$ (so $X \leqslant Y \leqslant 2^{\ell} P X$), the coefficients A_n satisfying

$$|A_n| \leqslant \sum_{\substack{n = p_1 \cdots p_{\ell} m \\ p_i \sim P \\ m \sim K}} |a_m|.$$

By the improved mean value theorem (Lemma 4), we see that (3·7) is bounded by

$$\ll M^{-2\alpha_2} P^{2\alpha_1 \ell} \ell^2 \left(T \sum_{n \sim Y} \left| \frac{A_n}{n} \right|^2 + T \sum_{1 \leqslant h \leqslant \frac{Y}{T}} \sum_{m-n=h} \frac{|A_m||A_n|}{mn} \right).$$

Note that $A_n \neq 0$ implies that n has at most $\ell + r$ prime factors from $[P, 2P]$ and that n is coprime to

$$\Pi = \prod_{\substack{p \leqslant X^{0.01} \\ p \notin \bigcup_{i=1}^{r} [Z_i, Z_i^2] \cup [P, 2P]}} p.$$

Consequently, $|A_n| \leqslant (\ell + r)!$, and so

$$\begin{aligned}\sum_{n \sim Y} \left| \frac{A_n}{n} \right|^2 &\leqslant \frac{1}{Y} \cdot (\ell + r)! \sum_{n \sim Y} \frac{|A_n|}{n} \\ &\ll \frac{1}{Y} (\ell!)^{1+o(1)} \sum_{m \sim K} \frac{|a_m|}{m} \sum_{p_1, \ldots, p_{\ell} \sim P} \frac{1}{p_1 \cdots p_{\ell}} \\ &\ll (\ell!)^{1+o(1)} \cdot \frac{1}{Y} \sum_{\substack{m \sim K \\ (m, \Pi) = 1}} \frac{|a_m|}{m} \\ &\ll (\ell!)^{1+o(1)} \cdot \frac{1}{X \log X}, \end{aligned}$$

where the last step comes from Brun's sieve and the facts that $Y \geqslant X$ and $K \gg X^{\varepsilon}$.

To deal with the second sum arising from the improved mean value theorem, notice that by Brun's sieve the number of $n \leqslant y$ with $(n(kn+h), \Pi) = 1$ is $\ll (y/\log^2 y)(hk/\varphi(hk))$ with an absolute implied constant. Since $\varphi(ab) \geqslant \varphi(a)\varphi(b)$ and $k/\varphi(k) \leqslant 2^{\ell}$ when k has ℓ

prime factors, we have

$$\begin{aligned}
&\sum_{1\leqslant h\leqslant \frac{Y}{T}} \sum_{n\sim Y} \frac{|A_n||A_{n+h}|}{n(n+h)} \\
&\leqslant \frac{1}{Y^2}\cdot(\ell+r)! \sum_{1\leqslant h\leqslant \frac{Y}{T}} \sum_{p_1,\dots,p_\ell\sim P} \sum_{\substack{(m,\Pi)=1\\(p_1\cdots p_\ell m+h,\Pi)=1\\ m\leqslant \frac{2Y}{p_1\cdots p_\ell}}} 1 \\
&\ll \frac{1}{Y^2}\cdot(\ell!)^{1+o(1)} \sum_{1\leqslant h\leqslant \frac{Y}{T}} \sum_{p_1,\dots,p_\ell\sim P} \frac{Y}{p_1\cdots p_\ell \log^2 \frac{Y}{p_1\cdots p_\ell}} \frac{p_1\cdots p_\ell h}{\varphi(p_1\cdots p_\ell h)} \\
&\ll \frac{1}{Y\log^2 Y}(\ell!)^{1+o(1)} \sum_{1\leqslant h\leqslant \frac{Y}{T}} \frac{h}{\varphi(h)} \sum_{p_1,\dots,p_\ell\sim P} \frac{1}{p_1\cdots p_\ell} \\
&\ll \frac{1}{T}(\ell!)^{1+o(1)} \frac{1}{\log^2 X},
\end{aligned}$$

as desired.

PROPOSITION 3. *Let $X^{1+o(1)} \geqslant T \geqslant T_0 = X^{0.01}$ and $0 \leqslant \alpha_1 \leqslant 1$. Furthermore, let*

$$P(s) = \sum_{p\sim P} \frac{a_p}{p^s}, \quad \text{and} \quad M(s) = \sum_{M\leqslant q\leqslant M'} \frac{1}{q^s},$$

with $|a_p| \leqslant 1$, $M' \in [M + \frac{M}{\log P}, 2M]$, $\log X \leqslant P \ll X^{o(1)}$ and $PM = X^{1+o(1)}$, and let

$$\mathcal{U} = \{t \in [T_0, T] : |P(1+it)| \geqslant P^{-\alpha_1}\}.$$

Then, for $\ell = \lfloor \varepsilon(\log X/\log P)\rfloor$,

$$\begin{aligned}
&\int_{\mathcal{U}} |P(1+it)M(1+it)|^2 dt \\
&\quad \ll (P^{2\alpha_1-1}\log^2 X)^{(1+o(1))\ell} X^{o(1)} + (\log X)^{-100}\left(1 + \frac{|\mathcal{U}'|T^{\frac{1}{2}}}{X^{\frac{2}{3}-o(1)}}\right),
\end{aligned}$$

for some well-spaced set $\mathcal{U}' \subset \mathcal{U}$.

Proof. Heath–Brown's decomposition (Lemma 10) with $k = 3$ allows us to write, for some $C > 0$,

$$|M(1+it)| \ll (\log^C X)(|G_1(1+it)| + \cdots + |G_L(1+it)|),$$

with $L \leqslant \log^C X$. Here each $G_j(s)$ is either of the form

$$G_j(s) = M_1(s)M_2(s)M_3(s), \ M_1M_2M_3 = X^{1+o(1)},$$
$$M_1 \geqslant M_2 \geqslant M_3, \ M_3 \geqslant \exp\left(\frac{\log X}{2\log\log X}\right)$$

with $M_i(s)$ prime-factored polynomials, or of the form

$$G_j(s) = N_1(s)N_2(s), \ N_1N_2 = X^{1+o(1)}, \ N_1 \geqslant N_2,$$

with $N_i(s)$ zeta sums (it is possible that $N_2(s)$ is the constant polynomial 1^{-s}). It suffices to bound the contributions of the zeta sums and the prime-factored polynomials separately.

We look at the zeta sums first. We split the integration domain into dyadic intervals $[T_1, 2T_1]$ with $T_0 \leqslant T_1 \leqslant T$. Keeping in mind that $N_1 \geqslant X^{\frac{1}{2}-o(1)}$, $P^\ell = X^{\varepsilon+o(1)}$, and $|P(1+it)P^{\alpha_1}|^{2\ell} \geqslant 1$ for $t \in \mathcal{U}$, Cauchy–Schwarz and Watt's theorem (Lemma 9) yield

$$\begin{aligned}
&\int_{\mathcal{U}\cap[T_1,2T_1]} |P(1+it)N_1(1+it)N_2(1+it)|^2 dt \\
&\ll P^{2\alpha_1\ell} \int_{\mathcal{U}\cap[T_1,2T_1]} |N_1(1+it)N_2(1+it)P(1+it)^\ell|^2 dt \\
&\ll P^{2\alpha_1\ell} \left(\int_{T_1}^{2T} |N_1(1+it)|^4 |P(1+it)|^{4\ell} dt\right)^{\frac{1}{2}} \cdot \left(\int_{T_1}^{2T_1} |N_2(1+it)|^4 dt\right)^{\frac{1}{2}} \\
&\ll P^{2\alpha_1\ell} X^{o(1)} \left(\left(\frac{T_1 + T_1^{\frac{1}{2}} P^{4\ell}}{N_1^2 P^{2\ell}} + \frac{1}{T_1^3}\right)(2\ell)!^2\right)^{\frac{1}{2}} \cdot \left(\frac{T_1 + N_2^2}{N_2^2}\right)^{\frac{1}{2}} \\
&\ll P^{(2\alpha_1-1)\ell} X^{o(1)} \cdot (\ell!)^{2+o(1)} + \frac{P^{2\alpha_1\ell} X^{o(1)} (\ell!)^{2+o(1)}}{T_0} \\
&\ll (P^{2\alpha_1-1} \log^2 X)^{(1+o(1))\ell} X^{o(1)} + X^{-\varepsilon}.
\end{aligned}$$

Combining the contributions of the dyadic intervals simply multiplies this bound by $\log X$.

To bound the contribution of the prime-factored polynomials, we first observe that

$$\int_{\mathcal{U}} |P(1+it)M(1+it)|^2 dt \ll \sum_{t\in\mathcal{U}'} |P(1+it)M(1+it)|^2,$$

for some well-spaced $\mathcal{U}' \subset \mathcal{U}$. We make use of the Halász–Montgomery inequality (Lemma 5), and of the prime-factored property applied to the polynomial M_3 with length $M_3 \in [\exp(\log X/2\log\log X), X^{\frac{1}{3}+o(1)}]$, finding that

$$\begin{aligned}
&\sum_{t\in\mathcal{U}'} |P(1+it)M_1(1+it)M_2(1+it)M_3(1+it)|^2 \\
&\quad\ll (\log X)^{-100-D} \sum_{t\in\mathcal{U}'} |P(1+it)M_1(1+it)M_2(1+it)|^2 \\
&\quad\ll (\log X)^{-100-2C} \left(1 + \frac{T^{\frac{1}{2}}|\mathcal{U}'|}{X^{\frac{2}{3}-o(1)}}\right),
\end{aligned}$$

where D is so large that $D - 2C - 1$ exceeds the power of logarithm arising from the mean square of the coefficients of the divisor-bounded polynomial $P(s)M_1(s)M_2(s)$. Now the statement is proved.

4. *Proof of theorem 4*

The following proposition yields Theorem 4 (and hence Theorems 1 and 2) immediately, in view of the remarks of Subsection 2·1

PROPOSITION 4. *Let $k \geqslant 3$ be a fixed integer, $\varepsilon > 0$ be small enough and $T_0 = X^{0.01}$, as before. Define*

$$F(s) = \sum_{\substack{p_1\cdots p_k \sim X \\ P_i \leqslant p_i \leqslant P_i^{1+\varepsilon} \\ i\leqslant k-1}} (p_1\cdots p_k)^{-s},$$

where P_i are as in theorem 4. *Then, for $T \geqslant T_0$, we have*

$$\int_{T_0}^{T} |F(1+it)|^2 dt \ll \left(\frac{T P_1 \log X}{X} + 1\right) \cdot \frac{1}{(\log^2 X)(\log_k X)^3}. \tag{4·8}$$

Proof. We make use of the ideas introduced in the paper [**15**] by Matomäki and Radziwiłł. Trivially, we may assume $T \leqslant X^{1+o(1)}$. Let $H = (\log_k X)^3$,

$$Q_{v,H}(s) = \sum_{e^{\frac{v}{H}} \leqslant p < e^{\frac{v+1}{H}}} p^{-s}$$

and, for each $j = 1, ..., k$,

$$F_{v,H,j}(s) = \sum_{\substack{p_1 \cdots p_{j-1} p_{j+1} \cdots p_k \sim X e^{-\frac{v}{H}} \\ P_i \leqslant p_i \leqslant P_i^{1+\varepsilon},\, i \neq j,\, i \leqslant k-1}} (p_1 \cdots p_{j-1} p_{j+1} \cdots p_k)^{-s}.$$

Define $\alpha_1, ..., \alpha_{k-1}$ by $\alpha_j = 10 j\varepsilon$ for $j \leqslant k-2$, and $\alpha_{k-1} = 1/12 - \varepsilon$, with ε so small that $\alpha_{k-2} \leqslant \sqrt{\varepsilon}/10$. We split the domain of integration as $[T_0, T] = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \cdots \cup \mathcal{T}_{k-1} \cup \mathcal{T}$. We write $t \in \mathcal{T}_1$ if

$$|Q_{v,H}(1+it)| \leqslant e^{-\frac{\alpha_1 v}{H}},$$

for all $v \in I_1 = [H \log P_1, (1+\varepsilon) H \log P_1]$. We define recursively $t \in \mathcal{T}_j$ for $j = 2, ..., k-1$ if $t \notin \bigcup_{j' \leqslant j-1} \mathcal{T}_{j'}$ but

$$|Q_{v,H}(1+it)| \leqslant e^{-\frac{\alpha_j v}{H}},$$

for all $v \in I_j = [H \log P_j, (1+\varepsilon) H \log P_j]$. Finally, we write

$$\mathcal{T} = [T_0, T] \setminus \bigcup_{j=1}^{k-1} \mathcal{T}_j.$$

Lemma 2, with the notation of Subsection 2·5, yields

$$\begin{aligned} \int_{\mathcal{S}} |F(1+it)|^2 dt \ll H^2 (\log^2 P_j) \int_{\mathcal{S}} |Q_{v_j,H}(1+it) F_{v_j,H,j}(1+it)|^2 dt \\ + \frac{T}{HX}(S_1(X, (c_n)) + S_2(X, (c_n))), \end{aligned} \tag{4·9}$$

for some $v_j \in I_j$, and any $\mathcal{S} \subset [T_0, T]$. The coefficients c_n in the definitions of S_1 and S_2 are naturally the convolution of the absolute values of the coefficients of the polynomials $Q_{v_j,H}(s)$ and $F_{v_j,H,j}(s)$. By Lemma 11 and the remark related to it, the last two terms above contribute

$$\begin{aligned} &\ll \frac{T}{X} \cdot \frac{1}{H \log X} + \frac{1}{H \log^2 X} \\ &\ll \left(\frac{T P_1 \log X}{X} + 1\right) \cdot \frac{1}{H \log^2 X}. \end{aligned}$$

We choose $\mathcal{S} = \mathcal{T}_1, ..., \mathcal{T}_{k-1}, \mathcal{T}$ in (4·9). Summarizing, it suffices to estimate for each $j = 1, ..., k-1$ the quantity

$$B_j := H^2 (\log^2 P_j) \int_{\mathcal{T}_j} |Q_{v_j,H}(1+it) F_{v_j,H,j}(1+it)|^2 dt,$$

where $v_j \in [H \log P_j, (1+\varepsilon) H \log P_j]$ is chosen so that the integral is maximal, and additionally the quantity

$$B := H^2(\log^2 X) \int_{\mathcal{T}} |Q_{v_k,H}(1+it) F_{v_k,H,k}(1+it)|^2 dt,$$

where $v_k \in [H \log X/(P_1 \cdots P_{k-1})^{1+\varepsilon}, H \log 2X/P_1 \cdots P_{k-1}]$ is also picked so that the integral is maximised.

The integral over $\mathcal{T}_1$ is bounded with the help of Proposition 1. We take $K(s) = F_{v_1,H,1}(s)$ and $P(s) = Q_{v_1,H}(s)$. Now Lemma 11 and Proposition 1 result in

$$\begin{aligned} B_1 &\ll H^2(\log^2 P_1) P_1^{1+\varepsilon-2\alpha_1} \frac{T}{X} \left(\frac{1}{\log X} + \frac{X}{P_1 T} \cdot \frac{1}{\log^2 X} \right) \\ &\ll \left(\frac{T P_1 \log X}{X} + 1 \right) \cdot \frac{P_1^{10\varepsilon-2\alpha_1}}{\log^2 X}, \end{aligned}$$

and this is an admissible bound, since $\alpha_1 = 10\varepsilon$ and $P_1 \gg (\log_k X)^{\varepsilon^{-1}}$.

For the integral over $\mathcal{T}_j$ with $2 \leqslant j \leqslant k-1$ we use Proposition 2, with $K(s) = F_{v_j,H,j}(s)$, $M(s) = Q_{v_j,H}(s)$ and $P(s) = Q_{v_{j-1},H}(s)$, and for $\ell = \lceil \log P_j / \log P_{j-1} \rceil$ deduce

$$\begin{aligned} B_j &\ll H^2(\log^2 P_j) P_j^{-2\alpha_j} \cdot P_{j-1}^{(2+10\varepsilon)\alpha_{j-1}\ell} \\ &\quad \cdot (\ell!)^{1+o(1)} \cdot \left(\frac{T}{X \log X} + \frac{1}{\log^2 X} \right) \\ &\ll P_{j-1}^{10} P_j^{2(\alpha_{j-1}-\alpha_j)+10\varepsilon+(1+\varepsilon)\frac{\log\log P_j}{\log P_{j-1}}} \left(\frac{T P_1 \log X}{X} + 1 \right) \frac{1}{\log^2 X}. \end{aligned} \tag{4·10}$$

For $2 \leqslant j \leqslant k-2$, we have $\log\log P_j / \log P_{j-1} \leqslant 2\varepsilon$ and $\alpha_j - \alpha_{j-1} = 10\varepsilon$, so the definitions of P_{j-1} and P_j result in

$$B_j \ll \left(\frac{T P_1 \log X}{X} + 1 \right) \frac{1}{\log^2 X} (\log_k X)^{-3},$$

as wanted. For $j = k-1$, we have $\alpha_{k-2} \leqslant \sqrt{\varepsilon}/10$, $\alpha_{k-1} = 1/12 - \varepsilon$ and $P_{k-1} = (\log X)^{\varepsilon^{-2}}$, so taking $j = k-1$ in the above computation gives

$$B_{k-1} \ll P_{k-1}^{-\frac{1}{6}+\frac{1}{4}\sqrt{\varepsilon}+\frac{1+\varepsilon}{6+10\sqrt{\varepsilon}}} \ll P_{k-1}^{-\varepsilon} \ll (\log X)^{-\varepsilon^{-1}},$$

and therefore the case of $\mathcal{T}_{k-1}$ has been dealt with.

Finally, the integral over $\mathcal{T}$ is estimated using Proposition 3 with $P(s) = Q_{v_{k-1},H}(s)$ and $M(s) = Q_{v_k,H}(s)$. Denoting $\ell = \lfloor \varepsilon(\log X/\log P_{k-1}) \rfloor$ and separating by Perron's formula the variable p_{k-1} from the rest of the variables in $F_{v_k,H,k}(s)$ (and bounding the polynomial corresponding to the variables $p_1, \ldots, p_{k-2}$ by $\ll 1$), we see that

$$\begin{aligned} B &\ll H^2(\log^4 X) \int_{\mathcal{T}} |Q_{v_{k-1},H}(1+it) Q_{v_k,H}(1+it)|^2 dt \\ &\ll H^2(\log^4 X)(P_{k-1}^{-\frac{5}{6}+2\varepsilon} \log^2 X)^{(1+o(1))\ell} X^{o(1)} + (\log X)^{-95} \left(1 + \frac{|\mathcal{T}'| T^{\frac{1}{2}}}{X^{\frac{2}{3}-o(1)}} \right), \end{aligned}$$

for some well-spaced set $\mathcal{T}' \subset \mathcal{T}$. Since $P_{k-1} = (\log X)^{\varepsilon^{-2}}$, the first term is $\ll X^{-\frac{\varepsilon}{3}}$. In

addition, Lemma 6 allows us to bound the size of $\mathcal{T}'$ by

$$|\mathcal{T}'| \ll T^{2\alpha_{k-1}} P_{k-1}^2 X^{(\varepsilon^2+o(1))} \ll X^{\frac{1}{6}-\frac{\varepsilon}{2}},$$

because $\alpha_{k-1} = 1/12 - \varepsilon$. Therefore, the integral over $\mathcal{T}$ is $\ll (\log X)^{-95}$. In conclusion, we deduced the bound

$$B_1 + \cdots + B_{k-1} + B \ll \left(\frac{T P_1 \log X}{X} + 1\right) \cdot \frac{1}{H \log^2 X},$$

which finishes the proof of this proposition and of Theorem 4.

4·1. *A corollary on products of two primes*

As a byproduct of the methods above, we arrive at the exponent $c = 5 + \varepsilon$ for products of two primes, which already replicates Mikawa's exponent for P_2 numbers[2]. Similarly as for products of three or more primes, it suffices to prove

$$\int_{T_0}^{T} |F(1+it)|^2 dt = o\left(\left(\frac{T P_1 \log X}{X} + 1\right) \cdot \frac{1}{(\log X)^{2+\varepsilon}}\right),$$

where

$$F(s) = \sum_{\substack{p_1 p_2 \sim X \\ P_1 \leqslant p_1 < P_1^{1+\varepsilon}}} (p_1 p_2)^{-s}$$

and $P_1 = \log^a X$ with $a = 4 + \varepsilon$. We may again suppose $T \leqslant X^{1+o(1)}$.

We can redefine the set $\mathcal{T}_1$ in the proof of Proposition 4 with the new values $P_1 = \log^a X$, $H = (\log X)^{3\varepsilon}$, keeping the value $\alpha_1 = 10\varepsilon$, and we see again from Proposition 1 that the mean square of $F(1+it)$ over $\mathcal{T}_1$ is suitably small. For applying Propositions 2 and 3, we need more polynomials than the two that correspond to the variables p_1 and p_2 in (1·2). Indeed, Heath–Brown's decomposition (Lemma 10) enables splitting the polynomial corresponding to p_2 as $(\log X)^{O(1)}$ sums of the form $|M_1(s)M_2(s)| + |N_1(s)N_2(s)|$, where $M_1(s)$ and $M_2(s)$ are prime-factored Dirichlet polynomials with $M_1 M_2 = X^{1+o(1)}$, $\exp(\log X/2 \log\log X) \ll M_1 \ll X^{\frac{1}{3}+o(1)}$ and $N_1(s)$ and $N_2(s)$ zeta sums with $N_1 N_2 = X^{1+o(1)}$. The contribution of the zeta sums over the complement of $\mathcal{T}_1$ can be managed easily with Watt's theorem, similarly as in the proof of Proposition 3.

To estimate the contribution of the prime-factored polynomials $M_i(s)$, we redefine the set $\mathcal{T}_2$ as $\{t \in [T_0, T] : |M_1(1+it)| \leqslant M_1^{-\alpha_2}\} \setminus \mathcal{T}_1$, and Proposition 2 (with $P(s)$ corresponding to p_1 and $K(s) = M_1(s)M_2(s)$) produces a valid bound[3] in the $\mathcal{T}_2$ case, as long as $a \geqslant 1/2(\alpha_2 - \alpha_1) + 100\varepsilon$. We take $\alpha_2 = 1/8 - \varepsilon$, which turns out to be the best choice here.

Finally, when considering the integral over the complement of $\mathcal{T}_1 \cup \mathcal{T}_2$, instead of Proposition 3, we apply the simple inequality

$$\int_{\mathcal{T}} |M_1(1+it)M_2(1+it)|^2 dt \ll (\log X)^{-100} \left(1 + \frac{|\mathcal{T}'| T^{\frac{1}{2}}}{M_2}\right),$$

for some well-spaced $\mathcal{T}' \subset \mathcal{T}$, with $\mathcal{T} \subset [T_0, T]$ arbitrary. This inequality follows just from

[2] Adding to the argument a small refinement from Subsection 5·1, as well as Proposition 5, which is rather similar to Proposition 3, would already give c somewhat smaller than 5.

[3] This bound for a arises by inserting $P_{j-1} = \log^a X$ and $P_j = X^{1+o(1)}$ into formula (4·10).

the prime-factored property of $M_1(s)$ combined with the Halász–Montgomery inequality (Lemma 5). Now, denoting $M_1 = X^{\nu+o(1)}$, we need to have $|\mathcal{T}'| \ll X^{\frac{1}{2}-\nu-\varepsilon^2}$ whenever

$$\mathcal{T}' \subset \{t \in [T_0, T] : |M_1(1+it)| \geqslant M_1^{-\alpha_2}\}$$

is well spaced. Jutila's large values theorem (Lemma 7) applied with $F(s) = M_1(s)^\ell$, $V = M_1^{-(\frac{1}{8}-\varepsilon)\ell}$ and $k = 2$, $\ell \in \{2, 3\}$ tells that

$$|\mathcal{T}'| \ll \begin{cases} X^{\max\{\frac{\nu}{2},\ -\frac{11}{4}\nu+1,\ 1-4\nu\}-2\varepsilon^2} \\ X^{\max\{\frac{3}{4}\nu,\ -\frac{33}{8}\nu+1,\ 1-6\nu\}-2\varepsilon^2}. \end{cases}$$

We know that $\nu \leqslant 1/3 + o(1)$, and for $\frac{2}{7} \leqslant \nu \leqslant \frac{1}{3}$ the first bound is $\ll X^{\frac{1}{2}-\nu-\varepsilon^2}$, while for $4/25 \leqslant \nu \leqslant 2/7$ the second bound is small enough.

In the case $\nu \leqslant 4/25$, we may simply appeal to Lemma 6 to bound $|\mathcal{T}'|$ (with $V = M_1^{-\alpha_2}$), and get

$$|\mathcal{T}'| \ll T^{2\alpha_2} X^{2\nu\alpha_2+o(1)} \ll X^{0.29+100\varepsilon} \ll X^{\frac{1}{2}-\nu-\varepsilon},$$

for $\alpha_2 = \frac{1}{8} - \varepsilon$. This proves that $\alpha_2 = 1/8 - \varepsilon$ was permissible, leading to $a = 1/2\alpha_2 + C_1\varepsilon$, so the admissible exponent becomes $c = a + 1 \leqslant 5 + 2C_1\varepsilon$ (and $\varepsilon > 0$ was arbitrary). The rest of the paper therefore deals with improving the value $c = 5 + \varepsilon$ to $c = 3.51$, which will require several further ideas, along with the ones already introduced.

5. *Lemmas for theorem 5*

5·1. *Exponent pairs*

In the proof of Theorem 5, several zeta sums arise, and in some instances it is useful to have a smallish, pointwise power saving in these sums. This is given by the theory of exponent pairs. We could compute a long list of exponent pairs and choose the optimal estimate depending on the length of the zeta sum, but it turns out that using a single suitable exponent pair improves the exponent c for E_2 numbers by approximately 0.02, while having more of them would have very little additional advantage, and would complicate the calculations. Therefore, instead of formulating the general definition of exponent pairs (found in [**17**, chapter 3]), we write down the estimate coming from this specific pair.

LEMMA 12. *Let*

$$\sigma(\nu) = -\min\left\{\frac{1-\nu}{126} - \frac{\nu}{21}, 0\right\}.$$

Then we have

$$\sum_{n\in I} n^{-1-it} \ll t^{-\sigma(\nu)+o(1)},$$

for each $I = [N_1, N_2]$ *with* $t^\nu \leqslant N_1 \leqslant N_2 \ll t^{\nu+o(1)}$.

Proof. This follows immediately from the fact that $(1/126, 20/21)$ is an exponent pair. For the proof of this, see Montgomery's book [**17**, chapter 3].

5·2. *Lemmas on sieve weights*

For finding products of two primes on short intervals, we need some lemmas concerning sieve weights. In the cases of sums $\Sigma_1(h)$ and $\Sigma_2(h)$ in Subsection 6·2, there will be too few

variables for finding cancellation in the mean square of the corresponding Dirichlet polynomials. However, introducing sieve weights to these sums, we get an additional variable which is summed over all integers in a certain range, and separating that variable gives a long zeta sum (because there are few variables), and Watt's theorem can be applied to this sum. Also in the case of these sums, we need to make use of an additional saving of a logarithm in the mean value theorem. However, here the coefficients are not supported on almost primes but are closely related to the Dirichlet convolution $\lambda_n * 1$, where λ_n are the sieve weights. The sieve weights λ_n can be taken to be those of Brun's pure sieve. Specifically, we take

$$\lambda_d^+ = \begin{cases} \mu(d), & \nu(d) \leqslant R, d \mid \mathcal{P}(w) \\ 0, & \text{otherwise,} \end{cases} \qquad \lambda_d^- = \begin{cases} \mu(d), & \nu(d) \leqslant R+1, d \mid \mathcal{P}(w) \\ 0, & \text{otherwise,} \end{cases}$$

where the notations are as in Subsection 1·1, and

$$w = \exp\left(\frac{\log X}{(\log\log X)^3}\right) \quad \text{and} \quad R = 2\left\lfloor (\log\log X)^{\frac{3}{2}} \right\rfloor.$$

Since the support of $\lambda_n * 1$ contains in addition to almost primes only numbers having exceptionally many prime factors, we are able to save one logarithm factor in the mean values. This is done in the following lemma.

LEMMA 13. *Let λ_d^+ and λ_d^- be the sieve weights of Brun's pure sieve with the above notations. Let $k \geqslant 0$ be a fixed integer, $R_1, \ldots, R_k \geqslant 1$ and*

$$a_n = \sum_{\substack{p_1\cdots p_k \mid n \\ R_i \leqslant p_i \leqslant R_i^{1+\varepsilon}}} \left| \sum_{n = p_1\cdots p_k d m} \lambda_d^{\pm} \right|,$$

where either the sign $+$ or $-$ is chosen throughout (for $k = 0$, we define $p_1 \cdots p_k = 1$). Then for $y \gg_A (x/\log^A x)$ and $x \sim X$ we have

$$\sum_{x \leqslant n \leqslant x+y} |a_n|^2 \ll_A (\log\log X)^{O_k(1)} \frac{y}{\log X} \tag{5·11}$$

$$\sum_{1 \leqslant h \leqslant \frac{x}{T}} \sum_{\substack{m-n=h \\ m,n \in [x, x+y]}} |a_m||a_n| \ll_A (\log\log X)^{O_k(1)} \frac{X}{T} \cdot \frac{y^2}{\log^2 X}. \tag{5·12}$$

For the proof of this lemma, we need the follows two lemmas.

LEMMA 14. *For $x \geqslant 2$ and positive integer ℓ, let*

$$\pi_\ell(x) = |\{n \in [1, x] : \nu(n) = \ell\}|.$$

There exist absolute constants K and C such that

$$\pi_\ell(x) < \frac{Kx}{\log x} \frac{(\log\log x + C)^{\ell-1}}{(\ell-1)!}$$

for all ℓ and $x \geqslant 2$.

Proof. This is an elementary result of Hardy and Ramanujan from [**8**].

LEMMA 15. *Let $a \geqslant 1$ be fixed, and let $R = 2\lfloor(\log\log X)^{\frac{3}{2}}\rfloor$ as before. Then, for any $A > 0$,*

$$\sum_{\substack{n\sim X\\ \nu(n)\geqslant R}} a^{\nu(n)} \ll_{a,A} \frac{X}{\log^A X}.$$

Proof. The sum in question can be written as

$$\sum_{\ell\geqslant R} a^\ell |\{n\sim X : \nu(n) = \ell\}|,$$

and, by Lemma 14, this is

$$\begin{aligned}&\ll \frac{X}{\log X}\sum_{\ell\geqslant R}\left(\frac{ae(\log\log X + C)}{\ell-1}\right)^{\ell-1}\\ &\ll_a X\cdot 2^{-R} \ll_A \frac{X}{\log^A X}\end{aligned}$$

by the definition of R.

We can now proceed to proving Lemma 13.

Proof of Lemma 13. It suffices to consider the lower bound sieve weights. We assume $k \geqslant 1$, as the case $k = 0$ is similar but a little simpler. Define $\theta_n = 1 * \lambda_n^-$. We have

$$\begin{aligned}\theta_n &= \sum_{\substack{d|n\\ \nu(d)\leqslant R\\ d|\mathcal{P}(w)}} \mu(d)\\ &= \sum_{d|(n,\mathcal{P}(w))} \mu(d) + O\left(\sum_{\substack{d|n\\ \nu(d)>R}} |\mu(d)|\right)\\ &= 1_{(n,\mathcal{P}(w))=1} + O(2^{\nu(n)}1_{\nu(n)>R}).\end{aligned}$$

Using this, we bound the sum (5·11). Denoting by Π the product of all the primes in $\bigcup_{i=1}^k [R_i, R_i^{1+\varepsilon}] \cap [1, w]$, we observe that

$$a_n = \sum_{\substack{p_1\cdots p_k | n\\ R_i\leqslant p_i\leqslant R_i^{1+\varepsilon}}} |\theta_{\frac{n}{p_1\cdots p_k}}| \leqslant \nu(n)^k(1_{(n,\frac{\mathcal{P}(w)}{\Pi})=1} + 2^{\nu(n)}1_{\nu(n)>R}). \tag{5·13}$$

The contribution of the first term on the right-hand side of (5·13) to the sum (5·11) is

$$\ll \sum_{\substack{x\leqslant n\leqslant x+y\\ (n,\frac{\mathcal{P}(w)}{\Pi})=1}} \nu(n)^{2k} \ll (\log\log X)^{O_k(1)} \sum_{\substack{x\leqslant n\leqslant x+y\\ (n,\frac{\mathcal{P}(w)}{\Pi})=1}} 1 \ll (\log\log X)^{O_k(1)}\frac{y}{\log X}$$

by Brun's sieve and the fact that $\nu(n) \ll (\log\log X)^3$ when $(n, \mathcal{P}(w)) = 1$. On the other hand, the the second term on the right-hand side of (5·13) contributes to (5·11) at most

$$\ll \sum_{\substack{x\leqslant n\leqslant x+y\\ \nu(n)\geqslant R}} \nu(n)^{2k}4^{\nu(n)} \ll_k \sum_{\substack{x\leqslant n\leqslant x+y\\ \nu(n)\geqslant R}} 5^{\nu(n)} \ll_{A,k} \frac{X}{\log^A X} \tag{5·14}$$

by Lemma 15. This proves the first bound in Lemma 13.

The second bound in Lemma 13 is proved analogously. The two terms in (5·13) can be combined in four ways into products of two terms (two of these are symmetric). One of the cases contributes to (5·12) at most

$$\ll \sum_{1\leqslant h\leqslant \frac{x}{T}} \sum_{\substack{m-n=h\\ m,n\in[x,x+y]}} \nu(m)^k \nu(n)^k 1_{(m,\frac{\mathcal{P}(w)}{\Pi})=1} 1_{(n,\frac{\mathcal{P}(w)}{\Pi})=1} \ll (\log\log X)^{O_k(1)} \frac{X}{T}\cdot\frac{y^2}{\log^2 X}$$

by Brun's sieve. The two symmetric terms obtained by multiplying terms in (5·13) have an impact of

$$\ll \sum_{1\leqslant h\leqslant \frac{x}{T}} \sum_{\substack{m-n=h\\ m,n\in[x,x+y]}} \nu(m)^k \nu(n)^k 1_{(m,\frac{\mathcal{P}(w)}{\Pi})=1} 2^{\nu(n)} 1_{\nu(n)>R},$$

where the coefficients depending on m can be bounded trivially, while the coefficients depending on n save an arbitrary power of logarithm, as in formula (5·14). Finally, the fourth term arising from multiplication of (5·13) also saves an arbitrary power of logarithm by the same argument.

6. *Proof of theorem 5*

Before proving Theorem 5, we need some preparation. Define

$$S_h(x) = \sum_{\substack{x\leqslant p_1 p\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}} 1, \quad S_X = S_X(X)$$

and set

$$w = \exp\left(\frac{\log X}{(\log\log X)^3}\right).$$

We use Buchstab's identity twice to decompose

$$\begin{aligned} S_h(x) &= \sum_{\substack{x\leqslant p_1 n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 - \sum_{\substack{x\leqslant p_1q_1 n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ w\leqslant q_1<\sqrt{x}\\ (n,\mathcal{P}(q_1))=1\\ n>1}} 1 \\ &= \sum_{\substack{x\leqslant p_1 n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 - \sum_{\substack{x\leqslant p_1q_1 n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ w\leqslant q_1<\sqrt{x}\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 + \sum_{\substack{x\leqslant p_1q_1q_2 n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ w\leqslant q_2<q_1<\sqrt{x}\\ (n,\mathcal{P}(q_2))=1\\ n>1}} 1. \end{aligned}$$

Call these sums $\Sigma_1(h)$, $\Sigma_2(h)$ and $\Sigma_3(h)$, respectively, and call the corresponding dyadic sums $\Sigma_1(X)$, $\Sigma_2(X)$ and $\Sigma_3(X)$, respectively. We will divide $\Sigma_3(h)$ into two parts $\Sigma_3'(h)$ and $\Sigma_3''(h)$ in such a way that $\Sigma_1(h)$, $\Sigma_2(h)$ and $\Sigma_3'(h)$ can be evaluated asymptotically,

while the error from $\Sigma_3''(h)$ is manageable. To be precise, we will prove that

$$\begin{aligned}\frac{1}{h}S_h(x) &= \frac{1}{h}(\Sigma_1(h) - \Sigma_2(h) + \Sigma_3'(h) + \Sigma_3''(h)) \\ &= \frac{1}{X}(\Sigma_1(X) - \Sigma_2(X) + \Sigma_3'(X)) + \frac{1}{h}\Sigma_3''(h) + o\left(\frac{1}{\log X}\right) \qquad (6{\cdot}15) \\ &= \frac{1}{X}S_X + \frac{1}{h}\Sigma_3''(h) - \frac{1}{X}\Sigma_3''(X) + o\left(\frac{1}{\log X}\right) \\ &\geqslant \frac{1}{X}S_X - \frac{1}{X}\Sigma_3''(X) + o\left(\frac{1}{\log X}\right) \\ &\geqslant \varepsilon \cdot \frac{1}{X}S_X \qquad (6{\cdot}16)\end{aligned}$$

almost always, with the steps (6·15) and (6·16) being the nontrivial ones. This estimate will then immediately lead to Theorem 5. To prove these statements, we require some auxiliary results for the cases of $\Sigma_1(h)$, $\Sigma_2(h)$ and $\Sigma_3(h)$.

6·1. *Mean square bounds related to Theorem 5*

We need three additional mean square bounds to deal with the sums $\Sigma_1(h)$, $\Sigma_2(h)$ and $\Sigma_3(h)$. The first is a relative of Proposition 3 and would already improve slightly the exponent $c = 5 + \varepsilon$ obtained from the proof of Theorem 4. It will not be applied directly in the proof of Theorem 5, but instead as an ingredient in the proof of Proposition 7.

PROPOSITION 5. *Let $X^{1+o(1)} \geqslant T \geqslant T_0 = X^{0.01}$, and $0 \leqslant \alpha_1 \leqslant 1$. Furthermore, let*

$$P(s) = \sum_{P \leqslant p \leqslant P'} \frac{1}{p^s}, \quad M(s) = \sum_{m \sim M} \frac{b_m}{m^s},$$

with $P = X^{\nu+o(1)}$, $P' \in \left[P + \frac{P}{\log X}, 2P\right]$, $0 < \nu \leqslant 1/2$, $|b_m| \leqslant d_r(m)$ for fixed r, and $PM = X^{1+o(1)}$. Also let

$$\mathcal{U} = \{t \in [T_0, T] : |P(1+it)| \geqslant P^{-\alpha_1}\}.$$

Then,

$$\int_{\mathcal{U}} |P(1+it)M(1+it)|^2 dt \ll (\log X)^{-100} + X^{\frac{1}{2}-\min\{2\sigma(\nu), \frac{\nu}{2}\}+o(1)} \cdot \frac{|\mathcal{U}'|P}{X}$$

for some well-spaced $\mathcal{U}' \subset \mathcal{U}$.

Proof. Note that Heath–Brown's decomposition (Lemma 10) gives

$$|P(1+it)| \ll (\log^C X)(|G_1(1+it)| + \cdots + |G_L(1+it)|),$$

with $L \leqslant \log^C X$ and each $G_j(s)$ either of the form $G_j(s) = N(s)$ with $N(s)$ a zeta sum of length $P^{1-o(1)}$, or $G_j(s) = M_1(s)M_2(s)$ with M_1 and M_2 prime-factored polynomials of length $M_1 \geqslant M_2 \geqslant \exp(\log X/\log\log X)$, $M_1M_2 = P^{1-o(1)}$. To bound the contribution of the zeta sum, we divide the integral over $\mathcal{U}$ into integrals over dyadic intervals $[T_1, 2T_1]$ with $T_1 \in [T_0, T]$, and write $N = T_1^{\mu+o(1)}$ with $\mu \geqslant \nu$. If $\mu > 1$, we know that $|N(1+it)| \ll \log t/t$ and $M(1+it) \ll (\log X)^{O(1)}$, so

$$\int_{\mathcal{U}\cap[T_1,2T_1]} |M(1+it)N(1+it)|^2 dt \ll \frac{(\log X)^{O(1)}}{T_0}.$$

If $\mu \leqslant 1$, we first pick a well-spaced $\mathcal{U}' \subset \mathcal{U}$ such that

$$\int_{\mathcal{U}} |M(1+it)N(1+it)|^2 dt \ll \sum_{t \in \mathcal{U}'} |M(1+it)N(1+it)|^2.$$

Now the Halász–Montgomery inequality and the the fact that $N(s)$ is a zeta sum give

$$\begin{aligned}\sum_{t \in \mathcal{U}' \cap [T_1, 2T_1]} |M(1+it)N(1+it)|^2 &\ll T^{-2\sigma(\nu)+o(1)} \sum_{t \in \mathcal{U}' \cap [T_1, 2T_1]} |M(1+it)|^2 \\ &\ll T^{-2\sigma(\nu)+o(1)} \left(1 + \frac{|\mathcal{U}'| T_1^{\frac{1}{2}+o(1)}}{\frac{X}{P}}\right).\end{aligned}$$

To deal with the contribution of the prime-factored polynomials $M_i(s)$, we may use the Halász–Montgomery inequality in a manner analogous to the above to obtain the estimate

$$\int_{\mathcal{U}} |M(1+it)M_1(1+it)M_2(1+it)|^2 dt \ll (\log X)^{-100} \left(1 + \frac{|\mathcal{U}'| T^{\frac{1}{2}+o(1)}}{\frac{X}{P^{\frac{1}{2}}}}\right),$$

since $MM_1 \gg X^{1+o(1)}/P^{\frac{1}{2}}$ Taking the maximum of these two results produces the claimed bound.

Our second mean square bound is a type I estimate where we exploit a long zeta sum with the help of Watt's theorem. In the cases of $\Sigma_1(h)$ and $\Sigma_2(h)$, this is necessary, and in the case $\Sigma_3(h)$ it improves our exponent for Theorem 5. A closely related estimate can be found for example in [**10**, chapter 9].

PROPOSITION 6. *Let $X^{1+o(1)} \gg T \geqslant T_0$, and let $M(s), N(s), P(s)$ be Dirichlet polynomials with coefficients bounded by $X^{o(1)}$ and supported on the intervals $[M, 2M], [N, 2N], [P, 2P]$, respectively. Denote $Q(s) = \sum_{m \sim Q} \frac{a_m}{m^s}$, and let $N(s)$ be a zeta sum. Suppose in addition that*

$$MNP = X^{1+o(1)}, \ PQ^2 \leqslant X^{\frac{1}{4}}, \ M^2 P \ll X^{1+o(1)}.$$

Then

$$\int_{T_0}^{T} |M(1+it)N(1+it)P(1+it)Q(1+it)|^2 dt \ll X^{o(1)} \left(Q^{-1} + \frac{1}{T_0}\right) \max_{m \sim Q} |a_m|^2.$$

Remark 9. In all our applications, the polynomial $Q(s)$ has length essentially X^{ε}, and it is used to win by X^{ε^2}, say, in our estimates.

Proof. We will reduce the proposition to Watt's theorem (Lemma 9). Divide the integration domain into dyadic intervals $[T_1, 2T_1]$. By Cauchy–Schwarz, the mean value theorem

and Watt's theorem, we see that

$$\begin{aligned}
&\int_{T_1}^{2T1} |M(1+it)N(1+it)P(1+it)Q(1+it)|^2 dt \\
&\ll \left(\int_{T_1}^{2T1} |N(1+it)|^4 |P(1+it)Q(1+it)^2|^2 dt\right)^{\frac{1}{2}} \\
&\quad \cdot \left(\int_{T_1}^{2T_1} |M(1+it)|^4 |P(1+it)|^2 dt\right)^{\frac{1}{2}} \\
&\ll \left(\left(\frac{T_1^{o(1)}(T_1 + T_1^{\frac{1}{2}}P^2Q^4)}{N^2PQ^2} + \frac{T^{o(1)}}{T_1^3}\right)\max_{m\sim Q}|a_m|^4\right)^{\frac{1}{2}} \left(\frac{T_1 + M^2P}{M^2P}\right)^{\frac{1}{2}} \\
&\ll \left(\left(\frac{T_1^{o(1)}(T_1 + T_1^{\frac{1}{2}}P^2Q^4)}{N^2PQ^2}\right)\max_{m\sim Q}|a_m|^4\right)^{\frac{1}{2}} \left(\frac{T_1 + M^2P}{M^2P}\right)^{\frac{1}{2}} + \frac{X^{o(1)}}{T_0}\max_{m\sim Q}|a_m|^2.
\end{aligned}$$

Hence, we need

$$(X + X^{\frac{1}{2}}P^2Q^4)(X + M^2P) \ll (MNPX^{o(1)})^2$$

and this is guaranteed by our conditions.

For the $\Sigma_3(h)$ case in Subsection 6·3, we also need the following mean square bound, which is somewhat analogous to Proposition 4 and is based on Propositions 1, 2 and 5, but it will be clear only later how it is crucial for proving Theorem 5.

PROPOSITION 7. *Let $0 \leqslant \nu \leqslant 1/2$, $0 < \alpha_2 \leqslant 1$, $a = 1/2\alpha_2 + C_2\varepsilon$, $P_1 = \log^a X$, $X^{1+o(1)} \geqslant T \geqslant T_0 = X^{0.01}$, and $w \leqslant P_2 = X^{\nu+o(1)}$ with $w = \exp(\log X/(\log\log X)^3)$. Also let*

$$G(s) = \sum_{\substack{p_1p_2p_3n\sim X \\ P_i\leqslant p_i\leqslant P_i^{1+\varepsilon},\, i\leqslant 2 \\ p_2<p_3 \\ (n,P(p_2))=1 \\ n>1}} a_n(p_1p_2p_3n)^{-s},$$

where $|a_n| \ll (\log X)^{\varepsilon}$. Suppose that for every Dirichlet polynomial $M(s) = \sum_{m\sim M} b_m/m^s$ with $|b_m| \leqslant d_r(m)$ for fixed r and $M = X^{\nu+o(1)}$ any well-spaced set

$$\mathcal{U}' \subset \{t \in [0,T] : |M(1+it)| \geqslant M^{-\alpha_2}\}$$

satisfies $|\mathcal{U}'| \ll X^{\frac{1}{2}-\nu+\min\{2\sigma(\nu),\frac{\nu}{2}\}-\varepsilon}$. Then we have

$$\int_{T_0}^{T} |G(1+it)|^2 dt \ll \left(\frac{TP_1\log X}{X} + 1\right)\frac{1}{\log^{2+\varepsilon} X}.$$

Proof. Let $\alpha_1 = 100\varepsilon$ and define $H = \log^{10\varepsilon} X$. Let

$$Q_{v,H,1}(s) = \sum_{e^{\frac{v}{H}}\leqslant p_1<e^{\frac{v+1}{H}}} p_1^{-s}, \quad Q_{v,H,2}(s) = \sum_{e^{\frac{v}{H}}\leqslant p_2<e^{\frac{v+1}{H}}} p_2^{-s}$$

and

$$G_{v,H,1}(s) = \sum_{\substack{p_2p_3p_4m\sim Xe^{-\frac{v}{H}}\\ P_2\leqslant p_2\leqslant P_2^{1+\varepsilon}\\ p_2<p_3,\, p_2\leqslant p_4\\ (m,\mathcal{P}(p_4))=1}} a_{p_4m}(p_2p_3p_4m)^{-s},$$

$$G_{v,H,2}(s) = \sum_{\substack{p_1p_3p_4m\sim Xe^{-\frac{v}{H}}\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ (m,\mathcal{P}(p_4))=1}} a_{p_4m}(p_1p_3p_4m)^{-s}.$$

For $j = 1, 2$, we have

$$\int_{\mathcal{S}} |G(1+it)|^2dt \ll \left(\frac{TP_1\log X}{X}+1\right)\frac{1}{\log^{2+\varepsilon}X} + H^2(\log^2 P_j)(\log^{10(j-1)}X)\int_{\mathcal{S}}|Q_{v_j,H,j}(1+it)G_{v_j,H,j}(1+it)|^2dt$$

for some $v_j \in [H\log P_j, (1+\varepsilon)H\log P_j]$ and any measurable $\mathcal{S}\subset[T_0,T]$. In the case $j = 1$, this follows from Lemmas 2 and 11, while in the case $j = 2$, we use Perron's formula to separate the variables in $G(s)$. We partition $[T_0, T]$ as $\mathcal{T}_1\cup\mathcal{T}_2\cup\mathcal{T}$ with

$$\mathcal{T}_1 = \{t\in[T_0,T] : |Q_{v_1,H,1}(1+it)| \leqslant P_1^{-\alpha_1}\},$$
$$\mathcal{T}_2 = \{t\in[T_0,T] : |Q_{v_2,H,2}(1+it)| \leqslant P_2^{-\alpha_2}\}\setminus\mathcal{T}_1,$$

and $\mathcal{T} = [T_0,T]\setminus(\mathcal{T}_1\cup\mathcal{T}_2)$.

What remains to be done is estimating the integrals

$$B_j = H^2(\log^2 P_j)(\log^{10(j-1)}X)\int_{\mathcal{T}_j}|Q_{v_j,H,j}(1+it)G_{v_j,H,j}(1+it)|^2dt,$$

for $j = 1, 2$, as well as

$$B = H^2(\log^{10}X)\int_{\mathcal{T}}|Q_{v_2,H,2}(1+it)G_{v_2,H,2}(1+it)|^2dt.$$

We have $B_1 \ll (TP_1\log X/X+1)(P_1^{10\varepsilon-\alpha_1}/\log^2 X)$ by Proposition 1 and Lemma 11, and this is small enough since $\alpha_1 = 100\varepsilon$. We also have, by Proposition 2 with $\ell = \lceil\log P_2/\log P_1\rceil$,

$$\begin{aligned} B_2 &\ll H^2(\log^{20}X)P_2^{-2\alpha_2}P_1^{(2+10\varepsilon)\alpha_1\ell}\ell^{(1+o(1))\ell}\\ &\ll P_2^{2(\alpha_1-\alpha_2)+20\varepsilon+\frac{1+2\varepsilon}{a}}\\ &\ll P_2^{-\varepsilon} \ll (\log X)^{-100}, \end{aligned}$$

as long as $a\geqslant 1/2(\alpha_2-\alpha_1)+C_2\varepsilon/2$, say. Lastly, Proposition 5 gives, for some well-spaced $\mathcal{U}'$ of the type mentioned in the proposition,

$$B \ll (\log X)^{-50} + X^{\frac{1}{2}-\min\{2\sigma(v),\frac{v}{2}\}+o(1)}\frac{|\mathcal{U}'|X^{v+o(1)}}{X} \ll (\log X)^{-50}$$

by our assumption on $\mathcal{U}'$. Now the proof is complete.

6·2. *Cases of* $\Sigma_1(h)$ *and* $\Sigma_2(h)$

Let λ_d^+ and λ_d^- be the sieve weights of Brun's pure sieve with $R = 2\lfloor(\log\log X)^{\frac{3}{2}}\rfloor$ and sieving parameter $w = \exp(\log X/(\log\log X)^3)$. We have

$$\sum_{\substack{x\leqslant p_1dn\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^- \leqslant \sum_{\substack{\frac{x}{p_1}\leqslant n\leqslant\frac{x+h}{p_1}\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 = \Sigma_1(h) \leqslant \sum_{\substack{x\leqslant p_1dn\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^+.$$

We consider the lower bound; the upper bound can be considered similarly. Letting $X_1 = X/T_0^3$ with $T_0 = X^{0.01}$, we have

$$\frac{h}{X_1}\sum_{\substack{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ d|\mathcal{P}(w)}}\lambda_d^- \sum_{\frac{X}{p_1d}\leqslant n\leqslant\frac{X+X_1}{p_1d}} 1 = h\sum_{\substack{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ d|\mathcal{P}(w)}}\frac{\lambda_d^-}{p_1d} + O\left(\frac{h}{X_1}w^R P_1^{1+\varepsilon}\right),$$

so

$$\Sigma_1(h) \geqslant \sum_{\substack{d|\mathcal{P}(w)\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^-\frac{h}{p_1d} + \left(\sum_{\substack{x\leqslant p_1dn\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^- - \frac{h}{X_1}\sum_{\substack{X\leqslant p_1dn\leqslant X+X_1\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^-\right) \tag{6·17}$$
$$+ O\left(\frac{1}{\log^{100}X}\right).$$

By the fundamental lemma of the sieve (see e.g. [**4**, chapter 6]), we further deduce that

$$\sum_{\substack{d|\mathcal{P}(w)\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^-\frac{h}{p_1d} = (1+O((\log X)^{-100}))\sum_{\substack{d|\mathcal{P}(w)\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^+\frac{h}{p_1d}$$
$$\geqslant \frac{h}{X}\Sigma_1(X) + O\left(\frac{h}{\log^{100}X}\right).$$

Therefore, we may concentrate on the expression in the parentheses in (6·17), which is a difference between a short and long average. By Lemma 1, it is $o(h/\log X)$ for $h \geqslant P_1\log X$ and for almost all $x \leqslant X$, provided that

$$\int_{T_0}^{T}|F(1+it)|^2dt = o\left(\left(\frac{TP_1\log X}{X}+1\right)\frac{1}{\log^2 X}\right),$$

for all $T \geqslant T_0$, where $T_0 = X^{0.01}$, and

$$F(s) = \sum_{\substack{p_1dn\sim X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^-(p_1dn)^{-s}.$$

Such an estimate is given by the following proposition, which is invoked again in the case of the sum $\Sigma_2(h)$.

PROPOSITION 8. *Let* $\varepsilon > 0$, $P_1 = \log^a X$ *with* $a \geqslant 2 + C_3\varepsilon$ *and*

$$F(s) = \sum_{\substack{p_1dn\sim X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^{\pm}(p_1dn)^{-s} \quad \text{or} \quad F(s) = \sum_{\substack{p_1pdn\sim X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ M\leqslant p\leqslant M^{1+\varepsilon}}}\lambda_d^{\pm}(p_1pdn)^{-s},$$

with $M \ll X^{\frac{1}{2}+o(1)}$, $X^{1+o(1)} \gg T \geqslant T_0 = X^{0.01}$ as before, and either $+$ or $-$ sign chosen throughout. Then,

$$\int_{T_0}^{T} |F(1+it)|^2 dt \ll \left(\frac{T P_1 \log X}{X} + 1\right) \frac{1}{\log^{2+\varepsilon} X}.$$

Proof. Let D be a large constant, and for positive integer v and $H = \log^{10\varepsilon} X$ denote

$$P_{v,H}(s) = \sum_{e^{\frac{v}{H}} \leqslant p < e^{\frac{v+1}{H}}} p^{-s}$$

and

$$F_{v,H}(s) = \sum_{dn \sim X e^{-\frac{v}{H}}} \lambda_d^{\pm} (dn)^{-s} \quad \text{or} \quad F_{v,H}(s) = \sum_{\substack{pdn \sim X e^{-\frac{v}{H}} \\ M \leqslant p \leqslant M^{1+\varepsilon}}} \lambda_d^{\pm} (pdn)^{-s}.$$

Lemma 2 gives

$$\int_{T_0}^{T} |F(1+it)|^2 dt \ll H^2 (\log\log X)^2 \int_{T_0}^{T} |P_{v_0,H}(1+it) F_{v_0,H}(1+it)|^2 dt + T \sum_{n \in [X e^{-\frac{1}{H}}, X e^{\frac{1}{H}}] \text{ or } n \in [2X, 2X e^{\frac{1}{H}}]} |a_n|^2 + T \sum_{1 \leqslant h \leqslant \frac{X}{T}} \sum_{\substack{m-n=h \\ m,n \in [X e^{-\frac{1}{H}}, X e^{\frac{1}{H}}] \text{ or} \\ m,n \in [2X, 2X e^{\frac{1}{H}}]}} |a_m||a_n|, \quad (6{\cdot}18)$$

for some $v_0 \in I_0$, where $I_0 = [H \log P_1, H \log P_1^{1+\varepsilon}]$ and

$$a_m = \sum_{\substack{p_1 | m \\ P_1 \leqslant p_1 \leqslant P_1^{1+\varepsilon}}} \left| \sum_{m = p_1 d n} \lambda_d^{\pm} \right| \quad \text{or} \quad a_m = \sum_{\substack{p_1 | m \\ P_1 \leqslant p_1 \leqslant P_1^{1+\varepsilon}}} \left| \sum_{\substack{m = p_1 pdn \\ M \leqslant p \leqslant M^{1+\varepsilon}}} \lambda_d^{\pm} \right|. \quad (6{\cdot}19)$$

Lemma 13 tells that the last two terms in (6·18) contribute, for some constant $C > 0$,

$$\ll \frac{T}{X} \left(\frac{(\log\log X)^C}{H} \cdot \frac{1}{\log X} + \frac{(\log\log X)^C}{H} \cdot \frac{X}{T} \cdot \frac{1}{\log^2 X} \right)$$
$$\ll \left(\frac{T P_1 \log X}{X} + 1 \right) \cdot \frac{1}{\log^{2+\varepsilon} X}$$

by the definition of H. We are now left with estimating the integral in (6·18). We consider the integrals in two parts, namely the part over $\mathcal{T}_1$ and its complement, with

$$\mathcal{T}_1 = \{t \in [T_0, T] : |P_{v_0,H}(1+it)| \leqslant P_1^{-100\varepsilon}\}.$$

The case of $\mathcal{T}_1$ is dealt with Proposition 1 and Lemma 13, and it contributes

$$\ll H^2 (\log\log X)^2 \frac{T}{X} P_1^{1-200\varepsilon} \left(S_1\left(\frac{X}{P_1}, (a_n)\right) + S_2\left(\frac{X}{P_1}, (a_n)\right) \right)$$
$$\ll (\log\log X)^C \left(\frac{T}{X} \cdot \frac{1}{\log X} + \frac{1}{P_1} \cdot \frac{1}{\log^2 X} \right) \cdot P_1^{1-100\varepsilon}$$
$$\ll \left(\frac{T P_1 \log X}{X} + 1 \right) \cdot \frac{1}{\log^{2+\varepsilon} X},$$

where the coefficients a_n involved in definition of $S_i(X, (a_n))$ are given by (6·19).

We turn to the integral over the complement of $\mathcal{T}_1$ and resort to the Watt-type Proposition 6. Let ℓ be a large positive integer such that $P_1^\ell = X^{\varepsilon+o(1)}$. Letting $N_a(s) = \sum_{n\sim Xe^{-a}} n^{-s}$ and

$$M_{v,H}(s) = \sum_{\substack{e^{\frac{v}{H}}\leqslant p_1 d<e^{\frac{v+1}{H}}\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}} \lambda_d^{\pm}(p_1 d)^{-s} \quad \text{or} \quad M_{v,H}(s) = \sum_{\substack{e^{\frac{v}{H}}\leqslant p_1 pd<e^{\frac{v+1}{H}}\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ M\leqslant p\leqslant M^{1+\varepsilon}}} \lambda_d^{\pm}(p_1 pd)^{-s},$$

an application of Perron's formula to separate variables, along with Lemma 13 and $|P_{v_1,H}(1+it)P_1^{100\varepsilon}|^{2\ell}\geqslant 1$, yields

$$\begin{aligned}\int_{[T_0,T]\setminus\mathcal{T}_1} &|F(1+it)|^2dt\\ &\ll H^2(\log^{10}X)P_1^{200\varepsilon\ell}\int_{T_0}^{T}|P_{v_0,H}(1+it)^\ell M_{v_1,H}(1+it)N_{\frac{v_1}{H}}(1+it)|^2dt \qquad (6{\cdot}20)\\ &\quad+\left(\frac{TP_1\log X}{X}+1\right)\cdot\frac{1}{\log^{2+\varepsilon}X},\end{aligned}$$

for some $v_1\in I_1$, where $I_1=[H\log M, H\log(M^{1+\varepsilon}w^R)]$. Now Proposition 6 with $N(s)=N_{\frac{v_1}{H}}(s)$, $M(s)=M_{v_1,H}(s)$, $P(s)\equiv 1$, $Q(s)=P_{v_0,H}(s)^\ell$ and $\ell=\lfloor\varepsilon\log X/\log P_1\rfloor$ bounds (6·20) with

$$X^{o(1)}P_1^{200\varepsilon\ell}\left(Q^{-1}+\frac{1}{T_0}\right)(\ell!)^2\ll(P_1^{-1}(\log^2X))^{(1+o(1))\ell}+X^{-\varepsilon}\ll X^{-\varepsilon^2} \qquad (6{\cdot}21)$$

for $a\geqslant 2+C_3\varepsilon$, since the condition $M^2P\ll X^{1+o(1)}$ certainly holds.

Note that Proposition 8 immediately shows that

$$\frac{1}{h}\Sigma_1(h)-\frac{1}{X_1}\Sigma_1(X_1)\geqslant o\left(\frac{1}{\log X}\right)$$

for almost all $x\leqslant X$, where $X_1=X/T_0^3$. Taking into account formula (6·17) and repeating the above argument with lower bound sieve weights replaced with upper bound sieve weights, we see that the reverse inequality holds, so $(1/h)\Sigma_1(h)$ can be replaced with its dyadic counterpart $(1/X)\Sigma_1(X)$ almost always.

Now we deal with $\Sigma_2(h)$. We use the same strategy, so that for example for the lower bound we start with

$$\Sigma_2\geqslant\sum_{\substack{x\leqslant p_1 pdn\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}}\lambda_d^-,$$

an inequality that is valid even when the interval $[x/p_1p,(x+h)/p_1p]$ contains no integers. This leads us to study the Dirichlet polynomial

$$F^*(s)=\sum_{\substack{p_1pdn\sim X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ w\leqslant p<\sqrt{x}}}\lambda_d^-(p_1pdn)^{-s},$$

where the variable p can be divided into $\ll\log\log X$ intervals of the form $[M,M^{1+\varepsilon}]$ with $M\ll X^{\frac12+o(1)}$ (the value of ε may be varied so that the division becomes exact). For each of these Dirichlet polynomials where p is restricted, Proposition 8 gives a bound

of $(TP_1 \log X/X + 1)(\log X)^{-2-\varepsilon}$ for their second moment. Now by the same argument as for $\Sigma_1(h)$, we infer that $(1/h)\Sigma_2(h)$ can also be replaced with its dyadic counterpart $(1/X)\Sigma_2(X)$ almost always.

6·3. *Case of* $\Sigma_3(h)$

We are left with the sum $\Sigma_3(h)$. This is the case that determines which value of a we obtain (and hence the value of c, which is just $a+1$), since so far in all cases $a \geqslant 2 + C_4\varepsilon$ has been a sufficient assumption. We will establish the value $a = 2.51$.

Let $\beta_1, \beta_2, \beta \in (1/6, 1/2)$ be parameters which are given the values

$$\beta_1 = 0.1680, \quad \beta_2 = 0.1803, \quad \beta = 0.1950$$

to optimise various subsequent conditions. We split $\Sigma_3(h)$ into three parts $\Sigma_3^{(1)}(h)$, $\Sigma_3^{(2)}(h)$ and $\Sigma_3^{(3)}(h)$, say, the first sum being a type II sum that can be evaluated asymptotically, the second being a type I sum (after Buchstab's identity) that can mostly be evaluated asymptotically, and the third being a type II sum that can be transformed into Buchstab integrals whose value is suitably small. Explicitly, let

$$\Sigma_3^{(i)}(h) = \sum_{\substack{x \leqslant p_1q_1q_2n \leqslant x+h \\ P_1 \leqslant p_1 \leqslant P_1^{1+\varepsilon} \\ (q_1,q_2)\in A_i \\ (n,\mathcal{P}(q_2))=1 \\ n>1}} 1, \quad i = 1, 2, 3$$

with

$$A_1 = \{(q_1,q_2) : \ w \leqslant q_2 < q_1, \ \text{one of } q_1, q_2 \in [w, X^{\beta_1}] \cup [X^{\beta_2}, X^{\beta}]\},$$

$$A_2 = \{(q_1,q_2) : \ w \leqslant q_2 < q_1, \ \text{either } q_1^2q_2^3 \leqslant X \text{ or } q_1q_2^4 \leqslant X, \ q_1 \leqslant X^{\frac{1}{4}-2\varepsilon}\} \setminus A_1,$$

$$A_3 = \{(q_1,q_2) : \ w \leqslant q_2 < q_1 \leqslant X^{\frac{1}{2}}\} \setminus (A_1 \cup A_2).$$

The underlying idea is that the small variable in A_1 enables efficient use of large values theorems, the conditions in A_2 make it possible to apply Watt's theorem (after two applications of Buchstab's identity), and the remaining set A_3 can be shown to contribute not too much. We study the sums $\Sigma_3^{(i)}(h)$ separately, starting with $\Sigma_3^{(1)}(h)$.

6·3·1. *Type* II *sums*

We consider the Type II sum $\Sigma_3^{(1)}(h)$. In order to prove that $(1/h)\Sigma_3^{(1)}(h)$ is asymptotically $(1/X)\Sigma_3^{(1)}(X)$ almost always, it suffices to prove that $(1/h)\Sigma_3^{(1)}(h)$ is asymptotically $(1/X_1)\Sigma_3^{(1)}(X_1)$ almost always with $X_1 = X/T_0^3$, and then apply the prime number theorem in short intervals. For this latter asymptotic equivalence, it suffices to show that the Dirichlet polynomial

$$G(s) = \sum_{\substack{p_1q_1q_2n \sim X \\ P_1 \leqslant p_1 \leqslant P_1^{1+\varepsilon} \\ Q_i \leqslant q_i \leqslant P_i^{1+\varepsilon}, i \leqslant 2 \\ q_2 < q_1 \\ (n,\mathcal{P}(q_2))=1 \\ n>1}} (p_1q_1q_2n)^{-s}$$

satisfies

$$\int_{T_0}^{T} |G(1+it)|^2 dt \ll \left(\frac{TP_1 \log X}{X} + 1\right) \frac{1}{\log^{2+\varepsilon} X}$$

with $T \leqslant X^{1+o(1)}$, $T_0 = X^{0.01}$, $P_1 = \log^a X$ and $Q_1, Q_2 \geqslant w$ otherwise arbitrary, but either Q_1 or Q_2 is of size $X^{\nu+o(1)}$ with $\nu \in [0, \beta_1] \bigcup [\beta_2, \beta]$. These cases are similar, so assume $Q_2 = X^{\nu+o(1)}$ with ν as above.

This is the setting of Proposition 7. Therefore, if for every polynomial of the form

$$M(s) = \sum_{m \sim M} \frac{b_m}{m^s},$$

with $M = X^{\nu+o(1)}$ and $|b_m| \leqslant d_r(n)$ for fixed r, any well-spaced set

$$\mathcal{U}' \subset \{t \in [0, T] : \ |M(1 + it)| \geqslant M^{-\alpha_2}\}$$

satisfies

$$|\mathcal{U}'| \ll X^{\frac{1}{2}-\nu+\min\{2\sigma(\nu), \frac{\nu}{2}\}-\varepsilon},$$

the sum $\Sigma_3^{(1)}(h)$ has the anticipated asymptotic for $a \geqslant 1/2\alpha_2 + C_5\varepsilon$. Of course, we fix $\alpha_2 = 1/(2{\cdot}2.51) + C_6\varepsilon$.

We are left with estimating $|\mathcal{U}'|$, and to this end we utilise Jutila's large values theorem. Jutila's large values theorem (Lemma 7) applied to the ℓth moment of $M(s)$ can be reformulated to say that if

$$\mathcal{R}(\nu, \alpha_2, k, \ell) = \max\left\{2\nu\alpha_2\ell, \left(6 - \frac{2}{k}\right)\nu\alpha_2\ell + 1 - 2\nu\ell, \ 1 + 8k\ell\nu\alpha_2 - 2k\ell\nu\right\}$$

and

$$\overline{\mathcal{R}}(\nu, \alpha_2) = \min_{k,\ell \in \{1,2,\ldots\}} \mathcal{R}(\nu, \alpha_2, k, \ell),$$

then $|\mathcal{U}| \ll X^{\tilde{\mathcal{R}}(\nu,\alpha_2)+o(1)}$. It turns out that the case $k = 3$ is always optimal for us, and it suffices to restrict to $4 \leqslant \ell \leqslant 12$ (so our upper bound for $\overline{\mathcal{R}}(\nu, \alpha_2)$ is a minimum of 9 piecewise linear functions). Now we check that, with our choices of β_1, β_2, β and α_2,

$$\overline{\mathcal{R}}(\nu, \alpha_2) \leqslant \frac{1}{2} - \nu + \min\left\{2\sigma(\nu), \frac{\nu}{2}\right\} - \varepsilon$$

for $\nu \in [0.05, \beta_1] \bigcup [\beta_1, \beta_2]$. Verifying this is straightforward, because both sides are piecewise linear functions.[4]

We must also prove the desired estimate for $|\mathcal{U}'|$ in the range $\nu \in [0, 0.05)$. In this case, we do not appeal to Jutila's large values theorem, but to Lemma 6 (along with its remark), which tells us that

$$|\mathcal{U}'| \ll T^{2\alpha_2} X^{2\alpha_2\nu+o(1)} \ll X^{0.42} < X^{\frac{1}{2}-\nu-\varepsilon}$$

for the same value $\alpha_2 = 1/(2{\cdot}2.51) + C_6\varepsilon$. This means that for $c = 3.51$, $(1/h)\Sigma_3^{(1)}(h)$ can be replaced with its dyadic counterpart almost always.

[4] These computations can be carried out by hand with a bit of patience. For example, the case $\ell = 4$ in Jutila's bound is good enough in the range $\nu \in [16315/90496, 15311/78512]$, and the bound for $\ell = 5$ is good enough when $\nu \in [753/5554, 15311/91112]$. These intervals are $[\beta_2, \beta]$ and $[0.1356, \beta_1]$, up to rounding.

6·3·2. *Type* I *sums*

We turn to the sum $\Sigma_3^{(2)}(h)$. By applying Buchstab's identity twice, we find that

$$\Sigma_3^{(2)}(h) = \sum_{\substack{x\leqslant p_1q_1q_2n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ (q_1,q_2)\in A_2\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 - \sum_{\substack{x\leqslant p_1q_1q_2q_3n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ (q_1,q_2)\in A_2\\ w\leqslant q_3<q_2\\ (n,\mathcal{P}(w))=1\\ n>1}} 1 + \sum_{\substack{x\leqslant p_1q_1q_2q_3q_4n\leqslant x+h\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ (q_1,q_2)\in A_2\\ w\leqslant q_4<q_3<q_2\\ (n,\mathcal{P}(q_4))=1\\ n>1}} 1.$$

Call these sums $\Sigma_3^{(2,1)}(h)$, $\Sigma_3^{(2,2)}(h)$ and $\Sigma_3^{(2,3)}(h)$, respectively. We show that $(1/h)\Sigma_3^{(2,1)}(h)$ and $(1/h)\Sigma_3^{(2,2)}(h)$ can be replaced with their dyadic counterparts almost always. We confine to studying $\Sigma_3^{(2,2)}(h)$, as $\Sigma_3^{(2,1)}(h)$ is easier to handle.

We may make in $\Sigma_3^{(2,2)}(h)$ the additional assumption that all the variables except P_1 are in the intervals $[X^{\beta_1}, X^{\beta_2}]\cup[X^{\beta}, X]$, since otherwise the sum can be dealt with in the same way as $\Sigma_3^{(1)}(h)$. We may also assume that $q_i\in[Q_i, Q_i^{1+\varepsilon}]$ for some Q_i. Defining

$$F(s) = \sum_{\substack{p_1q_1q_2q_3dn\sim X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ Q_i\leqslant q_i\leqslant Q_i^{1+\varepsilon}\\ (q_1,q_2)\in A}} \lambda_d^{\pm}(p_1q_1q_2q_3dn)^{-s},$$

with $\lambda_d^{\pm}$ the same Brun's sieve weights as before (the sign being the same throughout), and taking into account the prime number theorem in short intervals and Lemmas 1 and 2, it suffices to show that

$$\int_{T_0}^{T}|F(1+it)|^2dt \ll \left(\frac{TP_1\log X}{X}+1\right)\frac{1}{\log^{2+\varepsilon}X}.$$

This bound is achieved similarly as in Proposition 8. Indeed, if $\mathcal{T}_1$ is defined as in the proof of that proposition, the integral over $\mathcal{T}_1$ can be estimated in the same way as in that proposition. In the complementary case, we separate all the variables, and it remains to show that

$$\int_{[T_0,T]\setminus\mathcal{T}_1}|P_1(1+it)Q_1(1+it)Q_2(1+it)Q_3(1+it)D(1+it)N(1+it)|^2dt \ll (\log X)^{-100},$$

where $N(s)$ is a zeta sum, $P_1(s)$ and $Q_i(s)$ are polynomials supported on primes, and $D(s)$ has the sieve weights λ_d as its coefficients (actually, $D(s)$ can be neglected by simply estimating it pointwise). Moreover, the lengths P_1, Q_i, D and N are from the same intervals as p_1, q_i, d and n, respectively (in particular, $d\leqslant\exp(\log X/\log\log X)$). We appeal to Proposition 6 with $Q(s)=P_1(s)^{\ell}$, $P_1^{\ell}=X^{\varepsilon}$ and with $M(s)$ either $Q_1(s)Q_3(s)$ or $Q_2(s)Q_3(s)$. If $M(s)=Q_1(s)Q_3(s)$, the condition for Proposition 6 is $Q_2\leqslant X^{\frac{1}{4}-2\varepsilon}$, $(Q_1Q_3)^2Q_2\leqslant X$. If in turn $M(s)=Q_2(s)Q_3(s)$, the condition for Proposition 6 is $Q_1\leqslant X^{\frac{1}{4}-2\varepsilon}$, $Q_1(Q_2Q_3)^2\leqslant X$, and one of these conditions is always satisfied in our domain A_2, since $Q_3\leqslant Q_2$ and automatically $Q_2\leqslant X^{\frac{1}{5}}$. Now it follows from (6·21) that for $a\geqslant 2+C_7\varepsilon$, $\Sigma_3^{(2,2)}(h)$ has the desired asymptotic, and $\Sigma_3^{(2,1)}(h)$ can be evaluated similarly.

In the sum $\Sigma_3^{(2,3)}(h)$, we may again assume that all the variables lie in the intervals $[X^{\beta_1}, X^{\beta_2}]\cup[X^{\beta}, X]$, as otherwise we can use the type II sum argument. Let $\Sigma_3^{(2,4)}(h)$ be what remains of $\Sigma_3^{(2,3)}(h)$ after this reduction. The sum $\Sigma_3^{(2,4)}$ results in a Buchstab integral, and hence is postponed to Subsection 6·3·3.

6·3·3. *Buchstab integrals*

We are left with the sums $\Sigma_3^{(3)}(h)$ and $\Sigma_3^{(2,4)}(h)$, for which no asymptotic was found. We want to show that

$$\frac{1}{X}\Sigma_3^{(3)}(X) + \frac{1}{X}\Sigma_3^{(2,4)}(X) \leqslant (1-\varepsilon)\frac{1}{X}S_X,$$

which would complete the proof of Theorem 5, taking into account the estimates (6·15) and (6·16). The following lemma allows us to transform our sums into Buchstab integrals.

LEMMA 16. *Let a positive integer k and $\eta > 0$ be fixed. Let*

$$A \subset \{(u_1, \cdots, u_k) \in \mathbb{R}^k : \ u_1, \cdots, u_k \geqslant \eta, \ u_1 + \cdots + u_k \leqslant 1 - \eta\}$$

be any set such that 1_A is Riemann integrable. For a point $q = (q_1, ..., q_k) \in \mathbb{R}^k$ and $X \geqslant 2$, define $\mathcal{L}(q) = (\log q_1/\log X, ..., \log q_k/\log X)$. Then

$$\sum_{\substack{p_1q_1\cdots q_kn\sim X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ \mathcal{L}(q_1,\ldots,q_k)\in A\\ (n,\mathcal{P}(q_k))=1}} 1$$

$$= (1+o(1))\log(1+\varepsilon)\frac{X}{\log X}\int_{(u_1,\ldots,u_k)\in A}\omega\left(\frac{1-u_1-\cdots-u_k}{u_k}\right)\frac{du}{u_1\cdots u_{k-1}u_k^2},$$

where $\omega(\cdot)$ is Buchstab's function.

Proof. It suffices to prove the statement in the case that A is a box, that is, a set of the form $I_1 \times \cdots \times I_k$ with I_i intervals. Indeed, if the statement holds for boxes, then it holds for finite unions of boxes. Moreover, since 1_A is Riemann integrable, for every $\delta > 0$ there is a finite union $\mathcal{B}$ of boxes such that $A \setminus \mathcal{B}$ has measure at most δ. The part of A not contained in $\mathcal{B}$ contributes at most $\eta^{-k-1}\delta$ to the integral, and as $\delta \to 0$, this becomes arbitrarily small.

Now let A be a box. Using the connection between Buchstab's function and the sieving function (see the Appendix of Harman's book [**10**]), summing partially, and using the change of variables $u_i = \log v_i/\log X$, we see that

$$\sum_{\substack{p_1q_1\cdots q_kn\sim X\\ P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ \mathcal{L}(q_1,\ldots,q_k)\in A\\ (n,\mathcal{P}(q_k))=1}} 1 = \sum_{\substack{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ \mathcal{L}(q_1,\ldots,q_k)\in A}} S\left(\left[\frac{X}{p_1q_1\cdots q_k}, \frac{2X}{p_1q_1\cdots q_k}\right], \mathbb{P}, q_k\right)$$

$$= (1+o(1))\sum_{\substack{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}\\ \mathcal{L}(q_1,\ldots,q_k)\in A}} \frac{X}{p_1q_1\cdots q_k\log q_k}\omega\left(\frac{\log\frac{X}{p_1q_1\cdots q_k}}{\log q_k}\right)$$

$$= (1+o(1))\sum_{P_1\leqslant p_1\leqslant P_1^{1+\varepsilon}}\frac{1}{p_1}\sum_{\mathcal{L}(q_1,\ldots,q_k)\in A}\frac{X}{q_1\ldots q_k\log q_k}\omega\left(\frac{\log\frac{X}{q_1\ldots q_k}}{\log q_k}\right)$$

$$= (b+o(1))\int_{\mathcal{L}(v_1,\ldots,v_k)\in A}\frac{X}{v_1\cdots v_k\log v_1\cdots\log^2 v_k}\omega\left(\frac{\log\frac{X}{v_1\cdots v_k}}{\log v_k}\right)dv$$

$$= (b+o(1))\frac{X}{\log X}\int_{(u_1,\ldots,u_k)\in A}\frac{1}{u_1\cdots u_k^2}\omega\left(\frac{1-u_1-\cdots u_k}{u_k}\right)du$$

with $b = \log(1+\varepsilon)$, as wanted.

Let

$$A_3^* = \{(u_1,u_2) :\ u_2 < u_1,\ u_1,u_2 \in [\beta_1,\beta_2] \cup [\beta, \frac{1}{2}],\ 2u_1+3u_2 \geqslant 1,$$
$$\max\{u_1+4u_2, 4u_1-10\varepsilon\} \geqslant 1\},$$
$$A_2^* = \{(u_1,u_2,u_3,u_4) :\ \beta_1 \leqslant u_4 < u_3 < u_2 < u_1,\ u_1,u_2,u_3,u_4 \notin [\beta_2,\beta],\ (u_1,u_2) \in A_2\}$$

be the sets corresponding to the summation conditions in $\Sigma_3^{(3)}(X)$ and $\Sigma_3^{(2,4)}(X)$, respectively. The lemma above directly implies that

$$\frac{1}{X}\Sigma_3^{(3)}(X) = \frac{(1+o(1))\log(1+\varepsilon)}{\log X} J_1,$$
$$\frac{1}{X}\Sigma_3^{(2,4)}(X) = \frac{(1+o(1))\log(1+\varepsilon)}{\log X} J_2,$$
$$\frac{1}{X} S_X = \frac{(1+o(1))\log(1+\varepsilon)}{\log X},$$

where J_1 and J_2 are given by

$$J_1 = \int\limits_{(u_1,u_2)\in A_3^*} \omega\left(\frac{1-u_1-u_2}{u_2}\right)\frac{du}{u_1u_2^2},$$
$$J_2 = \int\limits_{(u_1,u_2,u_3,u_4)\in A_2^*} \omega\left(\frac{1-u_1-u_2-u_3-u_4}{u_4}\right)\frac{du}{u_1u_2u_3u_4^2}.$$

To compute J_1, we approximate Buchstab's function by

$$\omega(u) \leqslant \begin{cases} 0, & u < 1 \\ \frac{1}{u}, & 1 \leqslant u \leqslant 2 \\ \frac{1+\log(u-1)}{u}, & 2 \leqslant u \leqslant 3 \\ \frac{1+\log 2}{3}, & u > 3. \end{cases}$$

For $u \leqslant 3$ this is an equality, and for $u > 3$ the bound very sharp (it differs from the limiting value $e^{-\gamma}$, where γ is Euler's constant, by less than 0.003), but we only need the fact that it is an upper bound. We compute with Mathematica that $J_1 < 0.988$ (when ε in the definition of A_3^* is small enough).[5] The integral J_2 only gives a minor contribution, and hence can be estimated crudely as

$$J_2 \leqslant \beta_1^{-5} \int\limits_{\substack{(u_1,u_2,u_3,u_4)\in A_2^* \\ u_1+u_2+u_3+2u_4\leqslant 1}} du$$
$$< \beta_1^{-5} \int\limits_{\substack{\beta_1<u_4<u_3<u_2<u_1 \\ u_1+u_2+u_3+2u_4\leqslant 1}} du < 0.007$$

[5] The Mathematica code can be found at `http://codepad.org/XCqx2iH3` . There is also a Python code for computing the integral at `http://codepad.org/cVx065z5`, where the integration method is a rigorous computation of an upper Riemann sum.

with Mathematica (the last integral could actually be evaluated exactly). To sum up, we have $J_1 + J_2 < 0.995 < 1 - \varepsilon$, and this means, in view of (6·16), that with our parameter choices β_1, β_2, β, the sums $\Sigma_3^{(3)}(X)$ and $\Sigma_3^{(2,4)}(X)$ can be discarded. Now, from (6·15) and (6·16) we have $(1/h)S_h(x) \geqslant \varepsilon \cdot (1/X)S_X$, so Theorem 5 is proved. □

Remark 10. We can now observe that $c = 3+\varepsilon$ is the limit of this method. Indeed, we are forced to take $\alpha_2 \leqslant 1/4$ in the type II case, because nothing nontrivial is known about the large values of Dirichlet polynomials beyond this region, and consequently $a = 1/2\alpha_2+\varepsilon \geqslant 2 + \varepsilon$ and $c \geqslant 3 + \varepsilon$.

Acknowledgement. The author is grateful to his supervisor Kaisa Matomäki for various useful comments and discussions. The author thanks the referee for careful reading of the paper and for useful comments. While working on this project, the author was supported by the Vilho, Yrjö and Kalle Vaisälä foundation of the Finnish Academy of Science and Letters.

REFERENCES

[**1**] R. C. BAKER, G. HARMAN and J. PINTZ. The difference between consecutive primes. II. *Proc. London Math. Soc.* (3), **83**(3) (2001), 532–562.
[**2**] J. BOURGAIN. On large values estimates for Dirichlet polynomials and the density hypothesis for the Riemann zeta function. *Internat. Math. Res. Notices* (3) (2000), 133–146.
[**3**] T. FREIBERG. Short intervals with a given number of primes. *J. Number Theory*. **163** (2016), 159–171.
[**4**] J. FRIEDLANDER and H. IWANIEC. *Opera de cribro*. Amer. Math. Soc. Colloq. Pub. vol. 57 (American Mathematical Society, Providence, RI, 2010).
[**5**] P. X. GALLAGHER. On the distribution of primes in short intervals. *Mathematika*. **23**(1) (1976), 4–9.
[**6**] D. A. GOLDSTON, J. PINTZ and C. Y. YILDIRIM. Positive proportion of small gaps between consecutive primes. *Publ. Math. Debrecen*. **79**(3-4) (2011), 433–444.
[**7**] D. A. GOLDSTON, J. PINTZ and C. Y. YILDIRIM. Primes in tuples IV: Density of small gaps between consecutive primes. *Acta Arith.* **160**(1) (2013), 37–53.
[**8**] G. H. HARDY and S. RAMANUJAN. The normal number of prime factors of a number n [*Quart. J. Math.* **48** (1917), 76–92]. In *Collected papers of Srinivasa Ramanujan*. (AMS Chelsea Publ., Providence, RI, 2000), pages 262–275.
[**9**] G. HARMAN. Almost-primes in short intervals. *Math. Ann.* **258**(1) (1981/82), 107–112.
[**10**] G. HARMAN. *Prime-detecting sieves*. London Math. Soc. Monog. Series. vol. 33 (Princeton University Press, Princeton, NJ, 2007).
[**11**] D. R. HEATH–BROWN. Prime numbers in short intervals and a generalised Vaughan identity. *Canad. J. Math.* **34**(6) (1982), 1365–1377.
[**12**] H. IWANIEC and E. KOWALSKI. *Analytic number theory*. Amer. Math. Soc. Colloq. Pub. vol. 53 (American Mathematical Society, Providence, RI, 2004).
[**13**] C. JIA. Almost all short intervals containing prime numbers. *Acta Arith.* **76**(1) (1996), 21–84.
[**14**] M. JUTILA. Zero-density estimates for L-functions. *Acta Arith.* **32**(1) (1977), 55–62.
[**15**] K. MATOMÄKI and M. RADZIWIŁŁ. Multiplicative functions in short intervals. To appear in *Ann. of Math.*
[**16**] H. MIKAWA. Almost-primes in arithmetic progressions and short intervals. *Tsukuba J. Math.* **13**(2) (1989), 387–401.
[**17**] H. L. MONTGOMERY. *Ten lectures on the interface between analytic number theory and harmonic analysis*. CBMS Regional Conference Series in Mathematics. vol. 84 Published for the Conference Board of the Mathematical Sciences, Washington, DC (American Mathematical Society, Providence, RI, 1994).
[**18**] A. SELBERG. On the normal density of primes in small intervals, and the difference between consecutive primes. *Arch. Math. Naturvid.* **47**(6) (1943), 87–105.
[**19**] N. WATT. Kloosterman sums and a mean value for Dirichlet polynomials. *J. Number Theory* **53**(1) (1995), 179–210.
[**20**] N. WATT. Short intervals almost all containing primes. *Acta Arith.* **72**(2) (1995), 131–167.
[**21**] D. WOLKE. Fast-Primzahlen in kurzen Intervallen. *Math. Ann.* **244**(3) (1979), 233–242.

Publication II

J. Teräväinen: The Goldbach problem for primes that are sums of two squares plus one. *Mathematika*, 64(1):20–70, 2018. DOI: 10.1112/S0025579317000341

II

Mathematika 64 (2018) 20–70
doi:10.1112/S0025579317000341

THE GOLDBACH PROBLEM FOR PRIMES THAT ARE SUMS OF TWO SQUARES PLUS ONE

JONI TERÄVÄINEN

Abstract. We study the Goldbach problem for primes represented by the polynomial $x^2 + y^2 + 1$. The set of such primes is sparse in the set of all primes, but the infinitude of such primes was established by Linnik. We prove that almost all even integers n satisfying certain necessary local conditions are representable as the sum of two primes of the form $x^2 + y^2 + 1$. This improves a result of Matomäki, which tells us that almost all even n satisfying a local condition are the sum of one prime of the form $x^2 + y^2 + 1$ and one generic prime. We also solve the analogous ternary Goldbach problem, stating that every large odd n is the sum of three primes represented by our polynomial. As a byproduct of the proof, we show that the primes of the form $x^2 + y^2 + 1$ contain infinitely many three-term arithmetic progressions, and that the numbers $\alpha p \pmod 1$, with α irrational and p running through primes of the form $x^2 + y^2 + 1$, are distributed rather uniformly.

§1. *Introduction.* Let $\mathscr{P}$ be the set of primes represented by the quadratic polynomial $x^2 + y^2 + 1$. We consider the Goldbach problem for the set $\mathscr{P}$, our main result being the following.

THEOREM 1.1. *Almost all even positive integers $n \not\equiv 5, 8 \pmod 9$ can be represented as $n = p + q$ with $p, q \in \mathscr{P}$.*

By "almost all" we mean that the number of exceptional $n \leqslant N$ is $o(N)$. The local condition $n \not\equiv 5, 8 \pmod 9$ is necessary (unless p or q equals 3 in which case we can only represent $o(N)$ integers), as is easily seen by considering primes of the form $x^2 + y^2 + 1$ modulo 9. A result of Matomäki [**13**], using a somewhat different method, showed that one of the primes p and q can be taken to be from $\mathscr{P}$, the other one being a generic prime. A few years later, Tolev [**21**] gave an asymptotic formula for a weighted count of the representations $n = p+q$ with $p \in \mathscr{P}$ and q a generic prime for almost all even n. Naturally, there is a close connection between the almost-all version of the binary Goldbach problem and the ternary Goldbach problem, so we can also solve the ternary problem for the primes $x^2 + y^2 + 1$.

THEOREM 1.2. *All large enough odd positive integers n can be represented as $n = p + q + r$ with $p, q, r \in \mathscr{P}$.*

Received 7 December 2016, published online 25 January 2018.
MSC (2010): 11P32 (primary), 11N32, 11N36 (secondary).

We remark that Tolev [**22**] established an asymptotic formula for the weighted count of the representations of n as $n = p+q+r$ with $p, q \in \mathscr{P}$ but r a generic prime. The proof of Theorem 1.2 is very similar to that of Theorem 1.1, and is remarked on in §2.

As a byproduct of the method for proving Theorem 1.1, we will obtain an analog of Roth's theorem for the set of primes of the form $x^2 + y^2 + 1$, so that in particular the set $\mathscr{P}$ contains infinitely many three-term arithmetic progressions.

THEOREM 1.3. *Any subset of* $\mathscr{P}^* = \{x^2 + y^2 + 1 : x, y \text{ coprime}\} \cap \mathbb{P}$ *having a positive upper density with respect to* $\mathscr{P}^*$ *contains infinitely many non-trivial three-term arithmetic progressions.*

We will also conclude from the proof of Theorem 1.1 that for any irrational ξ, there is some uniformity in the distribution of the fractional parts of the numbers ξp with $p \in \mathscr{P}$.

THEOREM 1.4. *Let* ξ *be irrational and* $\kappa \in \mathbb{R}$. *Then there are infinitely many primes* $p \in \mathscr{P}$ *such that* $\|\xi p + \kappa\| \leqslant p^{-\theta}$, *where* $\theta = \frac{1}{80} - \varepsilon = 0.0125 - \varepsilon$ *and* $\varepsilon > 0$ *is arbitrary. Here* $\|\cdot\|$ *stands for the distance to the nearest integer.*

Theorems 1.3 and 1.4 are proved in §§4 and 11, respectively. In Theorem 1.4, we have not pursued maximizing the value of θ, and the main message is that θ can be taken to be positive.

It should be remarked that the distribution of $\xi p \pmod 1$ has been studied also for some other subsets of the primes, such as for Chen primes [**14**, **19**] and very recently for Gaussian primes [**1**] and Piatetski–Shapiro primes [**6**]. In the case of Chen primes the analog of Theorem 1.4 with $\theta > 0$ was obtained in [**14**] (and improved in [**19**] to $\theta = \frac{3}{200} = 0.015$).

The proof of Theorem 1.1 is based on a recent paper of Matomäki and Shao [**15**], where a transference-type theorem for additive problems of Goldbach type was established, allowing one to deduce from certain desirable properties of a set A the conclusion that $A + A + A$ contains all large enough integers. One should mention that a closely related transference principle for translation-invariant additive problems was famously introduced by Green [**3**] and Green and Tao [**4**, **5**] to find arithmetic progressions in the primes, their principle stating that a set A with certain desirable properties contains infinitely many three-term arithmetic progressions (or k-term arithmetic progressions if one assumes stronger conditions). The hypotheses of the transference-type result for Goldbach-type equations [**15**, Theorem 2.3] resemble the ones of the transference principle for translation-invariant equations [**4**, Proposition 5.1], but include an additional assumption. An additional assumption is evidently needed, since for example the primes p satisfying $\|\sqrt{2}p\| < \frac{1}{100}$ contain a lot of arithmetic progressions, but most odd integers are not the sum of three such primes.

The first property required from a set A in the transference-type result of [**15**] is "well-distribution" in *Bohr sets*, meaning that for $\xi, \kappa \in \mathbb{R}$ and $\eta > 0$ the

sets $\{n : \|\xi n + \kappa\| \leqslant \eta\}$ and their intersections contain a fair proportion of the elements of A. The second property, which is present in [**4**] as well, is that A is "Fourier bounded", in the sense that the Fourier transform $\widehat{1_A}$ is small in ℓ^r norm for $r > 2$. The last and simplest to check condition is that there should be a lower bound of the correct order of magnitude for the number of elements in A up to N. In [**15**], the transference-type result was applied to solve the ternary Goldbach problem with three Chen primes or with three primes p such that $[p, p + C]$ contains at least two primes for some large constant C.

We employ a variant of the transference-type result of [**15**] in this paper, the conditions for the principle being nearly identical, but with the conclusion that $A + A$ contains almost all positive integers (in the sense that there are $o(N)$ integers $n \leqslant N$ not representable in this form). This modification is easy to implement, so the main part of our proof is devoted to verifying the conditions involved in the transference-type result in the context of the set $\mathscr{P}$. The lower bound condition follows essentially from earlier work, so we are mostly concerned with proving two requirements.

The Fourier boundedness requirement follows from the restriction theory of the primes, in the form developed by Green and Tao in [**4**]. However, the "enveloping sieve" $\beta(n)$ (which is a pseudorandom majorant of a subset of the primes and enjoys certain pleasant Fourier properties) has to be modified. It turns out that the necessary modification is available in a paper of Ramaré and Ruzsa [**18**], where the enveloping sieve was developed for purposes related to additive bases, and actually the results in that paper imply that $\mathscr{P}$ is an additive basis of finite (but large and unspecified) order.

Proving the well-distribution of the set $\mathscr{P}$ in Bohr sets requires more work and occupies the majority of this paper. We use a strategy similar to the one that was used in [**15**] to deal with Chen's primes or with primes p with $[p, p + C]$ containing two primes for some large constant C, but we must use a different sieve to detect primes of the form $x^2 + y^2 + 1$. The sieve suitable for this purpose is a combination of the linear sieve and the semilinear sieve (also called the half-dimensional sieve), developed by Iwaniec in [**9**] and used by him in [**8**] to prove that the number of primes in $\mathscr{P}$ up to N is $\gg N(\log N)^{-3/2}$ (the infinitude of the primes in $\mathscr{P}$ was established earlier by Linnik [**11**] in 1960, using his dispersion method). An upper bound for $|\mathscr{P} \cap [1, N]|$ of the same order of magnitude follows from the Selberg sieve, so $\mathscr{P}$ is a sparse set of primes.

When it comes to the sieve-theoretic part of the argument, we proceed along the lines of [**12**] and [**24**] that consider the problem of finding primes from $\mathscr{P}$ in short intervals. However, unlike in these works, one cannot apply the Bombieri–Vinogradov theorem for the prime counting function, but one has to resort to a Bombieri–Vinogradov-type result for exponential sums $\sum_{n \leqslant N} \Lambda(n) e(\alpha n)$ over primes. Such average results for exponential sums appeared for instance in [**14**, **16**, **20**], but the level of distribution achieved in these works when the weight sequence is not well-factorable (in the sense defined in [**2**, Ch. 12]) is $\frac{1}{3} - \varepsilon$, which is not good enough for our purposes. We derive a combinatorial factorization for the semilinear sieve weights and apply [**15**, Lemma 8.4] (closely

related to the estimates in [**16**]) on Bombieri–Vinogradov-type averages for $\sum_{n\leqslant N}\Lambda(n)e(\alpha n)$ to increase the level of distribution sufficiently and hence obtain Theorem 1.1. In particular, the results of §§8, 9 and 10 imply the following Bombieri–Vinogradov-type bound.

THEOREM 1.5. *Let $N \geqslant 1$ be large and $\varepsilon > 0$, $C \geqslant 10$ fixed and let $\lambda_d^{+,\mathrm{SEM}}$ and $\lambda_d^{-,\mathrm{SEM}}$ be the upper and lower bound semilinear sieve weights defined by restricting the Möbius function $\mu(d)$ to the sets*

$$\begin{aligned}\mathcal{D}^{+,\mathrm{SEM}} = \{&p_1\cdots p_r \leqslant N^{\rho_+} : z_+ \geqslant p_1 > \cdots > p_r,\\ &p_1\cdots p_{2k-2}p_{2k-1}^2 \leqslant N^{\rho_+} \text{ for all } k \geqslant 1\},\\ \mathcal{D}^{-,\mathrm{SEM}} = \{&p_1\cdots p_r \leqslant N^{\rho_+} : z_- \geqslant p_1 > \cdots > p_r,\\ &p_1\cdots p_{2k-1}p_{2k}^2 \leqslant N^{\rho} \text{ for all } k \geqslant 1\}\end{aligned}$$

with the choices $\rho_+ = \frac{2}{5} - \varepsilon$, $\rho_- = \frac{3}{7} - \varepsilon$, $z_+ \leqslant N^{1/2}$ and $z_- \leqslant N^{1/3-\varepsilon}$. Let α be a real number with $|\alpha - a/q| \leqslant 1/q^2$ for some coprime integers a and q with $q \in [(\log N)^{1000C}, N(\log N)^{-1000C}]$. Then for any integer $b \neq 0$ we have (choosing either the + or the − sign throughout)

$$\sum_{\substack{d\leqslant N^{\rho_\pm}\\(d,b)=1}}\left|\lambda_d^{\pm,\mathrm{SEM}}\sum_{\substack{n\sim N\\ n\equiv b \pmod d}}\Lambda(n)e(\alpha n)\right| \ll \frac{N}{(\log N)^C}.$$

We remark that the arguments of this paper would easily generalize to primes of the form $x^2 + y^2 + a$, where $a \neq 0$ is any integer. We also note that since for all the primes of the form $x^2 + y^2 + 1$ appearing in the rest of the paper the only possible common prime factors of x and y are 2 and 3, Theorem 1.1 could be stated in the form that almost all even $n \not\equiv 5, 8 \pmod 9$ are representable as $n = p + q$ with p and q primes and neither $p - 1$ nor $q - 1$ having any prime factors greater than 3 that are $\equiv -1 \pmod 4$. One should also mention that we did not get an asymptotic formula for the number of representations of n as sums of two or three primes from $\mathscr{P}$ (unlike in the work of Tolev [**21**, **22**] on related problems), nor did we show that the number of exceptional n in Theorem 1.1 is $\ll N/(\log N)^A$ instead of merely $o(N)$. We can nevertheless get a lower bound of $cn(\log n)^{-3}$ for the number of representations in Theorem 1.1 for almost all n for some small $c > 0$, and this is the correct order of magnitude.

1.1. *Structure of the proofs.* We give a brief outline of the dependences between different theorems and propositions. The proof of Theorem 1.1 is deduced from the transference-type theorem (Proposition 2.1) in §3, provided that the two key conditions in the transference-type theorem are satisfied. One condition is the well-distribution of the set $\mathscr{P}$ in Bohr sets and the other one is a Fourier uniformity result for $\mathscr{P}$ (Propositions 3.2 and 3.3, respectively). The proof of Proposition 3.3 is presented in §4, and in §3 it is shown that Propositions 3.2 and 3.3 immediately imply Theorem 1.3.

The largest part of the paper is then devoted to proving Proposition 3.2 using sieve theory. The purpose of §5 is to show that Proposition 3.2 follows from Proposition 5.1, which involves more notation but is easier to approach. In §6, a weighted sieve for finding primes of the form $x^2 + y^2 + 1$ is presented, in the form of Theorem 6.5. Section 7 constructs the weighted sequence (ω_n) to which Theorem 6.5 is applied, as well as sets up the circle method. Section 10 is then devoted to proving Hypothesis 6.4 for (ω_n), since this hypothesis is the requirement for applying Theorem 6.5. Section 10, which finishes the proofs of Theorems 1.1 and 1.5, involves bounding Bombieri–Vinogradov sums related to either semilinear or linear sieve coefficients and weighted by additive characters that lie on either minor or major arcs. The type I and II inputs required in §10 come from §8, while the required combinatorial input comes from §9. As Remark 3.6 tells us, the only difference in the proofs of Theorems 1.2 and 1.1 is the form of the transference-type result being used. Finally, when it comes to proving Theorem 1.4, one needs the sections from §6 onwards, the last of which, §11, is required only for this purpose. We also remark that none of the §§2–6, 8 and 9 depend on each other.

1.2. *Notation.* The symbols j, k, ℓ, m, n and q always denote integers, and p is a prime number. We denote by $e(\alpha) = \mathrm{e}^{2\pi \mathrm{i}\alpha}$ the complex exponential, by $\mathrm{Li}(x) = \int_2^x (dt/\log t)$ the logarithmic integral and by $\pi(x; q, a)$ the number of primes up to x in the residue class $a \pmod q$. We denote by $\|\cdot\|$ the distance to the nearest integer function, by $(\cdot, \cdot)$ the greatest common divisor and by $[\cdot, \cdot]$ the least common multiple. We denote by $\mathbb{Z}_q$ the set of integers $\pmod q$, sometimes interpreting functions defined on this set as q-periodic functions on $\mathbb{Z}$ and vice versa. The expression $m^{-1} \pmod q$ stands for the inverse of m in $\mathbb{Z}_q$.

Starting from §3, there are various symbols that have been reserved a specific meaning. The integer $\mathcal{C}$ is given by (2.2), the function $s(n)$ by (3.1), the set $\mathcal{S}$ by (3.2), the integer b by Definition 3.1, the numbers U, J and W by (3.3), the set $\mathcal{Q}$ by (5.1), the product $\mathfrak{S}(L)$ by Definition 6.1, the function $g(\ell)$ by Definition 6.2 and lastly the parameter Q by Lemma 7.1. When it comes to sieve-theoretic notation, λ_d are sieve weights and, for sets $\mathcal{A}$ of integers and $\mathcal{P}$ of primes, $S(\mathcal{A}, \mathcal{P}, z)$ counts the elements of $\mathcal{A}$ that are coprime to all the primes in $\mathcal{P} \cap [2, z)$, with each integer n weighted by $\omega_n \geqslant 0$, where (ω_n) will be clear from context. The arithmetic functions $\Lambda(n)$, $\mu(n)$ and $\varphi(n)$ are the von Mangoldt, Möbius and Euler functions, as usual, and the functions $\tau(n)$ and $\nu(n)$ count the number of divisors and distinct prime factors of n, respectively.

The parameters $\varepsilon, \eta > 0$ are always assumed to be small enough, but fixed. The variables N and x tend to infinity and, in §§7 and 10, A, B and C are large enough constants (say greater than 10^{10}). The numbers $\mathcal{C}$, W and J are $\ll 1$, but may be large. The expression 1_S is the indicator function of a set S, so that $1_S(n) = 1$ when $n \in S$ and $1_S(n) = 0$ otherwise. We use the usual Landau and Vinogradov asymptotic notations $o(\cdot)$, $O(\cdot)$, $\ll$, $\gg$. When we write $n \sim X$ in a summation, we mean $X \leqslant n < 2X$. By $n \asymp X$, in turn, we mean $X \ll n \ll X$.

§2. *A transference-type result.* We need a transference-type result for binary Goldbach-type problems for proving Theorem 1.1. We begin with some definitions.

Let $\Omega \subset \mathbb{Z}_N$ and $\eta \in (0, \frac{1}{2})$, and write

$$B(\Omega, \eta) = \left\{ n \in \mathbb{Z}_N : \left\| \frac{\xi n}{N} \right\| \leqslant \eta \text{ for all } \xi \in \Omega \right\}$$

for the *Bohr set* associated to these parameters. We will need a function $\chi = \chi_{\Omega,\eta} : \mathbb{Z} \to \mathbb{R}_{\geqslant 0}$ that is a smoothed version of the characteristic function of the Bohr set $B(\Omega, \eta)$. The exact construction of χ is not necessary, and we just list the properties of χ we use, found in [**15**, Lemma 3.1]. We have

$$\begin{gathered} 0 \leqslant \chi(n) \ll_{|\Omega|} 1, \qquad \chi(n) = \chi(-n) \quad \text{and} \quad \chi(n+N) = \chi(n), \\ \chi(n) \geqslant 1 \quad \text{for } n \in B(\Omega, \eta), \qquad \chi(n) \leqslant \left(\frac{\eta^2}{8}\right)^{|\Omega|} \quad \text{for } n \notin B(\Omega, 2\eta), \\ \frac{1}{N} \sum_{n \in \mathbb{Z}_N} \chi(n) := \|\chi\|_1 \geqslant \left(\frac{\eta}{2}\right)^{|\Omega|}. \end{gathered} \tag{2.1}$$

Also from [**15**], we know that χ has *Fourier complexity* $\mathcal{C} \ll_{|\Omega|,\eta} 1$, where the Fourier complexity is defined as the smallest integer $\mathcal{C}$ for which we have a Fourier representation

$$\chi(n) = \sum_{k=1}^{\mathcal{C}} c_k e(\alpha_k n), \qquad |c_k| \leqslant \mathcal{C} \text{ and } \alpha_k \in \mathbb{R}/\mathbb{Z}. \tag{2.2}$$

The formulation of the transference-type result requires harmonic analysis, so we should state which normalization of the Fourier transform we use. For functions $f, g : \mathbb{Z}_N \to \mathbb{C}$ we define the Fourier transform and the convolution as

$$\hat{f}(\xi) = \frac{1}{N} \sum_{n \in \mathbb{Z}_N} f(n) e\left(-\frac{\xi n}{N}\right) \quad \text{and} \quad f * g(n) = \frac{1}{N} \sum_{k \in \mathbb{Z}_N} f(k) g(n-k),$$

so that Parseval's identity and the convolution formula of the Fourier transform take the forms

$$\sum_{n \in \mathbb{Z}_N} |f(n)|^2 = N \sum_{\xi \in \mathbb{Z}_N} |\hat{f}(\xi)|^2 \quad \text{and} \quad \widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi).$$

PROPOSITION 2.1. *Let functions $f_1, f_2 : \mathbb{Z}_N \to \mathbb{R}_{\geqslant 0}$ and parameters $K_0 \geqslant 1$, $\delta > 0$, $\varepsilon > 0$ be given. Then there exist $\eta = \eta(K_0, \delta, \varepsilon) > 0$ and $\Omega \subset \mathbb{Z}_N$, $|\Omega| \ll_{K_0,\delta,\varepsilon} 1$ with $1 \in \Omega$ such that the following holds. Assume that, for a function $\chi = \chi_{\Omega,\eta} : \mathbb{Z} \to \mathbb{R}_{\geqslant 0}$ obeying (2.1), we have:*

(i) $f_2 * \chi(t) \geqslant \delta \|\chi\|_1$ *for all* $t \in (N/3, 2N/3)$;
(ii) $\sum_{N/3 < n < N/2} f_1(n) \geqslant \delta N$;

(iii) $\sum_{\xi\in\mathbb{Z}_N}|\widehat{f_j}(\xi)|^r \leqslant K_0$ *for* $j\in\{1,2\}$ *and* $r\in\{3,4\}$.

Then

(iv) $f_1*f_2(n)\geqslant \delta^2/3$ *for all but* $\leqslant \varepsilon N$ *values of* $n\in[0.9N,N]$.

Proof. This is inspired by and similar to [**15**, Theorem 2.3] of Matomäki and Shao. See also [**4**, Proposition 5.1], where similar ideas were applied for Roth-type problems. Take $\Omega=\{\xi\in\mathbb{Z}_N : |\widehat{f_1}(\xi)|\geqslant\varepsilon_0\}\cup\{1\}$, where ε_0 will be chosen small enough in terms of δ, ε and K_0. Condition (iii) tells us that $|\Omega|\leqslant K_0\varepsilon_0^{-3}+1$. Let $\chi=\chi_{\Omega,\eta}:\mathbb{Z}\to\mathbb{R}_{\geqslant 0}$ be as in the proposition (so that χ fulfills (2.1)). We will later choose η to be small enough in terms of δ, ε and K_0. Introduce the functions

$$g_2=\frac{1}{\|\chi\|_1}f_2*\chi \quad\text{and}\quad h_2=f_2-g_2.$$

We have

$$\widehat{g_2}=\frac{1}{\|\chi\|_1}\widehat{f_2}\widehat{\chi} \quad\text{and}\quad \widehat{h_2}=\widehat{f_2}\bigg(1-\frac{\widehat{\chi}}{\|\chi\|_1}\bigg),$$

so that in particular $|\widehat{h_2}(\xi)|\leqslant 2|\widehat{f_2}(\xi)|$.

Next we estimate from above and below the average $(1/N)\sum_{n\in\mathbb{Z}_N}|f_1*h_2(n)|^2$, starting with the lower bound. Owing to conditions (i) and (ii), for $n\in[0.9N,N]$, we have

$$f_1*g_2(n)=\frac{1}{\|\chi\|_1}f_2*\chi*f_1(n)\geqslant\frac{\delta}{N}\sum_{\substack{n-2N/3<k<n-N/3\\ k\in\mathbb{Z}_N}}f_1(k)\geqslant\delta^2 \tag{2.3}$$

since $(N/3,N/2)\subset(n-2N/3,n-N/3)$ for $n\in[0.9N,N]$. Denoting $T=\{n\in[0.9N,N]: f_1*f_2(n)<\delta^2/3\}$ and using the simple inequality $|a-b|^2\geqslant a^2/2-b^2$ and (2.3), we infer that

$$\begin{aligned}\frac{1}{N}\sum_{n\in\mathbb{Z}_N}|f_1*h_2(n)|^2&\geqslant\frac{1}{N}\sum_{n\in T}\bigg(\frac{1}{2}|f_1*g_2(n)|^2-|f_1*f_2(n)|^2\bigg)\\&\geqslant\bigg(\frac{\delta^4}{2}-\bigg(\frac{\delta^2}{3}\bigg)^2\bigg)\frac{|T|}{N}\geqslant\frac{\delta^4}{10}\frac{|T|}{N}.\end{aligned}\tag{2.4}$$

When it comes to an upper bound, Parseval's identity gives

$$\begin{aligned}\frac{1}{N}\sum_{n\in\mathbb{Z}_N}|f_1*h_2(n)|^2&=\sum_{\xi\in\mathbb{Z}_N}|\widehat{f_1*h_2}(\xi)|^2\\&=\sum_{\xi\in\mathbb{Z}_N}|\widehat{f_1}(\xi)\widehat{h_2}(\xi)|^2\\&\leqslant\varepsilon_0^{1/2}\sum_{\xi\notin\Omega}|\widehat{f_1}(\xi)|^{3/2}|\widehat{h_2}(\xi)|^2+\sum_{\xi\in\Omega}|\widehat{f_1}(\xi)|^2|\widehat{h_2}(\xi)|^2.\end{aligned}$$

Here the first sum can be bounded with the Cauchy–Schwarz inequality and (iii), implying that

$$\varepsilon_0^{1/2} \sum_{\xi \notin \Omega} |\widehat{f_1}(\xi)|^{3/2} |\widehat{h_2}(\xi)|^2 \leqslant \varepsilon_0^{1/2} \Big(\sum_{\xi \in \mathbb{Z}_N} |\widehat{f_1}(\xi)|^3 \Big)^{1/2} \Big(\sum_{\xi \in \mathbb{Z}_N} |\widehat{h_2}(\xi)|^4 \Big)^{1/2} \leqslant 8\varepsilon_0^{1/2} K_0.$$

The sum over $\xi \in \Omega$ in turn can be bounded by using the fact that

$$\left| 1 - \frac{\widehat{\chi}(\xi)}{\|\chi\|_1} \right| \leqslant 30\eta \quad \text{for every } \xi \in \Omega,$$

the proof of which is contained in the proof of Theorem 2.3 in [**15**, §4]. After this, we may again use the Cauchy–Schwarz inequality and (iii) to obtain

$$\sum_{\xi \in \Omega} |\widehat{f_1}(\xi)|^2 |\widehat{h_2}(\xi)|^2 \leqslant (30\eta)^2 \sum_{\xi \in \Omega} |\widehat{f_1}(\xi)|^2 |\widehat{f_2}(\xi)|^2 \leqslant 1000\eta^2 K_0.$$

At this stage, we fix the choices $\varepsilon_0 = \eta = \delta^8 \varepsilon^2 / 10^4 K_0^2$, so that

$$\frac{1}{N} \sum_{n \in \mathbb{Z}_N} |f_1 * h_2(n)|^2 \leqslant 8\varepsilon_0^{1/2} K_0 + 1000\eta^2 K_0 \leqslant \frac{1}{10} \delta^4 \varepsilon. \tag{2.5}$$

Combining (2.4) and (2.5) above, we discover that $|T| \leqslant 10\delta^{-4} \cdot \frac{1}{10}\delta^4 \varepsilon N = \varepsilon N$, which concludes the proof. □

§3. *Deducing Theorem 1.1 from the transference-type result.* We will apply the transference-type result (Proposition 2.1) to prove Theorem 1.1. This deduction is done in this section assuming the conditions (i)–(iii) of the transference-type result, and the rest of the paper is focused on verifying these conditions. Naturally, the functions f_1 and f_2 in the transference-type result are taken to be the characteristic functions of the primes of the form $x^2 + y^2 + 1$ (restricted to a residue class), normalized in such a way that they have mean comparable to 1. First, we introduce some notation.

Define the function

$$s(n) = \prod_{\substack{p \mid n \\ p \equiv -1 \pmod 4 \\ p \neq 3}} p, \tag{3.1}$$

which excludes from the prime factorization of n the primes 2, 3 and those primes that are $\equiv 1 \pmod 4$. Denote

$$\mathcal{S} = \{a^2 + b^2 : a, b \in \mathbb{Z}, (a, b) \mid 6^\infty\}. \tag{3.2}$$

We also define a property that we require from the linear functions we work with in what follows.

Definition 3.1. We say that a linear polynomial L with integer coefficients is *amenable* if $L(n) = Kn + b$ for some integers $K \geqslant 1$ and b, and:

(i) $6^3 \mid K$;
(ii) $(b, K) = (b-1, s(K)) = 1$;
(iii) $b - 1 = 2^j 3^{2t}(4h+1)$ for some $h \in \mathbb{Z}$, $3 \nmid 4h+1$ and $j, t \geqslant 0$ with $2^{j+2}3^{2t+1} \mid K$.

What these conditions imply is that there are no local obstructions (modulo divisors of K) to $L(n)$ being prime and $L(n) - 1$ belonging to $\mathcal{S}$ (in particular, $L(n) - 1$ crucially has an even number of prime factors $p \equiv -1 \pmod 4$ with multiplicities by (iii)). We note that it is essential that $b - 1$ is allowed to be divisible by a power of 3. Indeed, if $L_i(n) = Kn + b_i$ are two amenable linear functions with $3 \mid K$ and $3 \nmid b_1 - 1$, $3 \nmid b_2 - 1$, then $L_1(m) + L_2(n)$ can only represent numbers that are $\equiv 1 \bmod 3$. We also note that in our application we must allow K to be divisible by arbitrarily high powers of 2. This is due to the fact that if $L_i(n) = 2^s n + b_i$ are amenable, then $L_i(n) - 1 \equiv 2^{a_i} \pmod{2^{a_i+2}}$ for some integers $0 \leqslant a_i \leqslant s-2$, which implies that $L_1(m) + L_2(n)$ is never $\equiv 2 \pmod{2^s}$.

The majority of this paper is devoted to proving for functions f_i related to the characteristic function of $\mathscr{P}$ the following versions of the conditions (i) and (iii) of the transference-type result. Throughout the rest of the paper, we use the notation

$$\begin{aligned} &U = 2^J \cdot 3^3 \quad \text{with } 5 \leqslant J \ll 1, \\ &W = U \cdot \prod_{5 \leqslant p \leqslant w} p \quad \text{with } 10^{10^{10}} \leqslant w \ll 1. \end{aligned} \tag{3.3}$$

PROPOSITION 3.2. *Let $\chi : \mathbb{Z} \to \mathbb{R}_{\geqslant 0}$ have Fourier complexity $\mathcal{C} \ll 1$. Let W be as in (3.3) with $w \geqslant \mathcal{C}^{20}$, and suppose that the linear function $Wn + b$ is amenable. For an integer $N \geqslant 1$, set*

$$f(n) = (\log N)^{3/2}\left(\frac{\varphi(W)}{W}\right)^{3/2} 1_{Wn+b \in \mathbb{P}, Wn+b-1 \in \mathcal{S}} \quad \text{for } n \in \left(\frac{N}{3}, \frac{2N}{3}\right) \tag{3.4}$$

and $f(n) = 0$ for other values of $n \in [0, N)$. Then, for $N \geqslant N_0(w, \mathcal{C})$, we have

$$\sum_{n \sim N/3} f(n)\chi(t-n) \geqslant \delta_0\Bigg(\sum_{n \sim N/3} \chi(t-n) - \frac{CN}{w^{1/3}}\Bigg)$$

for $t \in (N/3, 2N/3)$ and some absolute constants $\delta_0 > 0$, $C > 0$.

PROPOSITION 3.3. *Suppose that the linear function $Wn+b$ is amenable with W as in (3.3). Let $N \geqslant 1$ be an integer and $g : \mathbb{Z}_N \to \mathbb{R}_{\geqslant 0}$ with $0 \leqslant g(n) \leqslant f(n)$ for $n \in [0, N)$ and f as in (3.4). Then, for all $r > 2$,*

$$\sum_{\xi \in \mathbb{Z}_N} |\widehat{g}(\xi)|^r \leqslant K_r$$

for some positive constant K_r depending only on r.

In this section, we show that Propositions 3.2 and 3.3 indeed imply Theorem 1.1. First we prove some lemmas about local representations of integers modulo powers of 2 and 3.

LEMMA 3.4. *Let $J \geqslant 5$ and $n \not\equiv 0 \pmod{2^{J-1}}$ be integers. Then we may write $n = a + b$ for some integers a and b with $a \equiv 2^i \pmod{2^{i+2}}$ and $b \equiv 2^j \pmod{2^{j+2}}$ for some integers $0 \leqslant i, j \leqslant J - 3$.*

Proof. Since $2^{J-1} \nmid n$, we may write $n = 2^g s$, where $0 \leqslant g \leqslant J - 5$ and $s \not\equiv 0 \pmod{16}$. It is easy to check that every such s may be written as $s = a' + b'$ with $a' \equiv 2^i \pmod{2^{i+2}}$, $b' \equiv 2^j \pmod{2^{j+2}}$ for some $0 \leqslant i, j \leqslant 3$. Then $n = a + b$ with $a = 2^g a'$, $b = 2^g b'$ is a representation of the desired form. □

LEMMA 3.5. *Let m' be any integer such that $m' \not\equiv 3, 6 \pmod 9$. Then there exist integers x_1, x_2, x_3 and x_4 such that*

$$\begin{gathered} m' \equiv x_1^2 + x_2^2 + x_3^2 + x_4^2 \pmod{3^3}, \\ x_1^2 + x_2^2, \quad x_3^2 + x_4^2 \not\equiv 1 \pmod 3, \\ x_1^2 + x_2^2, \quad x_3^2 + x_4^2 \not\equiv 0 \pmod{3^3}. \end{gathered}$$

Proof. One easily sees that $x^2 + y^2 \pmod{27}$ attains all residue classes except those that are $\equiv 3 \pmod 9$ or $\equiv 6 \pmod 9$ as x and y vary. Now the lemma only states that every $m' \not\equiv 3, 6 \pmod 9$ is the sum of two numbers, each of which is 0, 2, 5 or 8 (mod 9) and neither of which is 0 (mod 27). This can quickly be verified by hand. □

Proof of Theorem 1.1 assuming Propositions 3.2 and 3.3. Given any small $\varepsilon > 0$, we must show that once N is large enough, the interval $[0.9N, N]$ contains at most εN integers $m \equiv 0 \pmod 2$, $m \not\equiv 5, 8 \pmod 9$ that cannot be written as $m = p + q$ with p and q primes of the form $x^2 + y^2 + 1$.

Let U and W be given by (3.3) with $J = \lfloor 10/\varepsilon \rfloor$ and $w \ll 1$ large enough. We start by showing that for any $m \in [0.9N, N]$, $m \equiv 0 \pmod 2$, $m \not\equiv 5, 8 \pmod 9$, $m \not\equiv 2 \pmod{2^J}$, we may find integers $0 \leqslant B_1, B_2 \leqslant W - 1$ such that $m = B_1 + B_2$ and the linear functions $Wn + B_1$ and $Wn + B_2$ are amenable. The integers $m \equiv 2 \pmod{2^J}$ can be disposed of since there are $\leqslant (\varepsilon^2/10)N$ such integers up to N.

To see that B_1 and B_2 exist, write $m = 2m' + 2$, so that $m' \not\equiv 3, 6 \pmod 9$. Then $2^{J-1} \nmid m'$, so using Lemma 3.4 we may write $m' \equiv a_1 + a_2 \pmod{2^J}$ with $a_1 \equiv 2^i \pmod{2^{i+2}}$, $a_2 \equiv 2^j \pmod{2^{j+2}}$ for some $0 \leqslant i, j \leqslant J - 3$. Moreover, using Lemma 3.5, we may write $m' \equiv a_1' + a_2' \pmod{3^3}$ with a_1' and a_2' numbers such that $3^3 \nmid a_1', 3^3 \nmid a_2', 2a_1' + 1, 2a_2 + 1' \not\equiv 0 \pmod 3$ and the largest powers of 3 dividing a_1' and a_2' have even exponents (take $a_1' = x_1^2 + x_2^2$ and $a_2' = x_3^2 + x_4^2$ in that lemma and notice that the largest power of 3 dividing $x^2 + y^2$ has an even exponent).

Now pick numbers b_p for $5 \leqslant p \leqslant w$ such that $b_p \not\equiv 0, 1, m, m - 1 \pmod p$. By the Chinese remainder theorem, we can find an integer B such that

$B \equiv 2a_1 + 1 \pmod{2^J}$, $B \equiv 2a_1' + 1 \pmod{3^3}$ and $B \equiv b_p \pmod{p}$ for all $5 \leqslant p < w$. Therefore, we have found some integers $B_1 := B$ and $B_2 := m - B$ such that $m = B_1 + B_2$, $p \nmid B_i$, $p \nmid B_i - 1$ for $5 \leqslant p < w$, and $B_1 - 1$ and $B_2 - 1$ satisfy condition (iii) in the definition of amenability.

Therefore, we have a representation of any m of the form above as

$$m \equiv B_1(m) + B_2(m) \pmod{W}$$

with $Wn + B_1(m)$, $Wn + B_2(m)$ amenable linear functions and $0 \leqslant B_i(m) \leqslant W - 1$ (we use the notation $B_i(m)$ to emphasize that the B_i depend on $m \pmod{W}$). For each $0 \leqslant a \leqslant W - 1$, we denote

$$\mathcal{B}_a = \{m \in [0.9N, N] : m \equiv a \pmod{W}\}.$$

We will show that each $\mathcal{B}_a$ with $a \equiv 0 \pmod 2$, $a \not\equiv 5, 8 \pmod 9$, $a \not\equiv 2 \pmod{2^J}$ contains at most $\varepsilon(N/2W)$ values of $m \in [0.9N, N]$ that are not of the form $p + q$ with p and q primes of the form $x^2 + y^2 + 1$, and afterwards we sum this result over a.

If a satisfies the congruence conditions above, the polynomials $Wn + B_1(a)$ and $Wn + B_2(a)$ are amenable linear polynomials. Set $M' = \lfloor N/W \rfloor$ and, for $\ell \in \{1, 2\}$, set

$$f_\ell(n) = (\log N)^{3/2} \left(\frac{\varphi(W)}{W}\right)^{3/2} 1_{Wn+B_\ell(a)\in\mathbb{P},\ Wn+B_\ell(a)-1\in\mathcal{S}} \quad \text{for } n \in \left(\frac{M'}{3}, \frac{2M'}{3}\right)$$

with $\mathcal{S}$ as in (3.2) and let $f_\ell(n) = 0$ for $n \in [0, M') \setminus (M'/3, 2M'/3)$.

Concerning condition (ii) of the transference-type result, applying Proposition 3.2 to the function $\chi \equiv 1$, we see that

$$\sum_{M'/3 < n < 2M'/3} f_1(n) \geqslant \frac{\delta_0}{10} M',$$

but we evidently get the same outcome with summation over $M'/3 < n < M'/2$ (since one could clearly replace $n \sim N/3$ with $N/3 < n < N/2$ in Proposition 3.2). This takes care of condition (ii).

Next, by Proposition 3.3,

$$\sum_{\xi \in \mathbb{Z}_{M'}} |\widehat{f_\ell}(\xi)|^r \leqslant K_0$$

for some absolute constant K_0 when $r \in \{3, 4\}$, so condition (iii) also holds.

Let then $\chi = \chi_{\Omega,\eta} : \mathbb{Z}_{M'} \to \mathbb{R}_{\geqslant 0}$ be as in Proposition 2.1 (with χ depending on K_0 and δ_0 that appeared above), where $\Omega \subset \mathbb{Z}_{M'}$ satisfies $1 \in \Omega$, $|\Omega| \ll_\varepsilon 1$ and $1 \ll_\varepsilon \eta \leqslant 0.05$. According to (2.1), χ is symmetric around the origin and

$$\sum_{\substack{n \in [-M'/2, M'/2] \\ |n| \geqslant 0.1M'}} \chi(n) \leqslant \left(\frac{\eta^2}{8}\right)^{|\Omega|} M' \leqslant \eta \left(\frac{\eta}{2}\right)^{|\Omega|} M' \leqslant 0.05 \|\chi_1\| M'.$$

Keeping this in mind and using Proposition 3.2, for $t \in (M'/3, 2M'/3)$ we obtain

$$\begin{aligned}\sum_{n\sim M'/3} f_2(n)\chi(t-n) &\geqslant \delta_0\Big(\sum_{n\sim M'/3}\chi(t-n) - \frac{CM'}{w^{1/3}}\Big)\\ &\geqslant \frac{\delta_0}{10}\Big(\sum_{n\in\mathbb{Z}_{M'}}\chi(t-n) - \frac{CM'}{w^{1/3}}\Big)\\ &\geqslant \frac{\delta_0}{20}M'\|\chi\|_1\end{aligned}$$

for w large enough, the final step coming from (2.1), since

$$\|\chi\|_1 \geqslant \left(\frac{\eta}{2}\right)^{|\Omega|} \geqslant \frac{1}{w^{0.1}}$$

for w large enough. This means that condition (i) of the transference-type result holds with $\delta = \delta_0/20$.

From the transference-type result (Proposition 2.1), we conclude that $f_1 * f_2(n) > 0$ for all $n \in [0.9M', M']$, $n \notin T_a$, where T_a is some set of integers with $|T_a| \leqslant (\varepsilon/2)M' = \varepsilon(N/2W)$. This leads to $n \equiv n_1 + n_2 \pmod{M'}$ with

$$Wn_i + B_i(a) \in \mathbb{P}, \qquad Wn_i + B_i(a) - 1 \in \mathcal{S} \tag{3.5}$$

for $n \in [0.9M', M']$, $n \notin T_a$. Since $n_1, n_2 \in (M'/3, 2M'/3)$, we can actually say that $n = n_1 + n_2$. What we showed at the beginning of the proof is that any $m \in \mathcal{B}_a$, $m \in [0.9N + 2W, N]$ with $m \equiv 0 \pmod 2$, $m \not\equiv 5, 8 \pmod 9$ and $m \not\equiv 2 \pmod{2^J}$ can be written as $m = Wn + B_1(a) + B_2(a)$ with $n \in [0.9M', M']$ and $Wn + B_1(a)$ and $Wn + B_2(a)$ amenable (the interval $[0.9N, 0.9N + 2W]$ contains $\leqslant (\varepsilon^2/10)N$ numbers and can hence be ignored). Then

$$m = (Wn_1 + B_1(a)) + (Wn_2 + B_2(a))$$

for some n_1 and n_2 satisfying (3.5) whenever $m \in \mathcal{B}_a \setminus T'_a$, $m \in [0.9N + 2W, N]$, $m \equiv 0 \pmod 2$, $m \not\equiv 5, 8 \pmod 9$ and $m \not\equiv 2 \pmod{2^J}$, where $T'_a = \{a + W\tau : \tau \in T_a\}$ satisfies $|T'_a| \leqslant \varepsilon(N/2W)$. Since

$$\sum_{\substack{0\leqslant a\leqslant W-1\\ a\equiv 0 \pmod 2\\ a\not\equiv 5,8 \pmod 9\\ a\not\equiv 2 \pmod{2^J}}} |T_a| \leqslant W \cdot \varepsilon\frac{N}{2W} = \frac{\varepsilon}{2}N,$$

we conclude that all but $\leqslant (\varepsilon/2 + \varepsilon^2)N \leqslant \varepsilon N$ even integers $m \in [0.9N, N]$ satisfying $m \not\equiv 5, 8 \pmod 9$ can be written as $m = p + q$ with p, q primes of the form $x^2 + y^2 + 1$. □

Remark 3.6. The proof of the ternary result, Theorem 1.2, goes along very similar lines. One would replace Proposition 2.1 with the analogous

ternary transference-type result, namely [**15**, Theorem 2.3]. The premise in both transference-type results is essentially the same (except that [**15**, Theorem 2.3] has one additional function f_3), and therefore the differences in the proofs can only arise when showing that the transference-type theorem implies the additive result. In fact, these proofs are also very similar, and one would simply replace Lemma 3.4 with a version where we want to represent an arbitrary integer n as a sum of three numbers of the form $2^i \pmod{2^{i+2}}$, and one would replace Lemma 3.5 with a version where there is no restriction on m' and there are six variables x_i (and one would define f_3 analogously to f_1 and f_2).

§4. *Restriction theory for primes of the form x^2+y^2+1.* The objective of the current section is proving Proposition 3.3, after which proving Theorem 1.1 has been reduced to demonstrating Proposition 3.2. As a byproduct of the arguments, we will obtain Theorem 1.3. The proof of Proposition 3.3 is based on the Green–Tao approach [**4**] that offers a way to estimate the Fourier norms of prime-related functions and therefore to detect translation-invariant constellations within the primes. The Green–Tao approach is based on proving a restriction theorem for the Fourier transform from $\ell^r(\mathbb{Z}_N)$ to $\ell^2(\mathbb{Z}_N)$ weighted by a certain "enveloping sieve" that acts as a pseudorandom majorant for the characteristic function of the primes of the desired form. Therefore, we start by asserting that there is a suitable enveloping sieve $\beta(\cdot)$ for the primes of the form x^2+y^2+1.

PROPOSITION 4.1. *Let W and w be as in (3.3), and suppose that B is an integer for which $Wn+B$ is an amenable linear function. Then, for any large N, there exists a function $\beta : \mathbb{N} \to \mathbb{R}_{\geqslant 0}$ with the following properties (for some absolute constants $\kappa_1, \kappa_2 > 0$):*

(i) $\beta(n) \geqslant \kappa_1 (\log N)^{3/2} (\log w)^{-3/2}$ *for $n \sim N/3$ when $Wn+B \in \mathbb{P} \cap (\mathcal{S}+1)$;*

(ii) $\sum_{n \leqslant N} \beta(n) \leqslant \kappa_2 N$;

(iii) *for every fixed $\varepsilon > 0$, we have $\beta(n) \ll N^{\varepsilon}$;*

(iv) *we may write, for $z = N^{0.1}$,*

$$\beta(n) = \sum_{q \leqslant z^2} \sum_{a \in \mathbb{Z}_q^{\times}} v\Big(\frac{a}{q}\Big) e\Big(-\frac{an}{q}\Big), \tag{4.1}$$

where $v(a/q) \ll q^{\varepsilon - 1}$ (and $\mathbb{Z}_q^{\times}$ is the set of primitive residue classes $(\mathrm{mod}\ q)$);

(v) *we have $v(1) = 1$ and $v(a/q) = 0$ in (4.1) whenever q is not square-free or $q \mid W$, $q \neq 1$.*

The message of the previous proposition, which we will soon prove, is that $\beta(\cdot)$ is an upper bound for the normalized characteristic function of the primes x^2+y^2+1 in a residue class, $\beta(\cdot)$ has average comparable to 1 and $\beta(\cdot)$ has a Fourier expansion with small coefficients. The above result implies the following restriction theorem, which is identical to [**4**, Proposition 4.2], except that $\beta(\cdot)$ has a different definition.

PROPOSITION 4.2. *Let $\beta : \mathbb{N} \to \mathbb{R}_{\geqslant 0}$ be as in Proposition 4.1. Let $N \geqslant 1$ be large and let $(a_n)_{n \leqslant N}$ be any sequence of complex numbers. Given a real number $r > 2$, for some $C_r > 0$ we have*

$$\left(\sum_{\xi \in \mathbb{Z}_N} \left| \frac{1}{N} \sum_{n \leqslant N} a_n \beta(n) e\left(\frac{-\xi n}{N} \right) \right|^r \right)^{1/r} \leqslant C_r \left(\frac{1}{N} \sum_{n \leqslant N} |a_n|^2 \beta(n) \right)^{1/2}.$$

Proof of Proposition 4.2 assuming Proposition 4.1. Our function $\beta(\cdot)$ fulfills the same axioms as in the paper of Green and Tao (except the pointwise lower bound, which is not used for the proof of [**4**, Proposition 4.2]). Therefore, the proof of [**4**, Proposition 4.2] goes through in this setting. □

At this point, we show that Proposition 4.2 easily implies Proposition 3.3, which corresponds to condition (iii) in the transference-type result.

Proof of Proposition 3.3 assuming Proposition 4.1. We already know that if Proposition 4.1 is true, so is Proposition 4.2. We choose $a_n = g(n)/\beta(n)$ whenever $\beta(n) \neq 0$ and $a_n = 0$ otherwise. Since $0 \leqslant g(n) \leqslant f(n) \leqslant \kappa_1^{-1} \beta(n)$ in the notation of Proposition 3.3, from Proposition 4.2 we immediately derive

$$\begin{aligned} \left(\sum_{\xi \in \mathbb{Z}_N} |\widehat{g}(\xi)|^r \right)^{1/r} &\leqslant C_r \left(\frac{1}{N} \sum_{\substack{n \leqslant N \\ \beta(n) \neq 0}} \frac{g(n)^2}{\beta(n)} \right)^{1/2} \\ &\leqslant C_r \left(\frac{\kappa_1^{-2}}{N} \sum_{n \leqslant N} \beta(n) \right)^{1/2} \leqslant C_r \kappa_1^{-1} \kappa_2^{1/2} \end{aligned}$$

by part (ii) of Proposition 4.1. □

What remains to be shown is that the enveloping sieve promised by Proposition 4.1 exists. This is based on an argument of Ramaré and Ruzsa [**18**] (which incidentally developed the enveloping sieve for purposes unrelated to restriction theory). The enveloping sieve $\beta(n)$ turns out to be a normalized Selberg sieve corresponding to sifting primes of the form $p = x^2 + y^2 + 1$, $p \equiv B \pmod{W}$.

Proof of Proposition 4.1. We first introduce some notation. For a prime p, let $\mathcal{A}_p \subset \mathbb{Z}_p$ denote the residue classes (mod p) that are sifted away when looking for primes of the form $x^2 + y^2 + 1 \equiv B \pmod{W}$. In other words,

$$\mathcal{A}_p = \begin{cases} \emptyset & \text{for } p \leqslant w, \\ \{0\} & \text{for } p \equiv 1 \pmod 4,\ p > w, \\ \{0, 1\} & \text{for } p \equiv -1 \pmod 4,\ p > w. \end{cases}$$

Further, for square-free d, let

$$\mathcal{A}_d = \bigcap_{p \mid d} \mathcal{A}_p,$$

where $\mathcal{A}_d$ is interpreted as a subset of $\mathbb{Z}_d$. Set also $\mathcal{A}_1 = \mathbb{Z}_1$ and $\mathcal{A}_d = \emptyset$ when d is not square-free. For $d \geqslant 2$, we have $|\mathcal{A}_d| = \omega(d)$, where $\omega(\cdot)$ is a multiplicative function supported on the square-free integers and having the values

$$\omega(p) = \begin{cases} 0 & \text{for } p \leqslant w, \\ 1 & \text{for } p \equiv 1 \pmod 4,\ p > w, \\ 2 & \text{for } p \equiv -1 \pmod 4,\ p > w. \end{cases}$$

For later use, we also define

$$\mathcal{K}_1 = \mathbb{Z}_1, \qquad \mathcal{K}_p = \mathbb{Z}_p \setminus \mathcal{A}_p, \qquad \mathcal{K}_d = \bigcap_{p \mid d} \mathcal{K}_p \quad \text{for } \mu(d)^2 = 1 \tag{4.2}$$

and let $\mathcal{K}_d = \mathbb{Z}_d$ for $\mu(d) = 0$.

Let the Selberg sieve coefficients ρ_d (not the same as sieve weights) be given by

$$\rho_d = \mu(d)\frac{G_d(z)}{G_1(z)}, \quad \text{where } z = N^{0.1},\ G_d(z) = \sum_{\substack{\delta \leqslant z \\ [d,\delta] \leqslant z}} h(\delta),$$

$$h(\delta) = \prod_{p \mid \delta} h(p) \quad \text{and} \quad h(p) = \frac{\omega(p)}{p - \omega(p)}.$$

The above notations are otherwise the same as in [**18**, §4], except that λ_d there has been replaced with ρ_d and $\mathcal{L}_d$ with $\mathcal{A}_d$. We define

$$\beta(n) = G_1(z)\Bigg(\sum_{\substack{d \mid P(z) \\ Wn+B \in \mathcal{A}_d}} \rho_d \Bigg)^2, \tag{4.3}$$

where

$$P(z) = \prod_{w < p < z} p.$$

In [**18**], the factor $G_1(z)$ does not appear in their definition of $\beta(n)$, but this is just a normalization constant. In (4.3), the condition $m \in \mathcal{A}_d$ means that $m \pmod d \in \mathcal{A}_d$. Now we can check parts (i)–(v) of Proposition 4.1.

For part (i), first observe that if $Wn + B = x^2 + y^2 + 1 \in \mathbb{P} \cap (\mathcal{S} + 1)$ with $n \sim N/3$, then $x^2 + y^2 + 1 \not\equiv 0 \pmod p$ for $w < p < z = N^{0.1}$ and $x^2 + y^2 \not\equiv 0 \pmod p$ for $p \equiv -1 \pmod 4$, $w < p < z$, since $(x, y) \mid 6^J$. This means that if $Wn + B = x^2 + y^2 + 1 \in \mathbb{P} \cap (\mathcal{S} + 1)$ with $n \sim N/3$, then $\beta(n) = G_1(z)$. Now the assertion follows from

$$G_1(z) \geqslant 10^{-10} \prod_{w < p < z} \left(1 - \frac{\omega(p)}{p}\right)^{-1} \geqslant 10^{-20} (\log N)^{3/2} (\log w)^{-3/2}.$$

Part (ii) in turn follows by applying the Selberg sieve [**10**, Ch. 7] to estimate

$$\begin{aligned} G_1(z)\sum_{n\leqslant N}\Bigg(\sum_{\substack{d\mid P(z)\\ Wn+B\in\mathcal{A}_d}}\rho_d\Bigg)^2 &\leqslant 10^{10}(\log N)^{3/2}(\log w)^{-3/2}\\ &\quad\times\left(N\prod_{w<p<z}\left(1-\frac{\omega(p)}{p}\right)+z^3\right)\\ &\leqslant 10^{20}(\log N)^{3/2}(\log w)^{-3/2}\cdot N\left(\frac{\log w}{\log z}\right)^{3/2}\\ &\leqslant 10^{30}N. \end{aligned}$$

Part (iii) is verified as follows. From the definition of ρ_d, it is clear that $|\rho_d|\leqslant 1$, so that

$$\beta(n)\leqslant G_1(z)\Bigg(\sum_{\substack{d\mid P(z)\\ Wn+B\in\mathcal{A}_d}}1\Bigg)^2. \tag{4.4}$$

Note that if $Wn+B\in\mathcal{A}_p$ for some $w<p\leqslant z$, then $p\mid Wn+B$ or $p\mid Wn+B-1$, so that p can be chosen in at most $\nu(Wn+B)+\nu(Wn+B-1)$ ways, where $\nu(\cdot)$ is the number of distinct prime factors. Since d is square-free and a product of such primes p, d can be chosen in at most $2^{\nu(Wn+B)+\nu(Wn+B-1)}\ll N^{\varepsilon/3}$ ways in (4.4). Therefore, (4.4) is $\ll(\log N)^{3/2}N^{(2/3)\varepsilon}\ll N^{\varepsilon}$.

Part (iv), which is the most crucial part concerning pseudorandomness, was verified in [**18**]. Namely, our set of primes of the form $Wn+B=x^2+y^2+1$ is "sufficiently sifted" in the sense of the definition given on [**18**, pp. 1 and 2] (to see that, take in that paper A to be the set of primes of the form under consideration up to N and $\kappa=\frac{3}{2}$). This property is all that is needed to obtain (iv) with the bound $v(a/q)\ll q^{-1/2}$, by [**18**, formula (4.1.19)]. It is clear that this can be replaced with the stronger bound $v(a/q)\ll q^{\varepsilon-1}$, since we have defined the sets $\mathcal{K}_d$ in (4.2) so that [**18**, formula (4.1.18)] holds for $\xi=\varepsilon/2$, instead of just some $0<\xi<\frac{1}{2}$.

We are then left with part (v). Equations (4.1.13) and (4.1.21) of [**18**] reveal that (4.1) holds when $v(a/q)$ is defined for $(a,q)=1$ by

$$v\left(\frac{a}{q}\right)=G_1(z)\sum_{q\mid[d_1,d_2]}\frac{\rho_{d_1}^*\rho_{d_2}^*}{[d_1,d_2]}|\mathcal{K}_{[d_1,d_2]}|\cdot\frac{\sum_{b\in\mathcal{K}_q}e(ab/q)}{|\mathcal{K}_q|}$$

with

$$\rho_\ell^*=\sum_{d\equiv 0\ (\mathrm{mod}\ \ell)}\mu\left(\frac{d}{\ell}\right)\mu(d)\rho_d,$$

where the set $\mathcal{K}_d$ is given by (4.2). As in [**18**, formula (4.1.17)], we have

$$\left|\sum_{b\in\mathcal{K}_q}e\left(\frac{ab}{q}\right)\right|=\left|\sum_{b\in\mathbb{Z}_q\setminus\mathcal{K}_q}e\left(\frac{ab}{q}\right)\right|\leqslant|\mathbb{Z}_q\setminus\mathcal{K}_q|\leqslant\prod_{p^\alpha\|q}(p^\alpha-|\mathcal{K}_{p^\alpha}|),$$

which immediately gives $v(a/q) = 0$ unless q is square-free and $(q, W) = 1$. In addition, by formula (4.1.13) of the same paper (with the right-hand side multiplied by $G_1(z)$), we have

$$v\left(\frac{a}{q}\right) = G_1(z) w_q^{\#} \cdot \frac{\sum_{b \in \mathcal{K}_q} e(ab/q)}{|\mathcal{K}_q|}, \tag{4.5}$$

where by (4.1.14) we have

$$w_q^{\#} = \frac{1}{G_1(z)} \sum_{\delta \leqslant z} h(\delta) \rho_z(q, \delta),$$

and $\rho_z(q, \delta)$ satisfies (4.1.15). Putting $q = 1$ into (4.1.15), we clearly get $w_1^{\#} = 1/G_1(z)$, so that $v(1) = 1$ by (4.5). □

We have now proved Proposition 3.3, which will be needed in the proof of Theorem 1.1. As a consequence of the above considerations, we can now establish Theorem 1.3, that is, Roth's theorem for the subset $\mathscr{P}$ of primes.

Proof of Theorem 1.3. This is very similar to the proof of [**4**, Theorem 1.2]. Let $\mathcal{A} \subset \mathscr{P}^*$ have positive upper density in $\mathscr{P}^*$. Then there is $\delta > 0$ (which may be assumed small) such that $|\mathcal{A} \cap (N/3, 2N/3)| \geqslant \delta |\mathscr{P}^* \cap (N/3, 2N/3)|$ for $N \in \mathcal{N}$, where $\mathcal{N}$ is some infinite set of positive integers. Let W, w and J be as in (3.3) with $J = \lfloor 10/\delta \rfloor$.

Let $S_B = S \cap \{Wn + B : n \geqslant 1\}$ for any set S and integer B. Note that if $n = x^2 + y^2 + 1 \in (N/3, 2N/3)$ is a prime with $(x, y) = 1$ and $N \geqslant 10W$, then $(n, W) = (n - 1, s(W)) = 1$ and $(n - 1, 3) = 1, 4 \nmid n - 1$. Therefore,

$$\sum_{\substack{1 \leqslant B \leqslant W \\ Wn+B \text{ amenable}}} \left|\mathcal{A}_B \cap \left(\frac{N}{3}, \frac{2N}{3}\right)\right| = \left|\mathcal{A} \cap \left(\frac{N}{3}, \frac{2N}{3}\right)\right| \geqslant \delta \left|\mathscr{P}^* \cap \left(\frac{N}{3}, \frac{2N}{3}\right)\right|$$

for $N \geqslant 10W$ and $N \in \mathcal{N}$, so using the pigeonhole principle and the lower bound for $|\mathscr{P}^* \cap (N/3, 2N/3)|$ coming from Proposition 3.2 with $\chi \equiv 1$, we can find a value of $B \in [1, W]$ such that the polynomial $Wn + B$ is amenable and

$$\left|\mathcal{A}_B \cap \left(\frac{N}{3}, \frac{2N}{3}\right)\right| \geqslant \delta_1 \cdot \delta (\log w)^{3/2} \frac{N}{W (\log N)^{3/2}} \tag{4.6}$$

for $N \in \mathcal{N}'$ with $\mathcal{N}'$ an infinite set of positive integers and for some small absolute constant $\delta_1 > 0$, since the Chinese remainder theorem shows that there are $\leqslant 10^{10} W (\log w)^{-3/2}$ amenable functions $Wn + B$ with $1 \leqslant B \leqslant W$.

Next set

$$g(n) = \delta_2 (\log N)^{3/2} (\log w)^{-3/2} 1_{\mathcal{A}_B \cap (N/3, 2N/3)}(n)$$
$$\text{for } N \in \mathcal{N}' \text{ and } 1 \leqslant n \leqslant N$$

with $\delta_2 > 0$ small and extend g periodically to $\mathbb{Z}_N$. The assertion of the theorem will follow from the Green–Tao transference principle [**4**, Proposition 5.1] as soon as we check formulas (5.3)–(5.6) of that paper for the functions $g(n)$ and $\nu(n) = \beta(n)1_{[1,N]}(n)$ (extended periodically to $\mathbb{Z}_N$) with $\beta(\cdot)$ given by Proposition 4.1. We know (5.3) from Proposition 4.1 and (5.6) from Proposition 3.3. Formula (5.5) follows from the properties (i)–(v) of $\beta(n)$ just as in [**4**, Ch. 6]. We are left with (5.4), which follows (for a different value of δ) for $N \in \mathcal{N}'$ from (4.6). Now, as mentioned, [**4**, Proposition 5.1] yields the result, since any triple of the form $(a, a + d + j_1N, a + 2d + j_2N)$ is an arithmetic progression in $\mathbb{Z}$ if $a, a + 2d + j_1N, a + 2d + j_2N \in (N/3, 2N/3)$. □

§5. *Reductions for finding primes in Bohr sets.* The proof of Proposition 3.2 goes through an intermediate result (namely Proposition 5.1 below) that resembles it and is slightly more technical, but at the same time easier to approach. The proof of Proposition 5.1 uses among other things the circle method, Bombieri–Vinogradov-type estimates and ideas similar to Iwaniec's proof [**8**] of the infinitude of primes $x^2 + y^2 + 1$, and will occupy §§6–10.

PROPOSITION 5.1. *Let $\chi : \mathbb{Z} \to \mathbb{R}_{\geqslant 0}$ have Fourier complexity $\mathcal{C} \ll 1$. Let $N \geqslant 1$ be an integer and W be as in (3.3) with $w \geqslant \mathcal{C}^{20}$, and suppose that $Wn+b$ is an amenable linear function. There exists an integer $Q \leqslant (\log N)^B$, depending only on χ, with $B \ll_{\mathcal{C}} 1$, such that the following holds. For $N \geqslant N_0(w, \mathcal{C})$, $|t| \leqslant 5N$ and $c_0 \in \mathcal{Q}$, we have*

$$\sum_{\substack{n\sim N\\ n\equiv c_0 \pmod Q\\ Wn+b\in\mathbb{P}\\ Wn+b-1\in\mathcal{S}}} \chi(t-n) \geqslant \frac{\delta_1}{(\log N)^{3/2}}\left(\frac{W}{\varphi(W)}\right)^{3/2} \times \frac{Q}{|\mathcal{Q}|}\Bigg(\sum_{\substack{n\sim N\\ n\equiv c_0 \pmod Q}} \chi(t-n) + o\left(\frac{N}{Q}\right)\Bigg),$$

where $\delta_1 > 0$ is an absolute constant and

$$\mathcal{Q} = \{c_0 \pmod Q : (Wc_0 + b, Q) = (Wc_0 + b - 1, s(Q)) = 1\}. \quad (5.1)$$

We remark that, by the Chinese remainder theorem,

$$|\mathcal{Q}| = Q \prod_{\substack{p\mid Q\\ p\nmid W\\ p\equiv 1 \pmod 4}} \left(1 - \frac{1}{p}\right) \prod_{\substack{p\mid Q\\ p\nmid W\\ p\equiv -1 \pmod 4}} \left(1 - \frac{2}{p}\right), \quad (5.2)$$

considering that $(b, W) = (b - 1, s(W)) = 1$ by the definition of amenability.

In this section, we will show that Proposition 5.1 implies Proposition 3.2, by appealing to the following lemma.

LEMMA 5.2. *Let $\chi : \mathbb{Z} \to \mathbb{R}_{\geqslant 0}$ have Fourier complexity at most $\mathcal{C}$. Let N, $Q \geqslant 1$ be such that $N \geqslant 2Q^2$. Let $\mathcal{Q}$ be a collection of residue classes* (mod Q) *such that for all $q \mid Q$, $q \neq 1$ and for all $(a, q) = 1$, we have*

$$\left| \sum_{c_0 \in \mathcal{Q}} e\left(\frac{a}{q} c_0\right) \right| \leqslant \eta_0 |\mathcal{Q}|$$

for some $\eta_0 > 0$. Then, with the same notations as in Proposition 5.1, for some absolute constant $C' > 0$ and for all integers t, we have

$$\frac{Q}{|\mathcal{Q}|} \sum_{c_0 \in \mathcal{Q}} \sum_{\substack{n \sim N \\ n \equiv c_0 \ (\mathrm{mod}\ Q)}} \chi(t-n) \geqslant \sum_{n \sim N} \chi(t-n) - C'(\eta_0 \mathcal{C}^2 N + Q\mathcal{C}^2 N^{1/2}).$$

Proof. This is [**15**, Lemma 7.4]. □

Note that the conclusion of Proposition 3.2 (with $N/3$ replaced with N) can be rewritten as

$$\sum_{\substack{n \sim N \\ Wn+b \in \mathbb{P} \\ Wn+b-1 \in \mathcal{S}}} \chi(t-n) \geqslant \frac{\delta_0}{(\log N)^{3/2}} \left(\frac{W}{\varphi(W)}\right)^{3/2} \left(\sum_{n \sim N} \chi(t-n) - \frac{CN}{w^{1/3}} \right) \quad (5.3)$$

for $N \geqslant N_0(w, \mathcal{C})$ and $t \in (N, 3N)$, with $\delta_0 > 0$ and $C > 0$ absolute constants. In view of the previous lemma, Proposition 3.2 follows immediately from Proposition 5.1 by splitting in (5.3) the sum over n on the left-hand side to a sum over n in different residue classes (mod Q), provided that the premise of Lemma 5.2 is true for $\eta_0 = w^{-1/2}$. This is what we will prove in the remainder of this section.

LEMMA 5.3. *Let $Q \geqslant 1$ and let $\mathcal{Q}$ be defined by (5.1) (and W and w in the definition of $\mathcal{Q}$ given by (3.3)). Let a and $q \mid Q$ be positive integers with $(a, q) = 1$, $q \neq 1$. We have*

$$\left| \sum_{c_0 \in \mathcal{Q}} e\left(\frac{a}{q} c_0\right) \right| \leqslant w^{-1/2} |\mathcal{Q}|. \quad (5.4)$$

Before proving this, we present another lemma, which will be used to prove Lemma 5.3.

LEMMA 5.4. *Let a and q be positive integers, $q \neq 1$, $(a, q) = 1$, and let $Wn + b$ be an amenable linear polynomial with W and w as in (3.3). Let $V \geqslant 1$ be an integer with $(q, V) = 1$. Then*

$$\left| \sum_{\substack{n \ (\mathrm{mod}\ q) \\ (WVn+b, q)=1 \\ (WVn+b-1, s(q))=1}} e\left(\frac{a}{q} n\right) \right| \leqslant \tau(q) \cdot 1_{(q,W)=1}. \quad (5.5)$$

Proof. Using Möbius inversion, the sum in question (without absolute values) becomes

$$\sum_{d|q} \mu(d) \sum_{k|s(q)} \mu(k) \sum_{\substack{n \pmod q \\ WVn \equiv -b \pmod d \\ WVn \equiv -(b-1) \pmod k}} e\left(\frac{a}{q} n\right). \tag{5.6}$$

Now consider the sum

$$\sum_{\substack{n \pmod q \\ WVn \equiv -b \pmod d \\ WVn \equiv -(b-1) \pmod k}} e\left(\frac{a}{q} n\right). \tag{5.7}$$

Note that the sum is non-empty only if $(d, k) = 1$. Let $x_1, \ldots, x_{R(d,k)} \pmod{dk}$ be the pairwise incongruent solutions to the system $WVx \equiv -b \pmod d$, $WVx \equiv -(b-1) \pmod k$ (if there are none, the sum (5.7) is empty). Since $dk = [d, k] \mid q$, after writing $n = x_j + dkt$ for some $1 \leqslant j \leqslant R(d, k)$ and $1 \leqslant t \leqslant q/dk$, (5.7) transforms into

$$\sum_{j=1}^{R(d,k)} \sum_{\substack{n \pmod q \\ n \equiv x_j \pmod{dk}}} e\left(\frac{an}{q}\right) = \sum_{j=1}^{R(d,k)} e\left(\frac{ax_j}{q}\right) \sum_{t \pmod{q/dk}} e\left(\frac{at}{q/dk}\right). \tag{5.8}$$

The inner sum is non-zero only when $dk = q$, in which case it is 1. Taking these considerations into account, (5.6) has absolute value at most

$$\sum_{\substack{d|q \\ k|s(q) \\ dk=q}} R(d, k)|\mu(d)||\mu(k)|. \tag{5.9}$$

We estimate this differently depending on whether $(q, W) > 1$ or $(q, W) = 1$. In the former case, there is some prime p such that $p \mid q$, $p \mid W$, so $dk = q$ tells us that p divides either d or k. If $p \mid d$, then supposing that $R(d, k) \neq 0$, the congruence $WVx \equiv -b \pmod p$ must be solvable. It however is not solvable, since $p \nmid b$ for $p \mid W$ by the amenability of $Wn + b$. If $p \mid k$, then $k \mid s(q)$ implies that $p \equiv -1 \pmod 4$, $p \neq 3$. If $R(d, k) \neq 0$, the congruence $WVx \equiv -(b-1) \pmod p$ has a solution, but $p \nmid b - 1$ by amenability, so we have a contradiction. We deduce that all the summands in (5.9) vanish for $(q, W) > 1$.

Then let $(q, W) = 1$. As $d, k \mid q$ in (5.9), we also have $(d, W) = (k, W) = 1$ and $(d, V) = (k, V) = 1$. Now clearly both of the congruences $WVx \equiv -b \pmod d$, $WVx \equiv -(b-1) \pmod k$ have a unique solution, so if the two congruences are thought of as a simultaneous equation, it has at most one solution $\pmod{dk}$. Therefore, $R(d, k) \leqslant 1$, which leads to (5.9) being at most

$$\sum_{dk=q} 1 \leqslant \tau(q),$$

as asserted. □

Proof of Lemma 5.3. This is similar to the argument on [**15**, p. 21]. We can find unique q' and Q' such that $Q = qq'Q'$ and $(q, Q') = 1$ and all the prime divisors of q' divide q. Writing $c_0 = c_1 q + c_2 Q'$, c_0 runs through each residue class (mod Q) exactly once as c_1 runs through residue classes (mod $q'Q'$) and c_2 runs independently through residue classes (mod q). Now the left-hand side of (5.4) (without absolute values) becomes

$$\Sigma := \sum_{\substack{c_1 \pmod{q'Q'} \\ (Wqc_1+b, Q')=1 \\ (Wqc_1+b-1, s(Q'))=1}} \sum_{\substack{c_2 \pmod{q} \\ (WQ'c_2+b, q)=1 \\ (WQ'c_2+b-1, s(q))=1}} e\left(\frac{aQ'}{q} c_2\right). \tag{5.10}$$

Since $(aQ', q) = 1$, the inner sum is exactly of the form appearing in Lemma 5.4. Therefore,

$$|\Sigma| \leqslant \sum_{\substack{c_1 \pmod{q'Q'} \\ (Wqc_1+b, Q')=1 \\ (Wqc_1+b-1, s(Q'))=1}} \tau(q) \cdot 1_{q>w}.$$

Since $w \geqslant 10^{10^{10}}$, estimating the divisor function crudely yields

$$\begin{aligned} |\Sigma| \leqslant 1_{q>w} \cdot q^{0.1} \sum_{\substack{c_1 \pmod{q'Q'} \\ (Wqc_1+b, Q')=1 \\ (Wqc_1+b-1, s(Q'))=1}} 1 &= 1_{q>w} \cdot q'q^{0.1} \sum_{\substack{c_1 \pmod{Q'} \\ (Wqc_1+b, Q')=1 \\ (Wqc_1+b-1, s(Q'))=1}} 1 \\ &= 1_{q>w} \cdot q'q^{0.1} Q' \prod_{\substack{p \mid Q' \\ p>w}} \left(1 - \frac{\omega(p)}{p}\right), \end{aligned}$$

where $\omega(p) \in \{1, 2\}$ and $\omega(p) = 2$ precisely when $p \equiv -1 \pmod 4$. The previous expression is, for $q > w \geqslant 10^{10^{10}}$,

$$\begin{aligned} &\leqslant q'q^{0.2} \prod_{\substack{p \mid q \\ p>w}} \left(1 - \frac{\omega(p)}{p}\right) \cdot Q' \prod_{\substack{p \mid Q' \\ p>w}} \left(1 - \frac{\omega(p)}{p}\right) \\ &= \frac{Q}{q^{0.8}} \prod_{\substack{p \mid Q \\ p>w}} \left(1 - \frac{\omega(p)}{p}\right) \leqslant \frac{|\mathcal{Q}|}{w^{1/2}}, \end{aligned}$$

where the last step comes from (5.2). □

From Lemma 5.3, we conclude that proving Proposition 5.1 is enough for establishing Proposition 3.2 (and hence Theorem 1.1).

§6. *Weighted sieve for primes of the form $p = x^2 + y^2 + 1$.* Next we investigate primes of the form $x^2 + y^2 + 1$ in Bohr sets and prove Proposition 5.1 concerning these, from which Theorem 1.1 will follow. We will prove in this section

Theorem 6.5 about weighted counting of primes in the shifted set $\mathcal{S} + 1 = \{s + 1 : s \in \mathcal{S}\}$. The proof resembles Iwaniec's proof [**8**] of the infinitude of primes of the form $x^2 + y^2 + 1$, as well as the later works [**12**, **24**] on the same problem in short intervals, but the theorem involves a weighted version of the sieve procedure and hence requires a hypothesis about the weights. We will later verify the conditions of this hypothesis for a weight function related to the function $\chi(n)$ in Proposition 5.1, and this will imply Proposition 5.1 and consequently Theorem 1.1. To formulate Theorem 6.5, we first introduce the hypothesis regarding our weight coefficients. To this end, we need a couple of definitions.

Definition 6.1. Given a linear function L, let $\mathfrak{S}(L)$ be the *singular product*

$$\mathfrak{S}(L) = \prod_{\substack{p \equiv -1 \pmod 4 \\ p \neq 3}} \left(1 - \frac{|\{n \in \mathbb{Z}_p : L(n) \equiv 0 \text{ or } 1 \pmod p\}|}{p}\right)\left(1 - \frac{2}{p}\right)^{-1} \times \prod_{p \not\equiv -1 \pmod 4} \left(1 - \frac{|\{n \in \mathbb{Z}_p : L(n) \equiv 0 \pmod p\}|}{p}\right)\left(1 - \frac{1}{p}\right)^{-1}.$$

Definition 6.2. We say that a sequence $(g(\ell))_{\ell \geqslant 1}$ of complex numbers is of *convolution type* (for a given large integer N and constant $\sigma \in (3, 4)$) if

$$g(\ell) = \sum_{\substack{\ell = km \\ N^{1/\sigma} \leqslant k \leqslant N^{1-1/\sigma}}} \alpha_k \beta_m$$

for some complex numbers $|\alpha_k|, |\beta_k| \leqslant \tau(k)^2 \log k$.

Definition 6.3. For $\frac{1}{3} < \rho_2 < \rho_1 < \frac{1}{2}$ and $\sigma \in (3, 4)$, let $\mathrm{H}(\rho_1, \rho_2, \sigma)$ be the proposition

$$\frac{1}{2\sqrt{\rho_2}} \int_1^{\rho_2 \sigma} \frac{dt}{\sqrt{t(t-1)}} > \frac{1}{2\rho_1} \int_2^{\sigma} \frac{\log(t-1)}{t(1 - t/\sigma)^{1/2}}\, dt + 10^{-10}. \tag{6.1}$$

In the proof of Theorem 1.1, we will use the fact that

$$\mathrm{H}\left(\tfrac{1}{2} - \varepsilon, \tfrac{3}{7} - \varepsilon, 3 + \varepsilon\right) \quad \text{is true for small enough } \varepsilon > 0.$$

This holds for $\varepsilon = 0$ by a numerical computation and by continuity in a small neighborhood of 0. Indeed, the difference between the integrals in (6.1) is then $> 10^{-3}$. We are ready to state our Bombieri–Vinogradov-type hypothesis, whose validity depends on the weight sequence (ω_n), as well as on the parameters ρ_1, ρ_2 and σ.

HYPOTHESIS 6.4. *Let $L(n) = Kn + b$ be an amenable linear function with $K \ll (\log N)^{O(1)}$. Let $(\omega_n)_{n \sim N}$ be a non-negative sequence of real numbers*

and let $\delta = (b-1, K)$. *Let* $\varepsilon > 0$ *be any small number. Let* $\frac{1}{3} < \rho_2 < \rho_1 < \frac{1}{2} - \varepsilon$, $\sigma \in (3,4)$. *Then, for any sequence* $(g(\ell))_{\ell \leqslant N^{0.9}}$ *of convolution type (with parameter* σ*),*

$$\sum_{\substack{d \leqslant N^{\rho_1} \\ (d,K)=1}} \lambda_d^{+,\mathrm{LIN}} \sum_{\substack{\ell \leqslant N^{0.9} \\ (\ell,K)=\delta \\ (\ell,d)=1}} g(\ell) \Bigg(\sum_{\substack{n \sim N \\ L(n) = \ell p + 1 \\ L(n) \equiv 0 \pmod{d}}} \omega_n - \frac{1}{\varphi(d)} \frac{K}{\varphi(K/\delta)} \sum_{n \sim N} \frac{\omega_n}{\ell \log(Kn/\ell)} \Bigg)$$

$$\ll \frac{\sum_{n \sim N} \omega_n}{(\log N)^{100}},$$

$$\sum_{\substack{d \leqslant N^{\rho_2} \\ (d,K)=1}} \lambda_d^{-,\mathrm{SEM}} \Bigg(\sum_{\substack{n \sim N \\ L(n) \in \mathbb{P} \\ L(n) \equiv 1 \pmod{d}}} \omega_n - \frac{1}{\varphi(d)} \frac{K}{\varphi(K)} \sum_{n \sim N} \frac{\omega_n}{\log(Kn)} \Bigg)$$

$$\ll \frac{\sum_{n \sim N} \omega_n}{(\log N)^{100}},$$

where $\lambda_d^{+,\mathrm{LIN}}$ *are the upper bound linear sieve weights with sifting parameter* $z_1 = N^{1/5}$ *and* $\lambda_d^{-,\mathrm{SEM}}$ *are the lower bound semilinear sieve weights with sifting parameter* $z_2 = N^{1/\sigma}$ *(the weights* $\lambda_d^{\pm,\mathrm{SEM}}$ *were defined in Theorem 1.5, and the weights* $\lambda_d^{\pm,\mathrm{LIN}}$ *are defined analogously by replacing* $\beta = 1$ *by* $\beta = 2$ *in that definition).*

THEOREM 6.5. *Assume Hypothesis 6.4 for a linear form* $L(n)$*, sequence* $(\omega_n)_{n \sim N}$ *and parameters* ρ_1, ρ_2, σ *satisfying* $\mathrm{H}(\rho_1, \rho_2, \sigma)$. *Then*

$$\sum_{\substack{n \sim N \\ L(n) \in \mathbb{P} \\ L(n) - 1 \in \mathcal{S}}} \omega_n \geqslant \frac{\delta_0 \cdot \mathfrak{S}(L)}{(\log N)^{3/2}} \sum_{n \sim N} \omega_n + O(N^{1/2}),$$

where $\delta_0 > 0$ *is an absolute constant.*

Remark 6.6. We will be able to prove Hypothesis 6.4 in §10 for $\rho_1 = \frac{1}{2} - \varepsilon$, $\rho_2 = \frac{3}{7} - \varepsilon$ and $\sigma = 3 + \varepsilon$ when $L(n)$ is suitable and ω_n is of bounded Fourier complexity. It would suffice to prove the same with $\rho_2 = 0.385$ instead of $\rho_2 = \frac{3}{7} - \varepsilon = 0.428\ldots$ (since then $\mathrm{H}(\rho_1, \rho_2, \sigma)$ is true). On the other hand, existing Bombieri–Vinogradov estimates such as [**20**, Lemma 12] would only give us $\rho_2 = \frac{1}{3} - \varepsilon = 0.333\ldots$, which falls short of what we need.

Proof. Put

$$\mathcal{A} = \{L(n) - 1 : n \sim N, L(n) \in \mathbb{P}\},$$
$$\mathcal{P}_{4,-1} = \{p \in \mathbb{P} : p \equiv -1 \pmod{4}, p \neq 3\},$$

$$P(z) = \prod_{\substack{p<z \\ p\in\mathcal{P}_{4,-1}}} p,$$

$$\mathcal{P}^*_{4,1} = \{n \geqslant 1 : p \mid n \Rightarrow p \equiv 1 \pmod 4\}.$$

If we weight the elements of $\mathcal{A}$ by $\nu_n = \omega_{(L^{-1}(n+1))}$, where L^{-1} is the inverse function of L, the sifting function is

$$S(\mathcal{A}, \mathcal{P}_{4,-1}, z) = \sum_{\substack{n\sim N \\ L(n)\in\mathbb{P} \\ (L(n)-1, P(z))=1}} \omega_n.$$

Note that $L(n) - 1 \equiv 2^\beta \pmod{2^{\beta+2}}$ for some $\beta \geqslant 1$ by the definition of amenability, so that $L(n) - 1$ has an even number of prime factors that are $\equiv -1 \pmod 4$ (counted with multiplicity). We have

$$\sum_{\substack{n\sim N \\ L(n)\in\mathbb{P} \\ L(n)-1\in\mathcal{S}}} \omega_n = S(\mathcal{A}, \mathcal{P}_{4,-1}, (3KN)^{1/2}), \tag{6.2}$$

since the right-hand side counts with weight ω_n the numbers $L(n) - 1 = 2^{\alpha_1}3^{\alpha_2}k \in \mathcal{A}$ with $k \in \mathcal{P}^*_{4,1}$, and we claim that these numbers are precisely the numbers in $\mathcal{S} \cap \mathcal{A}$. We have $2^{\alpha_1}3^{\alpha_2}k = L(n) - 1$, so by amenability $\alpha_2 \equiv 0 \pmod 2$. It is a fact in elementary number theory that for $k \in \mathcal{P}^*_{4,1}$, both k and $2k$ can be expressed in the form $a^2 + b^2$ with $(a, b) = 1$, and additionally no number of the form $2^{\alpha_1}3^{\alpha_2}k$ with $(k, 6) = 1$ and α_2 odd or $k \notin \mathcal{P}^*_{4,1}$ is of the form $x^2 + y^2$ with $(x, y) \mid 6^\infty$. Hence, both sides of (6.2) indeed count the same integers.

Buchstab's identity reveals that

$$S(\mathcal{A}, \mathcal{P}_{4,-1}, (3KN)^{1/2}) = S(\mathcal{A}, \mathcal{P}_{4,-1}, N^{1/\sigma}) - \sum_{\substack{n\sim N \\ L(n)\in\mathbb{P}}} \sum_{\substack{p_2 \mid L(n)-1 \\ N^{1/\sigma}\leqslant p_2<(3KN)^{1/2} \\ (L(n)-1, P(p_2))=1 \\ p_2\in\mathcal{P}_{4,-1}}} \omega_n.$$

The condition $p_2 \mid L(n) - 1 \equiv 2^\beta \pmod{2^{\beta+2}}$ implies that $L(n) - 1$ has either exactly two prime divisors from $\mathcal{P}_{4,-1}$ or at least four such prime divisors (with multiplicities). The second case is impossible, since all the prime divisors of $L(n) - 1$ that are from $\mathcal{P}_{4,-1}$ are $\geqslant p_2$ and $p_2^4 \geqslant N^{4/\sigma} > L(2N) - 1$. This means that we may write $L(n) - 1 = p_1p_2m'$, $p_1 \geqslant p_2$, $p_1 \in \mathcal{P}_{4,-1}$, with m' having no prime divisors from $\mathcal{P}_{4,-1}$. Now $\delta \mid L(n) - 1 = Kn + b - 1$ with $\delta = (b - 1, K)$ and, since $p_1 \geqslant p_2 \geqslant N^{1/\sigma} > K$, we have $\delta \mid m'$. Hence, we may write $m' = \delta m$, where $m \in \mathcal{P}^*_{4,1}$ (we have $3 \nmid m$, since K is divisible by a larger power of 3 than $b - 1$ is, by the definition of amenability. Similarly, $2 \nmid m$). We claim that $(m, K/\delta) = 1$. Indeed, if $p \mid m$ and $p \mid K/\delta$, we must have $p \mid (b - 1)/\delta$, which is a contradiction to $(K, b - 1) = \delta$. Now we have

$$S(\mathcal{A}, \mathcal{P}_{4,-1}, (3KN)^{1/2}) = S - T. \tag{6.3}$$

Here

$$S = S(\mathcal{A}, \mathcal{P}_{4,-1}, N^{1/\sigma}),$$
$$T = \sum_{\substack{n\sim N\\ L(n)\in\mathbb{P}}} \sum_{\substack{L(n)-1=\delta p_1 p_2 m\\ p_1,p_2\in\mathcal{P}_{4,-1}\\ N^{1/\sigma}\leqslant p_2\leqslant p_1\\ m\in\mathcal{P}^*_{4,1}}} \omega_n \leqslant \sum_{\ell\in\mathcal{L}} S(\mathcal{M}(\ell), \mathcal{P}(\ell), N^{1/6})$$

with

$$\mathcal{L} = \Big\{\delta p_2 m : N^{1/\sigma} \leqslant p_2 \leqslant (3KNm^{-1})^{1/2},\ p_2 \in \mathcal{P}_{4,-1}, m \in \mathcal{P}^*_{4,1}, \Big(m, \frac{K}{\delta}\Big) = 1\Big\},$$
$$\mathcal{M}(\ell) = \{L(n) : L(n) = \ell p + 1 : n \sim N,\ p \in \mathbb{P}\},$$
$$\mathcal{P}(\ell) = \{p \in \mathbb{P} : (p, 2\ell) = 1\}, \qquad Q(z) = \prod_{\substack{p<z\\ p\in\mathcal{P}(\ell)}} p$$

and $M(\ell)$ has been assigned the weights $\nu_n = \omega_{L^{-1}(n)}$, so that

$$S(\mathcal{M}(\ell), \mathcal{P}(\ell), z) = \sum_{\substack{n\sim N\\ L(n)=\ell p+1\\ (L(n),Q(z))=1}} \omega_n.$$

We carry out bounding S from below and bounding T from above separately.

Bounding S. For $d \mid P(z)$, $(d, K) = 1$, let

$$r(\mathcal{A}, d) = \sum_{\substack{n\sim N\\ L(n)\in\mathbb{P}\\ L(n)-1\equiv 0 \pmod d}} \omega_n - \frac{1}{\varphi(d)}\frac{K}{\varphi(K)} \sum_{n\sim N} \frac{\omega_n}{\log(Kn)}$$

and, for $(d, K) > 1$, let $r(\mathcal{A}, d) = 0$ (since if $p \mid d$, $p \mid K$ and $p \in \mathcal{P}_{4,-1}$, then p does not divide any element of $\mathcal{A}$ by the amenability of $L(n)$). Let $\sigma \in (3, 4)$ be as in Hypothesis 6.4. The semilinear sieve [**2**, Theorem 11.13], with $\beta = 1$, sifting parameter $z = N^{1/\sigma}$ and level $D = z^s$, $1 \leqslant s \leqslant 2$, gives

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}_{4,-1}, N^{1/\sigma}) \geqslant{}& \frac{K}{\varphi(K)} \sum_{n\sim N} \frac{\omega_n}{\log(Kn)} V_K^{\mathrm{SEM}}(N^{1/\sigma}) \\ &\times (f(s) + O((\log N)^{-0.1})) \\ &+ \sum_{d\leqslant N^{s/\sigma}} \lambda_d^{-,\mathrm{SEM}} r(\mathcal{A}, d), \end{aligned} \tag{6.4}$$

where $\lambda_d^{-,\mathrm{SEM}}$ are the lower bound semilinear weights with sifting parameter $z = N^{1/\sigma}$ and we have introduced the quantities

$$f(s) = \sqrt{\frac{\mathrm{e}^\gamma}{\pi s}} \int_1^s \frac{dt}{\sqrt{t(t-1)}} \quad \text{and} \quad V_K^{\mathrm{SEM}}(z) = \prod_{\substack{p<z\\ p\equiv -1 \pmod 4\\ p\nmid K}} \Big(1 - \frac{1}{\varphi(p)}\Big).$$

We take $s = \rho_2\sigma \in [1, 2]$, where ρ_2 is as in Hypothesis 6.4. Now Hypothesis 6.4 permits replacing the last sum in (6.4) with an error of $\ll \sum_{n\sim N} \omega_n/(\log N)^{100}$ (since the terms of that sum in (6.4) vanish unless $(d, K) = 1$). Moreover, the term $V_K^{\mathrm{SEM}}(N^{1/\sigma})$ can be computed asymptotically using [**24**, Proposition 1], which implies that

$$V_K^{\mathrm{SEM}}(z) = (1 + o(1)) \prod_{\substack{p\mid K\\ p\equiv -1 \pmod 4}} \left(1 - \frac{1}{p-1}\right)^{-1} \cdot 2AC_{4,-1} \cdot \left(\frac{\pi \mathrm{e}^{-\gamma}}{\log z}\right)^{1/2},$$

where

$$A = \frac{1}{2\sqrt{2}} \prod_{p\equiv -1 \pmod 4} \left(1 - \frac{1}{p^2}\right)^{1/2}$$

and

$$C_{4,i} = \prod_{p\equiv i \pmod 4} \left(1 - \frac{1}{(p-1)^2}\right)$$

for $i \in \{-1, 1\}$. Therefore, we end up with the bound

$$\begin{aligned} S &\geqslant \frac{4AC_{4,-1} + o(1)}{(\log N)^{1/2}} \cdot I_1(\rho_2, \sigma) \frac{K}{\varphi(K)} \prod_{\substack{p\mid K\\ p\equiv -1 \pmod 4}} \left(1 - \frac{1}{p-1}\right)^{-1} \\ &\quad \times \sum_{n\sim N} \frac{\omega_n}{\log(Kn)} \\ &= \frac{4AC_{4,-1} + o(1)}{(\log N)^{3/2}} \cdot I_1(\rho_2, \sigma) \frac{K}{\varphi(K)} \prod_{\substack{p\mid K\\ p\equiv -1 \pmod 4}} \left(1 - \frac{1}{p-1}\right)^{-1} \\ &\quad \times \sum_{n\sim N} \omega_n, \end{aligned} \tag{6.5}$$

where

$$I_1(\rho_2, \sigma) = \frac{1}{2\sqrt{\rho_2}} \int_1^{\rho_2\sigma} \frac{dt}{\sqrt{t(t-1)}}.$$

Bounding T. Write, for $d \mid Q(z)$, $(d, K) = 1$, $(\ell, d) = 1$ and $(\ell, K) = \delta$,

$$r(\mathcal{M}(\ell), d) = \sum_{\substack{n\sim N\\ L(n)-1=\ell p\\ L(n)\equiv 0 \pmod d}} \omega_n - \frac{1}{\varphi(d)} \frac{K}{\varphi(K/\delta)} \sum_{n\sim N} \frac{\omega_n}{\ell \log(Kn/\ell)}.$$

For all other d such that $d \mid Q(z)$, let $r(\mathcal{M}(\ell), d) = 0$ (since if $(d, K) > 1$, then $L(n) - 1 = \ell p$, $L(n) \equiv 0 \pmod d$ is impossible). With these notations, for

$1 \leqslant s \leqslant 3$ the linear sieve [**2**, Theorem 11.13] with $\beta = 2$ provides the bound

$$S(\mathcal{M}(\ell), \mathcal{P}(\ell), N^{1/6}) \leqslant \frac{(1+o(1))K}{\varphi(K/\delta)} \sum_{n \sim N} \frac{\omega_n}{\ell \log(Kn/\ell)} V_K^{\mathrm{LIN}}(N^{1/5}, \ell) F(s) + \sum_{d \leqslant N^{s/5}} \lambda_d^{+,\mathrm{LIN}} r(\mathcal{M}(\ell), d), \tag{6.6}$$

where $\lambda_d^{+,\mathrm{LIN}}$ are the upper bound linear sieve coefficients with sifting parameter $z = N^{1/5}$, $F(s) = 2\mathrm{e}^{\gamma}/s$ and

$$V_K^{\mathrm{LIN}}(z, \ell) = \prod_{\substack{p \in \mathcal{P}(\ell) \\ p < z \\ p \nmid K}} \left(1 - \frac{1}{\varphi(p)}\right) = \prod_{2 < p < z} \left(1 - \frac{1}{p-1}\right) \prod_{\substack{p \mid K\ell \\ 2 < p < z}} \left(1 - \frac{1}{p-1}\right)^{-1}.$$

Applying [**24**, formula (4.6)], we get the asymptotic

$$V_K^{\mathrm{LIN}}(z, \ell) = (1+o(1)) \frac{2C_{4,1}C_{4,-1}\mathrm{e}^{-\gamma}\mathfrak{f}(K\ell)}{\log z}, \quad \text{where } \mathfrak{f}(d) = \prod_{\substack{p \mid d \\ p > 2}} \left(1 - \tfrac{1}{p-1}\right)^{-1}. \tag{6.7}$$

We take $s = 5\rho_1 \in [1, 3]$ in the linear sieve. Then we have

$$\sum_{\ell \in \mathcal{L}} \sum_{d \leqslant N^{\rho_1}} \lambda_d^{+,\mathrm{LIN}} r(\mathcal{M}(\ell), d) = \sum_{\substack{d \leqslant N^{\rho_1} \\ (d,K)=1}} \lambda_d^{+,\mathrm{LIN}} \sum_{\substack{\ell \leqslant N^{3/4+\varepsilon} \\ (\ell,d)=1 \\ (\ell,K)=\delta}} 1_{\mathcal{L}}(\ell) r(\mathcal{M}(\ell), d), \tag{6.8}$$

since $1_{\mathcal{L}}(\ell)$ is supported on $\ell \leqslant 3K^2N^{1-1/\sigma} \leqslant N^{3/4+\varepsilon}$. Concerning the error sum in (6.6), observe that

$$1_{\mathcal{L}}(\ell) = \sum_{\substack{\ell = k \cdot \delta m \\ N^{1/\sigma} \leqslant k \leqslant (3KN)^{1/2} \\ k \leqslant (3KN/m)^{1/2}}} 1_{\mathcal{P}_{4,-1}}(k) 1_{\mathcal{P}_{4,1}^*}(m) 1_{(m, K/\delta)=1},$$

so $1_{\mathcal{L}}(\ell)$ is of convolution type (for the value of σ we are considering), except for the cross condition $k \leqslant (3KN/m)^{1/2}$. We use Perron's formula in the form

$$\begin{aligned} 1_{(1,\infty)}(y) &= \frac{1}{\pi} \int_{-N^4}^{N^4} \frac{\sin(t \log y)}{t}\, dt + O\left(\frac{1}{N^4 |\log y|}\right) \\ &= \frac{2}{\pi} \int_{N^{-5}}^{N^4} \frac{\sin(t \log y)}{t}\, dt + O\left(\frac{1}{N^4 |\log y|} + \frac{|\log y|}{N^5}\right) \end{aligned}$$

for $N^{-3} < y \leqslant N^3$, $y \neq 1$ to dispose of the cross condition. We choose $y = 3KN/k^2m$, which satisfies $|y - 1| \geqslant 1/3KN^2$ after altering N by $\leqslant 1$ if necessary, so that the error term in Perron's formula becomes $O(K/N^2)$. According to the addition formula for sine, we have

$$\sin(t\log y) = \sin(t\log(3KN) - t\log k^2)\cos(t\log m) - \cos(t\log(3KN) - t\log k^2)\sin(t\log m),$$

which permits us to separate the variables k and m. Then we have

$$1_{\mathcal{L}}(\ell) = \frac{2}{\pi}\int_{N^{-4}}^{N^3}\frac{1}{t}\sum_{\substack{\ell = k\cdot\delta m\\ N^{1/\sigma}\leqslant k\leqslant (3KN)^{1/2}}}(\alpha_k^{(1)}(t)\beta_m^{(1)}(t) - \alpha_k^{(2)}(t)\beta_m^{(2)}(t))\,dt + O\Big(\frac{1}{N^{2-\varepsilon}}\Big),$$

where $|\alpha_k^{(j)}(t)|, |\beta_m^{(j)}(t)| \leqslant 1$ and $t \mapsto \alpha_k^{(j)}(t)$ and $t \mapsto \beta_m^{(j)}(t)$ are continuous and $\alpha_k^{(j)}(t)$ is supported on $N^{1/\sigma} \leqslant k \leqslant (3KN)^{1/2}$. Substituting this to (6.8), Hypothesis 6.4 tells us that

$$\sum_{\ell\in\mathcal{L}}\sum_{d\leqslant N^{\rho_1}}\lambda_d^{+,\mathrm{LIN}} r(\mathcal{M}(\ell), d) \ll \frac{\sum_{n\sim N}\omega_n}{(\log N)^{99}} + O(N^{1/2-\varepsilon}).$$

We sum (6.6) over $\ell \in \mathcal{L}$ and make use of (6.7), after which we have obtained

$$\begin{aligned}\sum_{\ell\in\mathcal{L}} S(\mathcal{M}(\ell), \mathcal{P}(\ell), N^{1/5}) &\leqslant (F(s) + o(1))\\ &\quad\times\frac{K}{\varphi(K/\delta)}\sum_{n\sim N}\sum_{\ell\in\mathcal{L}}\frac{\omega_n}{\ell\log(Kn/\ell)}V_K^{\mathrm{LIN}}(N^{1/6}, \ell)\\ &\quad + O\Big(\frac{\sum_{n\sim N}\omega_n}{(\log N)^{99}}\Big)\\ &= \Big(\frac{2\mathrm{e}^{\gamma}}{5\rho_1} + o(1)\Big)\cdot\frac{K}{\varphi(K/\delta)}\sum_{\ell\in\mathcal{L}}\frac{\mathfrak{f}(K\ell)}{\ell\log(KN/\ell)}\\ &\quad\times\sum_{n\sim N}\omega_n\cdot\frac{2C_{4,1}C_{4,-1}\mathrm{e}^{-\gamma}}{1/5\log N} + O\Big(\frac{\sum_{n\sim N}\omega_n}{(\log N)^{99}}\Big).\end{aligned}$$

We analyze the sum over $\mathcal{L}$ in the above formula. Denoting $\mathcal{L}' = \{\ell/\delta : \ell \in \mathcal{L}\}$, it is

$$\sum_{\ell\in\mathcal{L}}\frac{\mathfrak{f}(K\ell)}{\ell\log(KN/\ell)} = \Big(\frac{1}{\delta} + o(1)\Big)\sum_{\ell'\in\mathcal{L}'}\frac{\mathfrak{f}(K\ell')}{\ell'\log(KN/\ell')}1_{(\ell', K/\delta)=1},$$

since $\delta \mid K$. The previous sum can be written as

$$(1 + o(1))\sum_{m\leqslant N^{1-2/\sigma+\varepsilon}}\frac{u(m)\mathfrak{f}(Km)1_{(m,K/\delta)=1}}{m} \times\sum_{\substack{N^{1/\sigma}\leqslant p\leqslant (3KN/m)^{1/2}\\ p\equiv -1 \pmod 4}}\frac{1}{p\log(N/pm)}, \tag{6.9}$$

where $u(m)$ is the characteristic function of $\mathcal{P}_{4,1}^*$. To evaluate this sum, we study the sum

$$\sum_{m\leqslant x} u(m)\mathfrak{f}(Km)1_{(m,K/\delta)=1}. \tag{6.10}$$

The sum can be written as

$$\mathfrak{f}(K)\sum_{m\leqslant x} u(m)\mathfrak{f}(\psi_K(m))1_{(m,K/\delta)=1} \quad \text{where } \psi_K(m)=\prod_{\substack{p\mid m\\ p\nmid K}} p$$

and the advantage is that $\mathfrak{f}(\psi_K(m))$ is a multiplicative function. By Wirsing's theorem [**23**, Satz 1] applied to the non-negative multiplicative function $h(m)=u(m)\mathfrak{f}(\psi_K(m))1_{(m,K/\delta)=1}$ (which is bounded by 2 at prime powers and fulfills $\sum_{p\leqslant x} h(p)\log p=(\frac{1}{2}+o(1))x$), we see that (6.10) equals

$$\begin{aligned}&(\mathfrak{f}(K)+o(1))\frac{\mathrm{e}^{-\gamma/2}}{\sqrt{\pi}}\frac{x}{\log x}\prod_{\substack{p\leqslant x\\ p\nmid K/\delta\\ p\equiv 1 \pmod 4}}\left(1+\frac{h(p)}{p}+\frac{h(p^2)}{p^2}+\cdots\right)\\ &\quad=(\mathfrak{f}(K)+o(1))\frac{\mathrm{e}^{-\gamma/2}}{\sqrt{\pi}}\frac{x}{\log x}\prod_{\substack{p\leqslant x\\ p\nmid K\\ p\equiv 1 \pmod 4}}\left(1+\frac{1}{p-2}\right)\prod_{\substack{p\mid K\\ p\nmid K/\delta\\ p\equiv 1 \pmod 4}}\left(1-\frac{1}{p}\right)^{-1}.\end{aligned}$$

Applying Wirsing's theorem reversely, this is

$$(\mathfrak{f}(K)+o(1))\sum_{m\leqslant x} u(m)\mathfrak{f}(m)\cdot\prod_{\substack{p\mid K\\ p\equiv 1 \pmod 4}}\left(1+\frac{1}{p-2}\right)^{-1}\prod_{\substack{p\mid K\\ p\nmid K/\delta\\ p\equiv 1 \pmod 4}}\left(1-\frac{1}{p}\right)^{-1}.$$

By [**24**, Lemma 3], we have

$$\sum_{m\leqslant x} u(m)\mathfrak{f}(m)=(1+o(1))\frac{A}{C_{4,1}}\frac{x}{(\log x)^{1/2}}.$$

Now, using the same argument as in the proof of [**12**, Lemma 5], we compute that (6.9) equals

$$\begin{aligned}&\frac{A+o(1)}{C_{4,1}(\log N)^{1/2}}\cdot\frac{1}{2}\int_2^\sigma\frac{\log(t-1)}{t(1-t/\sigma)^{1/2}}\,dt\\ &\quad\times\frac{\mathfrak{f}(K)}{\delta}\prod_{\substack{p\mid K\\ p\equiv 1 \pmod 4}}\left(1+\frac{1}{p-2}\right)^{-1}\prod_{\substack{p\mid K\\ p\nmid K/\delta\\ p\equiv 1 \pmod 4}}\left(1-\frac{1}{p}\right)^{-1}.\end{aligned}$$

Concluding the proof. Now we have

$$T \leqslant \frac{4AC_{4,-1} + o(1)}{(\log N)^{3/2}} \frac{I_2(\rho_1, \sigma) K \mathfrak{f}(K)}{\delta \varphi(K/\delta)} \times \prod_{\substack{p \mid K \\ p \equiv 1 \pmod 4}} \left(1 + \frac{1}{p-2}\right)^{-1} \prod_{\substack{p \mid K \\ p \nmid K/\delta \\ p \equiv 1 \pmod 4}} \left(1 - \frac{1}{p}\right)^{-1} \sum_{n \sim N} \omega_n, \tag{6.11}$$

where

$$I_2(\rho_1, \sigma) = \frac{1}{2\rho_1} \int_2^\sigma \frac{\log(t-1)}{t(1 - t/\sigma)^{1/2}}\, dt.$$

We claim that the local factors in (6.5) and (6.11) are identical, or in other words that

$$\begin{aligned} &\prod_{p \mid K} \left(1 - \frac{1}{p}\right)^{-1} \prod_{\substack{p \mid K \\ p \equiv -1 \pmod 4}} \left(1 - \frac{1}{p-1}\right)^{-1} \\ &\quad = \prod_{p \mid K/\delta} \left(1 - \frac{1}{p}\right)^{-1} \prod_{\substack{p \mid K \\ p > 2}} \left(1 - \frac{1}{p-1}\right)^{-1} \\ &\qquad \times \prod_{\substack{p \mid K \\ p \equiv 1 \pmod 4}} \left(1 + \frac{1}{p-2}\right)^{-1} \prod_{\substack{p \mid K \\ p \nmid K/\delta \\ p \equiv 1 \pmod 4}} \left(1 - \frac{1}{p}\right)^{-1}. \end{aligned} \tag{6.12}$$

By the identity $(1 + 1/(p-2))^{-1} = 1 - 1/(p-1)$, (6.12) is equivalent to

$$\prod_{p \mid K} \left(1 - \frac{1}{p}\right)^{-1} = \prod_{p \mid K/\delta} \left(1 - \frac{1}{p}\right)^{-1} \prod_{\substack{p \mid K \\ p \nmid K/\delta \\ p \equiv 1 \pmod 4}} \left(1 - \frac{1}{p}\right)^{-1},$$

which in turn is equivalent to the non-existence of a prime $p \not\equiv 1 \pmod 4$ for which $p \mid K$, $p \nmid K/\delta$. If $p \geqslant 5$ were such a prime, we would have $p \mid \delta$, so $p \mid b - 1$, which contradicts the definition of amenability. We also cannot have $p = 2$ or $p = 3$, since $2 \mid K/\delta$ and $3 \mid K/\delta$ for $\delta = (b-1, K)$ by amenability.

Thus, no such p exists and (6.12) holds. Furthermore, it is clear that (6.12) is at least $0.01\mathfrak{S}(L)$. Consequently,

$$\begin{aligned} S - T \geqslant (0.01 + o(1)) 4AC_{4,-1} \mathfrak{S}(L) (I_1(\rho_2, \sigma) - I_2(\rho_1, \sigma)) \\ \times \frac{\sum_{n \sim N} \omega_n}{(\log N)^{3/2}} + O(N^{1/2}). \end{aligned}$$

Owing to the fact that $\mathrm{H}(\rho_1, \rho_2, \sigma)$ is assumed to be true, we have $I_1(\rho_2, \sigma) - I_2(\rho_1, \sigma) \geqslant 10^{-10}$, and this completes the proof of Theorem 6.5 in view of (6.2) and (6.3). □

§7. *Preparation for verifying the hypothesis.* The sequence (ω_n) to which we will apply Theorem 6.5 will be determined by a function $\chi(n)$ having a Fourier series of the form (2.2). In (2.2), it is natural to separate the phases α_i into major and minor arc parameters. This partition arises from the following lemma.

LEMMA 7.1. *Let $\alpha_1, \dots, \alpha_{\mathcal{C}}$ be real numbers with $\mathcal{C} \ll 1$ and let $W \ll 1$ be as in (3.3). Also let the constants $A, B \geqslant 1$ be related by $B = A(3\mathcal{C})^{\mathcal{C}}$. Then, for any large N, there exists a positive integer $Q \leqslant (\log N)^B$ such that each α_k may be written as*

$$\alpha_k = W\frac{a_k}{q_k} + \varepsilon_k, \quad (a_k, q_k) = 1,\ 1 \leqslant q_k \leqslant \frac{N}{(\log N)^{100B}},\ |\varepsilon_k| \leqslant \frac{(\log N)^{100B} W}{q_k N}$$

and either $q_k \mid Q$ or $q_k \geqslant q_k/(q_k, Q^2) \geqslant (\log N)^A$.

Proof. This is [**15**, Lemma 3.2]. □

From now on, A (and therefore also B) will be large enough quantities (say A, $B \geqslant 10^{10}$). Let us define the sequence (ω_n) to which we will apply Theorem 6.5 in order to prove Proposition 5.1. Let $\chi : \mathbb{Z} \to \mathbb{R}_{\geqslant 0}$ be any function with Fourier complexity $\leqslant \mathcal{C}$ (i.e., χ satisfies (2.2)). Given an integer t with $|t| \leqslant 5N$, we choose

$$(\omega_n)_{n\sim N/Q} = (\chi(t - (Qn + c_0)))_{n\sim N/Q},$$

where Q is determined by the α_i in (2.2) with the help of Lemma 7.1 and $c_0 \in \mathcal{Q}$ with

$$\mathcal{Q} = \{c_0 \pmod Q : (Wc_0 + b, Q) = (Wc_0 + b - 1, s(Q)) = 1\}.$$

Recall that $|\mathcal{Q}|$ is given by (5.2).

From now on, let

$$x = \frac{N}{Q}, \qquad L(n) = QWn + Wc_0 + b, \quad c_0 \in \mathcal{Q}.$$

To prove Proposition 5.1 and hence Proposition 3.2 and Theorem 1.1, it suffices to show that for W as in (3.3) and $\mathfrak{S}(L)$ as in Definition 6.1, we have

$$\sum_{\substack{n\sim x\\ L(n)\in\mathbb{P}\\ L(n)-1\in\mathcal{S}}} \chi(t - (Qn + c_0)) \geqslant \frac{\delta_0 \cdot \mathfrak{S}(L)}{(\log x)^{3/2}} \sum_{n\sim x} \chi(t - (Qn + c_0)) + o\left(\frac{x}{(\log x)^{3/2}}\right), \tag{7.1}$$

since $L(n)$ is amenable and since by (5.2)

$$\mathfrak{S}(L) \asymp \prod_{\substack{p\equiv -1 \pmod 4\\ p\mid QW\\ p\nmid W}} \left(1 - \frac{1}{p}\right)^{-2} \prod_{\substack{p\not\equiv -1 \pmod 4\\ p\mid QW\\ p\nmid W}} \left(1 - \frac{1}{p}\right)^{-1}$$

$$\times \prod_{\substack{p \equiv -1 \pmod 4 \\ p \mid W}} \left(1 - \frac{1}{p}\right)^{-2} \prod_{\substack{p \not\equiv -1 \pmod 4 \\ p \mid W}} \left(1 - \frac{1}{p}\right)^{-1}$$

$$\asymp \left(\frac{W}{\varphi(W)}\right)^{3/2} \frac{Q}{|\mathcal{Q}|}.$$

By Theorem 6.5 and the remark after it, formula (7.1) will follow once we have verified Hypothesis 6.4 for our sequence $(\chi(t - (Qn + c_0)))_{n \sim x}$ and linear function $L(n)$ and parameters

$$\rho_1 = \tfrac{1}{2} - 10\varepsilon, \qquad \rho_2 = \tfrac{3}{7} - 10\varepsilon \quad \text{and} \quad \sigma = 3 + \varepsilon. \tag{7.2}$$

By formula (2.2) for $\chi(n)$ and Lemma 7.1, it suffices to inspect Hypothesis 6.4 with the choices (7.2) for $(e(\xi n))_{n \sim x}$, where ξ is an arbitrary real number satisfying, for some $Q \leqslant 2(\log x)^B$,

$$\left|\xi - \frac{QWa}{q}\right| \leqslant \frac{2(\log x)^{102B}}{qx}$$
$$\text{for } (a, q) = 1,\ q \leqslant \frac{x}{(\log x)^{99B}} \text{ and } q \mid Q \text{ or } \frac{q}{(q, Q^2)} \geqslant (\log x)^A. \tag{7.3}$$

Moreover, we may assume in (7.1) that

$$\sum_{n \sim x} \chi(t - (Qn + c_0)) \gg \frac{x}{(\log x)\mathfrak{S}(L)},$$

since otherwise we have nothing to prove, and consequently it suffices to prove Hypothesis 6.4 for $(e(\xi n))_{n \sim x}$ with $(\sum_{n \sim x} \omega_n)(\log x)^{-100}$ replaced by $x(\log x)^{-200}$ in that hypothesis.

§8. *Bombieri–Vinogradov sums weighted by additive characters.* We will establish Hypothesis 6.4 in the setting of §7 subsequently in §10. For that purpose as well as for proving Theorem 1.4 in §11, we need the following Bombieri–Vinogradov-type estimates for type I and II exponential sums. We employ for positive integers q and v the notation

$$q_v = \frac{q}{(q, v^2)}.$$

LEMMA 8.1. *Let $M \leqslant N^{0.4}$, $R \leqslant N^{0.1}$ and $\rho \leqslant \frac{1}{2} - \varepsilon$ for some $\varepsilon \in (0, \frac{1}{6})$. Let ξ be a real number with $|\xi - a/q| \leqslant 1/(qv)^2$ for some coprime a and $q \in [1, N]$ and some positive integer $v \leqslant N^{0.1}$. Then, for any complex numbers $|\alpha_m| \leqslant \tau(m)^2 \log m$ and any $t \in [N, 2N]$, we have*

$$\sum_{0 < |r| \leqslant R} \sum_{d \leqslant N^\rho} \max_{(c, dv) = 1} \left| \sum_{\substack{N \leqslant mn \leqslant t \\ mn \equiv c \pmod{dv} \\ m \leqslant M}} \alpha_m e(\xi r m n) \right|$$
$$\ll \left(\frac{RN}{v}\right)^{1/2} \left(RMN^\rho + \frac{RN}{vq_v} + q_v\right)^{1/2} (\log N)^{1000}.$$

Proof. We follow the proof of [**15**, Lemma 8.3]. It suffices to consider the sum over $0 < r \leqslant R$. Our task is to estimate

$$S_r = \sum_{d \leqslant N^\rho} \max_{(c,dv)=1} \Bigg| \sum_{\substack{N \leqslant mn \leqslant t \\ mn \equiv c \pmod{dv} \\ m \leqslant M}} \alpha_m e(\xi r m n) \Bigg|$$

for $r \leqslant R$. The inner sum in the definition of S_r is a geometric sum in the variable n, so evaluating it provides the bound

$$S_r \ll \sum_{d \leqslant N^\rho} \sum_{m \leqslant M} |\alpha_m| \min\left\{ \frac{RN}{rm\,dv}, \frac{1}{\|r\xi m\,dv\|} \right\}.$$

Observe that $|v\xi - av/q| \leqslant 1/q^2$. Based on this, writing $d' = rm\,d$ and using a standard bound for sums over fractional parts [**15**, Lemma B.3] (taking $x = RN/v$ in that lemma), we get

$$\begin{aligned} \sum_{r \leqslant R} S_r &\ll \sum_{d' \leqslant RMN^\rho} \tau(d')^5 \min\left\{ \frac{RN}{d'v}, \frac{1}{\|v\xi d'\|} \right\} (\log N) \\ &\ll \left(\frac{RN}{vq_v^{1/2}} + \left(\frac{RN \cdot RMN^\rho}{v} \right)^{1/2} + \left(\frac{RN}{v} q_v \right)^{1/2} \right) (\log N)^{1000} \\ &\ll \left(\frac{RN}{v} \right)^{1/2} \left(RMN^\rho + \frac{RN}{vq_v} + q_v \right)^{1/2} (\log N)^{1000}, \end{aligned}$$

as wanted. □

LEMMA 8.2. *Let $M \in [N^{1/2}, N^{3/4}]$ and $\Delta_1, \Delta_2 \geqslant 1$, $\Delta_1\Delta_2 \leqslant N^{1/2}$, $\Delta_1\Delta_2^2 \leqslant M/v$ for some positive integer $v \leqslant N^{0.1}$. Let ξ be a real number with $|\xi - a/q| \leqslant 1/(qv)^2$ for some coprime a and $q \in [1, N]$. Then, for any complex numbers $|\alpha_m|, |\beta_m| \leqslant \tau(m)^2 \log m$ and any integer $c' \neq 0$ and number $t \in [N, 2N]$, we have*

$$\begin{aligned} &\sum_{0<|r| \leqslant R} \sum_{d_1 \sim \Delta_1} \sum_{\substack{d_2 \sim \Delta_2 \\ (d_2, c' d_1 v)=1}} \max_{(c,d_1 v)=1} \Bigg| \sum_{\substack{N \leqslant mn \leqslant t \\ mn \equiv c \pmod{d_1 v} \\ mn \equiv c' \pmod{d_2} \\ m \sim M}} \alpha_m \beta_n e(\xi r m n) \Bigg| \\ &\ll \frac{RN}{v} \min\{F_1, F_2\} (\log N)^{1000} \end{aligned}$$

with

$$F_1 = \left(\frac{\Delta_1 M v}{N} + \Delta_1 \Delta_2^2 \frac{v}{M} \right)^{1/2} + \left(\frac{1}{\Delta_1} + \frac{1}{q_v} + \frac{q_v v^2}{RN} \right)^{1/8},$$

$$F_2 = \Delta_1\Delta_2 \left(\frac{1}{q_v^{1/2}} + \frac{v}{M^{1/2}} + \frac{v^2 M}{N} + \frac{q_v^{1/2} v}{(RN)^{1/2}} \right)^{1/2}.$$

Remark 8.3. In §10, we will only need the case $R = 1$, while the dependence on v will be crucial. In §11, on the other hand, $v = 1$ but the dependence on R will be crucial.

Proof. We follow the proof of [**15**, Lemma 8.4], which in turn is based on an argument of Mikawa [**16**]. It suffices to consider the case $r > 0$. We will first prove the lemma in the case $F_1 = \min\{F_1, F_2\}$. Let us write

$$I_r = \sum_{d_1 \sim \Delta_1} \sum_{\substack{d_2 \sim \Delta_2 \\ (d_2, c' d_1 v) = 1}} \max_{(c, d_1 v) = 1} \left| \sum_{\substack{N \leqslant mn \leqslant t \\ mn \equiv c \pmod{d_1 v} \\ mn \equiv c' \pmod{d_2} \\ m \sim M}} \alpha_m \beta_n e(\xi r m n) \right|,$$

so that $\sum_{r \leqslant R} I_r$ is what we are interested in. Since $\Delta_1 \Delta_2^2 \leqslant M/v$, a formula on [**15**, p. 37] tells us (with $x = N$, $D = \Delta_1$, $\alpha = r\xi$) that

$$I_r^2 \ll N (\log N)^{100} \Bigg(\Delta_1 \sum_{d_1 \sim \Delta_1} \sum_{0 < |j| \leqslant 8\Delta_2^2 N / \Delta_1 M v} \tau_3(j) \times \min\left\{ \frac{RN}{r(d_1 v)^2 |j|}, \frac{1}{\| r\xi (d_1 v)^2 j \|} \right\} + \frac{\Delta_1 M}{v} \Bigg)$$

(since the term $(x^2/Q^2)(\log x)^{-C+10}$ present in that formula of [**15**] can be replaced with $(DMx/Q)(\log x)^{100}$ without changing anything in the proof). Using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \frac{1}{(\log N)^{200}} \sum_{r \leqslant R} I_r &\leqslant \frac{1}{(\log N)^{200}} R^{1/2} \Bigg(\sum_{r \leqslant R} I_r^2 \Bigg)^{1/2} \\ &\leqslant (RN)^{1/2} \Bigg(\Delta_1 \sum_{d_1 \sim \Delta_1} \sum_{0 < |j| \leqslant 8\Delta_2^2 N / \Delta_1 M v} \sum_{r \leqslant R} \tau_3(j) \\ &\qquad \times \min\left\{ \frac{RN}{r(d_1 v)^2 |j|}, \frac{1}{\| r\xi (d_1 v)^2 j \|} \right\} + \frac{\Delta_1 RM}{v} \Bigg)^{1/2} \\ &\ll (RN)^{1/2} \Bigg(\Delta_1 \sum_{d_1 \sim \Delta_1} \sum_{1 \leqslant \ell \leqslant 8\Delta_2^2 RN / \Delta_1 M v} \tau_4(\ell) \\ &\qquad \times \min\left\{ \frac{RN}{(d_1 v)^2 \ell}, \frac{1}{\| v^2 \xi d_1^2 \ell \|} \right\} + \frac{\Delta_1 RM}{v} \Bigg)^{1/2}, \end{aligned} \tag{8.1}$$

after writing $\ell = rj$. When it comes to the sum above, we can estimate it using the lemma on [**16**, p. 6] (with $\tau_3(\cdot)$ replaced by $\tau_4(\cdot)$), stating that

$$\begin{aligned} &\Delta_1 \sum_{d_1 \sim \Delta_1} \sum_{\ell \sim J} \tau_4(\ell) \min\left\{ \frac{x}{d_1^2 \ell}, \frac{1}{\| \xi' d_1^2 \ell \|} \right\} \\ &\ll \left(\Delta_1^2 J + x^{3/4} \left(q' + \frac{x}{q'} + \frac{x}{\Delta_1} \right)^{1/4} \right) (\log x)^{100} \end{aligned} \tag{8.2}$$

for $1 \leqslant J \leqslant 10x$ and any real number ξ' satisfying $|\xi' - a'/q'| \leqslant 1/q'^2$ for some coprime a' and $q' \leqslant x$. In the case $q' > x$, (8.2) continues to hold, by trivial estimates. We substitute (8.2) with $x = RN/v^2$, $\xi' = v^2\xi$ and $J \leqslant 8\Delta_2^2 RN/\Delta_1 Mv$ into (8.1) (we have $J \leqslant 10(RN/v^2)$, since $\Delta_1\Delta_2^2 \leqslant M/v$), making use of our assumption on ξ, which implies that $|v^2\xi - (av^2/(q, v^2))/q_v| \leqslant 1/q_v^2$. This results in the claimed bound.

Then let $F_2 = \min\{F_1, F_2\}$. In this situation, we use the orthogonality of characters to bound the sum in Lemma 8.2 with

$$\sum_{r\leqslant R}\sum_{d_1\sim\Delta_1}\sum_{d_2\sim\Delta_2}\sum_{\psi}\max_{(\mathrm{mod}\ d_1d_2)}\Bigg|\sum_{\substack{N\leqslant mn\leqslant t\\ mn\equiv c_v(d_1,d_2)\ (\mathrm{mod}\ v)\\ m\sim M}}\alpha_m\psi(m)\beta_n\psi(n)e(\xi rmn)\Bigg|, \tag{8.3}$$

where $c_v(d_1, d_2)$ is a suitably chosen integer coprime to v. Estimating the sums over d_1 and d_2 trivially and using the Cauchy–Schwarz inequality and expanding a square, we find that (8.3) is, for some $|\beta_n'| \leqslant \tau(n)^2 \log n$ and some c_v coprime to v,

$$\begin{aligned}
&\leqslant \Delta_1\Delta_2(RM)^{1/2}\Bigg(\sum_{r\leqslant R}\sum_{m\leqslant M}\Bigg|\sum_{\substack{N/m\leqslant n\leqslant t/m\\ n\equiv c_vm^{-1}\ (\mathrm{mod}\ v)}}\beta_n'e(\xi rmn)\Bigg|^2\Bigg)^{1/2}(\log M)^{100}\\
&= \Delta_1\Delta_2(RM)^{1/2}\Bigg(\sum_{r\leqslant R}\sum_{\substack{N/2M\leqslant n_i\leqslant 2N/M\\ n_1\equiv n_2\ (\mathrm{mod}\ v)\\ \text{for } i\in\{1,2\}}}\beta_{n_1}'\overline{\beta_{n_2}'}\\
&\quad\times\sum_{\substack{m\leqslant M\\ N/n_i\leqslant m\leqslant t/n_i\\ m\equiv c_vn_i^{-1}\ (\mathrm{mod}\ v)\\ \text{for } \in\{1,2\}}}e(\xi rm(n_1-n_2))\Bigg)^{1/2}(\log M)^{100}\\
&\ll \Delta_1\Delta_2(RN)^{1/2}\Bigg(RM+\sum_{r\leqslant R}\sum_{\substack{1\leqslant n\leqslant 2N/M\\ n\equiv 0\ (\mathrm{mod}\ v)}}T(n)\\
&\quad\times\min\Bigg\{\frac{RN}{rnv}+1,\frac{1}{\|v\xi rn\|}\Bigg\}\Bigg)^{1/2}(\log M)^{101},
\end{aligned} \tag{8.4}$$

where

$$T(n) = \frac{M}{N}\sum_{\substack{n=n_1-n_2\\ n_1,n_2\leqslant 2N/M}}\tau(n_1)^2\tau(n_2)^2.$$

We can write $n = kv$ and $\ell = kr$ to bound (8.4) with

$$\begin{aligned}
&\ll \Delta_1\Delta_2(RN)^{1/2}\Bigg(RM+\sum_{\ell\leqslant 2RN/Mv}U(\ell)\min\Bigg\{\frac{RN}{\ell v^2}+1,\frac{1}{\|v^2\xi\ell\|}\Bigg\}\Bigg)^{1/2}\\
&\quad\times(\log N)^{101},
\end{aligned} \tag{8.5}$$

where

$$U(\ell) = \sum_{\substack{\ell=\ell_1\ell_2\\ \ell_1\leqslant 2N/Mv}} T(\ell_1 v).$$

We apply [**15**, Lemma B.3] (with $k = 20$) to (8.5). The weight function $U(\ell)$ is not a divisor function, but the only property of the weight function needed in that lemma is a second-moment bound. Therefore, (8.5) can be bounded with

$$\ll \Delta_1\Delta_2(RN)^{1/2}\left(\frac{RN}{q_v^{1/2}v^2} + \frac{RN}{(v^2M)^{1/2}} + RM + \left(\frac{RNq_v}{v^2}\right)^{1/2}\right)^{1/2} \times (\log N)^{1000}, \tag{8.6}$$

once we prove that

$$\sum_{\ell\leqslant 2RN/Mv} U(\ell)^2 \ll \frac{RN}{Mv}(\log N)^{100}. \tag{8.7}$$

We calculate

$$\begin{aligned}
&\sum_{\ell\leqslant 2RN/Mv}\Bigg(\sum_{\substack{\ell=\ell_1\ell_2\\ \ell_1\leqslant 2N/Mv}} T(\ell_1 v)\Bigg)^2\\
&\ll \frac{RN}{Mv}\sum_{\substack{\ell_1\leqslant 2N/Mv\\ \ell_1'\leqslant 2N/Mv}}\frac{T(\ell_1 v)T(\ell_1' v)}{[\ell_1,\ell_1']}\\
&\ll \frac{RN}{Mv}\sum_{d\leqslant 2N/Mv}\frac{1}{d}\sum_{\substack{\ell_1\leqslant 2N/dMv\\ \ell_1'\leqslant 2N/dMv}}\frac{T(\ell_1 dv)T(\ell_1' dv)}{\ell_1\ell_1'}\\
&= \frac{RN}{Mv}\sum_{d\leqslant 2N/Mv}\frac{1}{d}\Bigg(\sum_{\ell\leqslant 2N/dMv}\frac{T(\ell dv)}{\ell}\Bigg)^2.
\end{aligned} \tag{8.8}$$

We can estimate the sum inside the square using

$$\begin{aligned}
\sum_{\substack{n\leqslant 2N/M\\ n\equiv 0 \pmod c}}\frac{T(n)}{n} &\ll \frac{M}{N}\sum_{\substack{n_1\leqslant 2N/M\\ n_2\leqslant 2N/M\\ n_1\equiv n_2 \pmod c\\ n_1>n_2}}\frac{\tau(n_1)^2\tau(n_2)^2}{n_1-n_2}\\
&\ll \frac{M}{Nc}\sum_{1\leqslant a\leqslant c}\sum_{\substack{n_1'\leqslant 2N/Mc\\ n_2'\leqslant 2N/Mc\\ n_1'>n_2'}}\frac{\tau(cn_1'+a)^2\tau(cn_2'+a)^2}{n_1'-n_2'}\\
&\ll \frac{M}{Nc}\sum_{1\leqslant a\leqslant c}\sum_{n\leqslant 2N/Mc}\tau(cn+a)^4\\
&\ll \frac{M}{Nc}\sum_{m\leqslant 2N/M+c}\tau(m)^4 \ll \frac{1}{c}(\log N)^{15}
\end{aligned}$$

for $c \leqslant 2N/M$, where we used Hilbert's inequality [**17**, Ch. 7] in the third last step. Taking $c = dv$, and substituting to (8.8), we see that (8.7) holds, as claimed. Therefore, we indeed have the bound (8.6) for (8.5) and that bound can be rewritten as the desired bound F_2. □

§9. *Factorizing sieve weights.* The linear and semilinear sieve weights will play a crucial role in verifying Hypothesis 6.4, since we aim to split the summation over $d \leqslant x^\rho$ in that hypothesis to summations over $d_1 \sim \Delta_1$, $d_2 \sim \Delta_2$ for various values of Δ_1 and Δ_2. If such a factorization can be done, it provides more flexibility in our Bombieri–Vinogradov sums and hence gives better bounds. This advantage can be seen from Lemma 8.2, which often produces better bounds when Δ_1 and Δ_2 are of somewhat similar size, as opposed to the choice $\Delta_1 = x^\rho$, $\Delta_2 = 1$. The following lemmas about the combinatorial structure of sieve weights have been tailored so that the estimate given by Lemma 8.2 will be $\ll Nv^{-1}(\log N)^{-1000}$ if Δ_1 and Δ_2 satisfy the conditions for d_1 and d_2 in Lemma 9.1 or 9.2 with $D = x^{1-\varepsilon^2}/M$, $\theta = 0$, $R = 1$ and q suitably large, and additionally $\rho = \frac{3}{7}(1-4\theta)-\varepsilon$ in the case of Lemma 9.1 or $\rho = \frac{1}{2}(1-4\theta)-\varepsilon$ in the case of Lemma 9.2. It should be remarked that in §10 we will only need the case $\theta = 0$ of the following lemmas, but for the proof of Theorem 1.4 we will choose $\theta = \frac{1}{80} - \varepsilon$.

9.1. *Linear sieve weights.*

LEMMA 9.1. *Let $\varepsilon > 0$ be small, $0 \leqslant \theta \leqslant \frac{1}{30}$ and $\rho = \frac{1}{2}(1-4\theta)-\varepsilon$. Let*

$$\mathcal{D}^{+,\mathrm{LIN}} = \{p_1 \cdots p_r \leqslant x^\rho : z_1 \geqslant p_1 > \cdots > p_r, \\ p_1 \cdots p_{2k-2} p_{2k-1}^3 \leqslant x^\rho \text{ for all } k \geqslant 1\}$$

be the support of the upper bound linear sieve weights with level x^ρ and sifting parameter $z_1 \leqslant x^{1/2}$. Then, for any $D \in [x^{1/5}, x^\rho]$, every $d \in \mathcal{D}^{+,\mathrm{LIN}}$ can be written as $d = d_1 d_2$, where the positive integers d_1 and d_2 satisfy $d_1 \leqslant D$, $d_1 d_2^2 \leqslant x^{1-4\theta-2\varepsilon^2}/D$. Moreover, we can take either $d_1 \geqslant x^{0.1}$ or $d_2 = 1$.

Proof. The proof is similar to the proof of [**2**, Lemma 12.16] (which essentially says that the linear sieve weights $\lambda_d^{+,\mathrm{LIN}}$ are well-factorable for any sifting parameter $z \leqslant x^{1/2-\varepsilon}$). We will actually show that any $d = p_1 \cdots p_r \in \mathcal{D}^{+,\mathrm{LIN}}$ can be written as $d = d_1 d_2$ with $d_1 \leqslant D$, $d_2 \leqslant x^\rho/D$ and either $d_1 \geqslant x^{0.1}$ or $d_2 = 1$. After that statement has been proved, we have proved the lemma, because then $d_1 d_2^2 \leqslant x^{2\rho}/D \leqslant x^{1-4\theta-2\varepsilon^2}/D$. We use induction on r to prove the existence of such d_1 and d_2. For $r = 1$, we can simply take $d_1 = p_1$ and $d_2 = 1$, since $p_1 \leqslant x^{\rho/3} \leqslant x^{1/6}$. If $r = 2$, we can take $d_1 = p_1 p_2$, $d_2 = 1$, unless $p_1 p_2 > D$. In the case $p_1 p_2 > D$, in turn, the choice $d_1 = p_1$, $d_2 = p_2$ works, since $p_1 \leqslant x^{1/6}$ and $p_2 \leqslant x^\rho/p_1 p_2 \leqslant x^\rho/D$. Suppose then that $r \geqslant 3$ and that case $r-1$ has been proved and consider the case r. We have $p_1 \cdots p_{r-1} \in \mathcal{D}^{+,\mathrm{LIN}}$, so by the induction assumption $p_1 \cdots p_{r-1} = d_1' d_2'$ with $d_1' \leqslant D$, $d_2' \leqslant x^\rho/D$ and either $d_1' \geqslant x^{0.1}$ or $d_2' = 1$. We claim that we can take

either $d_1 = d_1' p_r, d_2 = d_2'$ or $d_1' = d_1, d_2 = d_2' p_r$. Firstly, if $d_1' < x^{0.1}$, then $d_2' = 1$ and $d_1' = p_1 \cdots p_{r-1}$. Since $r \geqslant 3$, this yields $p_1 p_2 < x^{0.1}$, so $p_2 < x^{0.05}$. Now the choice $d_1 = d_1' p_r$, $d_2 = d_2' = 1$ works because $d_1 < x^{0.1} p_r \leqslant x^{0.15} \leqslant D$. Secondly, if in the opposite case $d_1' \geqslant x^{0.1}$ neither of the choices for (d_1, d_2) works, then $d_1' d_2' p_r^2 > x^\rho$. However, $d_1' d_2' p_r^2 = p_1 \cdots p_{r-1} p_r^2 \leqslant x^\rho$ by the definition of $\mathcal{D}^{+,\mathrm{LIN}}$, so we have a contradiction and the induction works. □

9.2. *Semilinear sieve weights.*

LEMMA 9.2. *Let $\varepsilon > 0$ be small, $0 \leqslant \theta \leqslant \frac{1}{30}$ and $\rho = \frac{3}{7}(1 - 4\theta) - \varepsilon$. Let*

$$\mathcal{D}^{-,\mathrm{SEM}} = \{p_1 \cdots p_r \leqslant x^\rho : z_2 \geqslant p_1 > \cdots > p_r, \\ p_1 \cdots p_{2k-1} p_{2k}^2 \leqslant x^\rho \text{ for all } k \geqslant 1\}$$

be the support of the lower bound semilinear sieve weights with level x^ρ and sifting parameter $z_2 \leqslant x^{1/3-2\theta-2\varepsilon^2}$. Then, for any $D \in [x^{1/3-2\theta-2\varepsilon^2}, x^\rho]$, every $d \in \mathcal{D}^{-,\mathrm{SEM}}$ can be written as $d = d_1 d_2$, where the positive integers d_1 and d_2 satisfy $d_1 \leqslant D$, $d_1 d_2^2 \leqslant x^{1-4\theta-2\varepsilon^2}/D$. Moreover, we can take either $d_1 \geqslant x^{0.1}$ or $d_2 = 1$.

Remark 9.3. The exponent $\rho = \frac{3}{7}(1 - 4\theta) - \varepsilon$ is optimal in Lemma 9.2. Namely, if $\rho = \frac{3}{7}(1-4\theta) + 3\varepsilon$, then the lemma is false for $D = x^{(3/7)(1-4\theta)}$ and $p_1 p_2 p_3 \in \mathcal{D}^{-,\mathrm{SEM}}$, $p_1, p_2, p_3 \sim \frac{1}{2} x^{(1/7)(1-4\theta)+\varepsilon}$.

Remark 9.4. We remark that an argument almost identical to the proof of Lemma 9.2 below shows that the lemma holds also for the set

$$\mathcal{D}^{+,\mathrm{SEM}} = \{p_1 \cdots p_r \leqslant x^\rho : x^{1/2} \geqslant p_1 \geqslant \cdots \geqslant p_r, \\ p_1 \cdots p_{2k-2} p_{2k-1}^2 \leqslant x^\rho \text{ for all } k \geqslant 1\},$$

which is the support of the upper bound semilinear weights, when $\rho = \frac{2}{5}(1 - 4\theta) - \varepsilon$, $\theta \leqslant \frac{1}{40}$ and all the other parameters are as before. This observation will be used in the proof of Theorem 1.5. This exponent is also optimal, as is seen by taking $\rho = \frac{2}{5}(1 - 4\theta) + 2\varepsilon$ and $D = x^{(2/5)(1-4\theta)}$, $p_1 p_2 \in \mathcal{D}^{+,\mathrm{SEM}}$, $p_1, p_2 \sim \frac{1}{2} x^{(1/5)(1-4\theta)+\varepsilon}$.

Proof of Lemma 9.2. The proof resembles some arguments related to Harman's sieve [**7**, Ch. 3]. Let $d = p_1 \cdots p_r \in \mathcal{D}^{-,\mathrm{SEM}}$. The claim is that the set $\{p_1, \ldots, p_r\}$ can be partitioned into two subsets S_1 and S_2 in such a way that the products P_1 and P_2 of the elements of S_1 and S_2 satisfy $P_1 \leqslant D$, $P_1 P_2^2 \leqslant x^{1-4\theta-2\varepsilon^2}/D$ and additionally $P_1 \geqslant x^{0.1}$ or $P_2 = 1$. Note that for $r = 1$, one can take $S_1 = \{p_1\}$ and $S_2 = \emptyset$. Assume then that $r \geqslant 2$. If $p_1 \cdots p_r \leqslant D$, we may take $S_1 = \{p_1, \ldots, p_r\}$, $S_2 = \emptyset$. Indeed, then $P_1 \leqslant D$, $P_2 = 1$ and $P_1 P_2^2 \leqslant D \leqslant x^{1-4\theta-2\varepsilon^2}/D$. Now we may assume that $p_1 \cdots p_r > D$. Since

$p_1 \leqslant D$, we can select the largest j for which $p_1 \cdots p_j \leqslant D$. We have $j \leqslant r-1$ and $p_{j+1} \leqslant p_2 \leqslant x^{\rho/3}$, so

$$p_1 \cdots p_j = \frac{p_1 \cdots p_{j+1}}{p_{j+1}} \geqslant \frac{D}{x^{\rho/3}}.$$

We claim that the choice $S_1 = \{p_1, \dots, p_j\}$, $S_2 = \{p_{j+1}, \dots, p_r\}$ works. First of all, we have $P_1 \geqslant D/x^{\rho/3} \geqslant x^{0.1}$. Supposing that the claim does not hold for S_1 and S_2, we have $(P_1P_2)^2 > P_1 x^{1-4\theta-2\varepsilon^2}/D$. Using $P_1P_2 \leqslant x^\rho$ and $P_1 \geqslant D/x^{\rho/3}$, this yields $x^{2\rho} > x^{1-4\theta-\rho/3-2\varepsilon^2}$, from which we solve $\rho > \frac{3}{7}(1-4\theta) - \frac{6}{7}\varepsilon^2$, which is a contradiction to our choice of ρ. □

§10. *Verifying the hypothesis.*

10.1. *Splitting variables.* Based on §7, the proof of Hypothesis 6.4 for the sequence $(\omega_n)_{n\sim x}$ and linear function $L(n)$ defined in that section has been reduced to showing that

$$\sum_{\substack{d\leqslant x^{\rho_2}\\(d,QW)=1}} \lambda_d^{-,\mathrm{SEM}} \Bigg(\sum_{\substack{n\sim x\\ L(n)\in\mathbb{P}\\ L(n)\equiv 1 \pmod d}} e(\xi n) - \frac{1}{\varphi(d)} \frac{QW}{\varphi(QW)} \sum_{n\sim x} \frac{e(\xi n)}{\log(QWn)} \Bigg) \tag{10.1}$$

and

$$\sum_{\substack{d\leqslant x^{\rho_1}\\(d,QW)=1}} \lambda_d^{+,\mathrm{LIN}} \sum_{\substack{\ell\leqslant x^{1-\varepsilon}\\(\ell,QW)=\delta\\(\ell,d)=1}} g(\ell) \Bigg(\sum_{\substack{n\sim x\\ L(n)=\ell p+1\\ L(n)\equiv 0 \pmod d}} e(\xi n) - \frac{1}{\varphi(d)} \frac{QW}{\varphi(QW/\delta)} \sum_{n\sim x} \frac{e(\xi n)}{\ell \log(QWn/\ell)} \Bigg) \tag{10.2}$$

are $\ll x(\log x)^{-200}$, where $\delta = (Wc_0 + b - 1, QW)$, $(g(\ell))_{\ell\geqslant 1}$ is a sequence of convolution type (with parameter σ), the sieve weights $\lambda_d^{+,\mathrm{LIN}}$, $\lambda_d^{-,\mathrm{SEM}}$ have respective sifting parameters $z_1 \leqslant x^{1/5+\varepsilon}$, $z_2 \leqslant x^{1/(3+\varepsilon/2)}$, ρ_1, ρ_2, σ are as in (7.2) and ξ is subject to (7.3). It would actually suffice to replace $\ell \leqslant x^{1-\varepsilon}$ by $\ell \leqslant x^{0.9+\varepsilon}$ above, but this would not simplify the argument.

As mentioned in §9, we wish to split the sum over d into a double sum. This is enabled by Lemmas 9.1 and 9.2. If D is as in Lemma 9.2 with $0 \leqslant \theta \leqslant \frac{1}{30}$, we may write

$$|\lambda_d^{-,\mathrm{SEM}}| \leqslant \min_D \sum_{\substack{d=d_1d_2\\ d_1\leqslant D\\ d_1d_2^2\leqslant x^{1-4\theta-2\varepsilon^2}/D\\ (d_1,d_2)=1\\ d_1\geqslant x^{0.1} \text{ or } d_2=1}} 1 \leqslant \left(\frac{\log x}{\log 2}\right)^2 \min_D \max_{\Delta_1,\Delta_2} \sum_{\substack{d=d_1d_2\\ d_1\sim\Delta_1\\ d_2\sim\Delta_2\\ (d_1,d_2)=1}} 1, \tag{10.3}$$

where the maximum and minimum are over those $\Delta_1, \Delta_2 \geqslant 1$ and $D \geqslant 1$ that satisfy

$$\begin{aligned} D \in [x^{1/3-2\theta-2\varepsilon^2}, x^{\rho_2}], \quad \Delta_1 \leqslant D, \quad \Delta_1\Delta_2^2 \leqslant \frac{x^{1-4\theta-2\varepsilon^2}}{D}, \\ \Delta_1\Delta_2 \leqslant x^{\rho_2}, \quad \text{and either } \Delta_1 \geqslant x^{0.1} \text{ or } \Delta_2 = 1. \end{aligned} \tag{10.4}$$

By Lemma 9.1, formula (10.3) continues to hold with $\lambda_d^{-,\mathrm{SEM}}$ replaced with $\lambda_d^{+,\mathrm{LIN}}$ and (10.4) replaced with

$$\begin{aligned} D \in [x^{1/5}, x^{\rho_1}], \quad \Delta_1 \leqslant D, \quad \Delta_1\Delta_2^2 \leqslant \frac{x^{1-4\theta-2\varepsilon^2}}{D}, \\ \Delta_1\Delta_2 \leqslant x^{\rho_1}, \quad \text{and either } \Delta_1 \geqslant x^{0.1} \text{ or } \Delta_2 = 1. \end{aligned} \tag{10.5}$$

We take $\theta = 0$ in this section, but in §11 we will employ the same formulas with $\theta > 0$. As a conclusion, we see that (10.1) and (10.2) are bounded by $((\log x)/(\log 2))^2$ times

$$\sum_{\substack{d_1\sim\Delta_1\\(d_1,QW)=1}} \sum_{\substack{d_2\sim\Delta_2\\(d_2,QW)=1\\(d_1,d_2)=1}} \Bigg| \sum_{\substack{n\sim x\\L(n)\in\mathbb{P}\\L(n)\equiv 1 \pmod{d_1d_2}}} e(\xi n) - \frac{QW}{\varphi(d_1d_2)\varphi(QW)} \sum_{n\sim x} \frac{e(\xi n)}{\log(QWn)} \Bigg| \tag{10.6}$$

and

$$\begin{aligned} \sum_{\substack{d_1\sim\Delta_1\\(d_1,QW)=1}} \sum_{\substack{d_2\sim\Delta_2\\(d_2,QW)=1\\(d_1,d_2)=1}} \Bigg| \sum_{\substack{\ell\leqslant x^{1-\varepsilon}\\(\ell,QW)=\delta\\(\ell,d_1d_2)=1}} g(\ell) \Bigg(\sum_{\substack{n\sim x\\L(n)=\ell p+1\\L(n)\equiv 0 \pmod{d_1d_2}}} e(\xi n) \\ - \frac{QW}{\varphi(d_1d_2)\varphi(QW/\delta)} \sum_{n\sim x} \frac{e(\xi n)}{\ell \log(QWn/\ell)} \Bigg) \Bigg|, \end{aligned} \tag{10.7}$$

respectively, where Δ_1 and Δ_2 are any numbers constrained by (10.4) or (10.5), depending on whether we consider (10.6) or (10.7). At this point, it is also natural to split into two cases depending on whether ξ lies on a major arc or a minor arc (that is, whether $q \mid Q$ or $q/(q, Q^2) \geqslant (\log x)^A$ holds in (7.3)).

10.2. *Major arcs for the semilinear sieve.* We first assume the major arc condition $q \mid Q$ in the definition of ξ in (7.3). By partial summation, (10.1) becomes

$$\begin{aligned} = \int_x^{2x} e(\pm\|\xi\|t)\, d\Bigg\{ \sum_{\substack{d\leqslant x^{\rho_2}\\(d,QW)=1}} \lambda_d^{-,\mathrm{SEM}} \\ \times \Bigg(\sum_{\substack{x\leqslant n\leqslant t\\L(n)\in\mathbb{P}\\L(n)\equiv 1 \pmod d}} 1 - \frac{QW}{\varphi(QW)} \frac{1}{\varphi(d)} \sum_{x\leqslant n\leqslant t} \frac{1}{\log(QWn)} \Bigg) \Bigg\}. \end{aligned}$$

Naming the function inside $d\{\dots\}$ as $G(t)$, partial integration tells us that the previous expression is

$$\begin{aligned} &= G(2x)e(\pm 2\|\xi\|x) \mp 2\pi \mathrm{i}\|\xi\| \int_x^{2x} e(\pm\|\xi\|t)G(t)\,dt \\ &\ll (1+\|\xi\|x) \max_{x\leqslant t\leqslant 2x} |G(t)|. \end{aligned} \tag{10.8}$$

Since $1/\log(QWn) = (1/QW)\int_{QWn}^{QW(n+1)} (du/\log u) + O(1/n)$, putting $c_1 = Wc_0 + b$ we have

$$\begin{aligned} G(t) &\leqslant \sum_{\substack{d\leqslant x^{\rho_2}\\(d,QW)=1}} |\lambda_d^{-,\mathrm{SEM}}| \Bigg| \sum_{\substack{QWx\leqslant p\leqslant QWt\\ p\equiv c_1 \pmod{QW}\\ p\equiv 1 \pmod d}} 1 - \frac{1}{\varphi(QWd)} \int_{QWx}^{QWt} \frac{du}{\log u} \Bigg| \\ &\quad + O(x^{1/2}) \\ &\leqslant \sum_{\substack{d\leqslant x^{\rho_2}\\(d,QW)=1}} \max_{(r,QWd)=1} \left| \pi(QWt;\, QWd, r) - \frac{1}{\varphi(QWd)} \operatorname{Li}(QWt) \right| \\ &\quad + \sum_{\substack{d\leqslant x^{\rho_2}\\(d,QW)=1}} \max_{(r,QWd)=1} \left| \pi(QWx;\, QWd, r) - \frac{1}{\varphi(QWd)} \operatorname{Li}(QWx) \right| \\ &\quad + O(x^{1/2}) \\ &\ll \frac{x}{(\log x)^{1000B}} \end{aligned}$$

by the Bombieri–Vinogradov theorem [**10**, Theorem 17.1]. As ξ is on a major arc, by (7.3) we have $\|\xi\| \leqslant 2(\log x)^{102B}/x$, so (10.8) is $\ll x(\log x)^{-1000}$. Therefore, the major arc case for the semilinear sieve has been dealt with.

10.3. *Major arcs for the linear sieve.* Again assume that $q \mid Q$ in (7.3). After applying partial summation, (10.2) takes the form

$$\begin{aligned} &\int_x^{2x} e(\pm\|\xi\|t)\, d\Bigg\{ \sum_{\substack{d\leqslant x^{\rho_1}\\(d,QW)=1}} \lambda_d^{+,\mathrm{LIN}} \\ &\quad \times \Bigg(\sum_{\substack{x\leqslant n\leqslant t\\ L(n)=\ell p+1\\ L(n)\equiv 0 \pmod d\\ \ell\leqslant x^{1-\varepsilon}\\ (\ell,QW)=\delta\\ (\ell,d)=1}} g(\ell) - \frac{QW}{\varphi(d)\varphi(QW/\delta)} \sum_{\substack{x\leqslant n\leqslant t\\ \ell\leqslant x^{1-\varepsilon}\\ (\ell,QW)=\delta\\ (\ell,d)=1}} \frac{g(\ell)}{\ell\log(QWn/\ell)} \Bigg) \Bigg\}, \end{aligned}$$

so we want this to be $\ll x(\log x)^{-202}$. Proceeding as in §10.2, it suffices to prove for $t \in [x, 2x]$ that

$$\sum_{\substack{d\leqslant x^{\rho_1}\\(d,QW)=1}}\Bigg|\sum_{\substack{x\leqslant n\leqslant t\\L(n)=\ell p+1\\L(n)\equiv 0 \pmod d\\\ell\leqslant x^{1-\varepsilon}}} g(\ell)1_{(\ell,QW)=\delta,\ (\ell,d)=1} - \frac{QW}{\varphi(d)\varphi(QW/\delta)}\sum_{\substack{x\leqslant n\leqslant t\\\ell\leqslant x^{1-\varepsilon}\\(\ell,QW)=\delta\\(\ell,d)=1}}\frac{g(\ell)}{\ell\log(QWn/\ell)}\Bigg|$$

is $\ll x(\log x)^{-1000B}$.

We start by analyzing the second sum inside the absolute values in the previous expression. Since $QW \ll (\log x)^{B+1}$ and $\ell \leqslant x^{1-\varepsilon}$, a change of variables and the prime number theorem give

$$\begin{aligned}
&\frac{QW}{\varphi(QW/\delta)}\sum_{x\leqslant n\leqslant t}\frac{1}{\ell\log(QWn/\ell)}\\
&\quad=\frac{QW}{\varphi(QW/\delta)}\int_x^t\frac{du}{\ell\log(QWu/\ell)}+O(QW)\\
&\quad=\frac{1}{\varphi(QW/\delta)}\int_{QWx/\ell}^{QWt/\ell}\frac{du}{\log u}+O(QW)\\
&\quad=\frac{1}{\varphi(QW/\delta)}\sum_{QWx\leqslant\ell p\leqslant QWt}1+O\left(\frac{x}{\ell}(\log x)^{-3000B}\right).
\end{aligned}$$

The error term remains still $\ll x(\log x)^{-2000B}$ after multiplying it by $|g(\ell)|/\varphi(d)$ and summing over $d \leqslant x^{\rho_1}$, $\ell \leqslant x^{1-\varepsilon}$. Hence, what we wish to show is that

$$\sum_{\substack{d\leqslant x^{\rho_1}\\(d,QW)=1}}\Bigg|\sum_{\substack{QWx\leqslant\ell p\leqslant QWt\\\ell p\equiv -1 \pmod d\\\ell p\equiv c_1-1 \pmod{QW}\\\ell\leqslant x^{1-\varepsilon}\\(\ell,QW)=\delta\\(\ell,d)=1}} g(\ell) - \frac{1}{\varphi(QWd/\delta)}\sum_{\substack{QWx\leqslant\ell p\leqslant QWt\\\ell\leqslant x^{1-\varepsilon}\\(\ell,QW)=\delta\\(\ell,d)=1}} g(\ell)\Bigg| \tag{10.9}$$

is $\ll x/(\log x)^{1000B}$ for $t \in [x, 2x]$ and $c_1 = Wc_0 + b$. Since $(\ell, QW) = \delta$, $(\ell, d) = 1$ and $(d, \delta) = 1$, the congruences $\ell p \equiv -1 \pmod d$, $\ell p \equiv c_1 - 1 \pmod{QW}$ can be rewritten as $\ell' p \equiv -\delta^{-1} \pmod d$, $\ell' p \equiv (c_1 - 1/\delta) \pmod{QW/\delta}$ with $\ell' = \ell/\delta$. By the Chinese remainder theorem, these congruences are equivalent to $\ell' p \equiv c \pmod{QWd/\delta}$ for some c depending on Q, W, d and δ and coprime to QWd/δ. Concerning the second sum inside absolute values in (10.9), we wish to add the constraint $(\ell' p, QWd/\delta) = 1$ to that summation (where again $\ell' = \ell/\delta$). We know that $(\ell', QW/\delta) = (\ell', d) = 1$ and clearly $p \geqslant x^{\varepsilon}$ in (10.9), so $(p, QW) = 1$. Therefore,

we have shown that we may insert the constraint $(\ell' p, QWd) = 1$ if the case $p \mid d$ has a small enough contribution to the aforementioned sum. That case contributes at most

$$\sum_{\substack{p \mid d \\ p \geqslant x^{\varepsilon}}} \sum_{\ell \leqslant 2QWx/p} |g(\ell)| \ll_{\varepsilon} x^{1-\varepsilon/2},$$

which is $\ll x^{1-\varepsilon^2}$ when multiplied by $1/\varphi(QWd/\delta)$ and summed over $d \leqslant x^{\rho_1}$. Summarizing, our aim has been reduced to showing that

$$\sum_{\substack{d \leqslant x^{\rho_1} \\ (d, QW)=1}} \max_{(c, QWd/\delta)=1} \Bigg| \sum_{\substack{QWx/\delta \leqslant \ell' p \leqslant QWt/\delta \\ \ell' p \equiv c \pmod{QWd/\delta} \\ \ell' \leqslant x^{1-\varepsilon}/\delta}} g(\delta \ell') - \frac{1}{\varphi(QWd/\delta)} \sum_{\substack{QWx/\delta \leqslant \ell' p \leqslant QWt/\delta \\ (\ell' p, QWd/\delta)=1 \\ \ell' \leqslant x^{1-\varepsilon}/\delta}} g(\delta \ell') \Bigg| \tag{10.10}$$

is $\ll x/(\log x)^{1000B}$ for $t \in [x, 2x]$.

To obtain this estimate, we apply [**10**, Theorem 17.4] to the sequences $(\alpha_{\ell'})_{\ell' \leqslant x^{1-\varepsilon}/\delta} = (g(\delta \ell'))_{\ell' \leqslant x^{1-\varepsilon}/\delta}$ and $(\beta_k)_{k \geqslant 1} = (1_{\mathbb{P}}(k))_{k \geqslant 1}$—that theorem is applicable since the sequence $(1_{\mathbb{P}}(k))_{k \geqslant 1}$ is well-distributed in the sense of [**10**, formula (17.13)] (with $\Delta = (\log x)^{-20000B}$ there) by the Siegel–Walfisz theorem. Now, since in (10.10) we have $\ell' \geqslant x^{\varepsilon/2}$, $p \geqslant x^{\varepsilon}$, $\rho_1 < \frac{1}{2}$ and $|\alpha_{\ell'}| \leqslant \tau(\ell')^2 \log \ell'$, the claimed Bombieri–Vinogradov-type estimate follows immediately from the theorem cited above.

10.4. *Minor arcs for the semilinear sieve.* We assume then that ξ is on a minor arc, meaning that $q/(q, Q^2) \geqslant (\log x)^A$ in (7.3). We study the sum (10.6). Using partial summation, we see that

$$\sum_{n \sim x} \frac{e(\xi n)}{\log(QWn)} \ll \max_{x \leqslant t \leqslant 2x} \Bigg| \sum_{x \leqslant n \leqslant t} e\Big(\xi n\Big) \Bigg| \ll \frac{1}{\|\xi\|}.$$

We have $(q, QW) \leqslant W(q, Q) \leqslant Wq/(\log x)^A < q$, so $q \nmid QW$. Taking this and (7.3) into account, $\|\xi\| \geqslant 1/q - 2(\log x)^{102B}/qx \geqslant 1/2q$, so the second expression inside absolute values in (10.6) is $\ll q/\varphi(d) \ll x/(\log x)^{99B}\varphi(d)$. Hence, it contributes $\ll x(\log x)^{-98B}$ when summing over d.

When it comes to the first expression inside absolute values in (10.6), it equals

$$\sum_{\substack{n \sim x \\ L(n) \in \mathbb{P} \\ L(n) \equiv 1 \pmod{d_1 d_2}}} e(\xi n) = e\Big(\frac{-\xi c_1}{QW}\Big) \sum_{\substack{p \sim QWx \\ p \equiv c_1 \pmod{QW} \\ p \equiv 1 \pmod{d_1 d_2}}} e\Big(\frac{\xi}{QW} p\Big) + O(QW),$$

where the error $O(QW)$ remains $\ll x^{1/2}$ when summed over $d \leqslant x^{\rho_2}$. With partial summation, we may bound the sum on the right-hand side by

$$\Big| \sum_{\substack{n\sim QWx \\ n\equiv c_1 \pmod{QW} \\ n\equiv 1 \pmod{d_1d_2}}} \Lambda(n) e\Big(\frac{\xi}{QW} n\Big) \Big| + \int_{QWx}^{2QWx} \sum_{\substack{QWx\leqslant n\leqslant t \\ n\equiv c_1 \pmod{QW} \\ n\equiv 1 \pmod{d_1d_2}}} \Lambda(n) e\Big(\frac{\xi}{QW} n\Big) \frac{dt}{t\log^2 t} + O(x^{1/2+\varepsilon}),$$

the error coming from the values of n that are prime powers, and the error being $\ll x^{1-\varepsilon^2}$ after summing over $d \leqslant x^{\rho_2}$. This means that it suffices to prove that

$$\sum_{\substack{d_1\sim\Delta_1 \\ (d_1,QW)=1}} \sum_{\substack{d_2\sim\Delta_2 \\ (d_2,QW)=1 \\ (d_1,d_2)=1}} \Big| \sum_{\substack{QWx\leqslant n\leqslant t \\ n\equiv c_1 \pmod{QW} \\ n\equiv 1 \pmod{d_1d_2}}} \Lambda(n) e\Big(\frac{\xi}{QW} n\Big) \Big| \ll \frac{x}{(\log x)^{1000}} \tag{10.11}$$

uniformly for $t \in [QWx, 2QWx]$. We may now apply Vaughan's identity (in the form of [**10**, Proposition 13.4] with $y = z = (QWx)^{1/3}$ there), which transforms the sum inside absolute values in (10.11) (up to error $O(x^{1/3+\varepsilon})$) into a sum of $\ll (\log x)^{10}$ type I and type II sums of the form

$$\widetilde{R}^{\mathrm{I}}_{d_1d_2}(t) = \sum_{\substack{QWx\leqslant mn\leqslant t \\ mn\equiv c_1 \pmod{QW} \\ mn\equiv 1 \pmod{d_1d_2} \\ m\asymp M}} \alpha_m e\Big(\frac{\xi mn}{QW}\Big)$$

and

$$\widetilde{R}^{\mathrm{II}}_{d_1d_2}(t) = \sum_{\substack{QWx\leqslant mn\leqslant t \\ mn\equiv c_1 \pmod{QW} \\ mn\equiv 1 \pmod{d_1d_2} \\ m\asymp M}} \alpha_m\beta_n e\Big(\frac{\xi mn}{QW}\Big)$$

with $|\alpha_m|, |\beta_m| \leqslant \tau(m)^2 \log m$ some complex numbers and $M \leqslant (2QWx)^{1/3}$ in the case of $\widetilde{R}^{\mathrm{I}}_{d_1d_2}(t)$, while $M \in [(QWx)^{1/3}, (2QWx)^{2/3}]$ in the case of $\widetilde{R}^{\mathrm{II}}_{d_1d_2}(t)$. Moreover, we may assume in the latter case that $M \in [(QWx)^{1/2}, (2QWx)^{2/3}]$ by flipping the roles of the variables if necessary. We may replace the type I and type II sums with the (possibly larger) sums

$$\begin{aligned} R^{\mathrm{I}}_{d_1d_2}(t) &= \max_{(c,d_1d_2QW)=1} \Big| \sum_{\substack{QWx\leqslant mn\leqslant t \\ mn\equiv c \pmod{d_1d_2QW} \\ m\asymp M}} \alpha_m e\Big(\frac{\xi}{QW} mn\Big) \Big| \\ R^{\mathrm{II}}_{d_1d_2}(t) &= \max_{(c,d_1QW)=1} \Big| \sum_{\substack{QWx\leqslant mn\leqslant t \\ mn\equiv c \pmod{d_1QW} \\ mn\equiv 1 \pmod{d_2} \\ m\asymp M}} \alpha_m\beta_n e\Big(\frac{\xi}{QW} mn\Big) \Big|. \end{aligned} \tag{10.12}$$

We are now in a position to apply the Bombieri–Vinogradov Lemmas 8.1 and 8.2. Note that, by (7.3), we either have $|\xi - QWa/q| \leqslant 1/(QWq)^2$ or $q > x/2(\log x)^{102B}(QW)^2$. If the latter happens, we have $|e((\xi/QW)mn) - e((a/q)mn)| \leqslant |\xi/QW - a/q|mn \leqslant 8(QW)^3(\log x)^{204B}/x$ for $mn \leqslant 2QWx$. This implies that $e((\xi/QW)mn)$ can be replaced by $e((a/q)mn)$ in the type I and II sums. In conclusion, we can assume in any case that $|\xi - QWa/q| \leqslant 1/(QWq)^2$.

The type I Bombieri–Vinogradov sums cause no problems, as Lemma 8.1 with the choices $R = 1$, $N = QWx$, $v = QW$, $M = x^{1/3+\varepsilon}$, $\rho \leqslant \frac{1}{2} - \varepsilon$ tells us at once that

$$\sum_{\substack{d_1 \sim \Delta_1 \\ (d_1, QW)=1}} \sum_{\substack{d_2 \sim \Delta_2 \\ (d_2, QW)=1 \\ (d_1, d_2)=1}} R^{\mathrm{I}}_{d_1 d_2}(t) \ll x/(\log x)^{A/10},$$

since $q/(q, (QW)^2) \geqslant W^{-2}(\log x)^A$ and $\Delta_1\Delta_2 \leqslant x^{\rho_2}$.

We know that $(QWx)^{1/2} \leqslant M \leqslant (2QWx)^{2/3}$ in the sum $R^{\mathrm{II}}_{d_1 d_2}(t)$. We divide the analysis of this sum into three cases.

Case 1. Assume that $M \geqslant x^{1-\rho_2-\varepsilon^2}$, $\Delta_1 \geqslant (\log x)^{A/10}$. Take $D = x^{1-\varepsilon^2}/M$. We know that $x^{1/3-\varepsilon^2}(\log x)^{-B} \leqslant D \leqslant x^{\rho_2}$ by the bound on M. In view of (10.4) with $\theta = 0$, this means in particular that $\Delta_1 \leqslant x^{1-\varepsilon^2}/M$ and $\Delta_1\Delta_2^2 \leqslant x^{1-2\varepsilon^2}/D = Mx^{-\varepsilon^2}$. Now we apply Lemma 8.2 (in the case of F_1) with $R = 1$, $N = QWx$, $v = QW$, $\rho = \rho_2 \leqslant \frac{3}{7} - \varepsilon$ to deduce that

$$\sum_{\substack{d_1 \sim \Delta_1 \\ (d_1, QW)=1}} \sum_{\substack{d_2 \sim \Delta_2 \\ (d_2, QW)=1 \\ (d_1, d_2)=1}} R^{\mathrm{II}}_{d_1 d_2}(t) \ll x\Bigg(\bigg(\frac{1}{\Delta_1} + \frac{W^2}{(\log x)^A} + (\log x)^{-99B}(QW)^2\bigg)^{1/8} + \bigg(\frac{\Delta_1 M}{x} + \Delta_1\Delta_2^2\frac{QW}{M}\bigg)^{1/2}\Bigg)(\log x)^{1000},$$

which is $\ll x/(\log x)^{A/100}$ for A large enough by the lower bound on Δ_1.

Case 2. Assume then that $M \geqslant x^{1-\rho_2-\varepsilon^2}$, $\Delta_1 < (\log x)^{A/10}$. Since $\Delta_1 < x^{0.1}$, we know that $\Delta_2 = 1$, so applying Lemma 8.2 (in the case of F_2) we obtain, for A large enough,

$$\sum_{\substack{d_1 \sim \Delta_1 \\ (d_1, QW)=1}} \sum_{\substack{d_2 \sim \Delta_2 \\ (d, QW)=1 \\ (d_1, d_2)=1}} R^{\mathrm{II}}_{d_1 d_2}(t) \ll x(\log x)^{A/5}\bigg(\frac{W}{(\log x)^{A/2}} + \frac{QW}{M^{1/2}} + \frac{(QW)^2 M}{x} + \frac{(QW)^{1/2}}{(\log x)^{99B/2}}\bigg)^{1/2}$$

and this is again $\ll x/(\log x)^{A/100}$ for A large.

Case 3. Lastly, assume that $M < x^{1-\rho_2-\varepsilon^2}$. Then we estimate (10.1) instead of (10.6). This amounts to just replacing $d_1 \sim \Delta_1$, $d_2 \sim \Delta_2$ with $d_1 \leqslant x^{\rho_2}$, $d_2 = 1$ throughout this subsection. We have $x^{\rho_2} \leqslant x^{1-\varepsilon^2}/M$ and $x^{\rho_2} \leqslant Mx^{-\varepsilon^2}$, so we can bound the type II sums in the same way as for $M \geqslant x^{1-\rho_2-\varepsilon^2}$ (considering again the cases $\Delta_1 \geqslant (\log x)^{A/10}$ and $\Delta_1 < (\log x)^{A/10}$ separately), so also Case 3 contributes $\ll x/(\log x)^{A/100}$.

Consequently, we have shown that the contribution of the minor arcs for the semilinear sieve is small enough.

10.5. *Minor arcs for the linear sieve.* We assume again that $q/(q, Q^2) \geqslant (\log x)^A$. We first look at the second expression inside absolute values in (10.7). We have by partial summation

$$\sum_{n\sim x} \frac{e(\xi n)}{\ell \log(QWn/\ell)} \ll \frac{1}{\ell\|\xi\|}$$

for $\ell \leqslant x^{1-\varepsilon}$ just as in §10.4. We showed earlier that $1/\|\xi\| \ll x/(\log x)^{99B}$ when $q/(q, Q^2) \geqslant (\log x)^A$, so the second expression inside absolute values in (10.7) is $\ll x/\ell\varphi(d)(\log x)^{98B}$, which is $\ll x(\log x)^{-97B}$ after summing over $d \leqslant x^{\rho_1}$ and over $\ell \leqslant x^{1-\varepsilon}$ weighted by $|g(\ell)|$.

We may write the first expression inside absolute values in (10.7) as

$$e\left(\frac{-(c_1-1)\xi}{QW}\right) \sum_{\substack{\ell p\sim QWx\\ \ell p\equiv c_1-1 \pmod{QW}\\ \ell p\equiv -1 \pmod d\\ \ell\leqslant x^{1-\varepsilon}}} g(\ell)e\left(\frac{\xi}{QW}\ell p\right) + O(QW) \qquad (10.13)$$

and the error $O(QW)$ is $\ll x^{1/2}$ after summing over $d \leqslant x^{\rho_1}$. We have ignored the conditions $(\ell, QW) = \delta$, $(\ell, d) = 1$ above, since if either of them fails, $\ell p \equiv c_1 - 1 \pmod{QW}$, $\ell p \equiv -1 \pmod d$ is impossible.

Crucially, our assumption is that the sequence $(g(\ell))_{\ell\geqslant 1}$ is of convolution type, so the sum in (10.13) can be rewritten as

$$\sum_{\substack{kmp\sim QWx\\ kmp\equiv c_1-1 \pmod{QW}\\ kmp\equiv -1 \pmod d\\ km\leqslant x^{1-\varepsilon}}} \alpha_k\beta_m e\left(\frac{\xi}{QW}kmp\right),$$

where (α_k) is supported on $x^{1/\sigma} \leqslant k \leqslant (Qx)^{1-1/\sigma}$ for $\sigma = 3 + \varepsilon$. Putting

$$\beta_r^* = \sum_{r=mp} \beta_m$$

and splitting the previous sum dyadically, it becomes $\ll \log x$ sums of the form

$$\sum_{\substack{kr\sim QWx\\ kr\equiv c_1-1 \pmod{QW}\\ kr\equiv -1 \pmod d\\ k\asymp M}} \alpha_k\beta_r^* e\left(\frac{\xi}{QW}kr\right),$$

where $x^{1/\sigma} \leqslant M \leqslant (Qx)^{1-1/\sigma}$ and, by changing the roles of the variables, we may further assume that $(QWx)^{1/2} \leqslant M \leqslant Qx^{1-1/\sigma}$. Now our bilinear sums are exactly of the same form as in (10.12) (but with different M). Furthermore, we may assume that $|\xi - QWa/q| \leqslant 1/(QWq)^2$ for the same reason as in §10.4. If $M \geqslant x^{1-\rho_1-\varepsilon^2}$, denoting $D = x^{1-\varepsilon^2}/M \in [x^{1/5}, x^{\rho_1}]$, we again see that $\Delta_1\Delta_2^2 \leqslant Mx^{-\varepsilon^2}$ in (10.5) (with $\theta = 0$). Therefore, we may apply the very same estimates as in the Cases 1 and 2 of §10.4. If $M < x^{1-\rho_1-\varepsilon^2}$, we can apply precisely the same argument as in Case 3 of the previous subsection, since $x^{\rho_1} \leqslant x^{1-\varepsilon^2}/M$ and $x^{\rho_1} \leqslant Mx^{-\varepsilon^2}$. Summarizing, we have showed that the minor arcs for the linear sieve contribute $\ll x(\log x)^{-A/100}$, which is small enough for large A.

We have now concluded the proof of Theorem 1.1, in view of Theorem 6.5 and Proposition 5.1. □

Proof of Theorem 1.5. We take $Q = W = 1$ and $L(n) = n$ in (10.1) and replace $L(n) \equiv 1 \pmod d$ by $L(n) \equiv b \pmod d$ (with $b \neq 0$ an arbitrary integer) there and note that the proof that (10.1) is $\ll_C x(\log x)^{-C}$ is verbatim the same as the minor arc argument for the semilinear sieve in this section, provided that ξ is any real number with $|\xi - a/q| \leqslant 1/q^2$ for some coprime a and $q \in [(\log x)^{1000C}, x(\log x)^{-1000C}]$. This proves Theorem 1.5 in the case of lower bound sieve weights. The case of upper bound sieve weights follows very similarly by replacing $\lambda_d^{-,\mathrm{SEM}}$ with $\lambda_d^{+,\mathrm{SEM}}$ and making use of a remark after Lemma 9.2 (which is where the value $\rho_+ = \frac{2}{5} - \varepsilon$ comes from). □

§11. *The distribution of ξp modulo* 1. We show that our considerations on primes $x^2 + y^2 + 1$ in Bohr sets imply a result about the distribution of irrational multiples of such primes, in the form of Theorem 1.4.

For proving Theorem 1.4, it suffices to prove that, given an irrational $\xi > 0$, there exist infinitely many integers $N \geqslant 1$ such that some prime $p \sim N$ of the form $x^2 + y^2 + 1$ satisfies $\|\xi p + \kappa\| \leqslant N^{-\theta}/2$. Let χ_0 be a 1-periodic function which is a lower bound for the characteristic function of $[-\eta/2, \eta/2]$ with $\eta = N^{-\theta}$. Specifically, as in [**14**], we choose χ_0 so that

$$0 \leqslant \chi_0(t) \leqslant 1, \quad \chi_0(t) = 0 \quad \text{when } t \notin \left[-\frac{\eta}{2}, \frac{\eta}{2}\right],$$
$$\chi_0(t) = \frac{\eta}{2} + \sum_{|r|>0} c(r)e(rt) \quad \text{with } c(r) \ll \eta$$
$$\text{and} \quad \sum_{|r|>R} |c(r)| \ll R^{-1} \quad \text{for } R = \eta^{-1}(\log \eta^{-1})^C$$

for some large constant C. This construction goes back to Vinogradov's work. What we want to show is that

$$\sum_{\substack{p\sim N \\ p\in\mathcal{S}+1}} \chi_0(\xi p + \kappa) \geqslant \delta_0 \frac{\eta N}{(\log N)^{3/2}} \tag{11.1}$$

for some absolute constant $\delta_0 > 0$ and infinitely many N. From now on, we choose a large integer q satisfying $|\xi - a/q| \leqslant 1/q^2$ for some a coprime to q (there are infinitely many such q) and take

$$N = q^2, \qquad R = \eta^{-1}(\log \eta^{-1})^C \asymp N^{\theta}(\log N^{\theta})^C. \tag{11.2}$$

Concerning the term on the right-hand side of (11.1), we note that

$$\begin{aligned}\sum_{n\sim N} \chi_0(\xi n + \kappa) - \frac{\eta}{2}N &\ll \eta \sum_{0<|r|\leqslant R}\left|\sum_{n\sim N} e(\xi r n)\right| + \frac{N}{R}\\ &\ll \eta \sum_{0<|r|\leqslant R} \frac{1}{\|\xi r\|} + \eta N(\log N)^{-C}\\ &\ll \eta q \log 2q + \eta N(\log N)^{-C}\\ &\ll \eta N(\log N)^{-C}\end{aligned}$$

for $2\varepsilon \leqslant \theta \leqslant \frac{1}{2} - \varepsilon$, so (11.1) takes the form

$$\sum_{\substack{p\sim N\\ p\in\mathcal{S}+1}} \chi_0(\xi p + \kappa) \geqslant \frac{\delta_1}{(\log N)^{3/2}} \sum_{n\sim N} \chi_0(\xi n + \kappa) \tag{11.3}$$

for some absolute constant $\delta_1 > 0$. This is what we set out to prove.

Proof of Theorem 1.4. Pick any amenable linear polynomial, such as $L(n) = Kn + 5$ with $K = 6^4$. By applying Theorem 6.5 to $\omega_n = \chi_0(K\xi n + \kappa + 5\xi)$ and $L(n)$, we see that (11.3) will follow (with N replaced by N/K) once we establish Hypothesis 6.4 (with $\delta = (K, 5-1) = 4$) for this sequence (ω_n) and some parameters satisfying $\mathrm{H}(\rho_1, \rho_2, \sigma)$ under the conditions (11.2). Taking the definition of $\chi_0(\cdot)$ into account and making use of the classical Bombieri–Vinogradov theorem, it suffices to prove Hypothesis 6.4 for $\omega_n' = \sum_{0<|r|<R} c(r)e(K\xi r n)$ (with the choices (11.2)). Hence, what we must show is that

$$\sum_{\substack{d\leqslant N^{\rho_2}\\ (d,K)=1}} |\lambda_d^{-,\mathrm{SEM}}| \sum_{0<|r|<R}\left|\sum_{\substack{n\sim N\\ Kn+5\in\mathbb{P}\\ Kn+4\equiv 0 \pmod d}} e(K\xi r n) - \frac{K}{\varphi(Kd)}\sum_{n\sim N}\frac{e(K\xi r n)}{\log(Kn)}\right|$$

and

$$\begin{aligned}&\sum_{\substack{d\leqslant N^{\rho_1}\\ (d,K)=1}} |\lambda_d^{+,\mathrm{LIN}}| \sum_{0<|r|<R}\\ &\quad\times\left|\sum_{\substack{\ell\leqslant N^{1-\varepsilon}\\ (\ell,d)=1\\ (\ell,K)=\delta}} g(\ell)\Bigg(\sum_{\substack{n\sim N\\ Kn+4=\ell p\\ Kn+5\equiv 0 \pmod d}} e(K\xi r n) - \frac{K}{\varphi(Kd)}\sum_{n\sim N}\frac{e(K\xi r n)}{\ell\log(Kn/\ell)}\Bigg)\right|\end{aligned}$$

are $\ll N/(\log N)^{100}$, where $\lambda_d^{-,\mathrm{SEM}}$ has sifting parameter $z_2 \ll N^{1/\sigma}$, while $\lambda_d^{+,\mathrm{LIN}}$ has sifting parameter $z_1 \ll N^{1/5}$. We know that $|K\xi - a'/q'| \leqslant 6^4/q'^2$ for some coprime a' and $q' \asymp N^{1/2}$, so the minor arc arguments from §10 allow replacing the previous Bombieri–Vinogradov sums (up to error $\ll N^{1-\varepsilon}$) with the sums

$$\begin{aligned}&\sum_{\substack{d\leqslant N^{\rho_2}\\(d,K)=1}} |\lambda_d^{-,\mathrm{SEM}}| \sum_{0<|r|<R} \Bigg| \sum_{\substack{n\sim N\\ Kn+5\in\mathbb{P}\\ Kn+4\equiv 0 \ (\mathrm{mod}\ d)}} e(K\xi rn)\Bigg| \quad \text{and}\\ &\sum_{\substack{d\leqslant N^{\rho_1}\\(d,K)=1}} |\lambda_d^{+,\mathrm{LIN}}| \sum_{0<|r|<R} \Bigg| \sum_{\substack{\ell\leqslant N^{1-\varepsilon}\\(\ell,d)=1\\(\ell,K)=\delta}} g(\ell) \sum_{\substack{n\sim N\\ Kn+4=\ell p\\ Kn+5\equiv 0 \ (\mathrm{mod}\ d)}} e(K\xi rn)\Bigg|.\end{aligned} \tag{11.4}$$

Splitting the variables as in §10.1 and again employing the minor arc arguments from §10, the sums in (11.4) reduce to $\ll (\log N)^{10}$ sums of the same form as in Lemmas 8.1 and 8.2 with

$$R \leqslant N^{\theta}(\log N)^C, \quad v = 1, \quad q \asymp N^{1/2}, \quad M \ll N^{1/3}$$

in the type I case, while

$$\begin{gathered}R \leqslant N^{\theta}(\log N)^C, \quad v = 1, \quad q \asymp N^{1/2}, \quad M \in [N^{1/2}, N^{2/3+\varepsilon^2}],\\ \Delta_1, \Delta_2 \quad \text{subject to (10.4)}\end{gathered}$$

(with x replaced by N in (10.4)) in the type II sums arising from the semilinear sieve weights and

$$\begin{gathered}R \leqslant N^{\theta}(\log N)^C, \quad v = 1, \quad q \asymp N^{1/2}, \quad M \in [N^{1/2}, N^{3/4-\varepsilon}],\\ \Delta_1, \Delta_2 \quad \text{subject to (10.5)}\end{gathered}$$

(with x replaced by N in (10.5)) in the type II sums arising from the linear sieve weights.

From now on, we fix the values

$$\rho_1 = \frac{1}{2}(1-4\theta) - \varepsilon, \qquad \rho_2 = \frac{3}{7}(1-4\theta) - \varepsilon, \qquad \sigma = \frac{1}{1/3 - 2\theta} + \varepsilon.$$

The bound offered by Lemma 8.1 for the type I sums we face is evidently $\ll N^{1-\varepsilon^2}$ for $\theta \leqslant \frac{1}{30}$. This takes care of the type I sums.

We turn to the type II sums that are of the same form as in Lemma 8.2. Utilizing Lemma 8.2, such Bombieri–Vinogradov sums are bounded by

$$\ll RN(\log N)^{1000}\left(\left(\frac{\Delta_1 M}{N} + \frac{\Delta_1\Delta_2^2}{M}\right)^{1/2} + \left(\frac{1}{\Delta_1} + \frac{1}{N^{1/2}}\right)^{1/8}\right) \tag{11.5}$$

when $\Delta_1\Delta_2 \leqslant N^{1/2}$ and $\Delta_1\Delta_2^2 \leqslant M$. For $R \leqslant N^\theta(\log N)^C$, the estimate (11.5) is $\ll N^{1-0.1\varepsilon^2}$, provided that

$$\Delta_1 \leqslant \frac{N^{1-2\theta-\varepsilon^2}}{M}, \quad \Delta_1\Delta_2^2 \leqslant MN^{-2\theta-\varepsilon^2},$$
$$\Delta_1 \geqslant N^{0.1}, \quad \theta \leqslant \frac{1}{80} - \varepsilon. \tag{11.6}$$

We deal with the type II sums in three cases. We will use ρ to denote either ρ_1 or ρ_2.

Case 1. Suppose that $M \geqslant N^{1-\rho-2\theta-\varepsilon^2}$, $\Delta_1 \geqslant N^{0.1}$. By taking $D = N^{1-2\theta-\varepsilon^2}/M$ in (10.4)–(10.5) and using the fact that $1/\sigma \leqslant \frac{1}{3} - 2\theta - 2\varepsilon^2$, we can indeed achieve (11.6) as long as $D \in [N^{1/5}, N^\rho]$ in the case of the linear sieve and $D \in [N^{1/3-2\theta-2\varepsilon^2}, N^\rho]$ in the case of the semilinear sieve. The inequality $D \leqslant N^\rho$ holds due to our lower bound on M. The inequality $D \geqslant N^{1/5}$ holds for $M \leqslant N^{3/4}$, which is true in the linear case. Similarly, the inequality $D \geqslant N^{1/3-2\theta-2\varepsilon^2}$ reduces to $M \leqslant N^{2/3+\varepsilon^2}$ and this holds in the semilinear case. Therefore, in this case (11.6) is always valid, which means that our type II sums are $\ll N^{1-0.1\varepsilon^2}$, which is what we wanted.

Case 2. Suppose that $M \geqslant N^{1-\rho-2\theta-\varepsilon^2}$, $\Delta_1 < N^{0.1}$. In this case we know that $\Delta_2 = 1$ from (10.4) and (10.5). Now, choosing F_2 in Lemma 8.2, we obtain for the type II Bombieri–Vinogradov sum the bound

$$\ll RN\Delta_1\left(\frac{1}{N^{1/4}} + \frac{1}{M^{1/2}} + \frac{M}{N} + \frac{N^{1/4}}{(RN)^{1/2}}\right)^{1/2} \ll RN\Delta_1 N^{-1/8} \ll N^{0.999}$$

when $\theta \leqslant \frac{1}{50}$.

Case 3. Suppose finally that $M < N^{1-\rho-2\theta-\varepsilon^2}$, $\Delta_1 \geqslant N^{0.1}$. Similarly as in Case 3 of §10.4, we may take $\Delta_1 = N^\rho$, $\Delta_2 = 1$. Again we require this choice to fulfill (11.6). The first constraint in (11.6) follows directly from our upper bound on M. Since $M \geqslant N^{1/2}$, the second constraint in (11.6) holds for $\rho \leqslant \frac{1}{2} - 2\theta - \varepsilon^2$, which certainly holds for our choices of ρ_1 and ρ_2. This means that also in Case 3 we get good enough bounds for the type II sums. Putting everything together, in each of the Cases 1–3 we get a good enough bound for the type II sums.

Combining the analyses of the Cases 1–3, we see that Theorem 1.4 will follow with exponent θ if $\mathrm{H}(\rho_1, \rho_2, \sigma)$ is true for $\sigma = 1/(1/3 - 2\theta) + \varepsilon$, $\rho_1 = \frac{1}{2}(1 - 4\theta) - \varepsilon$ and $\rho_2 = \frac{3}{7}(1-4\theta) - \varepsilon$, provided that $\theta \leqslant \frac{1}{80} - \varepsilon$. By continuity, it suffices to check $\mathrm{H}(\frac{1}{2}(1 - 4\theta), \frac{3}{7}(1 - 4\theta), 1/(1/3 - 2\theta))$ for $\theta = \frac{1}{80}$ and this holds by a numerical computation (the difference between the left- and right-hand sides of (6.1) is then $> 10^{-3}$). This completes the proof of Theorem 1.4. □

Acknowledgements. The author is grateful to his supervisor Kaisa Matomäki for various useful comments and discussions. The author thanks the referee for careful reading of the paper and for useful comments. While working on this project, the author was funded by UTUGS Graduate School and project number 293876 of the Academy of Finland.

References

1. S. Baier, A note on Diophantine approximation with Gaussian primes. *Preprint*, 2016, arXiv:1609.08745 [math.NT].
2. J. Friedlander and H. Iwaniec, *Opera de Cribro* (*American Mathematical Society Colloquium Publications* **57**), American Mathematical Society (Providence, RI, 2010).
3. B. Green, Roth's theorem in the primes. *Ann. of Math.* (2) **161**(3) (2005), 1609–1636.
4. B. Green and T. Tao, Restriction theory of the Selberg sieve, with applications. *J. Théor. Nombres Bordeaux* **18**(1) (2006), 147–182.
5. B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.* (2) **167**(2) (2008), 481–547.
6. V. Z. Guo, Piatetski–Shapiro primes in a Beatty sequence. *J. Number Theory* **156** (2015), 317–330.
7. G. Harman, *Prime-Detecting Sieves* (*London Mathematical Society Monographs Series* **33**), Princeton University Press (Princeton, NJ, 2007).
8. H. Iwaniec, Primes of the type $\phi(x, y) + A$ where ϕ is a quadratic form. *Acta Arith.* **21** (1972), 203–234.
9. H. Iwaniec, The half dimensional sieve. *Acta Arith.* **29**(1) (1976), 69–95.
10. H. Iwaniec and E. Kowalski, *Analytic Number Theory* (*American Mathematical Society Colloquium Publications* **53**), American Mathematical Society (Providence, RI, 2004).
11. J. V. Linnik, An asymptotic formula in an additive problem of Hardy–Littlewood. *Izv. Akad. Nauk SSSR Ser. Mat.* **24** (1960), 629–706.
12. K. Matomäki, Prime numbers of the form $p = m^2 + n^2 + 1$ in short intervals. *Acta Arith.* **128**(2) (2007), 193–200.
13. K. Matomäki, The binary Goldbach problem with one prime of the form $p = k^2 + l^2 + 1$. *J. Number Theory* **128**(5) (2008), 1195–1210.
14. K. Matomäki, A Bombieri–Vinogradov type exponential sum result with applications. *J. Number Theory* **129**(9) (2009), 2214–2225.
15. K. Matomäki and X. Shao, Vinogradov's three primes theorem with almost twin primes. *Compos. Math.* **153**(6) (2017), 1220–1256.
16. H. Mikawa, On exponential sums over primes in arithmetic progressions. *Tsukuba J. Math.* **24**(2) (2000), 351–360.
17. H. L. Montgomery, *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis* (*CBMS Regional Conference Series in Mathematics* **84**), American Mathematical Society (Providence, RI, 1994). Published for the Conference Board of the Mathematical Sciences, Washington, DC.
18. O. Ramaré and I. Z. Ruzsa, Additive properties of dense subsets of sifted sequences. *J. Théor. Nombres Bordeaux* **13**(2) (2001), 559–581.
19. S.-Y. Shi, On the distribution of αp modulo one for primes p of a special form. *Osaka J. Math.* **49**(4) (2012), 993–1004.
20. D. I. Tolev, Arithmetic progressions of prime-almost-prime twins. *Acta Arith.* **88**(1) (1999), 67–98.
21. D. I. Tolev, The binary Goldbach problem with arithmetic weights attached to one of the variables. *Acta Arith.* **142**(2) (2010), 169–178.
22. D. I. Tolev, The ternary Goldbach problem with arithmetic weights attached to two of the variables. *J. Number Theory* **130**(2) (2010), 439–457.
23. E. Wirsing, Das asymptotische Verhalten von Summen über multiplikative Funktionen. *Math. Ann.* **143** (1961), 75–102.
24. J. Wu, Primes of the form $p = 1 + m^2 + n^2$ in short intervals. *Proc. Amer. Math. Soc.* **126**(1) (1998), 1–8.

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

Publication III

T. Tao and J. Teräväinen: Odd order cases of the logarithmically averaged Chowla conjecture. To appear in *J. Théor. Nombres Bordeaux*. arXiv: 1710.02112 [Math.NT]

III

ODD ORDER CASES OF THE LOGARITHMICALLY AVERAGED CHOWLA CONJECTURE

TERENCE TAO AND JONI TERÄVÄINEN

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

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

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

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

for any $k \ge 1$, $a_1,\dots,a_k,b_1,\dots,b_k \in \mathbb{N}$ satisfying the non-degeneracy condition $a_ib_j - a_jb_i \neq 0$ for $1 \le i < j \le k$. The non-degeneracy condition may be omitted when k is odd, since a degenerate pair $\lambda(a_in+b_i)\lambda(a_jn+b_j)$ with $a_ib_j - a_jb_i = 0$ is constant in n and can therefore be deleted. One can of course extend this conjecture to the case where the $b_1,\dots,b_k$ 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) 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) 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) and a large regime of uniformity in the parameters $a_1,\dots,a_k,b_1,\dots,b_k$; 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

2010 *Mathematics Subject Classification.* 11N37.

Key words and phrases. Liouville function, Chowla's conjecture, Gowers uniformity norms.

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 [23, 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 $b_1, \dots, b_k$, 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

$$\frac{1}{\log x} \sum_{n \le x} \frac{\lambda(a_1 n + b_1) \cdots \lambda(a_k n + b_k)}{n} = o_{x \to \infty}(1), \tag{2}$$

provided again that $a_i b_j - a_j b_i \neq 0$ for $1 \le i < j \le k$. These logarithmically averaged correlations are certainly easier, since (1) implies (2) by partial summation. For the logarithmically averaged variant (2) of Chowla's conjecture, it was shown by the first author [21] that (2) is $o_{x\to\infty}(1)$ for $k = 2$, and we recently showed in [23] that the same conclusion holds for all odd k. Both of these works actually handle more general correlations of bounded multiplicative functions, with [21] having the same assumptions as in Elliott's conjecture, and [23] having a non-pretentious assumption for the product of the multiplicative functions (see [23, Corollary 1.4] for a precise statement). In addition, it was recently shown by Frantzikinakis and Host [7, Theorem 1.4] that if one replaces the weight $\frac{1}{n}$ in (2) with $\frac{e^{2\pi i \alpha n}}{n}$ for any irrational α, then the analogue of (2) holds for all k. When it comes to conditional results, Frantzikinakis [6] showed that the logarithmically averaged Chowla conjecture would follow from ergodicity of the measure preserving system associated with the Liouville function.

The proof in [23] 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 $k \ge 1$ be an odd natural number, and let $a_1, \dots, a_k, b_1, \dots, b_k$ be natural numbers. Then we have*

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

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 $a_1, \dots, a_k, b_1, \dots, b_k$.*

[1] If $1 \le \omega(x) \le x$ is any function tending to infinity, one could equally well consider (2) with a sum over $\frac{x}{\omega(x)} \le n \le x$, with the $\log x$ normalisation replaced by $\log \omega(x)$. In fact, this is what is done in [21], [23].

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*

$$\frac{1}{\log x}\sum_{n\leq x}\frac{\mu(a_1n+b_1)^{c_1}\cdots\mu(a_kn+b_k)^{c_k}}{n} = o_{x\to\infty}(1) \tag{3}$$

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 [23, Corollary 1.6]*).*[2]

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* [23] *the $k = 3$ case was no easier than the general case.*

It was shown by the first author in [22] that the logarithmically averaged Chowla conjecture (2) for all k is equivalent to two difficult conjectures, namely the logarithmically averaged Sarnak conjecture [22, Conjecture 1.5] and the (logarithmic) local Gowers uniformity of the Liouville function [22, 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 [23, Remark 1.7]). Another indication that the even order cases are more challenging is Elliott's result [3] that for odd k the lim sup of the absolute value of (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 [21] which relied crucially on this theorem.

1.1. **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

[2]The only small modification in the proof of (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.

$X = O(Y)$.

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

$$f_x(a) := \mathbb{E}^{\log}_{n\leq x} \lambda(a_1 n + ab_1) \cdots \lambda(a_k n + ab_k). \tag{4}$$

To prove Theorem 1.1, it will suffice to show that

$$f_x(1) \ll \varepsilon \tag{5}$$

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 (5), 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*

$$\mathbb{E}_{2^m < p \leq 2^{m+1}} |f_x(ap) - (-1)^k f_x(a)| \ll \varepsilon \tag{6}$$

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

$$\sum_{m\in\mathcal{M}} \frac{1}{m} \ll a\varepsilon^{-3}, \tag{7}$$

where the quantity $f_x(a)$ is defined in (4)*.*

Results similar to these appear in [7, Theorem 3.6], [23, Theorem 3.6]. As in these references, we will prove Theorem 3.1 in Section 4 via the entropy decrement argument introduced in [21]; we will use the modification of that argument in [23] to obtain the relatively strong bound (7).
From (6) 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 (4)*.*

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 (5) when k is odd and x is sufficiently large depending on ε.

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

$$\frac{1}{\varepsilon} < w < H_1 < H_2 < H_3 < H_4 < x \tag{8}$$

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

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

for all m in the range $H_1 \leq 2^m \leq H_2$, 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

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

Averaging over all such m and using the prime number theorem, we conclude (given the hypotheses on the parameters (8)) that

$$f_x(1) = -\mathbb{E}^{\log}_{H_1 < p_1 \leq H_2} f_x(p_1) + O(\varepsilon). \tag{9}$$

A similar application of Theorem 3.1 yields

$$f_x(1) = -\mathbb{E}^{\log}_{H_3<p\le H_4} f_x(p) + O(\varepsilon). \tag{10}$$

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

$$f_x(p_1) = -\mathbb{E}^{\log}_{H_3<p_2\le H_4} f_x(p_1p_2) + O(\varepsilon)$$

for all primes p_1 with $H_1 < p_1 \le H_2$; inserting this into (9), we obtain

$$f_x(1) = +\mathbb{E}^{\log}_{H_1<p_1\le H_2}\mathbb{E}^{\log}_{H_3<p_2\le H_4} f_x(p_1p_2) + O(\varepsilon). \tag{11}$$

Crucially, the sign in (11) is the opposite of the sign in (10). To conclude the proof of (5) from (10), (11), it will suffice to show that the average (10) involving primes p and the average (11) involving semiprimes p_1p_2 are comparable in the sense that

$$\mathbb{E}^{\log}_{H_3<p\le H_4} f_x(p) = \mathbb{E}^{\log}_{H_1<p_1\le H_2}\mathbb{E}^{\log}_{H_3<p_2\le H_4} f_x(p_1p_2) + O(\varepsilon). \tag{12}$$

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

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

Similarly, we have

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

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

$$\mathbb{E}^{\log}_{H_3<n\le H_4:(n,W)=1} f_x(n) = \mathbb{E}^{\log}_{H_1<p_1\le H_2}\mathbb{E}^{\log}_{H_3<n_2\le H_4:(n_2,W)=1} f_x(p_1n_2) + O(\varepsilon). \tag{13}$$

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

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

and one can modify the range $p_1H_3 < n \le p_1H_4$ to $H_3 < n \le H_4$ incurring a further error of $O(\varepsilon)$. We may thus rearrange (13) as

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

where g is the weight

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

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

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

which will follow in turn from the bounds

$$\mathbb{E}^{\log}_{H_3<n\le H_4:(n,W)=1} g(n) = 1 + O(\varepsilon^2) \tag{14}$$

and

$$\mathbb{E}^{\log}_{H_3<n\le H_4:(n,W)=1} g(n)^2 = 1 + O(\varepsilon^2). \tag{15}$$

The left-hand side of (14) 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 (14) 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 (15) can be rewritten as

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

and the claim (15) follows since $\mathbb{E}^{\log}_{H_3<n\leq H_4:(n,W)=1} 1_{p_1,p_1'|n}$ is equal to $\frac{1+O(\varepsilon^2)}{p_1 p_1'}$ 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

$$x \geq \exp\exp\exp(a\varepsilon^{-3}) \tag{16}$$

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 (6) for m satisfying

$$\exp(a\varepsilon^{-3}) \leq m \leq \frac{1}{100}\log\log x \tag{17}$$

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 (7). Finally, we can assume that $\varepsilon \leq 1/2$, since for $1/2 < \varepsilon \leq 1$ the bound (6) holds from the triangle inequality.

For any prime p, one has the identity

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

for any natural number n, and hence

$$\lambda(a_1 n + ab_1)\cdots\lambda(a_k n + ab_k) = (-1)^k \lambda(a_1 pn + apb_1)\cdots\lambda(a_k n + apb_k).$$

From (4) we thus have

$$f_x(a) = (-1)^k \mathbb{E}^{\log}_{n\leq x} \lambda(a_1 pn + apb_1)\cdots\lambda(a_k pn + apb_k).$$

If $p \leq \log x$, then (using (16)) 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

$$f_x(a) = (-1)^k \mathbb{E}^{\log}_{n\leq px} \lambda(a_1 pn + apb_1)\cdots\lambda(a_k pn + apb_k) + O(\varepsilon)$$

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

$$f_x(a) = (-1)^k \mathbb{E}^{\log}_{n'\leq x} \lambda(a_1 n' + apb_1)\cdots\lambda(a_k n' + apb_k) p 1_{p|n'} + O(\varepsilon).$$

[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 (4) with a replaced by ap, we conclude that

$$f_x(a) - (-1)^k f_x(ap) = (-1)^k \mathbb{E}^{\log}_{n\le x} \lambda(a_1 n + apb_1)\cdots\lambda(a_k n + apb_k)(p1_{p|n} - 1) + O(\varepsilon).$$

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

$$f_x(a) - (-1)^k f_x(ap) = (-1)^k \mathbb{E}^{\log}_{x^\varepsilon<n\le x} \lambda(a_1 n + apb_1)\cdots\lambda(a_k n + apb_k)(p1_{p|n} - 1) + O(\varepsilon)$$

for all $p \le \log x$. If we set $c_p \in \{-1, 0, +1\}$ to be the signum of $\mathbb{E}^{\log}_{n\le x} \lambda(a_1 n + apb_1)\cdots\lambda(a_k n + apb_k)(p1_{p|n} - 1)$, it will thus suffice to show that

$$\mathbb{E}_{2^m<p\le 2^{m+1}} c_p \mathbb{E}^{\log}_{x^\varepsilon<n\le x} \lambda(a_1 n + apb_1)\cdots\lambda(a_k n + apb_k)(p1_{p|n} - 1) = O(\varepsilon) \tag{18}$$

for all m obeying (17), outside of an exceptional set $\mathcal{M}$ obeying (7).

Let m obey (17). 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 (16), the error term is certainly of size $O(\varepsilon)$. In particular, the left-hand side of (18) can be written as

$$\mathbb{E}_{2^m\le p\le 2^{m+1}} c_p \mathbb{E}^{\log}_{x^\varepsilon<n\le x} \lambda(a_1 n + a_1 j + apb_1)\cdots\lambda(a_k n + a_k j + apb_k)(p1_{p|n+j} - 1) + O(\varepsilon)$$

for any $1 \le j \le 2^m$. Averaging in j and rearranging, we can thus write the left-hand side of (18) in probabilistic language[4] as

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

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

$$\mathbf{Z}_m := \mathbb{E}_{2^m<p\le 2^{m+1}} \mathbb{E}_{j\le 2^m} c_p \lambda(a_1 \mathbf{n} + a_1 j + apb_1)\cdots\lambda(a_k \mathbf{n} + a_k j + apb_k)(p1_{p|\mathbf{n}+j} - 1),$$

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

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

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

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 $\mathbf{Y}_m \in \prod_{2^m<p\le 2^{m+1}} \mathbb{Z}/p\mathbb{Z}$ be the random variables

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

and

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

[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.

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

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

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 (17), outside of an exceptional set $\mathcal{M}$ obeying (7).
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 (17), 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* (17) *and is such that*

$$\mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m}) \leq \varepsilon^3 \frac{2^m}{m}. \tag{19}$$

Then one has

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

Proof. We argue as in [23], which are in turn a modification of the arguments in [21]. 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

$$\mathbf{P}(|F_m(X_m, \mathbf{U}_m)| \geq \varepsilon) \ll \exp(-c\varepsilon^2 2^m/m) \tag{20}$$

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 (20) now follows from Hoeffding's inequality [14].

Applying the Pinsker-type inequality from [23, Lemma 3.4] (see also [21, 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 (19) we have

$$\mathbf{E}|F_m(\mathbf{X}_m, \mathbf{Y}_m)| \ll \frac{m}{\varepsilon^2 2^m}(\mathbf{H}(\mathbf{U}_m) - \mathbf{H}(\mathbf{Y}_m|\mathbf{Y}_{<m})) + \varepsilon. \tag{21}$$

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

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

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

$$\mathbf{P}(\mathbf{Y}_m = Y_m|\mathbf{Y}_{<m} = Y_{<m}) = \frac{1}{\prod_{2^m<p\leq 2^{m+1}} p} + O\left(\frac{\exp(O(2^m))}{x^\varepsilon}\right). \tag{22}$$

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

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

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

$$|\mathbf{H}(\mathbf{X}) - \mathbf{H}(\mathbf{X}')| \leq 2C \cdot \max_{x\in\mathcal{X}} |\mathbf{P}(\mathbf{X} = x) - \mathbf{P}(\mathbf{X}' = x)|^{\frac{1}{2}} \cdot |\mathcal{X}|. \tag{23}$$

From this and (22) we compute that

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

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

Inserting this into (21) and using (16), (17) 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*

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

Proof. For any m obeying (17), 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 (23) 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 (23) 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

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

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. □

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*

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

where C is the complex conjugation operator, $|\omega|$ is the number of ones in $\omega \in \{0,1\}^k$, $\mathbf{h} = (h_1,\dots,h_k)$, 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 $f : \{1,\dots,N\} \to \mathbb{C}$ 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 [20, 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 $\theta, \phi_1, \dots, \phi_k : \mathbb{Z} \to \mathbb{C}$ 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 $a_1,\dots,a_k,b_1,\dots,b_k$, but independent of W and $r_1,\dots,r_k$.

Without the W-aspect, this is standard; see for instance [4, Lemma 2]. However, the uniformity of the bounds in W (and $r_1,\dots,r_k$) will be crucial in our arguments.

Proof. We shall adapt the proof of [22, Proposition 3.3]. By splitting the variable n into residue classes $\pmod W$ and setting $N' := \frac{N}{W}$ it suffices to show that

$$\begin{aligned}&\left|\mathbb{E}_{d\leq N'}\mathbb{E}_{n\leq N'}\theta(d)\phi_1(W(a_1n+b_1d)+r'_1)\cdots\phi_k(W(a_kn+b_kd)+r'_k)\right|\\ &\leq C\|\theta\|_{U^k[N']} + o_{N'\to\infty}(1)\end{aligned}$$

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 $|\phi_j| \leq 1$ that

$$|\mathbb{E}_{d\leq N}\mathbb{E}_{n\leq N}\theta(d)\phi_1(a_1n+b_1d)\cdots\phi_k(a_kn+b_kd)| \leq C\|\theta\|_{U^k[N]} + o_{N\to\infty}(1). \tag{24}$$

Since the statement of (24) involves the values of the functions θ and ϕ_i only on $(-HN, HN)$, where $H = \max_{i\leq k}(|a_i|+|b_i|)+1$, 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

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

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 $d = d_1 + \cdots + d_k$, $n = n' - d_1b_1 - \cdots - d_kb_k$, we are left with showing that

$$\left|\mathbb{E}_{d_1,\dots,d_k\in\mathbb{Z}_{2HN}}\theta'(d_1+\cdots+d_k)\prod_{i=1}^k\phi_i\left(a_in' + \sum_{\ell=1}^k d_\ell(b_i-b_\ell)\right)\right| \le C''\|\theta\|_{U^k[N]} + o_{N\to\infty}(1), \tag{25}$$

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

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

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 (25), where each term involving ϕ_i is independent of the variable d_i, we see that

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

Then, by noting that

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

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 $W = \prod_{p\le w} p$, 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*

$$\Lambda_{b,W}(n) := \frac{\varphi(W)}{W}\Lambda(Wn+b) \tag{26}$$

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 $2^m < p \le 2^{m+1}$, and Λ is non-zero for only $O(2^{2m/3})$ (say) other integers in the interval $(2^m, 2^{m+1}]$, we have

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

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

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

It therefore suffices to show that

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

Partitioning into residue classes modulo W and using (26), 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 (4), and replacing the average $n \leq x$ with the average $x^\varepsilon < n \leq x$, we have

$$f_x(a(Wd+b)) = \mathbb{E}^{\log}_{x^\varepsilon<n\leq x}\lambda(a_1n+Wab_1d+abb_1)\dots\lambda(a_kn+Wab_kd+abb_k)+O(\varepsilon),$$

so it suffices to show that

$$\mathbb{E}_{d\leq H}\mathbb{E}^{\log}_{x^\varepsilon<n\leq x}(\Lambda_{b,W}(d)-1)\lambda(a_1n+Wab_1d+abb_1)\dots\lambda(a_kn+Wab_kd+abb_k) \ll \varepsilon. \tag{27}$$

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 (27) may be written as

$$\mathbb{E}^{\log}_{x^\varepsilon<n\leq x}\mathbb{E}_{d\leq H}\mathbb{E}_{n'\leq aHW}(\Lambda_{b,W}(d)-1)[\lambda(a_1n'+Wab_1d+a_1n+abb_1)\dots$$
$$\lambda(a_kn'+Wab_kd+a_1n+abb_k)]+O(\varepsilon).$$

Applying Lemma 5.2 with N replaced by aHW, W replaced by aW, n replaced by n', and the r_j replaced by a_jn+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*, volume 28 of Dev. Math., pp. 67–83. Springer, New York, 2013.

[2] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*. Mathematics and Its Applications, Vol. 4. Gordon and Breach Science Publishers, New York-London-Paris, 1965.

[3] P. D. T. A. Elliott, *On the correlation of multiplicative functions*, Notas Soc. Mat. Chile **11** (1992), 1–11.

[4] N. Frantzikinakis, B. Host, B. Kra, *Multiple recurrence and convergence for sequences related to the prime numbers*, J. Reine Angew. Math. **611** (2007), 131–144.

[5] N. Frantzikinakis, *An averaged Chowla and Elliott conjecture along independent polynomials*, To appear in Int. Math. Res. Not. IMRN.

[6] N. Frantzikinakis, *Ergodicity of the Liouville system implies the Chowla conjecture*, Discrete Anal. **19** (2017), 41pp.

[7] N. Frantzikinakis, B. Host, *The logarithmic Sarnak conjecture for ergodic weights*, To appear in Ann. of Math.

[8] J. Friedlander, H. Iwaniec, *Opera de cribro*, Vol. 57 of *American Mathematical Society Colloquium Publications*, American Mathematical Society, Providence, RI, 2010.

[9] B. Green, T. Tao, *An inverse theorem for the Gowers U3-norm, with applications*, Proc. Edinburgh Math. Soc. **51** (2008), no. 1, 73–153.

[10] B. Green, T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), 1753–1850.

[11] B. Green, T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566.

[12] B. Green, T. Tao, T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), no. 2, 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, 1–70.
[14] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.
[15] I. Kátai, *A remark on a theorem of H. Daboussi*, Acta Math. Hungar. **47** (1986), 223–225.
[16] A. Le, *Nilsequences and multiple correlations along subsequences*, preprint. arXiv:1708.01361
[17] A. Leibman, *Nilsequences, null-sequences, and multiple correlation sequences*, Ergodic Theory and Dynamical Systems **35** (2015), no. 1, 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. of Math. (2) **183** (2016), no. 3, 1015–1056.
[19] K. Matomäki, M. Radziwiłł, T. Tao, *An averaged form of Chowla's conjecture*, Algebra & Number Theory **9** (2015), 2167–2196.
[20] T. Tao, V. Vu, *Additive combinatorics*, In *Cambridge Studies in Advanced Mathematics*, Vol. 105, Cambridge University Press, Cambridge, 2006.
[21] T. Tao, *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations*, Forum Math. Pi **4** (2016), e8, 36 pp.
[22] T. Tao, *Equivalence of the logarithmically averaged Chowla and Sarnak conjectures*, In *Number theory—Diophantine problems, uniform distribution and applications*, pp. 391–421. Springer, Cham, 2017.
[23] T. Tao, J. Teräväinen, *The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures*, preprint. arXiv:1708.02610

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

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

Publication IV

J. Teräväinen: On binary correlations of multiplicative functions. *Forum Math. Sigma*, 6:e10, 41, 2018. DOI: 10.1017/fms.2018.10

Forum of Mathematics, Sigma (2018), 41 pages
doi:10.1017/fms.2018.10

ON BINARY CORRELATIONS OF MULTIPLICATIVE FUNCTIONS

JONI TERÄVÄINEN
Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland;
email: joni.p.teravainen@utu.fi

Received 11 October 2017; accepted 13 May 2018

Abstract

We study logarithmically averaged binary correlations of bounded multiplicative functions g_1 and g_2. A breakthrough on these correlations was made by Tao, who showed that the correlation average is negligibly small whenever g_1 or g_2 does not pretend to be any twisted Dirichlet character, in the sense of the pretentious distance for multiplicative functions. We consider a wider class of real-valued multiplicative functions g_j, namely those that are uniformly distributed in arithmetic progressions to fixed moduli. Under this assumption, we obtain a discorrelation estimate, showing that the correlation of g_1 and g_2 is asymptotic to the product of their mean values. We derive several applications, first showing that the numbers of large prime factors of n and $n+1$ are independent of each other with respect to logarithmic density. Secondly, we prove a logarithmic version of the conjecture of Erdős and Pomerance on two consecutive smooth numbers. Thirdly, we show that if Q is cube-free and belongs to the Burgess regime $Q \leqslant x^{4-\varepsilon}$, the logarithmic average around x of the real character $\chi \pmod{Q}$ over the values of a reducible quadratic polynomial is small.

2010 Mathematics Subject Classification: 11N37 (primary); 11N60, 11L40 (secondary)

1. Introduction

Let $\mathbb{D} = \{z \in \mathbb{C} : |z| \leqslant 1\}$ be the unit disc of the complex plane, and let g_1, $g_2 : \mathbb{N} \to \mathbb{D}$ be multiplicative functions. We consider the logarithmically averaged binary correlations

$$\frac{1}{\log x} \sum_{n \leqslant x} \frac{g_1(n) g_2(n+h)}{n}, \tag{1.1}$$

with $h \neq 0$ a fixed integer and x tending to infinity. If $h < 0$ in (1.1), we can extend g_1 and g_2 arbitrarily to the negative integers, since this affects (1.1) only by $o(1)$.

In a recent breakthrough work, Tao [30] showed that the correlation (1.1) is $o(1)$ as $x \to \infty$, provided that at least one of the two functions g_j does not *pretend* to be a twisted Dirichlet character, in the sense that

$$\liminf_{X\to\infty} \inf_{|t|\leqslant X} \mathbb{D}(g_j, \chi(n)n^{it}; X) = \infty, \tag{1.2}$$

for all fixed Dirichlet characters χ, with the pretentious distance $\mathbb{D}(\cdot)$ measured by

$$\mathbb{D}(f, g; X) := \left(\sum_{p\leqslant X} \frac{1 - \mathrm{Re}(f(p)\overline{g(p)})}{p}\right)^{1/2}. \tag{1.3}$$

The main theorem in [30] that (1.1) is $o(1)$ under the nonpretentiousness assumption (1.2) is a logarithmically averaged version of the binary case of a conjecture of Elliott. Elliott's original conjecture [7, 8] (in the slightly corrected form presented in [23]) states that for any integer $k \geqslant 1$, any multiplicative functions $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$ and any distinct integer shifts $h_1, \dots, h_k$ we have the discorrelation estimate

$$\frac{1}{x}\sum_{n\leqslant x} g_1(n + h_1) \cdots g_k(n + h_k) = o(1) \tag{1.4}$$

as $x \to \infty$, provided that at least one of the g_j satisfies the nonpretentiousness assumption (1.2). (In the case where the functions g_j are allowed to depend on x, one needs a slightly stronger pretentiousness hypothesis; see [29].) The $k = 1$ case of Elliott's conjecture is known as Halász's theorem [15]. Already for $k = 2$, there is not much progress towards the nonlogarithmic version of Elliott's conjecture (see though [7]). However, if one averages (1.4) over the shifts $h_1, \dots, h_k \in [1, H]$, with $H = H(x)$ tending to infinity with any speed, then Elliott's conjecture holds on average by the work of Matomäki, Radziwiłł and Tao [23]. This was generalized by Frantzikinakis [11] to averages along independent polynomials. In the case of logarithmically averaged correlations, there has been a lot of recent progress, initiated by [30]; see [12, 31, 33].

We study in this paper the same logarithmically averaged correlation (1.1) as Tao studied in [30], but for a wider class of real-valued multiplicative functions (in [30] one works also with complex-valued functions). The multiplicative functions $g_j : \mathbb{N} \to [-1, 1]$ that we consider are uniformly distributed in residue classes to fixed moduli. Many of the most interesting bounded multiplicative

functions have such a uniform distribution property; in particular, the Liouville function λ and the indicator function of x^a-smooth numbers up to x have that property. Also the real primitive Dirichlet character $\chi_Q \pmod Q$ will be seen to be uniformly distributed in arithmetic progressions on $[x, 2x]$, provided that the modulus Q grows neither too slowly nor too rapidly in terms of x. Indeed, many of the applications of our main theorem concern consecutive smooth (or friable) numbers or quadratic residues.

The uniformity assumption we require of multiplicative functions is as follows.

DEFINITION 1.1 (Uniformity assumption). Let $x \geqslant 1$, $1 \leqslant Q \leqslant x$ and $\eta > 0$. For a function $g : \mathbb{N} \to \mathbb{D}$, we write $g \in \mathcal{U}(x, Q, \eta)$ if we have the estimate

$$\left| \frac{1}{x} \sum_{\substack{x \leqslant n \leqslant 2x \\ n \equiv a \pmod q}} g(n) - \frac{1}{qx} \sum_{x \leqslant n \leqslant 2x} g(n) \right| \leqslant \frac{\eta}{q} \quad \text{for all } 1 \leqslant a \leqslant q \leqslant Q.$$

REMARK 1.2. Note that in this definition we do not send x to infinity (but naturally we want x to be large). The fact that Definition 1.1 is not an asymptotic relation is important, since later we shall to apply it to $g(n) = 1_{n \leqslant x, n \text{ is } x^a\text{-smooth}}$, which is a function dependent on x.

REMARK 1.3. Let $g : \mathbb{N} \to [-1, 1]$ be a nonpretentious multiplicative function, in the sense that for some small $\varepsilon > 0$ and some large x we have

$$\inf_{|t| \leqslant x} \mathbb{D}(g, \chi(n)n^{it}; x) \geqslant \varepsilon^{-10}$$

for all Dirichlet characters χ of modulus $\leqslant \varepsilon^{-10}$. By expressing the condition $n \equiv a \pmod q$ in Definition 1.1 in terms of Dirichlet characters (after reducing to a coprime residue class), and applying Halász's theorem, one sees that $g \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$. Therefore, the collection of uniformly distributed real-valued multiplicative functions $g : \mathbb{N} \to [-1, 1]$ contains all nonpretentious real functions.

We use the notation $o_{\varepsilon \to 0}(1)$ to denote a quantity depending on ε and tending to 0 as $\varepsilon \to 0$, uniformly with respect to all other parameters. With this notation, our main theorem asserts the following.

THEOREM 1.4. *Let a small real number $\varepsilon > 0$, a fixed integer $h \neq 0$, and a function $\omega : \mathbb{R}_{\geqslant 1} \to \mathbb{R}$ with $1 \leqslant \omega(X) \leqslant \log(3X)$ and $\omega(X) \xrightarrow{X \to \infty} \infty$ be given. Let $x \geqslant x_0(\varepsilon, h, \omega)$. Then, for any multiplicative functions $g_1, g_2 : \mathbb{N} \to [-1, 1]$*

such that $g_1 \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$, *we have*

$$\frac{1}{\log \omega(x)} \sum_{x/\omega(x) \leqslant n \leqslant x} \frac{g_1(n)g_2(n+h)}{n} = \left(\frac{1}{x} \sum_{x \leqslant n \leqslant 2x} g_1(n)\right)\left(\frac{1}{x} \sum_{x \leqslant n \leqslant 2x} g_2(n)\right) + o_{\varepsilon \to 0}(1).$$

REMARK 1.5. Theorem 1.4 can be viewed as stating that the functions g_1 and g_2 do not correlate with the shifts of each other. Note that in the case where the mean values of g_1 and g_2 on $[x, 2x]$ are not $o(1)$, Theorem 1.4 is not covered by the logarithmically averaged Elliott conjecture from [**30**].

REMARK 1.6. In the case where g_1 and g_2 are complex-valued, one does not always have the conclusion of Theorem 1.4. Namely, take $g_1(n) = n^{it}$ and $g_2(n) = n^{iu}$ for some $t, u \neq 0$ with $t + u \neq 0$. One easily sees that g_1 and g_2 are uniformly distributed in arithmetic progressions, and by partial summation the shifted product $g_1(n)g_2(n+1) = n^{i(t+u)} + o(1)$ has logarithmic mean value $o(1)$ on $[x/\omega(x), x]$. However, by the simple estimate $(1/x)\sum_{n \leqslant x} n^{it} = (x^{it}/(1+it)) + o(1)$, the product of the mean values of g_1 and g_2 on $[x, 2x]$ is an oscillating function.

REMARK 1.7. Although the statement of Theorem 1.4 does not hold for all complex-valued multiplicative functions, one could show that it continues to hold if $g_1, g_2 : \mathbb{N} \to \mathbb{D}$ take values in the roots of unity of fixed order. Indeed, the only places in the proof of the main theorem where real-valuedness plays a role are Lemmas 2.2, 2.5 and 3.4. The first two lemmas could be proved also for functions g_j taking values in the roots of unity of bounded order by applying a standard generalization of [**23**, Lemma C.1] to such functions. For Lemma 3.4, one would also apply this generalization of [**23**, Lemma C.1] together with an extension of [**22**, Theorem 3] to multiplicative functions taking a bounded number of complex values. For this last extension, one notes that the only place in the proof of [**22**, Proposition 1] where real-valuedness is used is [**22**, Lemma 3], and this lemma can also be made to work for functions taking values in the roots of unity of fixed order. We leave the details to the interested reader.

REMARK 1.8. The bound $\omega(X) \leqslant \log(3X)$ in Theorem 1.4 is not restrictive in reality, since if one wants an asymptotic formula for the logarithmic correlation over the interval $[1, x]$, say, one can sum together the asymptotics for the correlations over $[y/\log(3y), y]$ for various $y \leqslant x$. It is nevertheless necessary for technical reasons to have an upper bound on $\omega(X)$ in the main theorem, since otherwise the asymptotic would not be valid for example for the correlations of the indicator function of x^a-smooth numbers.

REMARK 1.9. One could prove the same correlation bound for the more general logarithmic averages of $g_1(a_1n+h_1)g_2(a_2n+h_2)$ with $(a_1, h_1) = (a_2, h_2) = 1$ and $a_1, a_2 \geqslant 1$ and h_1, h_2 fixed integers. This is due to the fact that the main theorem in [30] deals with such correlations. To avoid complicating the notations, however, we deal with the case $a_1 = a_2 = 1$ here.

One might wonder at first why in the asymptotic formula in Theorem 1.4 one side of the formula involves the values of the functions g_j on $[x/\omega(x), x]$, whereas the other side only involves the values on $[x, 2x]$. However, by a result we present in Appendix A, essentially due to Granville and Soundararajan [13], the mean value of a real-valued multiplicative function is almost the same over the intervals $[x/\omega(x), x]$ and $[x, 2x]$, explaining the phenomenon.

Owing to Remark 1.3, the main theorem contains as a special case the logarithmically averaged binary Elliott conjecture from [30]. This is not surprising, since we use the same proof method. Of course, our interest lies in those cases where the functions g_1 and g_2 are pretentious (in the sense that (1.2) fails) but still satisfy our uniformity assumption.

It was recently shown by Klurman [20] that one can obtain an asymptotic formula for the k-point correlations

$$\frac{1}{x}\sum_{n\leqslant x} f_1(n+h_1)\cdots f_k(n+h_k)$$

for any integers $h_1, \dots h_k$, when $f_1, \dots, f_k : \mathbb{N} \to \mathbb{D}$ are pretentious multiplicative functions, in the sense that $\mathbb{D}(f_j, \chi_j(n)n^{it_j}; x) \ll 1$ for some characters χ_j. This result does not imply Theorem 1.4, however, since our theorem is in a nonasymptotic form, allowing the multiplicative functions g_1 and g_2 to strongly depend on x. Indeed, allowing the multiplicative functions g_j to depend on x is crucial for applications to smooth numbers and to Burgess-type bounds. The asymptotic formula in [20] is a sieve-theoretic product of local mean values, but one cannot express the density of smooth numbers as such a product.

1.1. Applications of the main theorem. We have a number of corollaries to Theorem 1.4. To state them, we recall the notion of logarithmic density of a set of integers.

DEFINITION 1.10. The *logarithmic density* of a set $A \subset \mathbb{N}$ is

$$\delta(A) = \lim_{x\to\infty} \frac{1}{\log x}\sum_{\substack{n\leqslant x\\ n\in A}} \frac{1}{n},$$

whenever it exists.

We prove using Theorem 1.4 the following theorem about the largest prime factors of consecutive integers.

THEOREM 1.11 (Independence of the number of large prime factors of n and $n+1$). *Let $\omega_{>y}(n) := |\{p > y : p \mid n\}|$ be the number of prime factors of n that are larger than y. Then, for any real numbers $a, b \in (0,1)$ and any integers $0 \leqslant k < 1/a$, $0 \leqslant \ell < 1/b$, we have*

$$\begin{aligned}&\delta(\{n \in \mathbb{N} : \omega_{>n^a}(n) = k, \omega_{>n^b}(n+1) = \ell\})\\ &\quad = \delta(\{n \in \mathbb{N} : \omega_{>n^a}(n) = k\}) \cdot \delta(\{n \in \mathbb{N} : \omega_{>n^b}(n) = \ell\}).\end{aligned}$$

Moreover, under the same assumptions, the set $\{n \in \mathbb{N} : \omega_{>n^a}(n) = k, \omega_{>n^b}(n+1) = \ell\}$ has positive asymptotic lower density.

REMARK 1.12. From the proof of Theorem 1.11 in Section 4, we can easily deduce a discorrelation estimate for the 'truncated Liouville function' $\lambda_{>y}(n)$, which is a multiplicative function taking the value $+1$ at the primes $p \leqslant y$ and -1 at the primes $p > y$. This estimate takes the form

$$\frac{1}{\log x}\sum_{n\leqslant x}\frac{\lambda_{>x^\varepsilon}(n)\lambda_{>x^\varepsilon}(n+1)}{n} = o_{\varepsilon\to 0}(1), \tag{1.5}$$

for $\varepsilon \in (0,1)$ and $x \geqslant x_0(\varepsilon)$. This result may be compared with that of Daboussi and Sárkőzy [3] and Mangerel [21], which states that if we define $\lambda_{<y}(n)$ as the completely multiplicative function taking the value -1 at the primes $p < y$ and $+1$ at the primes $p \geqslant y$ (so that $\lambda_{<y}(p)$ has the opposite sign as $\lambda_{>y}(p)$), then

$$\frac{1}{x}\sum_{n\leqslant x}\lambda_{<x^\varepsilon}(n)\lambda_{<x^\varepsilon}(n+1) = o_{\varepsilon\to 0}(1); \tag{1.6}$$

moreover, they proved this in a quantitative form. The proof of (1.6) is based on sieve theory and is very different from the proof of (1.5).

Our next applications concern smooth numbers, so we introduce the function $P^+(n)$, whose value is the largest prime factor of the positive integer $n \geqslant 2$ (and $P^+(1) = 1$). We say that a number n is *y-smooth* if $P^+(n) \leqslant y$. The simultaneous distribution of the function $P^+(\cdot)$ at consecutive integers is the subject of several conjectures. There is for instance a conjecture of Erdős and Pomerance [10], asserting that the largest prime factors of n and $n+1$ are independent events.

CONJECTURE 1.13 (Erdős–Pomerance). For any $a, b \in (0, 1)$, the asymptotic density of the set

$$\{n \in \mathbb{N} : P^+(n) \leqslant n^a,\ P^+(n+1) \leqslant n^b\} \tag{1.7}$$

exists and equals $\rho(1/a)\rho(1/b)$, where $\rho(\cdot)$ is the *Dickmann function* (see [18, Section 1]).

What we are able to prove, taking $k = \ell = 0$ in Theorem 1.11, is a logarithmic version of the conjecture.

THEOREM 1.14. *Conjecture 1.13 holds when asymptotic density is replaced with logarithmic density; that is, for any $a, b \in (0, 1)$ we have*

$$\delta(\{n \in \mathbb{N} : P^+(n) \leqslant n^a,\ P^+(n+1) \leqslant n^b\}) = \rho\left(\frac{1}{a}\right)\rho\left(\frac{1}{b}\right).$$

A closely related conjecture, formulated in the correspondence of Erdős and Turán in the 1930s (see [28, pp. 100–101], [9], [26, Section 1]) is that the distribution of $(P^+(n), P^+(n+1))$ is symmetric.

CONJECTURE 1.15 (Erdős–Turán). The asymptotic density of the set

$$\{n \in \mathbb{N} : P^+(n) < P^+(n+1)\} \tag{1.8}$$

exists and equals $\frac{1}{2}$.

There has been some progress towards this conjecture. Erdős and Pomerance [10] showed that the lower asymptotic density of the set in (1.8) is positive (in fact, at least 0.0099). The lower bound for the density was improved to 0.05544 by de la Bretèche, Pomerance and Tenenbaum [5], to 0.1063 by Wang [35], and a further improvement to 0.1356 was given by Wang in [36].

We can prove Conjecture 1.15 if asymptotic density is again replaced with logarithmic density.

THEOREM 1.16. *Conjecture 1.15 holds when asymptotic density is replaced with logarithmic density; that is,*

$$\delta(\{n \in \mathbb{N} : P^+(n) < P^+(n+1)\}) = \tfrac{1}{2}.$$

In fact, Theorem 1.14 implies Theorem 1.16, via the following theorem, which was also conjectured by Erdős [9] in the case of asymptotic density. (Erdős conjectured the existence of the density of integers n for which $P^+(n+1) > P^+(n)\cdot n^\alpha$.)

THEOREM 1.17. *Let $\alpha \in [0,1]$ be a real number. Let $u(x) := \rho((1/x)-1)/x$ for $x \in (0,1)$, where ρ is the Dickmann function. Then we have*

$$\delta(\{n \in \mathbb{N} : P^+(n+1) > P^+(n)\cdot n^\alpha\}) = \int_{T_\alpha} u(x)u(y)\,dx\,dy, \tag{1.9}$$

where T_α is the triangular domain $\{(x,y) \in [0,1]^2 : y \geqslant x+\alpha\}$. In particular, the logarithmic density above exists.

REMARK 1.18. The appearance of the function $u(\cdot)$ is to be expected in Theorem 1.17, since u is the derivative of $x \mapsto \rho(1/x)$, with the latter function expressing the probability that $P^+(n) \leqslant n^x$.

We prove Theorem 1.11, and consequently Theorem 1.14, in Section 4, where we also see that Theorem 1.17 quickly follows from the latter theorem. With Theorem 1.17 available, Theorem 1.16 follows by taking $\alpha = 0$ and noting that then the integral in (1.9) is symmetric in x and y, implying that its value is $\frac{1}{2}$. For the details, see Section 4.

We can also prove another approximation to Conjecture 1.13. This was obtained earlier by Hildebrand [17], using a combinatorial method, in the special case $(a,b) = (c,d)$ (Hildebrand's proof also applies to so-called stable sets, with power-smooth numbers being an example of such a set). The following theorem also implies a result of Wang [36, Théorème 2] on the integers $n \leqslant x$ with $P_y^+(n) < P_y^+(n+1)$ having a positive density, where $P_y^+(n) = \max\{p \leqslant y : p \mid n\}$ and $y \geqslant x^\varepsilon$.

THEOREM 1.19. *Let $a,b,c,d \in (0,1)$ be real numbers with $a < b$ and $c < d$. Then the set*

$$\{n \in \mathbb{N} : n^a \leqslant P^+(n) \leqslant n^b, n^c \leqslant P^+(n+1) \leqslant n^d\}$$

has positive asymptotic lower density.

Note that Theorem 1.19 is not implied by Theorem 1.14, as there are sets of positive logarithmic density having zero asymptotic lower density. Nevertheless, the proof we use for the latter theorem also works for the former, owing to the presence of an arbitrarily slowly growing function $\omega(X)$ in Theorem 1.4.

Since we can prove satisfactory results for the distribution of the largest prime factor function $P^+(\cdot)$ at two consecutive integers, it is natural to ask about the distribution of $P^+(\cdot)$ also at longer strings of consecutive integers. A conjecture of De Koninck and Doyon [4] states the following.

CONJECTURE 1.20 (De Koninck and Doyon). Let $k \geqslant 2$ be an integer and $(a_1, \dots a_k)$ any permutation of the set $\{1, 2, \dots, k\}$. Then the set

$$\{n \in \mathbb{N} : P^+(n + a_1) < \cdots < P^+(n + a_k)\} \tag{1.10}$$

has an asymptotic density, and it equals $1/k!$.

The case $k = 2$ of this is the earlier mentioned Conjecture 1.15 of Erdős and Turán. Little is known about this conjecture for $k \geqslant 3$; it is not even known that the sets in (1.10) have positive asymptotic lower density. Recently, Wang [36] proved a result about orderings of $P^+(\cdot)$ at consecutive integers, showing that

$$P^+(n + i) < \min_{\substack{j \leqslant J \\ j \neq i}} P^+(n + j) \quad \text{and} \quad P^+(n + i) > \max_{\substack{j \leqslant J \\ j \neq i}} P^+(n + j) \tag{1.11}$$

hold with positive asymptotic lower density for any $J \geqslant 3$ and $1 \leqslant i \leqslant J$. The method of [36] is based on the linear sieve and Bombieri–Vinogradov type estimates for smooth numbers. Applying Theorem 1.4 together with the Matomäki–Radziwiłł theorem [22] on multiplicative functions in short intervals (and using the method of [24]), we can give a different proof of the $J = 3$ case of Wang's result. We leave the details of this special case of (1.11) to the interested reader.

As our last application, we study character sums along the values of a reducible quadratic polynomial $n(n + h)$. A famous result of Burgess [1] states that for any nonprincipal Dirichlet character χ modulo Q we have

$$\sum_{y \leqslant n \leqslant y+x} \chi(n) \ll_{r,\varepsilon} x^{1-1/r} Q^{(r+1)/4r^2+\varepsilon},$$

whenever $r \in \mathbb{N}$ and Q is cube-free (that is, $p^3 \nmid Q$ for all primes p). In particular, we have the important special case

$$\sum_{n \leqslant x} \chi_Q(n) = o(x), \quad 3 \leqslant Q \leqslant x^{4-\varepsilon} \tag{1.12}$$

for cube-free values of Q, where χ_Q is a real primitive Dirichlet character modulo Q. Using Theorem 1.4, we can prove that a variant of the estimate (1.12) continues to hold for character sums over the values of a reducible quadratic polynomial.

THEOREM 1.21 (Character sums over $n(n+h)$ in the Burgess regime). *Let $\varepsilon > 0$ be small, $h \neq 0$ a fixed integer, and $1 \leqslant \omega(X) \leqslant \log(3X)$ any function tending to infinity. For $x \geqslant x_0(\varepsilon, h, \omega)$, let $Q = Q(x) \leqslant x^{4-\varepsilon}$ be a cube-free natural number with $Q(x) \xrightarrow{x\to\infty} \infty$. Then, the real primitive Dirichlet character χ_Q modulo Q satisfies the estimate*

$$\frac{1}{\log \omega(x)} \sum_{x/\omega(x)\leqslant n\leqslant x} \frac{\chi_Q(n(n+h))}{n} = o(1).$$

Moreover, if Q is as before and QNR stands for quadratic nonresidue (that is, an integer n with $\chi_Q(n) = -1$), we have

$$\frac{1}{\log x} \sum_{\substack{n\leqslant x \\ n,n+1 \text{ QNR} \pmod Q}} \frac{1}{n} = \frac{1}{4} \prod_{p\mid Q} \left(1 - \frac{2}{p}\right) + o(1) \tag{1.13}$$

and

$$\frac{1}{x} |\{n \leqslant x : n \text{ and } n+1 \text{ QNR} \pmod Q\}| \gg \prod_{p\mid Q} \left(1 - \frac{2}{p}\right). \tag{1.14}$$

REMARK 1.22. In light of Remark 1.7, we could also prove Theorem 1.21 for primitive characters χ modulo Q whose order is bounded (that is, characters χ such that χ^k is principal for some $k \ll 1$).

This theorem is related to [27, Problem 11], although there one asks for cancellation in the ordinary average instead of the logarithmic one, and one wants to take a maximum over $h \leqslant Q$ (but there Q is restricted to primes and $Q \leqslant x^{2+\delta}$ for some small $\delta > 0$). We also remark that in the much smaller range $Q = o(x^2/(\log x))$ and with Q prime, one can use the Weil bound [19, Theorem 11.23] to prove the above estimate. In the same range $Q \leqslant x^{4-\varepsilon}$ as in Theorem 1.21, it was shown by Burgess [2] that

$$x - \sum_{y\leqslant n\leqslant y+x} \chi_Q(n(n+h)) \gg_{\varepsilon,h} x^{\varepsilon/2},$$

and the same estimate holds with $n(n+h)$ replaced by any polynomial that factorizes into linear factors and is not the square of another polynomial.

We note that Theorem 1.21 does not directly follow from the logarithmically averaged binary Elliott conjecture proved in [30], since if the Vinogradov quadratic nonresidue conjecture failed, it would be the case that

$$\mathbb{D}(\chi_Q, 1; x) \ll 1. \tag{1.15}$$

The Vinogradov conjecture states that for any $q \geqslant q(\varepsilon)$, there is a quadratic nonresidue $\pmod q$ on the interval $[1, q^{\varepsilon}]$. We of course do not expect (1.15) to hold, but it cannot be ruled out with current knowledge. Furthermore, the correlation asymptotic in [**20**] does not apply either to Theorem 1.21, since the function χ_Q depends heavily on the length x of the sum. Nevertheless, the function χ_Q has mean value $o(1)$ by the Burgess bound, and by a slight generalization of that, it also has mean $o(1)$ in fixed arithmetic progressions, which is what is required to apply Theorem 1.4. For the details of the proof of Theorem 1.21, see Section 4.

1.2. Structure of the paper. The main theorem, Theorem 1.4, will be proved in Sections 2 and 3. In the former of these sections, the entropy decrement argument from [**30**, **33**] is deployed to replace the correlation average with a simpler, bilinear average. The proof of one lemma in Section 2, concerning stability of mean values of multiplicative functions, is postponed to Appendix A. In Section 3, we use circle method estimates and a short exponential sum estimate for multiplicative functions to show that the bilinear average we mentioned has the anticipated asymptotic formula, concluding the proof. The proof of this exponential sum estimate, which is a slight modification of the one by Matomäki, Radziwiłł and Tao [**23**], is left to Appendix B. In Section 4, we apply Theorem 1.4 to deduce the applications mentioned in the Introduction. Theorem 1.11 will be proved first, and then Theorems 1.14 and 1.19 will be deduced from this. Theorems 1.17 and 1.16 will in turn follow from Theorem 1.14. Theorem 1.21 will be deduced from the main theorem and the Burgess bound.

1.3. Notation. The functions $g_1, g_2 : \mathbb{N} \to [-1, 1]$ are always multiplicative functions. The pretentious distance $\mathbb{D}(f, g; x)$ between two multiplicative functions is given by (1.3). We denote by $\mu(n)$ the Möbius function, by $\varphi(n)$ the Euler totient function, and by $P^+(n)$ the largest prime factor of n, with the convention that $P^+(1) = 1$. By (a, b), we denote the greatest common divisor of a and b. For a proposition $P(n)$, the indicator $1_{P(n)}$ is defined as 1 if $P(n)$ is true and as 0 if $P(n)$ is false. By $\delta(S)$ we denote the logarithmic density of $S \subset \mathbb{N}$, not to be confused with $\delta_1, \delta_2 \in [-1, 1]$, which are the mean values of g_1 and g_2, as defined in formula (2.1).

The variables $p, p_1, p_2, \dots$ will always be primes. We reserve various letters, such as d, k, ℓ, m, n, q for positive integer quantities. The variables x, y in turn will be understood to be large, whereas $\varepsilon > 0$ will tend to zero. The integer $h \neq 0$ is always fixed, and the function $\omega : \mathbb{R}_{\geqslant 1} \to \mathbb{R}$ is a growth function satisfying $1 \leqslant \omega(X) \leqslant \log(3X)$ and tending to infinity with X.

We use the standard Landau and Vinogradov asymptotic notations $O(\cdot)$, $o(\cdot)$, $\gg$, $\ll$, with the convention that the implied constants are absolute unless

otherwise indicated. Thus for instance $o_{\varepsilon\to 0}(1)$ denotes a quantity depending on ε and tending to 0 as $\varepsilon \to 0$, uniformly with respect to all other involved parameters. All the logarithms in the paper will be to base e, and the function $\log_j x$ is the jth iterate of the logarithm function. The function $\exp_j x$ is analogously the jth iterate of $x \mapsto e^x$.

2. The entropy decrement argument and some reductions

Given a function $\omega(X)$ having the same properties as in Theorem 1.4, we define for $a \in \mathbb{Z}$ the correlation sequence

$$f_{x,\omega}(a) := \frac{1}{\log \omega(x)} \sum_{x/\omega(x)\leqslant n\leqslant x} \frac{g_1(n)g_2(n+a)}{n}.$$

As was noted in the Introduction, one can define this equally well for $a < 0$. Our task is then to show that if

$$\delta_1 := \frac{1}{x}\sum_{x\leqslant n\leqslant 2x} g_1(n), \quad \delta_2 := \frac{1}{x}\sum_{x\leqslant n\leqslant 2x} g_2(n) \tag{2.1}$$

are the mean values of g_1 and g_2 (which depend on x), then $|f_{x,\omega}(h) - \delta_1\delta_2| = o_{\varepsilon\to 0}(1)$ under the assumptions of Theorem 1.4. By replacing ε with $1/\exp_2(\varepsilon^{-2})$ in Theorem 1.4, with $\exp_2$ the second iterated exponential, we may in fact assume that

$$g_1 \in \mathcal{U}(x, \exp_2(\varepsilon^{-2}), 1/\exp_2(\varepsilon^{-2})); \tag{2.2}$$

we do this for notational convenience. We may also assume that $|h| \leqslant \varepsilon^{-1}$, since h is fixed in Theorem 1.4 and ε is small.

We average $f_{x,\omega}(h)$ over the primes belonging to a small scale using multiplicativity, and then apply the entropy decrement argument to relate $f_{x,\omega}(h)$ to a bilinear analogue $(\log P/P)\sum_{p\sim P}(g_1(p)^{-1}g_2(p)^{-1}/p)f_{x,\omega}(ph)$ of the same sum (this is the same approach as in Tao's paper [30], and in the later works [31, 33]). Similarly to [30], we then apply the circle method and establish a slight variant of the short exponential sum estimate for multiplicative functions, due to Matomäki, Radziwiłł and Tao [23], to finish the proof. Since Theorem 1.4 involves both pretentious and nonpretentious functions g_j, we need to make a distinction between them in certain parts of the argument. We also separate the case where $|g_1(p)g_2(p)|$ is small for many primes p from the opposite case, since expressions such as $g_1(p)^{-1}g_2(p)^{-1}$ naturally appear in the proof. To deal with these distinctions for g_j, we need the fact that the entropy argument works not only in infinitely many dyadic scales $[2^m, 2^{m+1}]$, but in fact in almost all of them with respect to some measure. Such a strengthening was presented in [33]. We begin with this entropy decrement argument.

LEMMA 2.1 (Entropy decrement argument). *Let $\varepsilon > 0$ be small, $|h| \leqslant \varepsilon^{-1}$ an integer, $x \geqslant x_0(\varepsilon, h, \omega)$, and $\omega : \mathbb{R}_{\geqslant 1} \to \mathbb{R}$ a function with $1 \leqslant \omega(X) \leqslant X$ and $\omega(X) \xrightarrow{X\to\infty} \infty$. Let $g_1, g_2 : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions and $c_p \in \mathbb{D}$ any complex numbers. Then for all $m \in \mathcal{M} \cap [1, \log_2 \omega(x)]$ we have*

$$\frac{m\log 2}{2^m}\sum_{2^m\leqslant p<2^{m+1}} c_p g_1(p)g_2(p)\cdot f_{x,\omega}(h) = \frac{m\log 2}{2^m}\sum_{2^m\leqslant p<2^{m+1}} c_p f_{x,\omega}(ph) + o_{\varepsilon\to 0}(1),$$

with the set $\mathcal{M} \subset \mathbb{N}$ being independent of c_p and being large in the sense that

$$\sum_{\substack{m\geqslant 1\\ m\notin\mathcal{M}}} \frac{1}{m} \ll \varepsilon^{-10}. \tag{2.3}$$

Proof. This follows from the proof of [**33**, Theorem 3.6], but since that argument uses generalized limit functionals, we outline how it goes through without them. We also remark that, without the density bound (2.3), Lemma 2.1 follows from [**30**, Section 3], and that in [**32**, Theorem 3.1] the lemma was proved in the special case of the Liouville function.

We may assume that $m \geqslant \varepsilon^{-1}$ for all $m \in \mathcal{M}$, since removing the numbers $m < \varepsilon^{-1}$ from $\mathcal{M}$ alters the sum in (2.3) by $\sum_{m<\varepsilon^{-1}} 1/m \ll \varepsilon^{-1}$. We have the multiplicativity property $g_j(p)g_j(n) = g_j(pn) + O(1_{p|n})$ for any prime p, so for $2^m \leqslant p < 2^{m+1}$ with $\varepsilon^{-1} \leqslant m \leqslant \log_2 \omega(x)$ we have

$$\begin{aligned} g_1(p)g_2(p)\cdot f_{x,\omega}(h) &= \frac{1}{\log\omega(x)}\sum_{x/\omega(x)\leqslant n\leqslant x}\frac{g_1(pn)g_2(pn+ph)}{n} \\ &\quad + O\Bigg(\frac{1}{\log\omega(x)}\sum_{\substack{x/\omega(x)\leqslant n\leqslant x\\ p\mid n(n+h)}}\frac{1}{n}\Bigg) \\ &= \frac{1}{\log\omega(x)}\sum_{x/\omega(x)\leqslant n\leqslant x}\frac{g_1(pn)g_2(pn+ph)}{n} + O(\varepsilon) \\ &= \frac{1}{\log\omega(x)}\sum_{px/\omega(x)\leqslant n\leqslant px}\frac{g_1(n)g_2(n+ph)}{n}p1_{p|n} + O(\varepsilon) \\ &= \frac{1}{\log\omega(x)}\sum_{x/\omega(x)\leqslant n\leqslant x}\frac{g_1(n)g_2(n+ph)}{n}p1_{p|n} + O(\varepsilon), \end{aligned}$$

where the last step comes from estimating the terms $n \in [x/\omega(x), px/\omega(x)]$ and

$n \in [x, px]$ trivially. (This is the part of the argument where it is crucial to work with logarithmic averaging.)

Define the modified functions $g_j^{(\varepsilon)}(n)$ by rounding $g_j(n)$ to the nearest element of the Gaussian lattice $\varepsilon\mathbb{Z}[i]$. Then, averaging over p the above formula for $g_1(p)g_2(p) \cdot f_{x,\omega}(h)$, we get

$$\begin{aligned}&\frac{m\log 2}{2^m}\sum_{2^m\leqslant p<2^{m+1}} c_p g_1(p)g_2(p)\cdot f_{x,\omega}(h)\\ &\quad=\frac{m\log 2}{2^m\log\omega(x)}\sum_{2^m\leqslant p<2^{m+1}} c_p \sum_{x/\omega(x)\leqslant n\leqslant x}\frac{g_1^{(\varepsilon)}(n)g_2^{(\varepsilon)}(n+ph)}{n}p1_{p|n}+O(\varepsilon).\end{aligned} \tag{2.4}$$

The concentration of measure argument in [33] tells that we may replace $p1_{p|n}$ with $1+O(\varepsilon)$ in (2.4), provided that the random variables

$$\begin{aligned}\mathbf{X}_m &:= (g_r^{(\varepsilon)}(\mathbf{n}+j))_{1\leqslant r\leqslant 2,\,0\leqslant j\leqslant(1+|h|)2^{m+2}},\\ \mathbf{Y}_m &:= (\mathbf{n} \pmod p)_{2^m\leqslant p<2^{m+1}},\quad \mathbf{Y}_{<m} := (\mathbf{Y}_{m'})_{m'<m}\end{aligned}$$

enjoy the conditional mutual information bound

$$\mathbf{I}(\mathbf{X}_m : \mathbf{Y}_m|\mathbf{Y}_{<m}) \leqslant \varepsilon^4\cdot\frac{2^m}{m}. \tag{2.5}$$

(For the definition of conditional mutual information, see [33, Section 2].) We thus need to show that the set $\mathcal{M}$ of m for which (2.5) holds satisfies (2.3). But this was shown in [33, Proposition 3.5] (see also Remark 3.7 there), so we obtain the claim. □

Before utilizing Lemma 2.1, we show that the quantities δ_1 and δ_2 in (2.1) are the mean values of g_1 and g_2 also on many other intervals than $[x, 2x]$. For this we use a slight generalization of a lemma due to Elliott [6] and Granville and Soundararajan [13, Proposition 4.1]. Such results are also proved in Matthiesen's work [25] in a more general setting.

LEMMA 2.2 (Stability of mean values of multiplicative functions). *Let $g : \mathbb{N} \to [-1, 1]$ be a real-valued multiplicative function, $x \geqslant 10$, and $y \in [1, \log^{10} x]$ arbitrary. Then, for $a, q \in \mathbb{N}$, we have*

$$\left|\frac{1}{x}\sum_{\substack{x\leqslant n\leqslant 2x\\ n\equiv a \pmod q}} g(n) - \frac{1}{x/y}\sum_{\substack{x/y\leqslant n\leqslant 2x/y\\ n\equiv a \pmod q}} g(n)\right| \ll_q (\log x)^{-1/400}.$$

Proof. We prove this in Appendix A. □

Owing to the above lemma, we can show that the uniformity assumption on g_1 implies the seemingly stronger assumption that g_1 be uniformly distributed also on intervals $[x/\omega(x), x]$. For this purpose, we need the following definition.

DEFINITION 2.3 (Stronger uniformity assumption). Let $1 \leqslant Q \leqslant x$, $\eta > 0$, and $\delta \in \mathbb{C}$. Let $\omega : \mathbb{R}_{\geqslant 1} \to \mathbb{R}$ be a function with $1 \leqslant \omega(X) \leqslant X$ for all $X \geqslant 1$. For a function $g : \mathbb{N} \to \mathbb{C}$, we write $g \in \mathcal{U}_\omega(x, Q, \eta, \delta)$ if we have the estimate

$$\left|\frac{1}{y} \sum_{\substack{y \leqslant n \leqslant 2y \\ n \equiv a \pmod q}} g(n) - \frac{\delta}{q}\right| \leqslant \frac{\eta}{q} \quad \text{for all } 1 \leqslant a \leqslant q \leqslant Q \text{ and } \frac{x}{\omega(x)} \leqslant y \leqslant x.$$

With the above notation, if δ_1 and δ_2 are as in (2.1), by Lemma 2.2 we have

$$\begin{aligned} &g_1 \in \mathcal{U}_\omega(x, \exp_2(\varepsilon^{-2}), 2/\exp_2(\varepsilon^{-2}), \delta_1), \\ &g_2 \in \mathcal{U}_\omega(x, 1, 2/\exp_2(\varepsilon^{-2}), \delta_2) \quad \text{for } \omega(X) \leqslant \log^{10} X. \end{aligned} \tag{2.6}$$

This property will be used several times in the rest of the proof of the main theorem. In particular, we have for all $y \in [x(\log x)^{-10}, x]$ the estimate

$$\sum_{y \leqslant n \leqslant 2y} g_j(n) = (\delta_j + O(\varepsilon))y,$$

where, as always, the $O(\cdot)$ constant is absolute. Summing this over the dyadic intervals $[y/2^{j+1}, y/2^j]$ for $j \geqslant 0$ and assuming that $y \geqslant x(\log(3x))^{-1}$, say, we get

$$\sum_{n \leqslant y} g_j(n) = (\delta_j + O(\varepsilon))y.$$

Subtracting this formula for two different lengths of summation, we see that

$$\sum_{y \leqslant n \leqslant z} g_j(n) = \delta_j(z - y) + O(\varepsilon z)$$

for all $x(\log(3x))^{-1} \leqslant y \leqslant z \leqslant 2x$. From this and partial summation, we obtain

$$\frac{1}{\log \omega(x)} \sum_{x/\omega(x) \leqslant n \leqslant x} \frac{g_j(n)}{n} = \delta_j + O(\varepsilon) \tag{2.7}$$

for $1 \leqslant \omega(X) \leqslant \log(3X)$, which also will be utilized in what follows.

We return to applying the entropy argument. Defining the normalized correlation sequence

$$\tilde{f}_{x,\omega}(a) := \frac{1}{\log \omega(x)} \sum_{x/\omega(x)\leqslant n\leqslant x} \frac{(g_1(n)-\delta_1)(g_2(n+a)-\delta_2)}{n}$$

and using the simple identity $XY = \delta_1\delta_2+\delta_1(Y-\delta_2)+\delta_2(X-\delta_1)+(X-\delta_1)(Y-\delta_2)$, we deduce from Lemma 2.1 that

$$\begin{aligned}
&\frac{m\log 2}{2^m} \sum_{2^m\leqslant p<2^{m+1}} c_p g_1(p)g_2(p)\cdot f_{x,\omega}(h) \\
&\quad = \delta_1\delta_2 \frac{m\log 2}{2^m} \sum_{2^m\leqslant p<2^{m+1}} c_p + \frac{m\log 2}{2^m} \sum_{2^m\leqslant p<2^{m+1}} c_p \tilde{f}_{x,\omega}(ph) \\
&\qquad + O\Bigg(\max_{r\in\{1,2\}} \frac{1}{\log\omega(x)} \Bigg| \sum_{x/\omega(x)\leqslant n\leqslant x} \frac{g_r(n)-\delta_r}{n} \Bigg| \Bigg) + o_{\varepsilon\to 0}(1), \\
&\qquad m\in\mathcal{M}\cap[1,\log_2\omega(x)].
\end{aligned} \tag{2.8}$$

Formula (2.7) tells that the $O(\cdot)$ error term in (2.8) is $o_{\varepsilon\to 0}(1)$. Then (2.8) takes the form

$$\begin{aligned}
&\frac{m\log 2}{2^m} \sum_{2^m\leqslant p<2^{m+1}} c_p g_1(p)g_2(p)\cdot f_{x,\omega}(h) \\
&\quad = \delta_1\delta_2 \frac{m\log 2}{2^m} \sum_{2^m\leqslant p<2^{m+1}} c_p + \frac{m\log 2}{2^m} \sum_{2^m\leqslant p<2^{m+1}} c_p \tilde{f}_{x,\omega}(ph) + o_{\varepsilon\to 0}(1)
\end{aligned} \tag{2.9}$$

for $m\in\mathcal{M}\cap[1,\log_2\omega(x)]$.

It is natural to predict that the average of the normalized correlation $\tilde{f}_{x,\omega}(h)$ in (2.9) is small, and this is indeed what we prove in Section 3. Before we deal with that term, we consider the main term arising in (2.9). One would like to choose $c_p = g_1(p)^{-1}g_2(p)^{-1}$ there, since then the main term becomes just $\delta_1\delta_2+o_{\varepsilon\to 0}(1)$. However, it may be that $|g_j(p)|$ takes very small values (or even 0), in which case c_p would be unbounded. To avoid this, we prove two lemmas, the first of which tells that if the correlation average in Theorem 1.4 is not negligibly small, then $|g_1(p)g_2(p)|\geqslant\frac{1}{2}$ for most primes p. The second lemma in turn tells that if $|\delta_1|$ and $|\delta_2|$ are not negligibly small, then $g_1(p)g_2(p)$ behaves like 1 in most scales.

LEMMA 2.4 (Dealing with small values of $g_j(p)$). *Let the notations be as in Theorem 1.4. Suppose that*

$$\left|\frac{1}{\log\omega(x)}\sum_{x/\omega(x)\leqslant n\leqslant x}\frac{g_1(n)g_2(n+h)}{n}\right| > \varepsilon^2. \tag{2.10}$$

Let $\exp_2(\varepsilon^{-1}) \leqslant y \leqslant \log\log x$ *be arbitrary. Then there exists a set* $\mathcal{N} \subset [1, y]$ *such that for all* $m \in \mathcal{N}$ *we have*

$$\sum_{\substack{2^m\leqslant p<2^{m+1}\\ |g_1(p)g_2(p)|>1/2}} 1 \geqslant (1-\varepsilon)\cdot\frac{2^m}{m\log 2},$$

with $\mathcal{N}$ *being large in the sense that*

$$\frac{1}{\log y}\sum_{\substack{n\leqslant y\\ n\in\mathcal{N}}}\frac{1}{n} \geqslant 1-\varepsilon.$$

Proof. Suppose for the sake of contradiction that such a set $\mathcal{N}$ does not exist. Then by the prime number theorem we have

$$\sum_{\substack{2^m\leqslant p<2^{m+1}\\ |g_1(p)g_2(p)|\leqslant 1/2}} 1 \geqslant \frac{\varepsilon}{2}\cdot\frac{2^m}{m\log 2} \tag{2.11}$$

for all $m \in \mathcal{N}_1 \subset [1, y]$ with $\mathcal{N}_1$ being a set with the property

$$\sum_{m\in\mathcal{N}_1}\frac{1}{m} \geqslant \frac{\varepsilon}{2}\log y. \tag{2.12}$$

In particular, from (2.11) we have

$$\sum_{\substack{2^m\leqslant p<2^{m+1}\\ |g_1(p)g_2(p)|\leqslant 1/2}}\frac{1}{p} \geqslant \frac{\varepsilon}{8m}$$

for $m \in \mathcal{N}_1$. Summing over $m \in \mathcal{N}_1$ and using (2.12), we conclude that

$$\sum_{\substack{p\leqslant 2^{y+1}\\ |g_1(p)g_2(p)|\leqslant 1/2}}\frac{1}{p} \geqslant \frac{\varepsilon^2}{16}\log y.$$

Hence, for at least one of $j = 1$ and $j = 2$ we have

$$\sum_{\substack{p \leqslant 2^{y+1} \\ |g_j(p)| \leqslant 1/\sqrt{2}}} \frac{1}{p} \geqslant \frac{\varepsilon^2}{32} \log y. \tag{2.13}$$

Fix such $j \in \{1, 2\}$. Let

$$\mathcal{P} := \left\{ \varepsilon^{-10} \leqslant p \leqslant 2^{y+1} : |g_j(p)| \leqslant \frac{1}{\sqrt{2}} \right\},$$

and let $\mu_{\mathcal{P}}^2(n)$ be the indicator function of integers n that are not divisible by p^2 for any $p \in \mathcal{P}$. Note that if $\mu_{\mathcal{P}}^2(n) = 1$, then

$$|g_j(n)| \leqslant \left(\frac{1}{\sqrt{2}}\right)^{\omega_{\mathcal{P}}(n)},$$

where $\omega_{\mathcal{P}}(n)$ is the number of prime factors of n from $\mathcal{P}$. In particular, we have $|g_j(n)| \leqslant \varepsilon^{10}$ whenever $\omega_{\mathcal{P}}(n) \geqslant \varepsilon^{-1}$ (and still $\mu_{\mathcal{P}}^2(n) = 1$). In conclusion, if we show that

$$\frac{1}{\log \omega(x)} \sum_{\substack{x/\omega(x) \leqslant n \leqslant x \\ \mu_{\mathcal{P}}^2(n)=0 \\ \text{or } \omega_{\mathcal{P}}(n) < \varepsilon^{-1}}} \frac{1}{n} \leqslant \varepsilon^3, \tag{2.14}$$

then (2.10) is violated, giving the desired contradiction. We are now left with showing (2.14), and for this we use some basic sieve theory. Note that

$$\frac{1}{\log \omega(x)} \sum_{\substack{x/\omega(x) \leqslant n \leqslant x \\ \mu_{\mathcal{P}}^2(n)=0}} \frac{1}{n} \leqslant \sum_{p \in \mathcal{P}} \frac{1}{\log \omega(x)} \sum_{x/p^2\omega(x) \leqslant m \leqslant x/p^2} \frac{1}{p^2 m} \ll \sum_{p \in \mathcal{P}} \frac{1}{p^2} \ll \varepsilon^{10}.$$

Note also that if $\omega_{\mathcal{P}}(n) = M$ and $\mu_{\mathcal{P}}^2(n) = 1$, then we may write $n = p_1 \cdots p_M m$ with $p_i \in \mathcal{P}$ and $\omega_{\mathcal{P}}(m) = 0$. Hence, by the sieve of Eratosthenes and Mertens' theorem,

$$\begin{aligned}
\sum_{\substack{x/\omega(x) \leqslant n \leqslant x \\ \omega_{\mathcal{P}}(n) < \varepsilon^{-1} \\ \mu_{\mathcal{P}}^2(n)=1}} \frac{1}{n} &\leqslant \varepsilon^{-1} \max_{M < \varepsilon^{-1}} \sum_{p_1, \dots, p_M \leqslant 2^{y+1}} \sum_{\substack{x/(\omega(x) p_1 \cdots p_M) \leqslant m \leqslant x/(p_1 \cdots p_M) \\ \omega_{\mathcal{P}}(m)=0}} \frac{1}{p_1 \cdots p_M m} \\
&\ll \varepsilon^{-1} \max_{M < \varepsilon^{-1}} \sum_{p_1, \dots, p_M \leqslant 2^{y+1}} \frac{1}{p_1 \cdots p_M} \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right) \cdot \log \omega(x)
\end{aligned}$$

$$\ll \varepsilon^{-1} \max_{M<\varepsilon^{-1}} \sum_{p_1,\dots,p_M\leqslant 2^{y+1}} \frac{1}{p_1\cdots p_M} \exp\Big(-\sum_{p\in\mathcal{P}}\frac{1}{p}\Big)\cdot \log\omega(x)$$
$$\ll \varepsilon^{-1}(\log y)^{\varepsilon^{-1}} y^{-\varepsilon^2/32}\log\omega(x) \ll \varepsilon^{10}\log\omega(x)$$

by (2.13) and the fact that $y \geqslant \exp_2(\varepsilon^{-1})$. Combining the above estimates, we obtain (2.14), and hence also the statement of the lemma. □

LEMMA 2.5 (Dealing with pretentious functions). *Let the notations be as in Theorem 1.4. Suppose that $|\delta_1| > \varepsilon^2$ and $|\delta_2| > \varepsilon^2$, where δ_1 and δ_2 are as in (2.1). Let $\exp_2(\varepsilon^{-1}) \leqslant y \leqslant \log\log x$ be arbitrary. Then there exists a set $\mathcal{N}' \subset [1, y]$ such that for all $m \in \mathcal{N}'$ we have*

$$\frac{m\log 2}{2^m}\Big|\sum_{2^m\leqslant p<2^{m+1}}(1-g_1(p)g_2(p))\Big| < \varepsilon,$$

with $\mathcal{N}'$ large in the sense that

$$\frac{1}{\log y}\sum_{\substack{n\leqslant y\\ n\in\mathcal{N}'}}\frac{1}{n} \geqslant 1-\varepsilon.$$

Proof. Note that $1 - g_1(p)g_2(p) \geqslant 0$ always holds. Arguing just as in the proof of Lemma 2.4, we see that if the statement failed, we would have

$$\sum_{p\leqslant 2^{y+1}}\frac{1-g_1(p)g_2(p)}{p} \geqslant \frac{\varepsilon^2}{8}\log y.$$

In particular, by the inequality $(1-a)+(1-b) \geqslant 1-ab$ for $a, b \in [-1, 1]$, for at least one of $j = 1$ and $j = 2$ we would have

$$\sum_{p\leqslant 2^{y+1}}\frac{1-g_j(p)}{p} \geqslant \frac{\varepsilon^2}{16}\log y. \tag{2.15}$$

Now, by (2.15) and a version of Halász's theorem for real-valued multiplicative functions [**16**], we have

$$|\delta_j| = \Big|\frac{1}{x}\sum_{n\leqslant x} g_j(n)\Big| \ll \exp\Big(-\frac{1}{10}\sum_{p\leqslant 2^{y+1}}\frac{1-g_j(p)}{p}\Big)$$
$$\ll \exp\Big(-\frac{\varepsilon^2}{200}\log y\Big) \ll \varepsilon^{10}$$

for $y \geqslant \exp_2(\varepsilon^{-1})$, and this contradicts $|\delta_j| > \varepsilon^2$, proving the lemma. □

Now we return to (2.9) and consider two cases separately. Suppose first that $|\delta_1|, |\delta_2| > \varepsilon^2$. Let $y = \exp_2(\varepsilon^{-1})^2$. Then, if $\mathcal{N}' \subset [1, y]$ is the set in Lemma 2.5 and $\mathcal{M}$ is the set in Lemma 2.1 (which is independent of c_p), taking $c_p = 1$ we deduce from (2.9) and Lemma 2.5 that

$$\begin{aligned} f_{x,\omega}(h) &= \frac{m\log 2}{2^m} \sum_{2^m \leqslant p < 2^{m+1}} g_1(p)g_2(p) f_{x,\omega}(h) + o_{\varepsilon\to 0}(1) \\ &= \delta_1\delta_2 + \frac{m\log 2}{2^m} \sum_{2^m \leqslant p < 2^{m+1}} \tilde{f}_{x,\omega}(ph) + o_{\varepsilon\to 0}(1) \end{aligned} \tag{2.16}$$

for $m \in \mathcal{M} \cap \mathcal{N}'$. We can pick some $m \in \mathcal{M} \cap \mathcal{N}'$ with $m \in [\sqrt{y}, y]$, since we have the lower bound

$$\sum_{\substack{m\in\mathcal{M}\cap\mathcal{N}' \\ m\in[\sqrt{y},y]}} \frac{1}{m} \geqslant \log y - \frac{1}{2}\log y - \varepsilon^{-100} - \varepsilon \log y \geqslant \frac{1}{3}\log y$$

for $y = \exp_2(\varepsilon^{-1})^2$.

Consider then the case where either $|\delta_1| \leqslant \varepsilon^2$ or $|\delta_2| \leqslant \varepsilon^2$. We may suppose that (2.10) holds, since otherwise Theorem 1.4 holds by the fact that $\delta_1\delta_2 + o_{\varepsilon\to 0}(1) = o_{\varepsilon\to 0}(1)$ in this situation. Let $y = \exp_2(\varepsilon^{-1})^2$. Taking $m \in \mathcal{M} \cap \mathcal{N}$ (with $\mathcal{N} \subset [1, y]$ as in Lemma 2.4) and $c_p = g_1(p)^{-1}g_2(p)^{-1}1_{|g_1(p)g_2(p)|\geqslant 1/2}$ in (2.9), we see from Lemma 2.4 that

$$\begin{aligned} f_{x,\omega}(h) + o_{\varepsilon\to 0}(1) &= \frac{m\log 2}{2^m} \\ &\times \left(\delta_1\delta_2 \sum_{\substack{2^m\leqslant p<2^{m+1} \\ |g_1(p)g_2(p)|\geqslant 1/2}} (g_1(p)g_2(p))^{-1} + \sum_{\substack{2^m\leqslant p<2^{m+1} \\ |g_1(p)g_2(p)|\geqslant 1/2}} (g_1(p)g_2(p))^{-1}\tilde{f}_{x,\omega}(ph)\right). \end{aligned}$$

for $m \in \mathcal{M} \cap \mathcal{N}$, which again contains an element $m \in [\sqrt{y}, y]$ by the same argument as above. We know that $|(g_1(p)g_2(p))^{-1}| \leqslant 2$ for all $2^m \leqslant p < 2^{m+1}$, except for at most $10\varepsilon(2^m/m)$ exceptions. Since by assumption $\delta_1\delta_2 = O(\varepsilon)$, we deduce that

$$f_{x,\omega}(h) = \delta_1\delta_2 + \frac{m\log 2}{2^m} \sum_{2^m\leqslant p<2^{m+1}} 2a_p\tilde{f}_{x,\omega}(ph) + o_{\varepsilon\to 0}(1)$$

for $m \in \mathcal{M} \cap \mathcal{N}$, where $a_p := \frac{1}{2}(g_1(p)g_2(p))^{-1}1_{|g_1(p)g_2(p))|\geqslant 1/2}$. In conclusion, regardless of the values of δ_j, Theorem 1.4 will follow once we prove that

$$\frac{m}{2^m \log \omega(x)} \sum_{2^m \leqslant p < 2^{m+1}} a_p \sum_{x/\omega(x) \leqslant n \leqslant x} \frac{(g_1(n) - \delta_1)(g_2(n+ph) - \delta_2)}{n} = o_{\varepsilon \to 0}(1) \tag{2.17}$$

for arbitrary $a_p \in \mathbb{D}$ and $m \in [\exp_2(\varepsilon^{-1}), \exp_2(\varepsilon^{-1})^2]$.

3. Circle method estimates

We proceed to prove (2.17) by applying the circle method and (slightly modified versions of) the short exponential sum estimates for multiplicative functions due to Matomäki, Radziwiłł and Tao [23]. We start with two lemmas, the first of which reduces (2.17) to bounding a short exponential sum and the second of which shows that the set of large frequencies of the exponential sum has small cardinality.

LEMMA 3.1 (A circle method estimate). *Let $\eta > \varepsilon > 0$ be small, h an integer with $1 \leqslant |h| \leqslant \varepsilon^{-1}$, and $\exp_2(\varepsilon^{-1}) \leqslant H \leqslant \log y$. For any complex numbers $a_p \in \mathbb{D}$, introduce the exponential sum*

$$S_H(\theta) := \sum_{P \leqslant p < 2P} a_p e(p\theta),$$

where $P := \varepsilon^{10} H$. Let Ξ_H be the set of residue classes $\xi \in \mathbb{Z}/H\mathbb{Z}$ that satisfy

$$\left| S_H\left(-\frac{h\xi}{H}\right) \right| \geqslant \eta^2 \frac{P}{\log H}. \tag{3.1}$$

Then, for any functions $g_1', g_2' : \mathbb{N} \to \mathbb{C}$ with $|g_1'(n)|, |g_2'(n)| \leqslant 2$, we have

$$\begin{aligned} &\left| \frac{\log P}{P} \sum_{P \leqslant p < 2P} a_p \sum_{y \leqslant n \leqslant y+H} g_1'(n) g_2'(n+ph) \right| \\ &\qquad \leqslant \eta H + 10 \sum_{\xi \in \Xi_H} \left| \sum_{y \leqslant n \leqslant y+H} g_1'(n) e\left(-\frac{\xi n}{H}\right) \right|. \end{aligned}$$

Proof. This follows from [30, Lemma 3.6], writing it using different notation. □

In order to make use of Lemma 3.1, we must know that the exceptional set Ξ_H in that theorem is not too large. Indeed, we have the following bound.

LEMMA 3.2 (Cardinality of large Fourier coefficients). *Let the notations be as in Lemma 3.1, and assume that H is a prime. Then we have $|\Xi_H| \ll \eta^{-20}$.*

Proof. Since $1 \leqslant |h| \leqslant \varepsilon^{-1}$ and H is a prime, the number of those ξ that satisfy (3.1) remains unchanged when h is replaced by 1 in that formula. In [**30**, Lemma 3.7], it was proved using a fourth moment bound and the Selberg sieve that $|S_H(-\xi/H)| \geqslant \eta^2 P/\log H$ for $\ll_\eta 1$ values of $\xi \in \mathbb{Z}/H\mathbb{Z}$, but the same proof gives the claimed quantitative bound. □

To make use of the two lemmas above, we split in (2.17) the sum over n into sums of length H, where H is a prime belonging to $[\varepsilon^{-10} \cdot 2^m, 2\varepsilon^{-10} \cdot 2^m]$, and approximate the sum with an integral, after which (2.17) is reduced to

$$\frac{1}{\log \omega(x)} \int_{x/\omega(x)}^{x} \frac{m}{2^m} \sum_{2^m \leqslant p < 2^{m+1}} \frac{a_p}{H} \sum_{y \leqslant n \leqslant y+H} (g_1(n) - \delta_1)(g_2(n+ph) - \delta_2) \frac{dy}{y} = o_{\varepsilon \to 0}(1). \tag{3.2}$$

By Lemmas 3.1 and 3.2, it suffices to show that

$$\sup_{\alpha \in \mathbb{R}} \frac{1}{\log \omega(x)} \int_{x/\omega(x)}^{x} \frac{1}{y} \left| \frac{1}{H} \sum_{y \leqslant n \leqslant y+H} (g_1(n) - \delta_1) e(\alpha n) \right| dy = o_{\varepsilon \to 0}(1), \tag{3.3}$$

for $H \in [\exp_3(\frac{1}{2}\varepsilon^{-1}), \exp_3(2\varepsilon^{-1})]$, where $\exp_3$ is the third iterated exponential. Indeed, if the left-hand side of (3.3) is $\leqslant F(\varepsilon)$, where $F(u) \to 0$ as $u \to 0$ is a slowly decaying function, one can take $\eta = F(\varepsilon)^{0.01}$ in Lemma 3.2 to deduce (3.2). Covering the interval $[x/\omega(x), x]$ with dyadic intervals, (3.3) will follow from

$$\sup_{\alpha \in \mathbb{R}} \frac{1}{X} \int_{X}^{2X} \left| \frac{1}{H} \sum_{y \leqslant n \leqslant y+H} (g_1(n) - \delta_1) e(\alpha n) \right| dy = o_{\varepsilon \to 0}(1) \tag{3.4}$$

for all $X \in [x/\omega(x), x/2]$ and all $H \in [\exp_3(\frac{1}{2}\varepsilon^{-1}), \exp_3(2\varepsilon^{-1})]$. This is what we set out to prove, following [**23**].

It is natural to split the supremum over α in (3.4) to major and minor arcs, defined using Dirichlet's approximation theorem as

$$\mathfrak{M} := \left\{ \theta \in \mathbb{R} : \left| \theta - \frac{a}{q} \right| \leqslant \frac{W}{qH} \text{ with } a \in \mathbb{Z},\ q < W,\ (a,q) = 1 \right\} \quad \text{and}$$

$$\mathfrak{m} := \mathbb{R} \setminus \mathfrak{M} \subset \left\{ \theta \in \mathbb{R} : \left| \theta - \frac{a}{q} \right| \leqslant \frac{W}{qH} \text{ with } a \in \mathbb{Z}, q \in \left[W, \frac{H}{W} \right], (a,q) = 1 \right\},$$

with $W := \log^5 H \leqslant \exp(5\exp(2\varepsilon^{-1}))$. (3.5)

In the case of the major arcs, the exponential $e(\alpha n)$ can essentially be replaced with $e(an/q)$, and this will lead us to study the distribution of the multiplicative

function g_1 in arithmetic progressions over short intervals. For that purpose, we prove a lemma that is closely related to [22, Theorem 1] and [23, Theorem A.1]. For this lemma, we need to introduce the same 'nicely factorable' set as in [22, Section 2] and [23, Definition 2.1].

DEFINITION 3.3. Let $10 < P_1 < Q_1 \leqslant X$ and $\sqrt{X} \leqslant X_0 \leqslant X$, with $Q_1 \leqslant \exp(\sqrt{\log X_0})$. For $j > 1$, set

$$P_j := \exp(j^{4j}(\log Q_1)^{j-1} \log P_1), \quad Q_j := \exp(j^{4j+2}(\log Q_1)^j).$$

Letting J be the largest integer such that $Q_J \leqslant \exp(\sqrt{\log X_0})$, we define $\mathcal{S}_{P_1,Q_1,X_0,X}$ as the set of those $1 \leqslant n \leqslant X$ that have at least one prime factor from each of the intervals $[P_j, Q_j]$ for all $1 \leqslant j \leqslant J$.

For a specific choice of the parameters, present in the next lemma, we denote

$$\mathcal{S} := \mathcal{S}_{P_1,Q_1,X_0,X}, \quad \text{where } P_1 = W^{200},\ Q_1 = \frac{H}{W^3},\ X_0 = \sqrt{X}. \tag{3.6}$$

LEMMA 3.4 (Uniform distribution of multiplicative functions in short intervals). *Let $\varepsilon > 0$ be small, $X \geqslant 100$ large, and $H \in [\exp_2(\frac{1}{10}\varepsilon^{-1}), \log\log X]$. Let $g : \mathbb{N} \to [-1, 1]$ be a real-valued multiplicative function. Further, let $b, q \in \mathbb{N}$ with $1 \leqslant b \leqslant q \leqslant W \in [\log^5 H, \log^{10} H]$. Then, if $\mathcal{S}$ is as in (3.6), we have*

$$\frac{1}{X}\int_X^{2X} \left| \frac{1}{H} \sum_{\substack{y \leqslant n \leqslant y+H \\ n \equiv b \pmod q}} g(n)1_{\mathcal{S}}(n) - \frac{1}{X} \sum_{\substack{X \leqslant n \leqslant 2X \\ n \equiv b \pmod q}} g(n)1_{\mathcal{S}}(n) \right| dy \ll W^{-10}. \tag{3.7}$$

REMARK 3.5. If the bound on the right-hand side of (3.7) was replaced with $W^{-0.001}$, the proof of the lemma would work even when $g(n)1_{\mathcal{S}}(n)$ is replaced with $g(n)$. However, for larger values of q we need to introduce the nicely factorable set $\mathcal{S}$ to get better error terms.

Proof of Lemma 3.4. We first reduce to primitive residue classes $b \pmod q$. Let $d_0 = (b, q)$, $b_0 = b/(b, q)$ and $q_0 = q/(b, q)$. Then we have

$$\frac{1}{H} \sum_{\substack{y \leqslant n \leqslant y+H \\ n \equiv b \pmod q}} g(n)1_{\mathcal{S}}(n) = \frac{1}{H} \sum_{\substack{y/d_0 \leqslant n' \leqslant (y+H)/d_0 \\ n \equiv b_0 \pmod{q_0}}} g(d_0n')1_{\mathcal{S}}(n'), \tag{3.8}$$

since $1_{\mathcal{S}}(d_0n') = 1_{\mathcal{S}}(n')$ for $d_0 \leqslant q \leqslant W < P_1$. Since the residue class $b_0 \pmod{q_0}$ is primitive, we may use a Dirichlet character expansion to write the

right-hand side of (3.8) as

$$\frac{1}{H}\frac{1}{\varphi(q_0)}\sum_{\chi \pmod{q_0}}\bar{\chi}(b_0)\sum_{y/d_0\leqslant n'\leqslant (y+H)/d_0} g(d_0n')1_{\mathcal{S}}(n')\chi(n')$$
$$=\frac{1}{H}\frac{1}{\varphi(q_0)}\sum_{\chi \pmod{q_0}}\bar{\chi}(b_0)\sum_{t\mid d_0^\infty}\sum_{\substack{y/d_0\leqslant n'\leqslant (y+H)/d_0\\ t\mid n'\\ (n'/t,d_0)=1}} g(d_0n')1_{\mathcal{S}}(n')\chi(n'), \tag{3.9}$$

where $t \mid d_0^\infty$ means that $t \mid d_0^k$ for some k. Since we have the condition $(n'/t, d_0) = 1$, we may use multiplicativity to write this as

$$\frac{1}{\varphi(q_0)}\sum_{\chi \pmod{q_0}}\bar{\chi}(b_0)\sum_{t\mid d_0^\infty}\frac{g(d_0t)\chi(t)}{d_0t}\frac{d_0t}{H}\sum_{y/(d_0t)\leqslant m\leqslant (y+H)/(d_0t)} g(m)1_{\mathcal{S}}(m)\chi\psi_0(m), \tag{3.10}$$

where $\psi_0(m) = 1_{(m,d_0)=1}$ is the principal character $\pmod{d_0}$ and we used the fact that $1_{\mathcal{S}}(tm) = 1_{\mathcal{S}}(m)$ for t having no prime factors that are larger than $d_0 \leqslant q \leqslant W < P_1$. By crude estimation, the contribution of the terms $t \geqslant H^\varepsilon$ to (3.10) is $\ll H^{-\varepsilon}$, so we may assume that $t < H^\varepsilon$. We now wish to compare the short sums in (3.10) to the corresponding long sums.

Suppose first that χ is real-valued. Then we may apply the Matomäki–Radziwiłł theorem [22, Theorem 3] to the real-valued multiplicative function $g\chi\psi_0$ conclude that

$$\frac{d_0t}{H}\sum_{y/(d_0t)\leqslant m\leqslant (y+H)/(d_0t)} g(m)1_{\mathcal{S}}(m)\chi\psi_0(m)$$
$$=\frac{d_0t}{X}\sum_{X/(d_0t)\leqslant m\leqslant 2X/(d_0t)} g(m)1_{\mathcal{S}}(m)\chi\psi_0(m)+E_{\chi,H}(y), \tag{3.11}$$

for $y \in [X, 2X]$, with the error $E_{\chi,H}(y)$ satisfying the L^2 bound

$$\frac{1}{X}\int_X^{2X}|E_{\chi,H}(y)|^2\,dy \ll \frac{(\log H)^{1/3}}{P_1^{1/10}}+(\log X)^{-1/50} \ll W^{-19}, \tag{3.12}$$

since $W \in [\log^5 H, \log^{10} H]$, and $P_1 = W^{200}$ in our definition of $\mathcal{S}$.

Suppose then that χ is complex-valued. We again write (3.11), and want to obtain an L^2 bound for the error $E_{\chi,H}(y)$. By an argument of Granville and Soundararajan (see [23, Lemma C.1]), the fact that $g\psi_0$ is real and χ is complex (and that $q \leqslant (\log_3 X)^{10}$) leads to

$$\inf_{|t|\leqslant x}\mathbb{D}(g\chi\psi_0, n^{it}; x) \geqslant \tfrac{1}{10}\sqrt{\log\log x}. \tag{3.13}$$

Now we appeal to a variant of the Matomäki–Radziwiłł theorem, established by Matomäki, Radziwiłł and Tao in [23, Theorem A.2]. This result (applied with $h = H$ and $h = X$ separately) gives

$$\frac{1}{X}\int_X^{2X} |E_{\chi,H}(y)|^2\,dy \ll \exp\Big(-\inf_{|t|\leqslant X}\frac{\mathbb{D}(g\chi\psi_0, n^{it}; X)^2}{2}\Big) + \frac{(\log H)^{1/3}}{P_1^{1/10}} + (\log X)^{-1/50}, \tag{3.14}$$

which is $\ll W^{-19}$ by (3.13).

Now, for all characters $\chi \pmod{q_0}$, we have (3.11) with the error bound (3.12). Note also that $\sum_{t|d_0^\infty} 1/d_0 t = 1/d_0 \prod_{p|d_0}(1 + 1/p + 1/p^2 + \cdots) \ll \log d_0/d_0$. Hence, applying the triangle inequality, and summing over χ and $t \mid d_0^\infty$, we see that (3.10) equals

$$\frac{1}{\varphi(q_0)}\sum_{\chi \pmod{q_0}}\bar{\chi}(b_0)\sum_{t|d_0^\infty}\frac{g(d_0t)\chi(t)}{d_0t}\frac{d_0t}{X}\sum_{X/(d_0t)\leqslant m\leqslant 2X/(d_0t)} g(m)1_{\mathcal{S}}(m)\chi\psi_0(m) + E(y), \tag{3.15}$$

with the error term $E(y)$ satisfying

$$\frac{1}{X}\int_X^{2X}|E(y)|^2\,dy \ll W^{-10}.$$

We can then reverse the deduction that led to (3.10) to conclude that (3.15) (and hence (3.8)) equals

$$\frac{1}{X}\sum_{\substack{X\leqslant n\leqslant 2X\\ n\equiv b \pmod q}} g(n)1_{\mathcal{S}}(n) + E(y).$$

This completes the proof. □

The major arc case $\alpha \in \mathfrak{M}$ of (3.4) is dealt with the following Lemma, whose proof uses Lemma 3.4 as an ingredient.

LEMMA 3.6 (Major arc estimate). *Let $\varepsilon > 0$ be small, $x \geqslant 100$ large, $\omega(X)$ as in Theorem 1.4, and $H \in [\exp_3(\frac{1}{2}\varepsilon^{-1}), \exp_3(2\varepsilon^{-1})]$. Let $g_1 : \mathbb{N} \to [-1, 1]$ be a multiplicative function satisfying $g_1 \in \mathcal{U}_\omega(x, \exp_2(\varepsilon^{-2}), 2/\exp_2(\varepsilon^{-2}), \delta_1)$. Then we have*

$$\sup_{\alpha\in\mathfrak{M}}\frac{1}{X}\int_X^{2X}\Big|\frac{1}{H}\sum_{y\leqslant n\leqslant y+H}(g_1(n)-\delta_1)e(\alpha n)\Big|\,dy = o_{\varepsilon\to 0}(1)$$

for all $X \in [x/\omega(x), x/2]$, with the major arcs $\mathfrak{M}$ as in (3.5).

Proof. This is proved in Appendix B. □

The minor arc case $\alpha \in \mathfrak{m}$ of (3.4), in turn, is taken care of by the next lemma.

LEMMA 3.7 (Minor arc estimate). *Let $\varepsilon > 0$ be small, $x \geqslant 100$ large, and suppose that $H \in [\exp_3(\frac{1}{2}\varepsilon^{-1}), \log\log x]$. Then, for any multiplicative function $g_1 : \mathbb{N} \to [-1, 1]$ and for any $\delta_1 \in [-1, 1]$ we have*

$$\sup_{\alpha\in\mathfrak{m}} \frac{1}{X}\int_X^{2X} \left|\frac{1}{H}\sum_{y\leqslant n\leqslant y+H} (g_1(n) - \delta_1)e(\alpha n)\right| dy \ll (\log H)^{-1/10} \tag{3.16}$$

for all $X \in [\sqrt{x}, x]$, with the minor arcs $\mathfrak{m}$ as in (3.5).

Proof. This is proved in Appendix B. □

With these lemmas available, Theorem 1.4 quickly follows.

Proof of Theorem 1.4. We reduced the proof of the theorem to proving (3.4). As was observed after Lemma 2.2, we may assume that we have $g_1 \in \mathcal{U}_\omega(x, \exp_2(\varepsilon^{-2}), 2/\exp_2(-\varepsilon^{-2}), \delta_1)$. Now, if $\alpha \in \mathfrak{M}$ in the supremum present in that formula, we appeal to Lemma 3.6. In the opposite case $\alpha \in \mathfrak{m}$, we appeal to Lemma 3.7. In both cases, we get a bound of $o_{\varepsilon\to 0}(1)$ for the left-hand side of (3.4). This finishes the proof. □

4. Proofs of the applications

Proof of Theorem 1.11. Given any real numbers $z, w \in [-1, 1]$, define the multiplicative functions $g_1, g_2 : \mathbb{N} \to [-1, 1]$ by setting at prime powers

$$g_1(p^j) = \begin{cases} 1 & \text{if } p \leqslant x^a \\ z & \text{if } p > x^a, \end{cases} \qquad g_2(p^j) = \begin{cases} 1 & \text{if } p \leqslant x^b \\ w & \text{if } p > x^b. \end{cases}$$

We apply Theorem 1.4 to g_1 and g_2, and then use a generating function argument to deduce Theorem 1.11. In order to use Theorem 1.4, we must verify that $g_1 \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$ for all $x \geqslant x_0(\varepsilon)$.

First observe that $g_1(n) = z^{\omega_{>x^a}(n)}$, so for any $c, q \in \mathbb{N}$ we have

$$\frac{1}{x}\sum_{\substack{x\leqslant n\leqslant 2x\\ n\equiv c \pmod q}} g_1(n) = \sum_{0\leqslant k<1/a} z^k \cdot \frac{1}{x}\sum_{\substack{x\leqslant n\leqslant 2x\\ n\equiv c \pmod q}} 1_{\omega_{>x^a}(n)=k}.$$

From this we see that $g_1 \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$ for all $x \geqslant x_0(\varepsilon)$ will follow, once we show that

$$\frac{1}{x} \sum_{\substack{x \leqslant n \leqslant 2x \\ n \equiv c \pmod q}} 1_{\omega_{>x^a}(n)=k} = \frac{1}{qx} \sum_{x \leqslant n \leqslant 2x} 1_{\omega_{>x^a}(n)=k} + o_q(1)$$

as $x \to \infty$ for all fixed $c, q, k \in \mathbb{N}$. Write $d_0 = (c, q)$, $c' = c/d_0$, $q' = q/d_0$. Then we have

$$\frac{1}{x} \sum_{\substack{x \leqslant n \leqslant 2x \\ n \equiv c \pmod q}} 1_{\omega_{>x^a}(n)=k} = \frac{1}{x} \sum_{\substack{x/d_0 \leqslant n' \leqslant 2x/d_0 \\ n' \equiv c' \pmod{q'}}} 1_{\omega_{>x^a}(n')=k} := S_k, \tag{4.1}$$

because $\omega_{>x^a}(d_0 n') = \omega_{>x^a}(n')$ for all $d_0 < x^a$. Let $b^{-1} \pmod q$ denote the inverse of b modulo q. Using the fact that $1_{\omega_{>x^a}(n)=0} = 1_{P^+(n) \leqslant x^a}$, we have

$$S_k = \sum_{\substack{x^a < p_1 < \cdots < p_k \leqslant x \\ p_1 \cdots p_k \leqslant x}} \frac{1}{x} \sum_{\substack{x/d_0 p_1 \cdots p_k \leqslant m \leqslant 2x/d_0 p_1 \cdots p_k \\ m \equiv c'(p_1 \cdots p_k)^{-1} \pmod{q'}}} 1_{P^+(m) \leqslant x^a} + o_{q'}(1),$$

with the $o(1)$ term coming from those numbers $n' \leqslant x$ such that $p^2 \mid n'$ for some $p > x^a$. As is well known, smooth numbers are uniformly distributed in arithmetic progressions to fixed moduli (see for instance [**18**, Formula (6.1)]), in the sense that

$$\frac{1}{y} |\{y \leqslant n \leqslant 2y : P^+(n) \leqslant y^u, n \equiv c \pmod{q'}\}| = \frac{1}{q'} \rho\left(\frac{1}{u}\right) + o_{q'}(1), \tag{4.2}$$

for $u \in [0, 1]$ and $y \to \infty$, with $\rho(\cdot)$ being the Dickmann function. Therefore,

$$S_k = \frac{1}{q' d_0} \sum_{\substack{x^a < p_1 < \cdots < p_k \leqslant x \\ p_1 \cdots p_k \leqslant x}} \frac{1}{p_1 \cdots p_k} \left(\rho\left(\frac{\log \frac{x}{d_0 p_1 \cdots p_k}}{a \log x} \right) + o_{q'}(1) \right). \tag{4.3}$$

One easily sees that $x \mapsto \rho(x)$ is a Lipschitz function, so that $|\rho(u) - \rho(v)| \leqslant C|u - v|$ for all $u, v \geqslant 0$ with some constant $C > 0$. Hence, we can use the prime number theorem in the form that the nth prime is asymptotic to $n \log n$ and approximate the term involving $\rho(\cdot)$ in (4.3) to deduce that

$$S_k = \frac{1}{q' d_0 k!} \sum_{\substack{x^a < n_1, \dots, n_k \leqslant x \\ n_1 \cdots n_k \leqslant x}} \frac{1}{n_1 \cdots n_k (\log n_1) \cdots (\log n_k)} \rho\left(\frac{\log \frac{x}{n_1 \cdots n_k}}{a \log x} \right) + o_{q'}(1). \tag{4.4}$$

Here we have estimated trivially as $o_{q'}(1)$ the contribution of the tuples $(n_1, \dots, n_k)$ with two of the n_i equal, or with $n_i \in [x^a/2\log x, x^a] \cup [x/2\log x, x]$ for some i, as for them it is not necessarily the case that the n_ith prime belongs to $[x^a, x]$. Approximating the expression (4.4) with an integral, again using the fact that $\rho(\cdot)$ is Lipschitz, it equals

$$\begin{aligned} S_k &= \frac{1}{q'd_0} \cdot \frac{1}{k!} \int_{\substack{x^a \leqslant x_i \leqslant x \\ x_1 \cdots x_k \leqslant x}} \frac{\rho\left(\frac{\log \frac{x}{x_1 \cdots x_k}}{a \log x}\right)}{x_1 \cdots x_k (\log x_1) \cdots (\log x_k)} d\mathbf{x} + o_{q'}(1) \\ &= \frac{1}{q'd_0} \cdot \frac{1}{k!} \int_{\substack{a \leqslant u_1, \dots, u_k \leqslant 1 \\ u_1 + \cdots + u_k \leqslant 1}} \frac{\rho\left(\frac{1-u_1-\cdots-u_k}{a}\right)}{u_1 \cdots u_k} d\mathbf{u} + o_{q'}(1), \end{aligned}$$

where the last integral comes from a change of variables $u_i = \log x_i / \log x$. Combining (4.1) with the previous equation, we have shown that

$$\frac{1}{x} \sum_{\substack{x \leqslant n \leqslant 2x \\ n \equiv c \pmod q}} 1_{\omega_{>x^a}(n)=k} = \frac{I_{a,k} + o_q(1)}{qk!}, \quad \frac{1}{x} \sum_{\substack{x \leqslant n \leqslant 2x \\ n \equiv c \pmod q}} 1_{\omega_{>x^b}(n)=\ell} = \frac{I_{b,\ell} + o_q(1)}{q\ell!},$$

where

$$I_{\alpha,m} := \int_{\substack{\alpha \leqslant u_1, \dots, u_m \leqslant 1 \\ u_1 + \cdots + u_m \leqslant 1}} \frac{\rho\left(\frac{1-u_1-\cdots-u_m}{\alpha}\right)}{u_1 \cdots u_m} d\mathbf{u}. \tag{4.5}$$

This implies that $g_j \in \mathcal{U}(x, \varepsilon^{-1}, \varepsilon)$ for all $x \geqslant x_0(\varepsilon)$.

Now that we have shown that g_1 and g_2 satisfy our uniform distribution in arithmetic progressions assumption, Theorem 1.4 with $\omega(X) = \log(3X)$ gives

$$\begin{aligned} \frac{1}{\log_2 x} \sum_{x/\log x \leqslant n \leqslant x} \frac{g_1(n)g_2(n+1)}{n} &= \frac{1}{\log_2 x} \sum_{x/\log x \leqslant n \leqslant x} \frac{z^{\omega_{>x^a}(n)} w^{\omega_{>x^b}(n+1)}}{n} \\ &= \left(\frac{1}{x} \sum_{x \leqslant n \leqslant 2x} z^{\omega_{>x^a}(n)}\right)\left(\frac{1}{x} \sum_{x \leqslant n \leqslant 2x} w^{\omega_{>x^b}(n)}\right) \\ &\quad + o(1). \end{aligned} \tag{4.6}$$

Note that the numbers $n \in [x/\log x, x]$ with $\omega_{>x^a}(n) \neq \omega_{>n^a}(n)$ have a prime divisor on the interval $[(x/\log x)^a, x^a]$, so their contribution to the left-hand side of the above sum is bounded by

$$\sum_{(x/\log x)^a \leqslant p \leqslant x^a} \frac{1}{\log_2 x} \sum_{\substack{x/\log x \leqslant n \leqslant x \\ p \mid n}} \frac{1}{n} = o(1).$$

We can do a similar computation to exclude the terms with $\omega_{>x^a}(n) \neq \omega_{>n^a}(n)$ on the right-hand side of (4.6). Applying the same arguments also to $\omega_{>x^b}(n)$, (4.6) takes the form

$$\frac{1}{\log_2 x}\sum_{x/\log x\leqslant n\leqslant x}\frac{z^{\omega_{>n^a}(n)}w^{\omega_{>n^b}(n+1)}}{n} = \left(\frac{1}{x}\sum_{x\leqslant n\leqslant 2x} z^{\omega_{>n^a}(n)}\right)\left(\frac{1}{x}\sum_{x\leqslant n\leqslant 2x} w^{\omega_{>n^b}(n)}\right) + o(1). \tag{4.7}$$

By the preceding considerations,

$$\frac{1}{X}\sum_{X\leqslant n\leqslant 2X} z^{\omega_{>n^a}(n)} = \sum_{0\leqslant k<\frac{1}{a}} z^k\cdot\frac{I_{a,k}}{k!} + o(1) \tag{4.8}$$

as $X\to\infty$, with $I_{a,k}$ as in (4.5), so summing this dyadically we find that (4.8) also holds with the summation range being $1\leqslant n\leqslant X$. Thus, by partial summation,

$$\begin{aligned}\frac{1}{\log x}\sum_{n\leqslant x}\frac{z^{\omega_{>n^a}(n)}}{n} &= \sum_{0\leqslant k<1/a} z^k\frac{I_{a,k}+o(1)}{k!},\\ \frac{1}{\log x}\sum_{n\leqslant x}\frac{w^{\omega_{>n^b}(n)}}{n} &= \sum_{0\leqslant \ell<1/b} w^\ell\frac{I_{b,\ell}+o(1)}{\ell!}.\end{aligned} \tag{4.9}$$

Based on (4.7) and (4.9), if we put

$$c_{k,\ell}(x) := \frac{1}{\log_2 x}\sum_{x/\log x\leqslant n\leqslant x}\frac{1_{\omega_{>n^a}(n)=k}1_{\omega_{>n^b}(n+1)=\ell}}{n},$$

$$a_k(x) := \frac{1}{\log x}\sum_{n\leqslant x}\frac{1_{\omega_{>n^a}(n)=k}}{n},\quad b_\ell(x) := \frac{1}{\log x}\sum_{n\leqslant x}\frac{1_{\omega_{>n^b}(n)=\ell}}{n},$$

then we have

$$\sum_{\substack{0\leqslant k<1/a\\ 0\leqslant \ell<1/b}} c_{k,\ell}(x)z^k w^\ell = \left(\sum_{0\leqslant k<1/a} a_k(x)z^k\right)\left(\sum_{0\leqslant \ell<\frac{1}{b}} b_\ell(x)w^\ell\right) + o(1)$$

for all $z, w\in[-1,1]$. Expanding out, we see that

$$\sum_{\substack{0\leqslant k<1/a\\ 0\leqslant \ell<1/b}} (c_{k,\ell}(x) - a_k(x)b_\ell(x))z^k w^\ell = o(1). \tag{4.10}$$

We show that $c_{k,\ell}(x) = a_k(x)b_\ell(x) + o(1)$. Suppose for the sake of contradiction that this is not the case. Then, by compactness, we can find a sequence x_i tending to infinity such that the numbers $D_{k,\ell} := \lim_{i\to\infty}(c_{k,\ell}(x_i) - a_k(x_i)b_\ell(x_i))$ exist, and at least one of them is nonzero. Taking limits in (4.10), we infer

$$\sum_{\substack{0\leqslant k<1/a\\ 0\leqslant \ell<1/b}} D_{k,\ell} z^k w^\ell = 0$$

for all $z, w \in [-1, 1]$. We now have a polynomial in two variables vanishing in an open set, so its coefficients $D_{k,\ell}$ must all be zero, which is a contradiction. Thus we have

$$c_{k,\ell}(x) = a_k(x)b_\ell(x) + o(1) = \delta_1^*\delta_2^* + o(1) \tag{4.11}$$

for all $0 \leqslant k < 1/a$, $0 \leqslant \ell < 1/b$, with $\delta_1^* := I_{a,k}/k!$ and $\delta_2^* := I_{b,\ell}/\ell!$.

Using (4.11) for $x \in \{y_1, y_2, \dots, y_{J-1}\}$, where $y_1 = x$, $y_{j+1} = y_j/\log y_j$ and $y_J \in [\sqrt{\log x}, \log x]$, it follows that

$$\begin{aligned}\frac{1}{\log x}\sum_{n\leqslant x}\frac{1_{\omega_{>n^a}(n)=k}1_{\omega_{>n^b}(n+1)=\ell}}{n} &= \frac{1}{\log x}\sum_{j=1}^{J-1}\log\log y_j\cdot(\delta_1^*\delta_2^* + o(1))\\ &= (\delta_1^*\delta_2^* + o(1))\frac{1}{\log x}\sum_{j=1}^{J-1}\log\frac{y_j}{y_{j+1}}\\ &= \delta_1^*\delta_2^* + o(1) = a_k(x)b_\ell(x) + o(1)\end{aligned}$$

by telescopic summation. Taking limits as $x \to \infty$ from this, we reach the statement of the theorem about logarithmic densities.

For the part of the theorem involving asymptotic density, we apply the same argument as above, but with $1 \leqslant \omega(X) \leqslant \log(3X)$ an arbitrary function tending to infinity (instead of $\omega(X) = \log(3X)$). We again have

$$\frac{1}{\log\omega(x)}\sum_{x/\omega(x)\leqslant n\leqslant x}\frac{1_{\omega_{>n^a}(n)=k}1_{\omega_{>n^b}(n+1)=\ell}}{n} = \delta_1^*\delta_2^* + o(1). \tag{4.12}$$

In particular, we get

$$\frac{1}{x}\sum_{n\leqslant x}1_{\omega_{>n^a}(n)=k}1_{\omega_{>n^b}(n+1)=\ell} \geqslant \frac{1}{2}\delta_1^*\delta_2^*\frac{\log\omega(x)}{\omega(x)} \tag{4.13}$$

for all large enough x (where large enough depends on the function $\omega(X)$). Now, supposing that the part of Theorem 1.11 concerning asymptotic density fails, there

is a function $\psi(x)$ tending to infinity such that the left-hand side of (4.13) is $\leqslant 1/\psi(x)$ for infinitely many integers x. However, taking $\omega(x) = \psi(x)$ in (4.13), we get a contradiction as $x \to \infty$. Hence, there exists some $c_0 > 0$ such that the left-hand side of (4.13) is $\geqslant c_0$ for all large enough x, which was to be shown. □

Our theorems on smooth numbers follow rather quickly from Theorem 1.11. In fact, one could also deduce these applications directly from Theorem 1.4, using the fact that smooth numbers are uniformly distributed in arithmetic progressions. We leave the details of this alternative argument to the interested reader.

Proof of Theorem 1.14. It follows from (4.2) with $q' = 1$ and partial summation that the set $\{n \in \mathbb{N} : P^+(n) \leqslant n^a\}$ has logarithmic density $\rho(1/a)$. Taking $k = \ell = 0$ in Theorem 1.11 and noticing that $\omega_{>y}(n) = 0$ if and only if $P^+(n) \leqslant y$, the conclusion is immediate. □

Theorem 1.16 is a corollary to Theorem 1.14, as we see next.

Proof of Theorems 1.16 and 1.17. As mentioned in the introduction, taking $\alpha = 0$ in Theorem 1.17 implies Theorem 1.16, since by symmetry

$$\int_{\substack{(x,y)\in[0,1]^2\\ x\geqslant y}} u(x)u(y)\,dx\,dy = \frac{1}{2}\int_{(x,y)\in[0,1]^2} u(x)u(y)\,dx\,dy = \frac{1}{2},$$

where the last equality comes from the fundamental theorem of calculus and the fact that $u(x) = (d/dx)\rho(1/x)$. Thus it suffices to prove Theorem 1.17. Let $0 < a, b, c, d < 1$ be given real numbers with $a < c$ and $b < d$. Applying the inclusion–exclusion formula to the sets $\{n \in \mathbb{N} : P^+(n) \leqslant n^a\}, \dots, \{n \in \mathbb{N} : P^+(n) \leqslant n^d\}$ and employing Theorem 1.14 and the fundamental theorem of calculus, we see that

$$\begin{aligned}
&\delta(\{n \in \mathbb{N} : n^a < P^+(n) < n^b,\ n^c < P^+(n+1) < n^d\})\\
&\quad = \rho\left(\frac{1}{b}\right)\rho\left(\frac{1}{d}\right) - \rho\left(\frac{1}{a}\right)\rho\left(\frac{1}{d}\right) - \rho\left(\frac{1}{b}\right)\rho\left(\frac{1}{c}\right) + \rho\left(\frac{1}{a}\right)\rho\left(\frac{1}{c}\right)\\
&\quad = \left(\rho\left(\frac{1}{d}\right) - \rho\left(\frac{1}{c}\right)\right)\left(\rho\left(\frac{1}{b}\right) - \rho\left(\frac{1}{a}\right)\right)\\
&\quad = \int_a^b \int_c^d u(x)u(y)\,dx\,dy.
\end{aligned}$$

In other words, for any rectangle $\mathcal{R} \subset [0,1]^2$ parallel to the coordinate axes we have

$$\delta\left(\left\{n \in \mathbb{N} : \left(\frac{\log P^+(n)}{\log n}, \frac{\log P^+(n+1)}{\log n}\right) \in \mathcal{R}\right\}\right) = \int_{\mathcal{R}} u(x)u(y)\,dx\,dy. \tag{4.14}$$

Now, if $S \subset [0,1]^2$ is any set such that 1_S is Riemann integrable, we can approximate S from the inside and outside with finite unions of rectangles, so by the monotone convergence theorem we see that (4.14) continues to hold for such sets S. Taking $S = T_\alpha$, Theorem 1.17 is proved. □

Proof of Theorem 1.19. Let $1 \leqslant \omega(X) \leqslant \log(3X)$ be a function tending to infinity. Defining $F_u(n) := 1_{P^+(n) \leqslant n^u}$, by the inclusion–exclusion principle we have

$$\begin{aligned}&\sum_{x/\omega(x) \leqslant n \leqslant x} \frac{1_{P^+(n) \in [n^a, n^b]} 1_{P^+(n+1) \in [n^c, n^d]}}{n} \\ &= \sum_{x/\omega(x) \leqslant n \leqslant x} \frac{F_b(n)F_d(n+1) - F_a(n)F_d(n+1) - F_b(n)F_c(n+1) + F_a(n)F_c(n+1)}{n}.\end{aligned}$$

From (4.12) (with $k = \ell = 0$), it follows that the previous expression is $\log \omega(x)$ times

$$\begin{aligned}&\rho\left(\frac{1}{b}\right)\rho\left(\frac{1}{d}\right) - \rho\left(\frac{1}{a}\right)\rho\left(\frac{1}{d}\right) - \rho\left(\frac{1}{b}\right)\rho\left(\frac{1}{c}\right) + \rho\left(\frac{1}{a}\right)\rho\left(\frac{1}{c}\right) + o(1) \\ &= \left(\rho\left(\frac{1}{d}\right) - \rho\left(\frac{1}{c}\right)\right)\left(\rho\left(\frac{1}{b}\right) - \rho\left(\frac{1}{a}\right)\right) + o(1).\end{aligned} \tag{4.15}$$

In particular, as in (4.13), we get

$$\sum_{n \leqslant x} \frac{1_{P^+(n) \in [n^a, n^b]} 1_{P^+(n+1) \in [n^c, n^d]}}{n} \geqslant \frac{1}{2} c_0(a,b,c,d) \frac{\log \omega(x)}{\omega(x)}, \tag{4.16}$$

where $c_0(a,b,c,d) > 0$ is the constant in (4.15), and since $\omega(X)$ was allowed to tend to infinity as slowly as we please, the left-hand side of (4.16) is lower-bounded by some positive constant, as asserted. □

Lastly, we deduce our quadratic character sum bound from the main theorem.

Proof of Theorem 1.21. The first part of the theorem will follow directly from Theorem 1.4, once we show that for any fixed $a, q \in \mathbb{N}$ we have

$$\sum_{\substack{x \leqslant n \leqslant 2x \\ n \equiv a \pmod q}} \chi_Q(n) = o(x)$$

as $x \to \infty$. Denoting $d_0 = (a, q)$, $a' = a/d_0$, $q' = q/d_0$, and using complete multiplicativity, it suffices to show that

$$\sum_{\substack{x/d_0 \leqslant m \leqslant 2x/d_0 \\ m \equiv a' \pmod{q'}}} \chi_Q(m) = o(x).$$

Expanding the congruence condition in terms of Dirichlet characters, we are left with showing that

$$\sum_{x/d_0 \leqslant m \leqslant 2x/d_0} \chi_Q(m)\psi(m) = o(x) \tag{4.17}$$

for all Dirichlet characters $\psi \pmod{q'}$. Note that the character $\chi^* := \chi_Q\psi$ has modulus $Q^* := Qq' \leqslant x^{4-\varepsilon/2}$ if x is large enough. In addition, the character $\chi_Q\psi$ cannot be the principal character, since then χ_Q would be induced by ψ, which has modulus $q' < Q$ (since $Q(x)$ is assumed to tend to infinity with x), contradicting the assumption that χ_Q is primitive. The number Q^* is not necessarily cube-free, but we can apply a slight generalization of the Burgess bound from [**19**, formula (12.56)] to bound the left-hand side of (4.17) with

$$\ll_{r,\varepsilon} \left(\frac{x}{d_0}\right)^{1-(1/r)} q^{1/r}(Q^*)^{((r+1)/4r^2)+\varepsilon^2} = o(x)$$

for $r = 10\lfloor \varepsilon^{-2} \rfloor$, say. Now the first part of the theorem has been proved.

For the proof of (1.13), note that the quantity on the left-hand side of that formula is

$$\begin{aligned}
&\frac{1}{\log x} \sum_{\substack{n \leqslant x \\ (n(n+1), Q)=1}} \frac{1}{n} \cdot \frac{1-\chi_Q(n)}{2} \cdot \frac{1-\chi_Q(n+1)}{2} \\
&\quad = \frac{1}{4\log x} \sum_{\substack{n \leqslant x \\ (n(n+1), Q)=1}} \frac{1}{n} - \frac{1}{4\log x} \sum_{n \leqslant x} \frac{\chi_Q(n)\chi_0(n+1)}{n} \\
&\qquad - \frac{1}{4\log x} \sum_{n \leqslant x} \frac{\chi_0(n)\chi_Q(n+1)}{n} + \frac{1}{4\log x} \sum_{n \leqslant x} \frac{\chi_Q(n)\chi_Q(n+1)}{n},
\end{aligned}$$

where χ_0 stands for the principal character $\pmod{Q}$. Here the first term equals the right-hand side of (1.13) by elementary sieve theory. The other three terms are seen to be $o(1)$ just as in the first part of the theorem (in order to apply Theorem 1.4, it suffices that one of χ_Q and χ_0 is uniformly distributed in arithmetic progressions).

For the last part of the theorem, namely proving (1.14), we apply the same argument as in the second part to show that

$$\frac{1}{\log\omega(x)}\sum_{x/\omega(x)\leqslant n\leqslant x}\frac{1_{n,n+1\text{ QNR }\pmod Q}}{n}=\frac{1}{4}\prod_{p\mid Q}\left(1-\frac{2}{p}\right)+o(1).$$

Since $\omega(X)$ is any function tending to infinity slowly, we can apply exactly the same argument as at the end of the proof of Theorem 1.11 to conclude that (1.14) holds. □

Acknowledgements

The author is grateful to Kaisa Matomäki and Terence Tao for various useful comments and discussions. He thanks the referee for careful reading of the paper and for useful comments. The author also thanks Zhiwei Wang for showing his preprint [**36**] to the author. Thanks go also to Alexander Mangerel for a discussion on character sums. Part of this work was done while the author was visiting the Mathematical Sciences Research Institute (funded by NSF grant DMS-1440140) in spring 2017, and he thanks the institute for a stimulating atmosphere. The author was funded by UTUGS Graduate School and project number 293876 of the Academy of Finland.

Appendix A. Stability of mean values of multiplicative functions

We prove Lemma 2.2, which was used in the proof of Theorem 1.4 and tells that mean values of the functions g_j over the arithmetic progression $a \pmod q$ vary very slowly in terms of the interval over which the mean value is taken. The case $q = 1$ of the lemma was proved by Elliott [**6**] and refined by Granville and Soundararajan [**13**, Proposition 4.1] (see also [**14**, Theorem 4]). Also Matthiesen's work [**25**] contains estimates of the type of Lemma 2.2, but for the sake of completeness we give a proof here. We have not aimed to optimize the error terms in the lemma.

Proof of Lemma 2.2. By writing

$$\frac{1}{x}\sum_{\substack{x\leqslant n\leqslant 2x\\ n\equiv a\pmod q}}g(n)=\frac{1}{x}\sum_{\substack{n\leqslant 2x\\ n\equiv a\pmod q}}g(n)-\frac{1}{x}\sum_{\substack{n\leqslant x\\ n\equiv a\pmod q}}g(n)$$

and the same with x/y in place of x, we see that it suffices to show that

$$\left|\frac{1}{x}\sum_{\substack{n\leqslant x\\ n\equiv a \pmod q}} g(n) - \frac{1}{x/y}\sum_{\substack{n\leqslant x/y\\ n\equiv a \pmod q}} g(n)\right| \ll_q (\log x)^{-1/400}. \tag{A.1}$$

for $y \in [1, 2\log^{10} x]$. Putting $d_0 := (a, q)$, $a' := a/d_0$ and $q' := q/d_0$, (A.1) becomes

$$\left|\frac{1}{x}\sum_{\substack{n'\leqslant x/d_0\\ n'\equiv a' \pmod{q'}}} g(d_0n') - \frac{1}{x/y}\sum_{\substack{n'\leqslant x/d_0y\\ n'\equiv a' \pmod{q'}}} g(d_0n')\right| \ll_q (\log x)^{-1/400}.$$

Making use of the orthogonality of Dirichlet characters and the triangle inequality, it suffices to show that

$$\left|\frac{1}{x}\sum_{n'\leqslant x/d_0} g(d_0n')\chi(n') - \frac{1}{x/y}\sum_{n'\leqslant x/d_0y} g(d_0n')\chi(n')\right| \ll_q (\log x)^{-1/400}.$$

for all Dirichlet characters $\chi \pmod{q'}$. Writing $n' = rm$, where $(m, d_0) = 1$ and $r \mid d_0^\infty$ (meaning that $r \mid d_0^k$ for some k), and using the fact that $g(d_0rm) = g(d_0r)g(m)$, the previous bound will follow from

$$\sum_{r\mid d_0^\infty}\frac{1}{d_0r}\left|\frac{d_0r}{x}\sum_{\substack{m\leqslant x/d_0r\\ (m,d_0)=1}} g(m)\chi(m) - \frac{d_0r}{x/y}\sum_{\substack{m\leqslant x/d_0ry\\ (m,d_0)=1}} g(m)\chi(m)\right| \ll_q (\log x)^{-1/400}. \tag{A.2}$$

The terms $r > \log x$ can be discarded, since

$$\sum_{\substack{r\mid d_0^\infty\\ r>\log x}}\frac{1}{rd_0} \leqslant (\log x)^{-1/2}\frac{1}{d_0}\prod_{p\mid d_0}\left(1+\frac{1}{p^{1/2}}+\frac{1}{p}+\cdots\right) \ll_q (\log x)^{-1/2},$$

since $d_0 \leqslant q$. Writing $x' := x/d_0r \gg_q x/\log x$ and applying the triangle inequality to (A.2), together with the simple fact that $\sum_{r\mid d_0^\infty} 1/r \ll_q 1$, it suffices to show that

$$\left|\frac{1}{x'}\sum_{m\leqslant x'} g(m)\chi(m)\psi_0(m) - \frac{1}{x'/y'}\sum_{m\leqslant x'/y} g(m)\chi(m)\psi_0(m)\right| \ll_q (\log x)^{-1/400} \tag{A.3}$$

for all $x/\log x \ll_q x' \leqslant x$, $1 \leqslant y' \leqslant 2\log^{10} x$ and for all characters $\chi \pmod q$, with ψ_0 the principal character $\pmod{d_0}$. Note that

$$\mathbb{D}(g\chi\psi_0, f; x') = \mathbb{D}(g\chi, f; x') - O_q(1) \tag{A.4}$$

for any function $f : \mathbb{N} \to \mathbb{D}$, so we may replace ψ_0 with 1 in any computations involving the pretentious distance.

Consider the character $\chi \pmod q$ for which the left-hand side of (A.3) is maximal. If χ is complex, we may apply an argument of Granville and Soundararajan (see [**23**, Lemma C.1]) and the assumption that g is real-valued to obtain the pretentious distance bound

$$\begin{aligned}\sqrt{M} &:= \inf_{|t|\leqslant x} \mathbb{D}\left(g\chi\psi_0, n^{it}; \frac{x'}{y'}\right) = \inf_{|t|\leqslant x} \mathbb{D}(g\chi, n^{it}; x) - O_q(\log_3 x)\\ &\geqslant \frac{1}{10}\sqrt{\log\log x}\end{aligned}$$

by (A.4), since q is fixed and x is large enough. Thus by Halász's theorem [**34**, Ch. III.4], we may bound (A.3) by $\ll Me^{-M} \ll (\log x)^{-1/200}$.

In the opposite case that χ is real in (A.3), we appeal to [**13**, Proposition 4.1], provided that $\mathbb{D}(g\chi\psi_0; 1; x) \leqslant \frac{2}{3}\sqrt{\log\log x}$ holds. This gives a bound of $\ll (\log x)^{-1/10}$ for (A.3), so we may assume that $\mathbb{D}(g\chi\psi_0; 1; x) > \frac{2}{3}\sqrt{\log\log x}$. But since $g\chi$ is real-valued, again by [**23**, Lemma C.1] we have

$$\begin{aligned}\sqrt{M} &:= \inf_{|t|\leqslant x} \mathbb{D}\left(g\chi\psi_0, n^{it}; \frac{x'}{y'}\right) = \inf_{|t|\leqslant x} \mathbb{D}(g\chi, n^{it}; x) - O_q(\log_3 x)\\ &\geqslant \frac{1}{10}\mathbb{D}(g\chi, 1; x) - O_q(\log_3 x) \geqslant \frac{1}{16}\sqrt{\log\log x}\end{aligned}$$

for all large enough x. Now, again applying Halász's theorem, (A.3) is bounded by $\ll Me^{-M} \ll (\log x)^{-1/400}$. This shows that (A.3) always holds, which proves the lemma. □

Appendix B. Short exponential sum bounds for multiplicative functions

We prove the short exponential sum estimates over major and minor arcs that were employed in the proof of Theorem 1.4 in Section 3. The proofs of both lemmas follow the ideas of Matomäki, Radziwiłł and Tao [**23**] for estimating short exponential sum bounds weighted by a multiplicative function, but require some small modifications to the arguments.

Proof of Lemma 3.7. Since $\alpha \in \mathfrak{m}$, we have the trivial estimate

$$\Big|\sum_{y\leqslant n\leqslant y+H} e(\alpha n)\Big| \ll \frac{1}{\|\alpha\|} \ll \frac{H}{W} \ll H(\log H)^{-1/10},$$

so by the triangle inequality we may assume that $\delta_1 = 0$ in (3.16). We introduce the same nicely factorable set $\mathcal{S} := \mathcal{S}_{P_1,Q_1,X_0,X}$ as in (3.6). By a simple sieve estimate [**23**, Lemma 2.3], we have

$$\sum_{n\leqslant X+H} (1 - 1_{\mathcal{S}}(n)) \ll \frac{\log\log H}{\log H} X. \tag{B.1}$$

Hence, by the triangle inequality,

$$\frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{y\leqslant n\leqslant y+H} g_1(n)(1 - 1_{\mathcal{S}}(n))e(\alpha n)\Big| dy \ll \frac{\log\log H}{\log H} \ll (\log H)^{-1/10}.$$

This means that (3.16) has been reduced to

$$\sup_{\alpha\in\mathfrak{m}} \frac{1}{X}\int_X^{2X} \Big|\frac{1}{H}\sum_{y\leqslant n\leqslant y+H} g_1(n)1_{\mathcal{S}}(n)e(\alpha n)\Big| dy \ll (\log H)^{-1/10}. \tag{B.2}$$

This estimate would follow directly from [**23**, Section 3] (with $d = 1$ there), if the function g_1 was completely multiplicative, but we show that the argument goes through even without that assumption.

Let $\mathcal{S}'$ be the set of those $n \leqslant X$ that have a prime factor from each of the intervals $[P_j, Q_j]$ (defined in Definition 3.3) for $j \geqslant 2$. We have the Ramaré identity

$$\begin{aligned} g_1(n)1_{\mathcal{S}}(n) &= \sum_{\substack{n=mp\in\mathcal{S}\\ P_1\leqslant p\leqslant Q_1}} \frac{g_1(mp)1_{\mathcal{S}}(mp)}{|\{P_1 \leqslant p_1 \leqslant Q_1 : p_1 \mid n\}|} \\ &= \sum_{\substack{n=mp\\ p\nmid m\\ P_1\leqslant p\leqslant Q_1}} \frac{g_1(m)g_1(p)1_{\mathcal{S}'}(m)}{1+|\{P_1 \leqslant p_1 \leqslant Q_1 : p_1 \mid m\}|} + O\Big(\sum_{P_1\leqslant p\leqslant Q_1} 1_{p^2\mid n}\Big) \\ &= \sum_{\substack{n=mp\\ P_1\leqslant p\leqslant Q_1}} \frac{g_1(m)g_1(p)1_{\mathcal{S}'}(m)}{1+|\{P_1 \leqslant p_1 \leqslant Q_1 : p_1 \mid m\}|} + O\Big(\sum_{P_1\leqslant p\leqslant Q_1} 1_{p^2\mid n}\Big). \end{aligned}$$

By trivial estimation,

$$\sum_{n\leqslant X+H}\sum_{P_1\leqslant p\leqslant Q_1} 1_{p^2\mid n} \ll \sum_{p\geqslant P_1} \frac{X}{p^2} \ll XW^{-200} \ll X(\log H)^{-1/10},$$

so proving (B.2) has been reduced to proving

$$\sum_{P_1\leqslant p\leqslant Q_1}\sum_m \frac{1_{\mathcal{S}'}(m)g_1(m)g_1(p)e(mp\alpha)}{1+|\{P_1\leqslant p_1\leqslant Q_1 : p_1\mid m\}|}\int_{\mathbb{R}}\theta(x)1_{x\leqslant mp\leqslant x+H}\,dx \ll HX(\log H)^{-1/10}$$

for all measurable functions $|\theta(x)|\leqslant 1$ supported on $[0,X]$. This is same expression as in [**23**, Section 3], so the proof continues from here in an identical manner (since the rest of the argument does not use multiplicativity). □

Proof of Lemma 3.6. We follow the proof of the major arc exponential sum in [**23**, Section 4]. However, here we need to be a bit more careful when approximating the exponential $e(\alpha n)$ with $e(an/q)$, as we do not want to lose a factor of W/q that would come from a partial summation approximation of $e(\alpha n)$.

By our assumption $g_1\in\mathcal{U}_\omega(x,\exp_2(\varepsilon^{-2}),2/\exp_2(\varepsilon^{-2}),\delta_1)$ and formula (B.1), it suffices to show that

$$\frac{1}{X}\int_X^{2X}\left|\frac{1}{H}\sum_{x\leqslant n\leqslant x+H}(g_1(n)1_{\mathcal{S}}(n)-\delta_1')e(\alpha n)\right|dx = o_{\varepsilon\to 0}(1), \tag{B.3}$$

where the nicely factorable set $\mathcal{S}$ is as in (3.6) and

$$\delta_1' := \frac{1}{X}\sum_{X\leqslant m\leqslant 2X} g_1(n)1_{\mathcal{S}}(n).$$

Let $H' := H/W^3$. By exchanging the order of integration and summation, we have

$$\frac{1}{H}\sum_{x\leqslant n\leqslant x+H}a_n = \frac{1}{H}\int_x^{x+H}\frac{1}{H'}\sum_{y\leqslant n\leqslant y+H'}a_n\,dy + O\left(\frac{H'}{H}\right)$$

for any $a_n\in\mathbb{D}$. Applying this, we see that the left-hand side of (B.3) is

$$= \frac{1}{X}\int_X^{2X}\left|\frac{1}{H}\int_x^{x+H}\frac{1}{H'}\sum_{y\leqslant n\leqslant y+H'}(g_1(n)1_{\mathcal{S}}(n)-\delta_1')e(\alpha n)\,dy\right|dx + O\left(\frac{1}{W^2}\right)$$

$$\ll \frac{1}{HX}\int_X^{2X}\int_x^{x+H}\left|\frac{1}{H'}\sum_{y\leqslant n\leqslant y+H'}(g_1(n)1_{\mathcal{S}}(n)-\delta_1')e\left(\frac{an}{q}\right)\right|dy\,dx + O\left(\frac{1}{W^2}\right),$$

where we used the fact that any $n\in[y,y+H']$ obeys

$$\begin{aligned} e(\alpha n) &= e(\alpha y)\, e(\alpha(n-y)) \\ &= e(\alpha y)\, e\Big(\frac{a}{q}(n-y)\Big) + O\Big(\frac{1}{W^2}\Big) \\ &= e\Big(\Big(\alpha - \frac{a}{q}\Big)y\Big) e\Big(\frac{an}{q}\Big) + O\Big(\frac{1}{W^2}\Big) \end{aligned}$$

by the inequality

$$\Big|e(\alpha(n-y)) - e\Big(\frac{a}{q}(n-y)\Big)\Big| \leqslant 2\pi\Big|\alpha - \frac{a}{q}\Big||n-y| \leqslant \frac{2\pi W}{qH}\cdot H' \leqslant \frac{2\pi}{W^2}.$$

By exchanging the order of integration above, it suffices to show that

$$\frac{1}{X}\int_X^{2X}\Big|\frac{1}{H'}\sum_{x\leqslant n\leqslant x+H'}(g_1(n)1_{\mathcal{S}}(n) - \delta_1')e\Big(\frac{an}{q}\Big)\Big|\,dx \ll \varepsilon. \tag{B.4}$$

for all $1 \leqslant a \leqslant q \leqslant W = \log^5 H$, and with $H' = H/W^3$, as before. By splitting into residue classes $\pmod q$, (B.4) would follow from

$$\frac{1}{X}\int_X^{2X}\Big|\frac{1}{H'}\sum_{\substack{x\leqslant n\leqslant x+H'\\ n\equiv b \pmod q}} g_1(n)1_{\mathcal{S}}(n) - \frac{1}{qX}\sum_{X\leqslant n\leqslant 2X} g_1(n)1_{\mathcal{S}}(n)\Big|\,dx \ll \frac{\varepsilon}{q} \tag{B.5}$$

for all $1 \leqslant a \leqslant q \leqslant W$. Applying the triangle inequality and Lemma 3.4 (and the fact that $q \leqslant W$), it suffices to show that

$$\Big|\frac{1}{X}\sum_{\substack{X\leqslant n\leqslant 2X\\ n\equiv b \pmod q}} g_1(n)1_{\mathcal{S}}(n) - \frac{1}{qX}\sum_{X\leqslant n\leqslant 2X} g_1(n)1_{\mathcal{S}}(n)\Big| \ll \frac{\varepsilon}{q}. \tag{B.6}$$

As in [**23**, Section 2], the fundamental lemma of sieve theory gives for $q \leqslant W$ the estimate

$$\sum_{\substack{X\leqslant n\leqslant 2X\\ n\equiv a \pmod q}} (1 - 1_{\mathcal{S}}(n)) \ll \frac{X}{q}\cdot\frac{\log\log H}{\log H} \ll \frac{\varepsilon}{q}X.$$

Taking this into account on both sides of (B.6), that claim is reduced to

$$\Big|\frac{1}{X}\sum_{\substack{X\leqslant n\leqslant 2X\\ n\equiv b \pmod q}} g_1(n) - \frac{1}{qX}\sum_{X\leqslant n\leqslant 2X} g_1(n)\Big| \ll \frac{\varepsilon}{q}$$

for $X \in [x/\omega(x), x]$, and this follows immediately from our uniform distribution assumption $g_1 \in \mathcal{U}_\omega(x, \exp_2(\varepsilon^{-2}), 2/\exp_2(\varepsilon^{-2}), \delta_1)$ and the fact that $q \leqslant \log^5 H \leqslant \exp_2(10\varepsilon^{-1})$. The proof is complete. □

References

[1] D. A. Burgess, 'On character sums and L-series. II', *Proc. Lond. Math. Soc. (3)* **13** (1963), 524–536.

[2] D. A. Burgess, 'On Dirichlet characters of polynomials', *Proc. Lond. Math. Soc. (3)* **13** (1963), 537–548.

[3] H. Daboussi and A. Sárközy, 'On the correlation of the truncated Liouville function', *Acta Arith.* **108**(1) (2003), 61–76.

[4] J.-M. De Koninck and N. Doyon, 'On the distance between smooth numbers', *Integers* **11**(A25) (2011), 22.

[5] R. de la Bretèche, C. Pomerance and G. Tenenbaum, 'Products of ratios of consecutive integers', *Ramanujan J.* **9**(1–2) (2005), 131–138.

[6] P. D. T. A. Elliott, 'Extrapolating the mean-values of multiplicative functions', *Nederl. Akad. Wetensch. Indag. Math.* **51**(4) (1989), 409–420.

[7] P. D. T. A. Elliott, 'On the correlation of multiplicative functions', *Notas Soc. Math. Chile* **11**(1) (1992), 1–11.

[8] P. D. T. A. Elliott, 'On the correlation of multiplicative and the sum of additive arithmetic functions', *Mem. Amer. Math. Soc.* **112**(538) (1994), viii+88.

[9] P. Erdős, 'Some unconventional problems in number theory', in *Journées Arithmétiques de Luminy (Colloq. Internat. CNRS, Centre Univ. Luminy, Luminy, 1978)*, Astérisque, 61 (Soc. Math. France, Paris, 1979), 73–82.

[10] P. Erdős and C. Pomerance, 'On the largest prime factors of n and $n + 1$', *Aequationes Math.* **17**(2–3) (1978), 311–321.

[11] N. Frantzikinakis, 'An averaged Chowla and Elliott conjecture along independent polynomials', *Int. Math. Res. Not. IMRN* **2018**(12) (2018), 3721–3743.

[12] N. Frantzikinakis, 'Ergodicity of the Liouville system implies the Chowla conjecture', *Discrete Anal.* **19** (2017), 41pp.

[13] A. Granville and K. Soundararajan, 'The spectrum of multiplicative functions', *Ann. of Math. (2)* **153**(2) (2001), 407–470.

[14] A. Granville and K. Soundararajan, 'Decay of mean values of multiplicative functions', *Canad. J. Math.* **55**(6) (2003), 1191–1230.

[15] G. Halász, 'Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen', *Acta Math. Acad. Sci. Hungar.* **19** (1968), 365–403.

[16] R. R. Hall, 'A sharp inequality of Halász type for the mean value of a multiplicative arithmetic function', *Mathematika* **42**(1) (1995), 144–157.

[17] A. Hildebrand, 'On a conjecture of Balog', *Proc. Amer. Math. Soc.* **95**(4) (1985), 517–523.

[18] A. Hildebrand and G. Tenenbaum, 'Integers without large prime factors', *J. Théor. Nombres Bordeaux* **5**(2) (1993), 411–484.

[19] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, American Mathematical Society Colloquium Publications, 53 (American Mathematical Society, Providence, RI, 2004).

[20] O. Klurman, 'Correlations of multiplicative functions and applications', *Compos. Math.* **153**(8) (2017), 1622–1657.

[21] A. P. Mangerel, On the Bivariate Erdős–Kac Theorem and Correlations of the Möbius Function, ArXiv e-prints, 2016.

[22] K. Matomäki and M. Radziwiłł, 'Multiplicative functions in short intervals', *Ann. of Math. (2)* **183**(3) (2016), 1015–1056.

[23] K. Matomäki, M. Radziwiłł and T. Tao, 'An averaged form of Chowla's conjecture', *Algebra Number Theory* **9**(9) (2015), 2167–2196.

[24] K. Matomäki, M. Radziwiłł and T. Tao, 'Sign patterns of the Liouville and Möbius functions', *Forum Math. Sigma* **4**(e14) (2016), 44.
[25] L. Matthiesen, 'Generalized Fourier coefficients of multiplicative functions', *Algebra Number Theory*, to appear.
[26] P. Pollack, 'Arithmetic properties of polynomial specializations over finite fields', *Acta Arith.* **136**(1) (2009), 57–79.
[27] I. E. Shparlinski, 'Open problems on exponential and character sums', in *Number Theory*, Series on Number Theory and its Applications, 6 (World Scientific Publishing, Hackensack, NJ, 2010), 222–242.
[28] V. T. Sós, 'Turbulent years: Erdős in his correspondence with Turán from 1934 to 1940', in *Paul Erdős and his Mathematics, I (Budapest, 1999)*, Bolyai Soc. Math. Stud., 11 (János Bolyai Math. Soc., Budapest, 2002), 85–146.
[29] T. Tao, A small remark on the Elliott conjecture. http://terrytao.wordpress.com/2015/09/18.
[30] T. Tao, 'The logarithmically averaged Chowla and Elliott conjectures for two-point correlations', *Forum Math. Pi* **4**(e8) (2016), 36.
[31] T. Tao, 'Equivalence of the logarithmically averaged Chowla and Sarnak conjectures', in *Number Theory—Diophantine Problems, Uniform Distribution and Applications* (Springer, Cham, 2017), 391–421.
[32] T. Tao and J. Teräväinen, 'Odd order cases of the logarithmically averaged Chowla conjecture', *J. Théor. Nombres Bordeaux*, to appear.
[33] T. Tao and J. Teräväinen, The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures, ArXiv e-prints, 2017.
[34] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, 3rd edn, Graduate Studies in Mathematics, 163 (American Mathematical Society, Providence, RI, 2015), Translated from the 2008 French edition by Patrick D. F. Ion.
[35] Z. Wang, 'On the largest prime factors of consecutive integers in short intervals', *Proc. Amer. Math. Soc.* **145**(8) (2017), 3211–3220.
[36] Z. Wang, 'Sur les plus grands facteurs premiers d'entiers consécutifs', *Mathematika* **64**(2) (2018), 343–379.

Annales Universitatis Turkuensis

Turun yliopisto
University of Turku

Painosalama Oy, Turku, Finland 2018

ISBN 978-951-29-7339-2 (PRINT)
ISBN 978-951-29-7340-8 (PDF)
ISSN 0082-7002 (PRINT) | ISSN 2343-3175 (PDF)