Received: 15 September 2021 | Revised: 23 May 2022 | Accepted: 6 June 2022

DOI: 10.1112/jlms.12663

RESEARCH ARTICLE

Journal of the London Mathematical Society

The Hardy–Littlewood–Chowla conjecture in the presence of a Siegel zero

Terence Tao[1] | **Joni Teräväinen**[2,3]

[1]Department of Mathematics, University of California, Los Angeles, California, USA

[2]Mathematical Institute, University of Oxford, Oxford, UK

[3]Department of Mathematics and Statistics, University of Turku, Turku, Finland

Correspondence
Joni Teräväinen, Mathematical Institute, University of Oxford, Oxford, UK.
Email: joni.p.teravainen@gmail.com

[Correction added on 24 Nov 2022, after first online publication: CAUL funding statement has been added.]

Funding information
Simons Investigator; James and Carol Collins Chair; Mathematical Analysis & Application Research Fund Endowment; NSF, Grant/Award Number: DMS-1764034; Titchmarsh Fellowship; Academy of Finland, Grant/Award Number: 340098

Abstract

Assuming that Siegel zeros exist, we prove a hybrid version of the Chowla and Hardy–Littlewood prime tuples conjectures. Thus, for an infinite sequence of natural numbers x, and any distinct integers $h_1, \dots, h_k, h'_1, \dots, h'_\ell$, we establish an asymptotic formula for

$$\sum_{n \leqslant x} \Lambda(n+h_1) \cdots \Lambda(n+h_k) \lambda(n+h'_1) \cdots \lambda(n+h'_\ell)$$

for any $0 \leqslant k \leqslant 2$ and $\ell \geqslant 0$. Specializing to either $\ell = 0$ or $k = 0$, we deduce the previously known results on the Hardy–Littlewood (or twin primes) conjecture and the Chowla conjecture under the existence of Siegel zeros, due to Heath-Brown and Chinis, respectively. The range of validity of our asymptotic formula is wider than in these previous results.

MSC 2020
11N37 (primary), 11N36 (secondary)

1 | INTRODUCTION

1.1 | The Hardy–Littlewood–Chowla conjecture and Siegel zeroes

Let $\lambda : \mathbb{N} \to \{-1, +1\}$ denote the Liouville function. We have the following well known conjecture of Chowla [3]:

Conjecture 1.1 (Chowla's conjecture). Let $h_1', \dots, h_\ell'$ be distinct fixed natural numbers for some fixed $\ell \geqslant 1$. Then[†]

$$\mathbb{E}_{n \leqslant x} \lambda(n + h_1') \cdots \lambda(n + h_\ell') = o(1)$$

as $x \to \infty$.

Here and in the sequel, n is understood to range over natural numbers, and we use the averaging notation $\mathbb{E}_{n \in A} f(n) := \frac{1}{|A|} \sum_{n \in A} f(n)$ for any set A of a finite cardinality $|A|$. The reasons for the primes in the notation $h_1', \dots, h_\ell'$ is for compatibility with Conjecture 1.3.

For $\ell = 1$, Chowla's conjecture is equivalent to the prime number theorem, but the conjecture is open for all $\ell \geqslant 2$, although a slightly weaker 'logarithmically averaged' conjecture is known to hold for $\ell = 2$ [27] or for odd ℓ [28, 29]. All the discussion here concerning the Liouville function λ has a counterpart for the Möbius function μ, but for simplicity of exposition we restrict attention to the Liouville function here.

The analogous conjecture for the von Mangoldt function $\Lambda : \mathbb{N} \to \mathbb{R}^+$ is the well-known *prime tuples conjecture* of Hardy and Littlewood [10]:

Conjecture 1.2 (Hardy–Littlewood prime tuples conjecture). Let $h_1, \dots, h_k$ be distinct fixed natural numbers for some fixed $k \geqslant 0$. Then

$$\mathbb{E}_{n \leqslant x} \Lambda(n + h_1) \cdots \Lambda(n + h_k) = \mathfrak{S} + o(1)$$

as $x \to \infty$, where the *singular series* $\mathfrak{S}$ is defined by the formula

$$\mathfrak{S} := \prod_p \beta_p, \tag{1.1}$$

the local factors β_p are defined by

$$\beta_p := \mathbb{E}_{n \in \mathbb{Z}/p\mathbb{Z}} \Lambda_p(n + h_1) \cdots \Lambda_p(n + h_k) = \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{|\{h_1, \dots, h_k\} \pmod p|}{p}\right) \tag{1.2}$$

and $\Lambda_p : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}^+$ is the local von Mangoldt function $\Lambda_p(n) := \frac{p}{p-1} 1_{n \neq 0\,(p)}$. (In this paper, we adopt the convention that the empty product is equal to 1.)

It is not difficult to show the asymptotic

$$\beta_p = 1 + O\left(\frac{1}{p^2}\right), \tag{1.3}$$

so the product in (1.1) converges, though it could vanish if the $h_1, \dots, h_k$ cover a complete set of residues modulo p for some prime p. Conjecture 1.2 is trivial for $k = 0$ and equivalent to the prime number theorem for $k = 1$, but is open for all other values of k, with the $k = 2$ case already implying the notorious twin prime conjecture.

[†] See Section 2 for our conventions on asymptotic notation.

It is natural to unify Conjectures 1.1 and 1.2 as follows.

Conjecture 1.3 (Hardy–Littlewood–Chowla conjecture). Let $k, \ell \geqslant 0$, and let $h_1, \dots, h_k, h'_1, \dots, h'_\ell$ be distinct fixed natural numbers. Then

$$\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda(n+h'_1)\cdots\lambda(n+h'_\ell) = \mathfrak{S} + o(1)$$

as $x \to \infty$, where $\mathfrak{S}$ is defined by (1.1) when $\ell = 0$ and is equal to 0 otherwise.

Clearly Conjectures 1.1 and 1.2 correspond to the special cases $k = 0$ and $\ell = 0$, respectively, of Conjecture 1.3. One could also generalize this conjecture by replacing the forms $n + h_j, n + h'_{j'}$ by more general linear forms $a_j n + b_j, a'_{j'} n + b'_{j'}$, no two of which are scalar multiples of each other, but we do not do so here in order to simplify the notation.

Only the $k + \ell \leqslant 1$ cases of Conjecture 1.3 are currently known, even if one assumes the generalized Riemann hypothesis, though see [26] for some recent progress in the function field case, and the recent works [15], [16] for some progress on an averaged version of this conjecture. On the other hand, it turns out (perhaps surprisingly) that some progress on this conjecture can be made under an opposing hypothesis, namely the existence of a Siegel zero. We use the notational conventions from Heath-Brown's work [11]:

Definition 1.4 (Siegel zero). A *Siegel zero* β is a real number associated to a primitive quadratic Dirichlet character χ of conductor q_χ such that $L(\beta, \chi) = 0$ and

$$\beta = 1 - \frac{1}{\eta \log q_\chi}$$

for some $\eta \geqslant 10$ (which we call the *quality* of the zero).

The lower bound on η is mostly in order to ensure that $\log\log\eta$ is positive; the precise numerical value of the lower bound is not important. From Siegel's theorem, we have the (ineffective) upper bound

$$\eta \ll_\varepsilon q_\chi^\varepsilon \tag{1.4}$$

on the quality of a Siegel zero for any $\varepsilon > 0$.

There are prior results in the literature towards Conjecture 1.3 in the presence of a Siegel zero when only either the von Mangoldt function or the Liouville function appears in the correlation. These results are due to Heath-Brown [11] in the case of two-point correlations of the von Mangoldt function, and due to Chinis [2] in the case of the Chowla conjecture (with previous work by Germán and Katái [6] on the two-point case). We can summarize them as follows:

Theorem 1.5 (Prior results on Hardy–Littlewood–Chowla given a Siegel zero). *Suppose that one has a Siegel zero β with associated conductor q_χ and quality η.*

(i) [11, Theorem 1] *For any distinct fixed natural numbers h_1, h_2, one has*

$$\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\Lambda(n+h_2) = \mathfrak{S} + O\left(\frac{1}{\log\log\eta}\right)$$

uniformly for all $q_\chi^{250} \leqslant x \leqslant q_\chi^{300}$, where $\mathfrak{S}$ is defined by (1.1).

(ii) [6, Theorem 2] *One has*

$$\mathbb{E}_{n\leqslant x}\lambda(n)\lambda(n+1) \ll \frac{1}{\log\log\eta} + \epsilon(x)$$

for $q_\chi^{10} \leqslant x \leqslant q_\chi^{(\log\log\eta)/3}$, *where* $\epsilon(x)$ *is a quantity that goes to zero as* $x \to \infty$ *(uniformly in the choice of Siegel zero).*

(iii) [2, Theorem 1.2] *For any distinct fixed natural numbers* $h_1', \dots, h_\ell'$, *one has*

$$\mathbb{E}_{n\leqslant x}\lambda(n+h_1')\cdots\lambda(n+h_\ell') \ll \frac{1}{(\log\log\eta)^{1/2}\log^{1/12}\eta}$$

for $q_\chi^{10} \leqslant x \leqslant q_\chi^{(\log\log\eta)/3}$.

The main result of this paper is the following common generalization and strengthening of these results.

Theorem 1.6 (New results on Hardy–Littlewood–Chowla given a Siegel zero). *Let* $0 \leqslant k \leqslant 2$ *and* $\ell \geqslant 0$, *and let* $h_1, \dots, h_k$, $h_1', \dots, h_\ell'$ *be fixed distinct natural numbers. Suppose that one has a Siegel zero* β *with associated conductor* q_χ *and quality* η. *Let* $0 < \varepsilon_0 < 1$ *be fixed, and let* x *lie in the range*

$$q_\chi^{10k+\frac{1}{2}+\varepsilon_0} \leqslant x \leqslant q_\chi^{\eta^{1/2}}. \tag{1.5}$$

Then we have

$$\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda(n+h_1')\cdots\lambda(n+h_\ell') = \mathfrak{S} + O\left(\frac{1}{\log^{\frac{1}{10\max(1,k)}}\eta}\right), \tag{1.6}$$

where $\mathfrak{S}$ *is as in Conjecture 1.3.*

Remark 1.7. The k-dependent exponent of $10k$ in the range (1.5) can be improved somewhat, particularly when $k = 1$, but we will not attempt to optimize it here. On the other hand, in order to improve the exponent $\frac{1}{2}$ in (1.5) in the case $k = 0$, it seems necessary to be able to obtain non-trivial bounds on short character sums such as

$$\sum_{n\in I}\chi(n+h_1')\cdots\chi(n+h_\ell')$$

for intervals I of length less than $q_\chi^{1/2}$, which is beyond the range of direct application of the Weil bounds and completion of sums (and for $\ell > 1$ we were not able to adapt the Burgess argument [1] to such sums due to the lack of multiplicative structure). The exponent $\frac{1}{10\max(1,k)}$ in (1.6) can similarly be improved, but we will not attempt to do so here.

Note that Theorem 1.6 improves the dependence on the quality η of Siegel zero, and also allows for correlations that involve both the von Mangoldt function Λ and the Liouville function λ, so

long as the former function appears at most two times. This latter restriction is an inherent limitation of our current state of knowledge of correlations for functions like the divisor function $\tau := 1 * 1$; in particular, k-point correlations $\mathbb{E}_{n\leqslant x}\tau(n+h_1)\cdots\tau(n+h_k)$ are currently only well understood when $k \leqslant 2$.

As a direct corollary to Theorem 1.6, we can state the following strengthening of previous results.

Corollary 1.8. *Suppose that one has a Siegel zero β with associated conductor q_χ and quality η. Let $0 < \varepsilon_0 < 1$ be fixed.*

(i) *For any distinct fixed natural numbers h_1, h_2, one has*

$$\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\Lambda(n+h_2) = \mathfrak{S} + O\left(\frac{1}{\log^{1/20}\eta}\right)$$

uniformly for all $q_\chi^{41/2+\varepsilon_0} \leqslant x \leqslant q_\chi^{\eta^{1/2}}$, where $\mathfrak{S}$ is defined by (1.1).

(ii) *For any distinct fixed natural numbers $h_1', \dots, h_\ell'$, one has*

$$\mathbb{E}_{n\leqslant x}\lambda(n+h_1')\cdots\lambda(n+h_\ell') \ll \frac{1}{\log^{1/10}\eta}$$

uniformly for all $q_\chi^{1/2+\varepsilon_0} \leqslant x \leqslant q_\chi^{\eta^{1/2}}$.

(iii) *For any fixed integer $h \neq 0$, one has*

$$\mathbb{E}_{|h|<p\leqslant x}\lambda(p+h) \ll \frac{1}{\log^{1/10}\eta}$$

uniformly for all $q_\chi^{21/2+\varepsilon_0} \leqslant x \leqslant q_\chi^{\eta^{1/2}}$.

Corollary 1.8(ii) can further be applied to strengthen Chinis's result [2, Corollary 1.1] on Sarnak's conjecture on Möbius disjointness being true at infinitely many scales under the assumption of Siegel zeros. Applying Corollary 1.8 and Sarnak's argument for the implication from Chowla's conjecture to Sarnak's conjecture (as in [2]), we see that, under the hypotheses of Corollary 1.8, for any fixed deterministic $f : \mathbb{N} \to \mathbb{C}$, we have

$$\sum_{n\leqslant x}\lambda(n)f(n) = o(x)$$

in the range $q_\chi^{1/2+\varepsilon_0} \leqslant x \leqslant q_\chi^{\eta^{1/2}}$. This improves on the range $q_\chi^{10} \leqslant x \leqslant q_\chi^{\log\log\eta/3}$ in [2].

Corollary 1.8(iii) relates to the conjecture (considered in, for example, [23], [25], [15], [16]) that $\sum_{|h|<p\leqslant x}\lambda(p+h) = o(\pi(x))$, proving it for infinitely many x under the existence of infinitely many Siegel zeros (of arbitrarily high quality).

We lastly note that, after the submission of this paper, Matomäki and Merikoski [17] proved a quantitatively stronger version of Corollary 1.8(i).

1.2 | Overview of proof

The general strategy for proving results such as Theorem 1.6 is now well known: in the presence of a Siegel zero (and for x comparable in log-scale to q_χ), the function λ 'pretends'[†] to be like the Dirichlet character χ, and the von Mangoldt function $\Lambda = \mu * \log$ similarly 'pretends' to be like $\chi * \log$, so the correlation in (1.6) is of comparable complexity to the average

$$\mathbb{E}_{n\leqslant x}(\chi * \log)(n+h_1)\cdots(\chi * \log)(n+h_k)\chi(n+h_1')\cdots\chi(n+h_\ell')$$

(in practice we also have to insert some sieve weights to account for the fact that not all numbers are rough). This is a twisted and weighted version of the divisor correlation

$$\mathbb{E}_{n\leqslant x}\tau(n+h_1)\cdots\tau(n+h_k)$$

which, as previously mentioned, is well understood for $k \leqslant 2$, basically because the Weil bounds for Kloosterman sums ensure that τ has level of distribution at least $2/3$, the key point being that this is larger than $1/2$. The twist by χ introduces the need to estimate character sums such as

$$\mathbb{E}_{n\leqslant x}\chi(n+h_1)\cdots\chi(n+h_k)\chi(n+h_1')\cdots\chi(n+h_\ell'),$$

which can be adequately controlled by the Weil estimates for character sums since we are in the regime $x \gg q_\chi^{1/2}$.

To make this strategy rigorous, we will approximate the functions Λ, λ by a series of more tractable approximants that involve the exceptional character χ (as well as the scale x). We will do this by executing the following steps in order.

(i) Replace the Liouville function λ with an approximant $\lambda_{\mathrm{Siegel}}$, which is a completely multiplicative function that agrees with λ at small primes and agrees with χ at large primes. (This step was also performed in [2, 6].)

(ii) Replace the von Mangoldt function Λ with an approximant $\Lambda_{\mathrm{Siegel}}$, which is the Dirichlet convolution $\chi * \log$ multiplied by a Selberg sieve weight ν to essentially restrict that convolution to almost primes. (This step essentially also appears in [11].)

(iii) Replace $\lambda_{\mathrm{Siegel}}$ with a more complicated truncation $\lambda_{\mathrm{Siegel}}^\sharp$ which has the structure of a 'Type I sum', and which agrees with $\lambda_{\mathrm{Siegel}}$ on numbers that have a 'typical' factorization.

(iv) Replace the approximant $\Lambda_{\mathrm{Siegel}}$ with a more complicated approximant $\Lambda_{\mathrm{Siegel}}^\sharp$ which has the structure of a 'Type I sum'. (This step is inspired by a similar Type I approximation to the divisor function τ (and its higher order generalizations) recently introduced in [18, 19].)

(v) Now that all terms in the correlation have been replaced with tractable Type I sums, use standard Euler product calculations and Fourier analysis, similar in spirit to the proof of the pseudo-randomness of the Selberg sieve majorant for the primes in [9, Appendix D], to evaluate the correlation to high accuracy.

[†] Following [8], we informally say that one arithmetic function f 'pretends' to be another g if they are often close to each other when evaluated at rough numbers.

More succinctly, the proof of Theorem 1.6 proceeds by justifying all of the following approximations:

$$\begin{aligned}
&\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda(n+h_1')\cdots\lambda(n+h_\ell')\\
&\overset{(i)}{\approx}\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell')\\
&\overset{(ii)}{\approx}\mathbb{E}_{n\leqslant x}\Lambda_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda_{\mathrm{Siegel}}(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell')\\
&\overset{(iii)}{\approx}\mathbb{E}_{n\leqslant x}\Lambda_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda_{\mathrm{Siegel}}(n+h_k)\lambda^\sharp_{\mathrm{Siegel}}(n+h_1')\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h_\ell')\\
&\overset{(iv)}{\approx}\mathbb{E}_{n\leqslant x}\Lambda^\sharp_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda^\sharp_{\mathrm{Siegel}}(n+h_k)\lambda^\sharp_{\mathrm{Siegel}}(n+h_1')\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h_\ell')\\
&\overset{(v)}{\approx}\mathfrak{S},
\end{aligned}\tag{1.7}$$

where the precise meaning of the symbol $\approx$ is given in (2.11) below.

The steps (i)–(v) are executed in Sections 4–8, respectively. Interestingly, the hypothesis $k\leqslant 2$ is only used in step (iv) of this process.

Steps (i) and (ii) of the strategy rely ultimately on the well-known phenomenon that in the presence of a Siegel zero, one has $\chi(p)=-1$ for most primes p that are comparable to the conductor q_χ in log-scale. Traditionally, such phenomena are justified using complex-analytic methods, and in particular by exploiting the Deuring–Heilbronn phenomenon. It turns out that an alternate approach relying almost entirely on elementary methods leads instead to significantly superior dependence on the quality η of the zero; see Proposition 3.5. This eventually enables us to obtain a wider x range in Theorem 1.6 than in previous results.

Step (iii) involves splitting $\lambda_{\mathrm{Siegel}}$, which is a kind of character-twisted divisor sum, into two parts as $\lambda^\sharp_{\mathrm{Siegel}}+\lambda^\flat_{\mathrm{Siegel}}$, where $\lambda^\sharp_{\mathrm{Siegel}}$ accounts for the small divisors (with a smooth truncation) and $\lambda^\flat_{\mathrm{Siegel}}$ accounts for the large divisors. It turns out that $\lambda^\flat_{\mathrm{Siegel}}$ has a negligible contribution to the correlation (basically because smooth numbers become extremely rare at large scales). This is shown by first constructing a majorant for $\lambda^\flat_{\mathrm{Siegel}}$ (in Lemma 6.1) that after some Euler product computations is seen to be small 'on average' in a suitable sense.[†]

Steps (iv) and (v) morally speaking amount to computing correlations such as

$$\mathbb{E}_{n\leqslant x,n=a(q)}(\chi*\log(n))(\chi*\log(n+h))\tag{1.8}$$

with power-saving error term (for $1\leqslant a\leqslant q\leqslant x^\delta$ for a small $\delta>0$), as well as correlations of the form

$$\mathbb{E}_{n\leqslant x}f(n+h_1)\cdots f(n+h_k),\tag{1.9}$$

where $f(n)=\sum_{d|n,d\leqslant x^\delta}b_d$ is a Type I sum with explicit coefficients b_d. However, both of these tasks are rather tedious as such; the first correlation (1.8) has secondary main terms of order

[†] It would probably be possible to execute steps (ii) and (iii) in the opposite order, but that would offer no noteworthy simplifications, as we would still need to construct a majorant for $\lambda^\flat_{\mathrm{Siegel}}$.

$O(\frac{1}{\log x})$ times the main term (cf., [4]), and we would need a fully explicit asymptotic in terms of h, a, q; meanwhile, evaluating the second correlation (1.9) with the Goldston–Yıldırım approach [7] leads to some tricky contour integrals. We therefore smoothen Λ_{Siegel} by inserting a smooth partition of unity; the smoothness of the resulting functions makes handling error terms easier, just as in the smoothed approach to Goldston–Yıldırım type correlations in [9, Appendix D]. We can also avoid explicitly obtaining asymptotics for sums such as (1.8) by using the Dirichlet hyperbola method, although the main ingredient for evaluating such correlations (namely Kloosterman sum bounds) is still needed. Our use of smooth weights does still necessitate some lengthy yet standard Fourier-analytic computations, but the arithmetic input is easier than in a direct approach involving an evaluation of (1.8) and (1.9).

2 | NOTATION

2.1 | Asymptotic notation

For the rest of the paper, we let $k, \ell, h_1, \dots, h_k, h_1', \dots, h_\ell', \varepsilon_0, \beta, \chi, q_\chi, \eta, x$ be as in Theorem 1.6, save that we will not require the hypothesis $k \geqslant 2$ except in Section 5, and that we do not impose the restriction (1.5) on $x > 1$ before Section 4. We use the asymptotic notation $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote the bound $|X| \leqslant CY$ where C is a constant which is allowed to depend on the 'fixed' quantities $k, \ell, h_1, \dots, h_k, h_1', \dots, h_\ell', \varepsilon_0$; we permit the constants to be ineffective. Thus, for instance, the singular series $\mathfrak{S}$ in Conjecture 1.3 obeys the bound $\mathfrak{S} = O(1)$. If we need the constant C to depend on additional parameters, we will indicate this by subscripts, for instance, $X \ll_A Y$ denotes the bound $|X| \leqslant C_A Y$ where C_A depends on the parameter A as well as the fixed quantities. We write $X \asymp Y$ for $X \ll Y \ll X$.

By shrinking ε_0 if necessary, we may assume that ε_0 is sufficiently small depending on k, ℓ. We will also assume that η is sufficiently large depending on the fixed quantities, since otherwise the claim follows from standard upper bound sieves (such as Lemma 3.2). By (1.4), this also means that q_χ (and hence x) is also sufficiently large depending on the fixed quantities.

2.2 | Indicator and exponential functions

If S is a sentence, we use 1_S to denote its indicator, thus $1_S = 1$ when S is true and $1_S = 0$ otherwise. If E is a set, we use 1_E to denote the indicator function $1_E(n) := 1_{n \in E}$.

In addition to the notation $e(\theta) := e^{2\pi i \theta}$, we also write $e_q(a) := e(a/q) = e^{2\pi i a/q}$ for natural numbers q and $a \in \mathbb{Z}/q\mathbb{Z}$. We also write $\|\theta\|_{\mathbb{R}/\mathbb{Z}}$ for the distance of θ to the nearest integer.

2.3 | Primes and prime factorization

Unless otherwise specified, all sums and products will be over the natural numbers $\mathbb{N} = \{1, 2, \dots\}$, with the exception of sums and products involving the variable p (or p', p_1, etc.), which will be over primes. We define an *exceptional prime* to be a prime p^* such that $\chi(p^*) \neq -1$; sums over p^* (or p_1^*, etc.) will always be understood to be over exceptional primes.

If n is a natural number and p is a prime, we let $n_{(p)}$ denote the largest power of p dividing n, thus from the fundamental theorem of arithmetic

$$n = \prod_p n_{(p)}. \tag{2.1}$$

For any threshold $z > 1$, we may therefore factor a natural number n as

$$n = n_{(\leqslant z)} n_{(>z)}, \tag{2.2}$$

where the z-smooth and z-rough components $n_{(\leqslant z)}, n_{(>z)}$ of n are defined as

$$n_{(\leqslant z)} := \prod_{p \leqslant z} n_{(p)}$$

$$n_{(>z)} := \prod_{p > z} n_{(p)}.$$

For a prime p, we let

$$\mathbb{N}_{(p)} := \{n_{(p)} : n \in \mathbb{N}\} = \{1, p, p^2, \dots\}$$

denote the multiplicative semigroup generated by p, and similarly for a threshold $z > 1$ we write

$$\mathbb{N}_{(\leqslant z)} := \{n_{(\leqslant z)} : n \in \mathbb{N}\}$$

$$\mathbb{N}_{(>z)} := \{n_{(>z)} : n \in \mathbb{N}\}$$

for the multiplicative semigroups of z-smooth and z-rough numbers, respectively.

If $d_1, \dots, d_m$ are natural numbers, we use $(d_1, \dots, d_m)$ and $[d_1, \dots, d_m]$ to denote their greatest common divisor and least common multiple, respectively. We use $d\ (q)$ to denote the reduction of d to $\mathbb{Z}/q\mathbb{Z}$, and $q|d$ to denote the assertion that q divides d (or equivalently $d = 0\ (q)$).

A function $g : \mathbb{N}^m \to \mathbb{C}$ of m natural numbers $d_1, \dots, d_m$ is *multiplicative* if one has

$$g(d_1 d_1', \dots, d_m d_m') = g(d_1, \dots, d_m) g(d_1', \dots, d_m')$$

whenever $(d_1 \cdots d_m, d_1' \cdots d_m') = 1$. Observe the *Euler product identity*

$$\sum_{d_1, \dots, d_m} g(d_1, \dots, d_m) = \prod_p E_p \tag{2.3}$$

whenever the left-hand side is absolutely convergent, where the Euler factors E_p are defined as

$$E_p := \sum_{d_1, \dots, d_m \in \mathbb{N}_{(p)}} g(d_1, \dots, d_m).$$

We observe the localized form

$$\sum_{d_1, \dots, d_m \in \mathbb{N}_{(\leqslant z)}} g(d_1, \dots, d_m) = \prod_{p \leqslant z} E_p \tag{2.4}$$

of the Euler product identity for any threshold $z > 0$; in particular, if g is non-negative, then

$$\sum_{d_1,\dots,d_m \leqslant z} g(d_1,\dots,d_m) \leqslant \prod_{p \leqslant z} E_p. \tag{2.5}$$

We will frequently rely on Dirichlet convolution

$$f * g(n) := \sum_{d|n} f(d) g\Big(\frac{n}{d}\Big).$$

We let pointwise product take precedence over convolution, thus, for instance,

$$f_1 f_2 * f_3 f_4 = (f_1 f_2) * (f_3 f_4).$$

From (2.2), we observe the identity

$$f = f_{(\leqslant z)} * f_{(>z)} \tag{2.6}$$

for any multiplicative function f and any threshold $z > 1$, where

$$f_{(\leqslant z)} = f 1_{\mathbb{N}_{(\leqslant z)}}$$

$$f_{(>z)} = f 1_{\mathbb{N}_{(>z)}}$$

are the restrictions of f to z-smooth and z-rough numbers, respectively. Thus, for instance, $1_{(\leqslant z)} = 1_{\mathbb{N}_{(\leqslant z)}}$. Observe that this splitting respects Dirichlet convolutions, in the sense that

$$(f * g)_{(\leqslant z)} = f_{(\leqslant z)} * g_{(\leqslant z)}; \quad (f * g)_{(>z)} = f_{(>z)} * g_{(>z)} \tag{2.7}$$

for any $f, g : \mathbb{N} \to \mathbb{C}$.

2.4 | Scales

We will make frequent use of the scales

$$R := x^{1/\log^{\frac{1}{5\max(1,k)}} \eta} \tag{2.8}$$

and

$$D := x^{\frac{\varepsilon_0}{10(k+\ell)}}. \tag{2.9}$$

We will also occasionally need the auxiliary scale

$$R_0 := x^{1/\sqrt{\log \eta}}. \tag{2.10}$$

The reader may wish to keep in mind the hierarchy of scales

$$1 < \log\eta \ll \log q_\chi \ll \log x < R_0 < R < D < x.$$

which follows easily from (1.4). The conductor q_χ lies between $\log x$ and x^2 but can be either smaller or larger than R_0, R, or D.

We adopt the notation

$$\log_z y := \frac{\log y}{\log z}$$

for the logarithm of y to base z for any $y, z > 0$, and use the notation $X \approx Y$ as an abbreviation for

$$X = Y + O\left(\frac{1}{\log^{\frac{1}{10\max(1,k)}}\eta}\right). \tag{2.11}$$

Thus, for instance, the estimate (1.6) can be abbreviated to

$$\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda(n+h_1')\cdots\lambda(n+h_\ell') \approx \mathfrak{S}.$$

The scales R_0, R, D have been chosen so that certain combinations of these scales with x, η, q_χ that will arise in our calculations are negligible with respect to the relation $\approx$. More precisely, we observe for future reference that due to (1.5), (2.8), (2.9), (2.10), (1.4), we have relations

$$\log_D x \asymp \log_x D \asymp 1 \tag{2.12}$$

$$\log_D R \asymp \log_x R = \log^{-\frac{1}{5\max(1,k)}}\eta \approx 0 \tag{2.13}$$

$$(\log_R^k x)\log_R R_0 = \log^{\frac{k}{5\max(1,k)}-\frac{1}{2}}\eta \approx 0 \tag{2.14}$$

$$\frac{\log_R^k x}{\log^k \eta} = \log^{\frac{k}{5\max(1,k)}-k}\eta \approx 0 \tag{2.15}$$

$$\frac{R^{2k}D^{2(k+\ell)}q_\chi^{4k+1/2}x^{\varepsilon_0}}{x} \leqslant x^{\frac{2k}{\log^{1/(5\max(1,k))}\eta}+\frac{2(k+\ell)\varepsilon_0}{10(k+\ell)}+\frac{4k+1/2}{10k+\frac{1}{2}+\varepsilon_0}+\varepsilon_0-1} \approx 0 \tag{2.16}$$

$$(\log_R^{O(1)} x)\exp(-\sqrt{\log\eta}/2) = (\log^{O(1)}\eta)\exp(-\sqrt{\log\eta}/2) \approx 0 \tag{2.17}$$

$$q_\chi^{-\frac{\varepsilon_0}{4}}\log^{O(1)} x \leqslant x^{-\frac{\varepsilon_0}{4(10k+\frac{1}{2}+\varepsilon_0)}}\log^{O(1)} x \approx 0 \tag{2.18}$$

as well as the estimate

$$\exp\left(-\frac{1}{8}\log_R D\right) = \exp\left(-\frac{\varepsilon_0\log^{1/5\max(1,k)}\eta}{80(k+\ell)}\right) \ll_A \log^{-A}\eta \tag{2.19}$$

for all $A > 0$. Also, for $k = 1, 2$, we note for future reference that

$$\begin{aligned}\frac{(\sqrt{x}R^2)^{\frac{3}{2}(k-1)}D^{2(k+\ell)}q_\chi^{9/2}}{x^{1-2\varepsilon_0}} &\ll x^{\frac{3}{4}(k-1)+\frac{9}{2}\frac{1}{10k}+3\varepsilon_0-1}\\ &\ll x^{-\varepsilon_0}\\ &\approx 0\end{aligned} \tag{2.20}$$

since $\frac{3}{4}(k-1)+\frac{9}{2}\frac{1}{10k} \leqslant 1-\frac{1}{40} < 1$ for $k = 1, 2$.

2.5 | The Selberg sieve

We fix a smooth function $\psi : \mathbb{R} \to \mathbb{R}$ supported on $[-1, 1]$ that equals to 1 on $[-1/2, 1/2]$, and define the smooth cutoffs

$$\psi_{\leqslant z}(n) := \psi(\log_z n) \tag{2.21}$$

and

$$\psi_{>z}(n) := 1 - \psi(\log_z n) \tag{2.22}$$

for any $z > 1$. We then define the *Selberg sieve*[†]

$$\nu(n) := \left(\sum_{d|n} \mu(d)\psi_{\leqslant R}(d)\right)^2. \tag{2.23}$$

Note that ν is an upper bound sieve for $1_{(>R)}$, thus

$$1_{(>R)}(n) \leqslant \nu(n) \tag{2.24}$$

for all natural numbers n.

3 | TOOLS

In this section, we collect some (mostly standard) estimates on various arithmetic functions which will be used in our main argument.

[†] Here we use the Selberg sieve with smoothed coefficients, which was implicitly introduced by Goldston and Yıldırım; see, for instance, [9, Appendix D] for further discussion. Other sieve approximants to $1_{(>z)}$ could be used as a substitute for this sieve if desired; for instance, the beta sieve was used in place of a Selberg sieve in the recent work [17], which appeared subsequently to the initial release of this paper.

3.1 | Multiplicative number theory bounds

We recall the crude *divisor bound*

$$\tau(n) \ll_\varepsilon n^\varepsilon \tag{3.1}$$

for any $n \geqslant 1$ and $\varepsilon > 0$; see, for example, [20, (2.20)].

From the Euler product formula,

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}$$

and the fact that ζ has a simple pole at $s = 1$ with residue 1 and no zeroes in $\{s : |s-1| \leqslant \frac{1}{2}\}$, we see that

$$\prod_p \left(1 - \frac{1}{p^s}\right) = (1 + O(|s-1|))(s-1) \tag{3.2}$$

whenever s is a complex number with $\mathrm{Re} s > 1$ and $|s-1| < \frac{1}{2}$.

From Mertens' theorem, we easily verify that

$$\sum_{p \leqslant z} \frac{\min(\sigma \log_R p, 1)}{p} \ll \log(1 + \sigma \log_R z) \tag{3.3}$$

for any $\sigma > 0$ and $R, z \geqslant 1$, as can be seen by verifying the cases $\sigma \log_R z < 1$ and $\sigma \log_R z \geqslant 1$ separately; in exponential form, we thus have

$$\prod_{p \leqslant z} \left(1 + O\left(\frac{\min(\sigma \log_R p, 1)}{p}\right)\right) \leqslant (1 + \sigma \log_R z)^{O(1)}. \tag{3.4}$$

Mertens' theorem also gives (by dyadic decomposition) the bounds

$$\sum_{p \geqslant y} \frac{1}{p^{1+1/\log z}} = \sum_{p \geqslant y} \frac{\exp(-\log_z p)}{p} \asymp \frac{\exp(-\log_z y)}{\log_z y} \tag{3.5}$$

and

$$\prod_{p \leqslant z} \left(1 - \frac{m}{p}\right) \asymp_m \log^{-m} z \tag{3.6}$$

for any $y \geqslant z \geqslant 2$ and $m \geqslant 1$. In particular,

$$\prod_p \left(1 + O\left(\frac{1}{p^{1+1/\log z}}\right)\right) \ll \log^{O(1)} z. \tag{3.7}$$

We recall an elementary inequality of Landreau [14] that allows one to upper bound the divisor function τ by a Type I sum:

Lemma 3.1 (Landreau's inequality).

(i) *If n is a natural number and $y > z > 1$, then we can factor*

$$n = n_{(>z)} n_1 \cdots n_m, \tag{3.8}$$

where $n_1, \dots, n_m \leqslant y$ lie in $\mathbb{N}_{(\leqslant z)}$ and $0 \leqslant m \leqslant 1 + \log_{y/z} n$. Also, $n_{(>z)}$ is the product of at most $\log_z n$ primes.

(ii) *If $\varepsilon > 0$, then*

$$\tau(n) \ll_\varepsilon \sum_{d|n : d \leqslant n^\varepsilon} \tau(d)^{O_\varepsilon(1)} \tag{3.9}$$

for all $n \geqslant 1$. In particular, by (2.9), one has

$$\tau(n) \ll \sum_{d|n : d \leqslant D} \tau(d)^{O(1)} \tag{3.10}$$

for $n \ll x$.

Proof. Observe from the greedy algorithm that any number in $\mathbb{N}_{(\leqslant z)}$ is either greater than y, or contains a factor between y/z and y. Iterating this fact, we can factor $n_{(\leqslant z)} = n_1 \cdots n_m$ where $n_1, \dots, n_m \leqslant y$ and all but at most one of the $n_1, \dots, n_m$ are greater than or equal to y/z. This gives the bound $m \leqslant 1 + \log_{y/z} n$. Since $n_{(>z)}$ is the product of primes greater than z, the total number of primes is at most $\log_z n$. This gives (i).

For (ii), we apply (i) with $y = n^\varepsilon$ and $z = n^{\varepsilon/2}$ and use (2.12) to obtain the factorization (3.8) with $n_{(>n^{\varepsilon/2})}$ the product of $O_\varepsilon(1)$ primes, $m = O_\varepsilon(1)$, and $n_1, \dots, n_m \leqslant n^\varepsilon$. Using the elementary inequality $\tau(ab) \leqslant \tau(a)\tau(b)$, we conclude that

$$\tau(n) \ll \tau(n_1) \cdots \tau(n_m)$$

and hence by the pigeonhole principle

$$\tau(n) \ll \tau(d)^m$$

for d equal to one of the $n_1, \dots, n_m$. The claim (ii) follows. □

We also record a standard sieve upper bound, which can easily be deduced from the fundamental lemma of sieve theory (or the large sieve):

Lemma 3.2 (Sieve upper bound). *Suppose that for every prime $p \leqslant x$ there is a natural number $0 \leqslant \omega(p) \ll 1$, and let E be a subset of $\{n : n \leqslant x\}$ which avoids at least $\omega(p)$ residue classes modulo p for each $p \leqslant x$. Then we have*

$$\mathbb{E}_{n \leqslant x} 1_E(n) \ll \prod_{p \leqslant x} \left(1 - \frac{\omega(p)}{p}\right).$$

Proof. We may assume that $\omega(p) < p$ for all p, since otherwise E is empty and the claim is trivial (of course, this assumption is only non-trivial for the very small primes $p = O(1)$). By Mertens' theorem, the contribution of the primes $x^{1/100} < p \leqslant x$ to the right-hand is $\asymp 1$, so we may replace the product $\prod_{p \leqslant x}$ here with $\prod_{p \leqslant x^{1/100}}$.

Let g be the multiplicative function $g(d) := \prod_{p|d:\, p\leqslant x^{1/100}} \frac{\omega(p)}{p}$. By the fundamental lemma of sieve theory (see [12, Lemma 6.3]), we can find weights $\lambda_d^+ \in [-1,1]$ for all $d \leqslant \sqrt{x}$ such that

$$1_{n=1} \leqslant \sum_{d|n} \lambda_d^+ \tag{3.11}$$

for all natural numbers n, and

$$\sum_{d\leqslant x^{1/2}} \lambda_d^+ g(d) \ll \prod_{p\leqslant x^{1/100}} (1-g(p)). \tag{3.12}$$

For each $d \leqslant x^{1/2}$, let E_d be the set formed by removing the $\omega(p)$ residue classes modulo p from $\{n : n \leqslant x\}$ for all $p|d$. Then from (3.11) (with n replaced by $\prod_{p\leqslant x^{1/100}:n\notin E_p} p$), we have the pointwise bound

$$1_E \leqslant \sum_{d\leqslant x^{1/2}} \lambda_d^+ 1_{E_d}$$

and thus

$$\mathbb{E}_{n\leqslant x} 1_E(n) \leqslant \sum_{d\leqslant x^{1/2}} \lambda_d^+ \mathbb{E}_{n\leqslant x} 1_{E_d}(n).$$

From the Chinese remainder theorem, we have

$$\mathbb{E}_{n\leqslant x} 1_{E_d}(n) = g(d) + O\Big(\frac{1}{x}\Big)$$

and thus by (3.12)

$$\mathbb{E}_{n\leqslant x} 1_{E_d}(n) \ll \prod_{p\leqslant x^{1/100}} \Big(1-\frac{\omega(p)}{p}\Big) + \frac{1}{x^{1/2}}.$$

By Mertens' theorem (3.6), the second term on the right-hand side is certainly dominated by the former, and the claim follows. □

We also record the following easy consequence of the Chinese remainder theorem.

Lemma 3.3 (Chinese remainder theorem). *Let $d_1,\dots,d_{k'}$ be natural numbers for some $2 \leqslant k' \leqslant k+\ell$, and set $d := [d_1,\dots,d_{k'}]$.*

(i) *If (d_i,d_j) does not divide $h_i - h_j$ for some $1 \leqslant i < j \leqslant k'$, then $\prod_{j=1}^{k'} 1_{d_j|n+h_j}$ vanishes for all n.*
(ii) *If instead (d_i,d_j) divides $h_i - h_j$ for all $1 \leqslant i < j \leqslant k'$, then there is a unique residue class $a\ (d)$ such that $\prod_{j=1}^{k'} 1_{d_j|n+h_j} = 1_{n=a\ (d)}$. Furthermore, $a = -h_j\ (d_j)$ for all $j = 1,\dots,k'$, and $d \asymp d_1\cdots d_{k'}$.*

Proof. All the claims are immediate except for the existence of the residue class a in part (ii) (the final part of (ii) following from the general relation $\frac{d_1\cdots d_{k'}}{[d_1,\dots,d_{k'}]} \mid \prod_{1\leqslant i<j\leqslant k'}(d_i,d_j)$). By the Chinese remainder theorem, we may assume that the d_i are all powers of a single prime p. Then we have $d = d_i$ for some $1 \leqslant i \leqslant k'$, and the claim follows by setting $a := -h_i$. □

3.2 | Some fourier analysis

Recall the *Fourier inversion formula*: if $g : \mathbb{R} \to \mathbb{C}$ is a Schwartz function, then one has

$$g(u) = \int_{\mathbb{R}} e^{-itu} f(t)\, dt \tag{3.13}$$

for all $u \in \mathbb{R}$, where the Fourier transform $f : \mathbb{R} \to \mathbb{C}$ of g is another Schwartz function defined by the formula

$$f(t) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{itu} g(u)\, du.$$

As a special case of this, if $\varphi : \mathbb{R} \to \mathbb{C}$ is a function such that $u \mapsto e^u \varphi(u)$ is Schwartz, then

$$e^u \varphi(u) = \int_{\mathbb{R}} e^{-itu} f(t)\, dt$$

for all $u \in \mathbb{R}$, where

$$f(t) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{(1+it)u} \varphi(u)\, du.$$

In particular, for any real $n, z > 0$, we have

$$\varphi(\log_z n) = \int_{\mathbb{R}} \frac{1}{n^{\frac{1+it}{\log z}}} f(t)\, dt. \tag{3.14}$$

Evaluating this formula at $n = 1$, we conclude that

$$\varphi(0) = \int_{\mathbb{R}} f(t)\, dt \tag{3.15}$$

and if one differentiates at $n = 1$ instead one obtains the variant identity

$$-\varphi'(0) = \int_{\mathbb{R}} (1 + it) f(t)\, dt. \tag{3.16}$$

As an application of these Fourier representations, we give an analogue of Lemma 3.2 for the Selberg sieve ν (cf., [22, Lemma 3], [24, Proposition 14]):

Lemma 3.4 (Selberg sieve concentrates on almost primes). *Let $0 \leqslant \ell' \leqslant \ell$, and let $1 \leqslant d_1, \dots, d_k, d'_1, \dots, d'_{\ell'} \leqslant x$ be integers. Then*

$$\begin{aligned} &\mathbb{E}_{n \leqslant x} \prod_{j=1}^{k} \nu(n + h_j) 1_{d_j | n + h_j} \prod_{j'=1}^{\ell'} 1_{d'_{j'} | n + h'_{j'}} \\ &\ll_A \frac{\tau(d_1 \cdots d_k d'_1 \cdots d'_{\ell'})^{O(1)}}{d_1 \cdots d_k d'_1 \cdots d'_{\ell'} \log^k R} \int_1^\infty \left(\prod_{p | d_1 \cdots d_k} \min(\sigma \log_R p, 1) \right) \frac{d\sigma}{\sigma^A} + \frac{R^{2k}}{x} \end{aligned} \tag{3.17}$$

for any $A > 0$.

The $\frac{R^{2k}}{x}$ error term is negligible in practice. The σ variable of integration is technical and as a first approximation the reader is invited to replace σ with 1 (and delete the integral). The key feature of this estimate are the factors of $\min(\sigma \log_R p, 1)$, which make the left-hand side of (3.17) small when $d_1, \dots, d_k$ have one or more small prime factors. With further effort, one could obtain a more precise asymptotic for the left-hand side of (3.17) (in the spirit of [9, Theorem D.3]) but we will not need to do so here.

Proof. It is convenient to relabel by writing $k' := k + \ell'$ and $h_{k+j} := h'_j$, $d_{k+j} := d'_j$ for $j = 1, \dots, \ell'$. By Lemma 3.3, we may assume that $(d_i, d_j)|h_i - h_j$ for all $1 \leqslant i < j \leqslant k'$. In particular, if we set $d := [d_1, \dots, d_k]$ and $d' := [d_1, \dots, d_{k'}]$, then $d \asymp d_1 \cdots d_k$ and $d' \asymp d_1 \cdots d_{k'}$.

By (2.23), the left-hand side of (3.17) may be expanded as

$$\sum_{d'_1, d''_1, \dots, d'_k, d''_k} \mathbb{E}_{n \leqslant x} \prod_{j=1}^{k} \mu\psi_{\leqslant R}(d'_j)\mu\psi_{\leqslant R}(d''_j) \prod_{j=1}^{k'} 1_{d^*_j | n+h_j},$$

where $d^*_j := [d_j, d'_j, d''_j]$ for $j = 1, \dots, k$ and $d^*_j := d_j$ for $j = k+1, \dots, k'$. From Lemma 3.3, we see that the average $\mathbb{E}_{n \leqslant x} \prod_{j=1}^{k'} 1_{d^*_j | n+h_j}$ vanishes unless $(d^*_i, d^*_j)|h_i - h_j$ for all $1 \leqslant i < j \leqslant k'$, in which case it is equal to $\frac{1}{[d^*_1, \dots, d^*_{k'}]} + O(\frac{1}{x})$. The contribution of the error $O(\frac{1}{x})$ is of size $O(\frac{R^{2k}}{x})$, so it suffices to show that

$$\sum_{d'_1, d''_1, \dots, d'_k, d''_k} \frac{\prod_{j=1}^{k} \mu\psi_{\leqslant R}(d'_j)\mu\psi_{\leqslant R}(d'_j) \prod_{1 \leqslant i < j \leqslant k'} 1_{(d^*_i, d^*_j)|h_i - h_j}}{[d^*_1, \dots, d^*_{k'}]}$$
$$\ll_A \int_1^\infty \frac{\tau(d'')^{O(1)}}{d' \log^k R} \prod_{p|d} \min(\sigma \log_R p, 1) \frac{d\sigma}{\sigma^A}.$$

We can expand the left-hand side using (3.14) and Fubini's theorem as

$$\int_{\mathbb{R}^{2k}} \sum_{d'_1, d''_1, \dots, d'_k, d''_k} \frac{\prod_{j=1}^{k} \mu(d'_j)\mu(d''_j) \prod_{1 \leqslant i < j \leqslant k'} 1_{(d^*_i, d^*_j)|h_i - h_j}}{[d^*_1, \dots, d^*_{k'}] \prod_{j=1}^{k} (d'_j)^{\frac{1+it'_j}{\log R}} (d''_j)^{\frac{1+it''_j}{\log R}}} \prod_{j=1}^{k} f(t'_j) f(t''_j)\, dt'_j dt''_j$$

for some Schwartz function f. Changing variables using the substitution $\sigma := 1 + \sum_{j=1}^{k} |t'_j| + |t''_j|$, and using the rapid decay of f and the triangle inequality, it will suffice to establish the pointwise bound

$$\sum_{d'_1, d''_1, \dots, d'_k, d''_k} \frac{\prod_{j=1}^{k} \mu(d'_j)\mu(d''_j) \prod_{1 \leqslant i < j \leqslant k'} 1_{(d^*_i, d^*_j)|h_i - h_j}}{[d^*_1, \dots, d^*_{k'}] \prod_{j=1}^{k} (d'_j)^{\frac{1+it'_j}{\log R}} (d''_j)^{\frac{1+it''_j}{\log R}}} \ll \sigma^{O(1)} \frac{\tau(d')^{O(1)}}{d' \log^k R} \prod_{p|d} \min(\sigma \log_R p, 1)$$

for all $t'_1, \dots, t''_k \in \mathbb{R}$.

By (2.3) and the fact that μ is supported on square-free numbers, the left-hand side factors as an Euler product $\prod_p E_p$ where

$$E_p := F_p\left(\frac{1+it_1'}{\log R}, \frac{1+it_1''}{\log R}, \dots, \frac{1+it_k'}{\log R}, \frac{1+it_k''}{\log R}\right)$$

and

$$F_p(z_1', z_1'', \dots, z_k', z_k'') := \sum_{d_1', d_1'', \dots, d_k', d_k'' \in \{1,p\}} \frac{\prod_{j=1}^k \mu(d_j')\mu(d_j'') \prod_{1 \leqslant i < j \leqslant k'} 1_{((d_i^*)_{(p)}, (d_j^*)_{(p)})|h_i - h_j}}{[d_1^*, \dots, d_{k'}^*]_{(p)} \prod_{j=1}^k (d_j')^{z_j'} (d_j'')^{z_j''}}.$$

From the triangle inequality, we have

$$E_p = 1 + O\left(\frac{1}{p^{1+\frac{1}{\log R}}}\right) \tag{3.18}$$

when $d'_{(p)} = 1$ and

$$E_p \ll \frac{1}{d'_{(p)}} \tag{3.19}$$

when $d'_{(p)} > 1$, hence by (3.5) we have

$$\prod_{p>R} E_p d'_{(p)} \ll \tau(d')^{O(1)}.$$

Now let $p \leqslant R$ and $d'_{(p)} > 1$. Then from the triangle inequality, we have

$$F_p(z_1', z_1'', \dots, z_k', z_k'') \ll \frac{1}{d'_{(p)}}$$

whenever $z_1', z_1'', \dots, z_k', z_k''$ are complex numbers of size $O(\frac{1}{\log p})$, while from the cancellation in the Möbius coefficients $\mu(d_j'), \mu(d_j'')$ and the hypothesis $(d_i, d_j)|h_i - h_j$ for all $1 \leqslant i < j \leqslant k'$, we see that

$$F_p(0, \dots, 0) = 0.$$

From the Cauchy integral formula (in the case $\sigma \log_R p \leqslant 1$) or (3.19) (otherwise), we have

$$E_p d'_{(p)} \ll \min(\sigma \log_R p, 1).$$

Finally, suppose that $p \leqslant R$ and $d'_{(p)} = 1$. Then from the triangle inequality, we have

$$F_p(z_1', z_1'', \dots, z_k', z_k'') = 1 + O\left(\frac{1}{p}\right)$$

whenever $z_1', z_1'', \dots, z_k', z_k''$ are complex numbers of size $O(\frac{1}{\log p})$, while from noting that the conditions $((d_i^*)_{(p)}, (d_j^*)_{(p)})|h_i - h_j$ permit d_j', d_j'' to equal p for h_j in at most one residue class a (p),

we have

$$\begin{aligned} F_p(0,\dots,0) &= 1 + \sum_{a\in\mathbb{Z}/p\mathbb{Z}} \frac{\sum_{d_j',d_j''\in\{1,p\} \text{ when } h_j=a\ (p)} \prod_{h_j=a\ (p)} \mu(d_j')\mu(d_j'') - 1}{p} \\ &= 1 + \sum_{a\in\mathbb{Z}/p\mathbb{Z}} \frac{(1-1)^{2\#\{j=1,\dots,k\,:\,h_j=a\ (p)\}} - 1}{p} \\ &= 1 - \frac{\#\{h_j\ (p)\,:\,j=1,\dots,k\}}{p} \\ &= \beta_p\left(1-\frac{1}{p}\right)^k \end{aligned}$$

due to (1.2). From the Cauchy integral formula (in the case $\sigma \log p \leqslant \log R$) or (3.18) (otherwise), we thus have

$$E_p = E_p d'_{(p)} = \beta_p\left(1-\frac{1}{p}\right)^k + O\left(\frac{\min(\sigma \log_R p, 1)}{p}\right),$$

and hence by (1.3)

$$E_p d'_{(p)} \ll \left(1-\frac{k}{p}\right)\left(1 + O\left(\frac{\min(\sigma \log_R p, 1)}{p}\right) + O\left(\frac{1}{p^2}\right)\right).$$

From Mertens' theorem (3.4), we have

$$\prod_{p\leqslant R}\left(1 + O\left(\frac{\min(\sigma \log_R p, 1)}{p}\right) + O\left(\frac{1}{p^2}\right)\right) \ll \sigma^{O(1)}.$$

Putting all this together, we see that

$$\prod_p E_p d'_{(p)} \ll \sigma^{O(1)}\tau(d')^{O(1)} \prod_{p\leqslant R\,:\,p\nmid d'}\left(1-\frac{k}{p}\right) \prod_{p\leqslant R\,:\,p\mid d} \min(\sigma \log_R p, 1)$$

and hence by Mertens' theorem (3.6)

$$\prod_p E_p d'_{(p)} \ll \frac{\sigma^{O(1)}\tau(d'')^{O(1)}}{\log^k R} \prod_{p\leqslant R\,:\,p\mid d} \min(\sigma \log_R p, 1).$$

From (2.1), we have

$$\prod_p d''(p) = d' \asymp d_1 \cdots d_{k'},$$

and the claim follows. □

3.3 | Elementary consequences of a Siegel zero

Recall from Section 2 that we use p^* to denote primes that are *exceptional* in the sense that $\chi(p^*) \neq -1$. It is a well-known phenomenon that exceptional primes become rare at scales comparable in log-scale to q_χ. For instance, in [11, Lemma 3], it was shown that[†]

$$\sum_{p^* \leqslant q_\chi^{500}} \frac{\log p^*}{p^*} \ll \frac{\log q_\chi}{\sqrt{\log \eta}} \tag{3.20}$$

while in [6] it was shown more generally that

$$\sum_{p^* \leqslant x} \frac{\log p^*}{p^*} \ll \exp\left(\log_{q_\chi} x\right) \frac{\log q_\chi}{\sqrt{\log \eta}} \tag{3.21}$$

for $q_\chi^{10} \leqslant x \leqslant q_\chi^{\log\log \eta/3}$. In fact, we can do better:

Proposition 3.5. *Let $\varepsilon > 0$. Then for any $x \geqslant q_\chi^{\frac{1+\varepsilon}{2}}$, one has*

$$\sum_{q_\chi^{\frac{1+\varepsilon}{2}} < p^* \leqslant x} \frac{1}{p^*} \ll_\varepsilon \frac{\log_{q_\chi} x}{\eta} \tag{3.22}$$

and for any natural number $m \geqslant 2$, we have

$$\sum_{q_\chi^{\frac{1+\varepsilon}{2m}} < p^* \leqslant q_\chi^{\frac{1+\varepsilon}{2(m-1)}}} \frac{1}{p^*} \ll_\varepsilon \frac{m}{\eta^{1/m}}. \tag{3.23}$$

The first bound is non-trivial for x as large as $q_\chi^{\eta^{1-\varepsilon_0}}$, while the second bound is non-trivial for primes p^* as small as $q_\chi^{1/\log^{1-\varepsilon_0}\eta}$. It is not difficult to recover (3.21) (and hence (3.20)) from the above proposition by taking a suitable linear combination of (3.22) and (3.23) for $m \leqslant \sqrt{\log \eta}$, and using Mertens' theorem to control the contribution of exceptional primes $p^* \leqslant q_\chi^{10/\sqrt{\log\eta}}$ (say); we leave the details to the interested reader.

Proof. For any $x \geqslant q_\chi^{\frac{1+\varepsilon}{2}}$, we have from [20, Exercise 11.2.3(g)] that

$$\sum_{n \leqslant x} \frac{1 * \chi(n)}{n} = (\log x + \gamma) L(1,\chi) + L'(1,\chi) + O_\varepsilon(q_\chi^{-\varepsilon/10}).$$

From Siegel's theorem, we have $L(1,\chi) \gg_\varepsilon q_\chi^{-\varepsilon/10}$, and hence

$$\sum_{n \leqslant x} \frac{1 * \chi(n)}{n} = L(1,\chi)\left(\log x + \frac{L'}{L}(1,\chi) + O_\varepsilon(1)\right). \tag{3.24}$$

[†] Strictly speaking, these results only claim to control the set where $\chi(p^*) = 1$, ignoring the relatively small number of primes where $\chi(p^*) = 0$, but it is not difficult to modify the arguments to also include the latter set.

From [20, Theorem 11.4], we also have $\frac{L'}{L}(1,\chi) \asymp \eta \log q_\chi$. Thus, (3.24) gives

$$\sum_{n \leqslant q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1*\chi(n)}{n} \gg L(1,\chi)\eta \log q_\chi, \tag{3.25}$$

while applying (3.24) with x replaced by $q_\chi^{\frac{1+\varepsilon}{2}}, xq_\chi^{\frac{1+\varepsilon}{2}}$ and subtracting we obtain

$$\sum_{q_\chi^{\frac{1+\varepsilon}{2}} < n \leqslant x q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1*\chi(n)}{n} = L(1,\chi)(\log x + O_\varepsilon(1)). \tag{3.26}$$

On the other hand, from the non-negativity and multiplicativity of $1*\chi$, we have

$$\sum_{q_\chi^{\frac{1+\varepsilon}{2}} < n \leqslant x q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1*\chi(n)}{n} \geqslant \left(\sum_{n \leqslant q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1*\chi(n)}{n} \right) \left(\sum_{q_\chi^{\frac{1+\varepsilon}{2}} < p \leqslant x} \frac{1*\chi(p)}{p} \right).$$

Since $1*\chi(p)$ is non-negative and is at least one when p is exceptional, the first claim (3.22) follows.

In a similar vein, since any $n \leqslant q_\chi^{10}$ has $\leqslant \binom{20m}{m}$ representations in the form $n' p_1 \cdots p_m$ with $q_\chi^{(1+\varepsilon)/2} < p_1 < p_2 < \cdots < p_m$, we have for any natural number $m \geqslant 2$ that

$$\begin{aligned}
&\sum_{q_\chi^{\frac{1+\varepsilon}{2}} < n \leqslant q_\chi^{10}} \frac{1*\chi(n)}{n} \\
&\geqslant \binom{20m}{m}^{-1} \sum_{n < q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1*\chi(n)}{n} \sum_{q_\chi^{\frac{1+\varepsilon}{2m}} < p_1 < \cdots < p_m \leqslant q_\chi^{\frac{1+\varepsilon}{2(m-1)}} : p_1,\dots,p_m \nmid n} \frac{1*\chi(p_1)}{p_1} \cdots \frac{1*\chi(p_m)}{p_m} \\
&\geqslant \binom{20m}{m}^{-1} \sum_{n < q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1*\chi(n)}{n} \sum_{q_\chi^{\frac{1+\varepsilon}{2m}} < p_1^* < \cdots < p_m^* \leqslant q_\chi^{\frac{1+\varepsilon}{2(m-1)}} : p_1^*,\dots,p_m^* \nmid n} \frac{1}{p_1^* \cdots p_m^*} \\
&= \binom{20m}{m}^{-1} \sum_{n < q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1*\chi(n)}{n} \frac{\sum_{q_\chi^{\frac{1+\varepsilon}{2m}} < p_1^*,\dots,p_m^* \leqslant q_\chi^{\frac{1+\varepsilon}{2(m-1)}} : p_1^*,\dots,p_m^* \nmid n, \text{ distinct}} \frac{1}{p_1^* \cdots p_m^*}}{m!}.
\end{aligned}$$

Observe that once $n < q_\chi^{(1+\varepsilon)/2}$ and some of the exceptional primes $p_1^*, \dots, p_j^*$, $j < m$ have been chosen, the restrictions that the exceptional prime p_{j+1}^* be distinct from $p_1^*, \dots, p_j^*$ and not divide n only excludes at most $2m$ primes p_{j+1}^* from the range $q_\chi^{\frac{1+\varepsilon}{2m}} < p_{j+1}^* \leqslant q_\chi^{\frac{1+\varepsilon}{2(m-1)}}$, since n has at most

m factors in this range. Thus we have

$$\sum_{q_\chi^{\frac{1+\varepsilon}{2}} < n \leqslant q_\chi^{10}} \frac{1 * \chi(n)}{n} \geqslant \binom{20m}{m}^{-1} \left(\sum_{n < q_\chi^{\frac{1+\varepsilon}{2}}} \frac{1 * \chi(n)}{n} \right) \frac{\left(\sum^{*}_{q_\chi^{\frac{1+\varepsilon}{2m}} < p^* \leqslant q_\chi^{\frac{1+\varepsilon}{2(m-1)}}} \frac{1}{p^*} \right)^m}{m!},$$

where the asterisk in the sum means that we are allowed to delete the $2m$ largest terms from the sum (or delete the sum entirely, that is, replace it by zero, if there are fewer than $2m$ terms in all). The estimates (3.25) and (3.26) then give

$$\sum^{*}_{q_\chi^{\frac{1+\varepsilon}{2m}} < p^* \leqslant q_\chi^{\frac{1+\varepsilon}{2(m-1)}}} \frac{1}{p^*} \ll_\varepsilon \left(\frac{\binom{20m}{m} m! L(1,\chi) \log q_\chi}{L(1,\chi)\eta \log q_\chi} \right)^{1/m} \ll \frac{m}{\eta^{1/m}}.$$

One can reinstate the top $2m$ terms from the sum on the left-hand side, since their contribution is $\ll m/q_\chi^{1/(2m)} \ll m\eta^{-1/m}$ by the Siegel bound (1.4). The claim (3.23) follows. □

Corollary 3.6. *Let $q_\chi^{(1+\varepsilon)/2} \leqslant x \leqslant q_\chi^{\eta^{1/2}}$. We have*

$$\sum_{R_0 \leqslant p^* \leqslant x} \frac{1}{p^*} \ll \exp(-\sqrt{\log \eta}/2)$$

and

$$\sum_{p^*} \frac{\min(\log^{0.1} \eta \log_R p^*, 1)}{(p^*)^{1+\frac{1}{\log x}}} \ll \frac{1}{\log^{0.3} \eta}.$$

This bound will be used in steps (i) and (ii) of the main argument.

Proof. From (3.22) (with $\varepsilon = 1$) and (1.5), we have

$$\sum_{q_\chi < p^* \leqslant x} \frac{1}{p^*} \ll \frac{\sqrt{\eta}}{\eta} \ll \exp(-\sqrt{\log \eta}/2)$$

and from (3.23), we similarly have

$$\sum_{q_\chi^{\frac{1}{m}} < p^* \leqslant q_\chi^{\frac{1}{m-1}}} \frac{1}{p^*} \ll \frac{m}{\eta^{1/m}}$$

for all $m \geqslant 2$. Summing over $2 \leqslant m \leqslant \sqrt{\log \eta} + 1$, we obtain the first claim.

Now we prove the second claim. The contribution of those p^* with $p^* \geqslant x^{\log^{0.1} \eta}$ is acceptable by (3.5), while the contribution of those p^* with $p^* \leqslant R^{1/\log^{0.4} \eta}$ is also acceptable by (3.3). Thus

it remains to show that

$$\sum_{R^{1/\log^{0.4}\eta}<p^*<x^{\log^{0.1}\eta}} \frac{1}{p^*} \ll \frac{1}{\log^{0.3}\eta}.$$

The contribution of those p^* with $q_\chi < p^* < x^{\log^{0.1}\eta}$ is acceptable by (3.22) (for $\varepsilon = 1$), (1.5), while the contribution of those p^* with $R^{1/\log^{0.4}\eta} < p^* \leqslant q_\chi$ is acceptable by (3.23) (for $\varepsilon = 1$ and $2 \leqslant m \ll \log^{0.5}\eta$, say) and (2.8). □

3.4 | Consequences of the Weil bound for character sums

Let $f : \mathbb{Z} \to \mathbb{Z}$ be a polynomial of degree $O(1)$. If p is a prime, we have the standard Weil bounds

$$\sum_{n\in\mathbb{Z}/p\mathbb{Z}} \chi_p(f(n))e_p(an) \ll p^{1/2}$$

uniformly for all integers a whenever f is not a constant multiple of perfect square modulo p, where χ_p is the quadratic character modulo p; see [30] (or [21]). When f is a constant multiple of a perfect square, we can of course use the trivial bound of $O(p)$. Since the exceptional modulus q_χ is a fundamental discriminant, it is of the form $2^j p_1 \cdots p_m$ for some $j \leqslant 3$ and distinct odd primes $p_1, \dots, p_m$, and so from the Chinese remainder theorem, we obtain the bounds

$$\sum_{n\in\mathbb{Z}/q_\chi\mathbb{Z}} \chi(f(n))e_{q_\chi}(an) \ll \tau(q_\chi)^{O(1)} q_\chi^{1/2} d^{1/2}$$

uniformly in a, where d is the largest factor of q_χ for which f is a constant multiple of a perfect square modulo d. Applying (3.1) and completion of sums (see [12, Lemma 12.1]), we conclude that

$$\sum_{n\in I} \chi(f(n)) \ll_\varepsilon q_\chi^{1/2+\varepsilon} d^{1/2}$$

for any interval I of length at most q_χ and any $\varepsilon > 0$; by subdividing longer intervals into intervals of length q_χ, plus a remainder, we conclude that

$$\sum_{n\in I} \chi(f(n)) \ll_\varepsilon q_\chi^{1/2+\varepsilon} d^{1/2} \left(\frac{|I|}{q_\chi} + 1\right) \tag{3.27}$$

for any interval I and any $\varepsilon > 0$.

This gives us the following bounds:

Lemma 3.7. *Let $d_1, \dots, d_{k+\ell}$ be natural numbers. Let I be an interval in $[1, x]$. Let J be a non-empty subset of $\{1, \dots, k+\ell\}$, and for each $j \in J$, let d'_j be a factor of d_j. Then*

$$\mathbb{E}_{n\leqslant x} 1_I(n) \left(\prod_{j=1}^{k+\ell} 1_{d_j | n+h_j}\right) \prod_{j\in J} \chi\left(\frac{n+h_j}{d'_j}\right) \ll_\varepsilon q_\chi^{1/2+\varepsilon} (d_1 \cdots d_{k+\ell}, q_\chi)^{1/2} \left(\frac{1}{q_\chi d_1 \cdots d_{k+\ell}} + \frac{1}{x}\right)$$

for any $\varepsilon > 0$, where we use the notation $h_{k+j} := h'_j$ for $j = 1, \dots, \ell$.

This bound will be used in step (v) of the main argument, to dispose of any 'Type I sum' contributions that are twisted by one or more factors of the exceptional character χ.

Proof. By Lemma 3.3, we may assume that $(d_i, d_j)|h_i - h_j$ for all $1 \leqslant i < j \leqslant k+\ell$ and replace the conditions $d_j|n+h_j$ with $n = a\ (d)$ where

$$d := [d_1, \dots, d_{k+\ell}] \asymp d_1 \cdots d_{k+\ell}$$

and

$$a = -h_j\ (d_j)$$

for $j = 1, \dots, k+\ell$. Our task is now equivalent to showing that

$$\sum_{n: dn+a \in I} \prod_{j \in J} \chi\left(\frac{dn+a+h_j}{d'_j}\right) \ll_\varepsilon q_\chi^{1/2+\varepsilon}(d, q_\chi)^{1/2}\left(\frac{x}{q_\chi d}+1\right).$$

We can write the left-hand side as

$$\sum_{n: dn+a \in I} \chi(f(n))$$

where

$$f(n) := \prod_{j \in J} \frac{dn+a+h_j}{d'_j}.$$

Suppose that there is a prime p not dividing d such that f is a constant multiple of a square modulo p. Then the roots $-\frac{a+h_j}{d}\ (p)$ of f must experience a repetition, and hence p divides $h_i - h_j$ for some $1 \leqslant i < j \leqslant k+\ell$. Since the $h_1, \dots, h_{k+\ell}$ are fixed, this forces $p = O(1)$. From the Chinese remainder theorem (and the fact that q_χ is a fundamental discriminant), we conclude that the largest factor d' of q_χ for which f is a constant multiple of a square modulo d' is $O((d, q_\chi))$. The claim now follows from (3.27). □

3.5 | Consequences of Kloosterman sum bounds

We recall[†] Estermann's form [5]

$$\left|\sum_{x \in \mathbb{Z}/q\mathbb{Z}: (x,q)=1} e_q(u_1 x + u_2 x^*)\right| \leqslant \tau(q) q^{1/2} (u_1, u_2, q)^{1/2}$$

of the Weil bound for Kloosterman sums, where x^* is the inverse of x in $\mathbb{Z}/q\mathbb{Z}$ and u_1, u_2 are arbitrary integers. From this and a simple change of variables, we see that

$$|\mathbb{E}_{n_1, n_2 \in \mathbb{Z}} 1_{wn_1 n_2 = a\ (q)} e_q(u_1 n_1 + u_2 n_2)| \leqslant \tau(q) q^{-3/2} (u_1, u_2, q)^{1/2} \tag{3.28}$$

[†] For the applications in this paper, one could also proceed using the weaker but more elementary bounds of Kloosterman [13], as the important thing is that we gain a power savings over the trivial bound of q, at the cost of degrading the numerical exponent $10k$ in (1.5) somewhat. We leave the details of this variant of the argument to the interested reader.

for any natural number q and integers w, a, u_1, u_2 with $(w, q) = (a, q) = 1$, where we use the averaging notation

$$\mathbb{E}_{n_1,n_2\in\mathbb{Z}} f(n_1, n_2) := \frac{1}{L^2} \sum_{n_1=1}^{L} \sum_{n_2=1}^{L} f(n_1, n_2)$$

whenever $f : \mathbb{Z}^2 \to \mathbb{C}$ is a periodic function with some period L (thus, $f(n_1 + Lm_1, n_2 + Lm_2) = f(n_1, n_2)$ for all integers n_1, n_2, m_1, m_2).

We will need to extend the bound (3.28) to the case where a shares a common factor with q, and where we also insert a periodic weight:

Lemma 3.8 (Fourier coefficients on a hyperbola). *Let q be a natural number, and let a, u_1, u_2 be integers. Let q_0 be a factor of q such that $(a, q)|q_0$. Let $f : \mathbb{Z}^2 \to \mathbb{C}$ be a 1-bounded[†] function with period q_0. Then*

$$|\mathbb{E}_{n_1,n_2\in\mathbb{Z}} f(n_1, n_2)1_{n_1n_2=a\ (q)} e_q(u_1n_1 + u_2n_2)| \leqslant \tau(q_0)^2 q_0^{3/2} \tau(q) q^{-3/2} (u_1, u_2, q)^{1/2}.$$

The factor $\tau(q_0)^2 q_0^{3/2}$ can be improved somewhat, but we will not attempt to optimize it here. This bound will be needed in step (iv) of the main argument, in order to dispose of the non-Type I portion $\Lambda^{\flat}_{\text{Siegel}}$ to the Siegel approximant Λ_{Siegel}.

Proof. If $n_1n_2 = a\ (q)$, then from considering the prime factorizations of n_1, n_2, a, q we see that $(n_1, q_0), (n_2, q_0)$ must be factors of (a, q) and hence of q_0; also, we have $((n_1, q_0)(n_2, q_0), q) = (a, q_0) = (a, q)$. Thus there are at most $\tau(q_0)^2$ possible choices for $(n_1, q), (n_2, q)$, and by the triangle inequality it suffices to show that

$$|\mathbb{E}_{n_1,n_2\in\mathbb{Z}} f(n_1, n_2)1_{n_1n_2=a\ (q)} e_q(u_1n_1 + u_2n_2)| \leqslant q_0^{3/2} \tau(q) q^{-3/2} (u_1, u_2, q)^{1/2}. \tag{3.29}$$

under the additional hypothesis that f is supported in the region where $(n_1, q_0) = q_1, (n_2, q_0) = q_2$ for some factors q_1, q_2 of q_0 with

$$(q_1q_2, q) = (a, q). \tag{3.30}$$

In particular, if we write $q' := \frac{q}{(a,q)}$, then the quantity $w = \frac{q_1q_2}{(a,q)}$ is a primitive element of $\mathbb{Z}/q'\mathbb{Z}$. Making the change of variables $n_1 = q_1n_1'$, $n_2 = q_2n_2'$, we can now rewrite the left-hand side of (3.29) as

$$\frac{1}{q_1q_2} |\mathbb{E}_{n_1',n_2'\in\mathbb{Z}} f(q_1n_1', q_2n_2') 1_{wn_1'n_2'=\frac{a}{(a,q)}\ (q')} e_q(u_1q_1n_1 + u_2q_2n_2)|.$$

By Fourier inversion and the Plancherel formula, we have

$$f(q_1n_1', q_2n_2') = \sum_{k_1\in\mathbb{Z}/(q_0/q_1)\mathbb{Z}} \sum_{k_2\in\mathbb{Z}/(q_0/q_2)\mathbb{Z}} c_{k_1,k_2} e_{q_0}(k_1q_1n_1' + k_2q_2n_2'),$$

[†] A function f is 1-bounded if $|f(x)| \leqslant 1$ for all x in the domain of f.

where the coefficients c_{k_1,k_2} obey the bound

$$\sum_{k_1 \in \mathbb{Z}/(q_0/q_1)\mathbb{Z}} \sum_{k_2 \in \mathbb{Z}/(q_0/q_2)\mathbb{Z}} |c_{k_1,k_2}|^2 \leqslant 1,$$

and hence by Cauchy–Schwarz

$$\sum_{k_1 \in \mathbb{Z}/(q_0/q_1)\mathbb{Z}} \sum_{k_2 \in \mathbb{Z}/(q_0/q_2)\mathbb{Z}} |c_{k_1,k_2}| \leqslant \frac{q_0}{q_1^{1/2} q_2^{1/2}}.$$

Thus by the triangle inequality and pigeonhole principle, we can bound the left-hand side of (3.29) by

$$\frac{q_0}{q_1^{3/2} q_2^{3/2}} \left| \mathbb{E}_{n_1', n_2' \in \mathbb{Z}} 1_{w n_1' n_2' = \frac{a}{(a,q)} \ (q')} e_q\left(\left(u_1 + k_1 \frac{q}{q_0} \right) q_1 n_1 + \left(u_2 + k_2 \frac{q}{q_0} \right) q_2 n_2 \right) \right|$$

for some integers k_1, k_2. Since $1_{w n_1' n_2' = \frac{a}{(a,q)} \ (q')}$ is a q'-periodic function of n_1', n_2', this expression vanishes unless the integers $(u_1 + k_1 \frac{q}{q_0}) q_1$, $(u_2 + k_2 \frac{q}{q_0}) q_2$ are divisible by $q/q' = (a,q)$. Since w and $\frac{a}{(a,q)}$ are both primitive in $\mathbb{Z}/q'\mathbb{Z}$, we may then apply (3.28) and bound the left-hand side of (3.29) by

$$\frac{q_0}{q_1^{3/2} q_2^{3/2}} (q')^{-3/2} \left(\left(u_1 + k_1 \frac{q}{q_0} \right) \frac{q_1}{(a,q)}, \left(u_2 + k_2 \frac{q}{q_0} \right) \frac{q_2}{(a,q)}, q' \right)^{1/2},$$

which we can rewrite as

$$\frac{q_0}{q_1^{3/2} q_2^{3/2}} (a,q) q^{-3/2} d^{1/2},$$

where

$$d := \left(\left(u_1 + k_1 \frac{q}{q_0} \right) q_1, \left(u_2 + k_2 \frac{q}{q_0} \right) q_2, q \right).$$

By construction, we have

$$\left(u_1 + k_1 \frac{q}{q_0} \right) q_1 = \left(u_2 + k_2 \frac{q}{q_0} \right) q_2 = q = 0 \ (d)$$

and hence by taking suitable linear combinations

$$u_1 q_0 q_1 q_2 = u_2 q_0 q_1 q_2 = q q_0 q_1 q_2 = 0 \ (d).$$

We conclude in particular that $d | q_0 q_1 q_2 (u_1, u_2, q)$, and the claim follows (noting from (3.30) that $(a,q) \leqslant q_1 q_2$). □

From Lemma 3.8 and the Fourier inversion formula, one can express the periodic function $f(n_1, n_2) 1_{n_1 n_2 = a \ (q)}$ as a linear combination of Fourier phases $e_q(u_1 n_1 + u_2 n_2)$ with good bounds on the Fourier coefficients. However, the contribution of those terms in which one of u_1, u_2 is

divisible by q (or by a very large factor of q) will be inconvenient to handle. We therefore perform the following substitute expansion:

Lemma 3.9 (Modified Fourier expansion). *Let q be a natural number, and let a be an integer. Let q_0 be a factor of q such that $(a,q)|q_0$. Let $f : \mathbb{Z}^2 \to \mathbb{C}$ be a 1-bounded function with period q_0. Define $q_0' := (q_0(a,q),q)$. Then we have*

$$f(n_1,n_2)1_{n_1n_2=a\ (q)} = \frac{\alpha q_0'}{q} f(n_1,n_2)1_{n_1n_2=a\ (q_0')}1_{(n_1n_2,q)=(a,q)} + \sum_{u_1,u_2\in\mathbb{Z}/q\mathbb{Z}:\ \frac{q}{q_0}\nmid u_1,u_2} c_{u_1,u_2}e_q(u_1n_1+u_2n_2),$$

where α is the quantity

$$\alpha := \prod_{p\mid \frac{q}{q_0'};p\nmid \frac{q_0'}{(a,q)}} \frac{p}{p-1}$$

and the coefficients c_{u_1,u_2} obey the bounds

$$|c_{u_1,u_2}| \leqslant 2\tau(q_0)^2 q_0^{3/2}\tau(q)q^{-3/2}(u_1,u_2,q)^{1/2}.$$

Proof. We may assume without loss of generality that $1 \leqslant a \leqslant q$. Let $\mathcal{A}$ denote the collection of those $1 \leqslant a' \leqslant q$ such that $a' = a\ (q_0')$ and $(a',q) = (a,q)$. From Lemma 3.8, we see that the Fourier coefficient

$$\mathbb{E}_{n_1,n_2\in\mathbb{Z}} f(n_1,n_2)(1_{n_1n_2=a\ (q)} - 1_{n_1n_2=a'\ (q)})e_q(u_1n_1+u_2n_2) \tag{3.31}$$

for $u_1,u_2 \in \mathbb{Z}/q\mathbb{Z}$ is bounded in magnitude by $2\tau(q_0)^2q_0^{3/2}\tau(q)q^{-3/2}(u_1,u_2,q)^{1/2}$ for any $a' \in \mathcal{A}$. We claim furthermore that this Fourier coefficient vanishes whenever one of u_1,u_2 is divisible by q/q_0. Indeed, suppose, for instance, that u_2 is divisible by q/q_0, so that $n_2 \mapsto f(n_1n_2)e_q(u_1n_1 + u_2n_2)$ is q_0-periodic for any n_1. To obtain the vanishing of (3.31), it suffices to show that

$$\sum_{n_2\in\mathbb{Z}/q\mathbb{Z}:n_2=a_2\ (q_0)} 1_{n_1n_2=a\ (q)} = \sum_{n_2\in\mathbb{Z}/q\mathbb{Z}:n_2=a_2\ (q_0)} 1_{n_1n_2=a'\ (q)} \tag{3.32}$$

for any integers n_1,a_2. But since $(a',q) = (a,q)$, we can write $a' = wa\ (q)$ for some primitive $w \in \mathbb{Z}/q\mathbb{Z}$; since $a' = a\ ((q_0(a,q),q))$ we have $w = 1\ ((q_0,q/(a,q)))$; as we have the freedom to adjust w by an arbitrary multiple of $q/(a,q)$, we may in fact assume that $w = 1\ (q_0)$. The claim (3.32) then follows after applying the change of variables $n_2 \mapsto wn_2$ on the right-hand side. We argue similarly if u_1 is divisible by q/q_0 instead of u_2.

Averaging in a', we conclude that the Fourier coefficient

$$\mathbb{E}_{n_1,n_2\in\mathbb{Z}} f(n_1,n_2)(1_{n_1n_2=a\ (q)} - \mathbb{E}_{a'\in\mathcal{A}}1_{n_1n_2=a'\ (q)})e_q(u_1n_1+u_2n_2)$$

is bounded in magnitude by $2\tau(q_0)^2q_0^{3/2}\tau(q)q^{-3/2}(u_1,u_2,q)^{1/2}$, and vanishes whenever u_1 or u_2 vanish in $\mathbb{Z}/q\mathbb{Z}$. To establish the claim, it now suffices by the Fourier inversion formula to obtain

the identity

$$\mathbb{E}_{a'\in\mathcal{A}} 1_{n=a'\ (q)} = \frac{\alpha q_0'}{q} 1_{n=a\ (q_0')} 1_{(n,q)=(a,q)}$$

for any integer n. By the Chinese remainder theorem, it suffices to establish this identity at each prime p, that is to say it suffices to show that

$$\mathbb{E}_{a'=a\ (p^{j_0}):(a',p^j)=(a,p^j)} 1_{n=a'\ (p^j)} = \frac{\alpha_p p^{j_0}}{p^j} 1_{n=a\ (p^{j_0})} 1_{(n,p^j)=(a,p^j)}$$

whenever p is a prime, $0 \leqslant j_0 \leqslant j$, and a is an integer with $(a,p^j)|p^{j_0}$, where $\alpha_p := \frac{p}{p-1}$ if $j > j_0$ and $(a,p^j) = p^{j_0}$, and $\alpha_p = 1$ otherwise. But this follows by a direct case analysis.

- If $j = j_0$, then the conditions $(a',p^j) = (a,p^j)$ and $(n,p^j) = (a,p^j)$ are redundant, $\alpha_p = 1$, a' is restricted to a single residue class mod p^j, and both sides are equal to $1_{n=a\ (p^{j_0})}$.
- If $j < j_0$ and $(a,p^j) < p^{j_0}$, then the conditions $(a',p^j) = (a,p^j)$ and $(n,p^j) = (a,p^j)$ are redundant, $\alpha_p = 1$, a' is restricted to p^{j-j_0} residue classes mod p^j, and both sides are equal to $\frac{1}{p^{j-j_0}} 1_{n=a\ (p^{j_0})}$.
- If $j < j_0$ and $(a,p^j) = p^{j_0}$, then $\alpha_p = \frac{p}{p-1}$, a' is restricted to $\frac{p-1}{p} p^{j-j_0}$ residue classes mod p^j, and both sides are equal to $\frac{p}{p-1}\frac{1}{p^{j-j_0}} 1_{n=a\ (p^{j_0})}$. □

4 | FIRST STEP: REPLACING THE LIOUVILLE FUNCTION WITH A SIEGEL MODEL

We now execute step (i) of the strategy outlined in the introduction. From (2.6), we have the splitting

$$\lambda = \lambda_{(\leqslant R)} * \lambda_{(>R)}.$$

In view of Corollary 3.6, we expect λ to resemble the exceptional character χ on the rough numbers $\mathbb{N}_{(>R)}$. It is therefore natural to introduce the Siegel approximant

$$\lambda_{\text{Siegel}} := \lambda_{(\leqslant R)} * \chi_{(>R)}, \tag{4.1}$$

thus λ_{Siegel} is the completely multiplicative function that agrees with λ for primes $p \leqslant R$ and agrees with χ for primes $p > R$. Similar approximants were also introduced in [2, 6]. Clearly $\lambda, \lambda_{\text{Siegel}}$ are both bounded by 1:

$$|\lambda(n)|, |\lambda_{\text{Siegel}}(n)| \leqslant 1. \tag{4.2}$$

The error between λ and λ_{Siegel} can be controlled by exceptional primes and by rough numbers:

Lemma 4.1 (Error bound between λ and λ_{Siegel}). *For any natural number $n \leqslant 2x$, one has*

$$\lambda(n) - \lambda_{\text{Siegel}}(n) \ll \sum_{p^*|n, R<p^*\leqslant x/R} 1 + \sum_{d\leqslant 2R\,:\,d|n} 1_{(\geqslant x/R)}(n/d). \tag{4.3}$$

Proof. If n is not divisible by any exceptional prime $p^* > R$, then we have $\lambda(n) = \lambda_{\text{Siegel}}(n)$ since $\lambda, \lambda_{\text{Siegel}}$ agree on every prime dividing n. Clearly (4.3) holds in this case. If n is divisible by an exceptional $R < p^* \leqslant x/R$, then the first term on the right-hand side of (4.3) is at least one, and the claim (4.3) then follows from (4.2).

The only remaining case is if n is divisible by an exceptional prime $p^* \geqslant x/R$, so $n = p^* d$ for some $d \leqslant 2R$. Since $n/d = p^* \geqslant x/R$ is prime, the second term on the right-hand side of (4.3) is at least one, and the claim (4.3) again follows from (4.2). □

In this section, we establish

Proposition 4.2 (Replacing λ with a Siegel model). *We have*

$$\begin{aligned}&\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda(n+h_1')\cdots\lambda(n+h_\ell')\\&\quad\approx \mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda_{\text{Siegel}}(n+h_1')\cdots\lambda_{\text{Siegel}}(n+h_\ell').\end{aligned}$$

From (4.2) and the triangle inequality, it suffices to show that

$$\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)|\lambda(n+h_{j'}')-\lambda_{\text{Siegel}}(n+h_{j'}')|\approx 0$$

for each $1 \leqslant j' \leqslant \ell$. Applying Lemma 4.1, it suffices to show the bounds

$$\sum_{R<p^*\leqslant x/R}\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)1_{p^*|n+h_{j'}'}\approx 0 \tag{4.4}$$

and

$$\sum_{d\leqslant 2R}\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)1_{d|n+h_{j'}'}1_{(\geqslant x/R)}\left(\frac{n+h_{j'}'}{d}\right)\approx 0. \tag{4.5}$$

We begin with (4.4). For $n \leqslant x$ and $1 \leqslant j \leqslant k$, the quantity $\Lambda(n+h_j)$ is bounded by $\log(2x)1_{(\geqslant\sqrt{2x})}(n+h_j)$ unless we are in the exceptional case where $n+h_j$ is of the form p^i for some prime $p < \sqrt{2x}$ (cf. the sieve of Eratosthenes). The contribution of such exceptional cases can easily be shown to be ≈ 0, so it suffices to show that

$$(\log^k x)\sum_{R<p^*\leqslant x/R}\mathbb{E}_{n\leqslant x}1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{p^*|n+h_{j'}'}\approx 0.$$

Let p^* be as in the above sum. Changing variables, we have

$$\begin{aligned}&\mathbb{E}_{n\leqslant x}1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{p^*|n+h_{j'}'}\\&\ll\frac{1}{p^*}\mathbb{E}_{n\leqslant 2x/p^*}1_{(\geqslant\sqrt{2x})}(p^*n+h_1-h_{j'}')\cdots 1_{(\geqslant\sqrt{2x})}(p^*n+h_k-h_{j'}').\end{aligned}$$

Let C_0 be a sufficiently large constant depending on $h_1,\dots,h_k,h_{j'}'$. Then for any prime $C_0 < p \leqslant \sqrt{2x}$ other than p^*, the support set of $1_{(\geqslant\sqrt{2x})}(p^*n+h_1-h_{j'}')\cdots 1_{(\geqslant\sqrt{2x})}(p^*n+h_k-h_{j'}')$ excludes

k residue classes modulo p. Thus by[†] Lemma 3.2, we have

$$\mathbb{E}_{n\leqslant x} 1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{p|n+h'_{j'}} \ll \frac{1}{p^*}\prod_{C_0<p\leqslant \min(2x/p^*,\sqrt{2x}):\, p\neq p^*}\left(1-\frac{k}{p}\right)$$

and hence by Mertens' theorem (3.6) and the bound $p_* \leqslant x/R$

$$(\log^k x)\mathbb{E}_{n\leqslant x} 1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{p^*|n+h'_{j'}} \ll \frac{\log_R^k x}{p^*}.$$

The claim (4.4) now follows from Corollary 3.6 and (2.17).

Now we prove (4.5). Arguing as in the proof of (4.4), it suffices to show that

$$(\log^k x)\sum_{d\leqslant 2R}\mathbb{E}_{n\leqslant x} 1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{d|n+h'_{j'}}1_{(\geqslant x/R)}\left(\frac{n+h'_{j'}}{d}\right)\approx 0. \tag{4.6}$$

For $d\leqslant 2R$, we have after change of variables that

$$\begin{aligned}&\mathbb{E}_{n\leqslant x} 1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{d|n+h'_{j'}}1_{(\geqslant x/R)}\left(\frac{n+h'_{j'}}{d}\right)\\ &\ll \frac{1}{d}\mathbb{E}_{n\leqslant 2x/d} 1_{(\geqslant\sqrt{2x})}(dn+h_1-h'_{j'})\cdots 1_{(\geqslant\sqrt{2x})}(dn+h_k-h'_{j'})1_{(\geqslant x/R)}(n).\end{aligned}$$

With C_0 as before, we see that for any prime $C_0\leqslant p<\sqrt{2x}$ not dividing d, we are excluding $k+1$ residue classes modulo p (since $h'_{j'}$ is distinct from $h_1,\dots,h_k$), hence by Lemma 3.2

$$\mathbb{E}_{n\leqslant x} 1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{d|n+h'_{j'}}1_{(\geqslant x/R)}\left(\frac{n+h'_{j'}}{d}\right)\ll\frac{1}{d}\prod_{C_0\leqslant p<\sqrt{2x}:\, p\nmid d}\left(1-\frac{k+1}{p}\right)$$

and hence by Mertens' theorem (3.6)

$$\begin{aligned}&(\log^k x)\mathbb{E}_{n\leqslant x} 1_{(\geqslant\sqrt{2x})}(n+h_1)\cdots 1_{(\geqslant\sqrt{2x})}(n+h_k)1_{d|n+h'_{j'}}1_{(\geqslant x/R)}\left(\frac{n+h'_{j'}}{d}\right)\\ &\ll \frac{1}{d\log x}\prod_{p|d}\left(1+O\left(\frac{1}{p}\right)\right).\end{aligned}$$

By (2.5), we may therefore bound the left-hand side of (4.6) by

$$\frac{1}{\log x}\prod_{p\leqslant 2R}\left(1+\frac{1}{p}+O\left(\frac{1}{p^2}\right)\right).$$

By (3.6). this latter quantity is $O(\log_x R)$, and the claim follows from (2.13).

[†] One could also use Lemma 3.4 here instead if desired to give a comparable estimate.

5 | SECOND STEP: REPLACING THE VON MANGOLDT FUNCTION WITH A SIEGEL MODEL

We now execute step (ii) of the strategy outlined in the introduction. In order to (mostly) restrict to rough numbers, we will insert the Selberg sieve ν defined in (2.23). Namely, observe that

$$\Lambda - \Lambda\nu$$

is supported on prime powers p^j with $p \leqslant R$ and can be crudely bounded by $O(\log^2 x)$ on such powers. Since the number of such powers of size $O(x)$ is crudely bounded by $O(R \log x)$, one easily sees from the triangle inequality that

$$\begin{aligned}&\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell')\\ &\quad\approx \mathbb{E}_{n\leqslant x}\Lambda\nu(n+h_1)\cdots\Lambda\nu(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell')\end{aligned} \tag{5.1}$$

(with plenty of room to spare in the error term). Next, we expand

$$\Lambda\nu = (\mu * \log)\nu.$$

Since μ is expected to be close to χ on rough numbers, and the Selberg sieve ν is mostly restricted to such numbers, it is then natural to introduce the Siegel approximant

$$\Lambda_{\mathrm{Siegel}} := (\chi * \log)\nu.$$

From the triangle inequality, we have the crude bounds

$$\Lambda\nu(n), \Lambda_{\mathrm{Siegel}}(n) \ll \tau\nu(n)\log x. \tag{5.2}$$

We also have the following bound for the error between $\Lambda\nu$ and $\Lambda_{\mathrm{Siegel}}$:

Lemma 5.1. *For $n \leqslant 2x$, we have the bounds*

$$\Lambda\nu(n) - \Lambda_{\mathrm{Siegel}}(n) \ll E(n) + F(n) + G(n), \tag{5.3}$$

where

$$E(n) := \left(\sum_{R_0<p^*\leqslant\sqrt{2x}} 1_{p^*|n} + \sum_{R_0<p\leqslant\sqrt{2x}} 1_{p^2|n}\right)\tau\nu(n) \tag{5.4}$$

$$\log x.F(n) := \left(\sum_{1<d\leqslant D\,:\,d|n} \tau_{(\leqslant R_0)}(d)^{O(1)}\right)\nu(n)\log x. \tag{5.5}$$

$$G(n) := \sum_{\sqrt{2x}<p^*\leqslant 2x/R^{1/2}} 1_{p^*|n}\Lambda\nu(n/p^*). \tag{5.6}$$

Proof. If n is divisible by an exceptional $R_0 < p^* \leqslant \sqrt{2x}$, then $E \gg \tau(n)\nu(n)\log x$, and (5.3) then follows from (5.2) and (5.4). Similarly if n is divisible by the square of a prime $p > R_0$ (which must then necessarily be at most $\sqrt{2x}$).

Next, suppose that n is not divisible by any exceptional prime $p^* > R_0$, nor by any square p^2 of a prime $p > R_0$. We write

$$\chi * \log = (1 * \chi) * \mu * \log = (1 * \chi) * \Lambda.$$

Note that $1 * \chi(d)$ is only non-zero when d is the product of exceptional primes times a perfect square, so if $d|n$ and n is as above, then d must be the product of some primes less than or equal to R_0. Also $\sum_{d|n} \Lambda(d) = \log n$. Thus, for n as above, we have

$$\chi * \log(n) \leqslant \tau(n_{(\leqslant R_0)}) \log n,$$

where we recall that $n_{(\leqslant R_0)}$ is the largest factor of n that is the product of primes less than or equal to R_0. Applying (3.10) (with n replaced by $n_{(\leqslant R_0)}$), we have

$$\tau(n_{(\leqslant R_0)}) \ll \sum_{1<d\leqslant D : d|n} \tau_{(\leqslant R_0)}(d)^{O(1)}$$

and the claim (5.3) now follows in this case from (5.5).

We are left with the case where n is divisible by an exceptional prime $p^* > \sqrt{2x}$. Then $n = dp^*$ for some $d < \sqrt{2x}$. The only factors of n that are less than or equal to R are factors of d as well, thus $\nu(n) = \nu(d)$. Since

$$\chi * \log(n) = \chi * 1 * \mu * \log(n) = \chi * 1 * \Lambda(n)$$

and $\chi * 1$ vanishes at all factors of n except for 1 and p^*, we have

$$\chi * \log(n) = \Lambda(n) + (1 + \chi(p^*))\Lambda(d)$$

and thus

$$\Lambda(n) - \Lambda_{\mathrm{Siegel}}(n) \ll \Lambda\nu(d).$$

If $p^* > 2x/R^{1/2}$, then $d \leqslant R^{1/2}$, and hence $\nu(d)$ vanishes by (2.23). The claim (5.3) now follows in this case from (5.6). □

Now we can prove

Proposition 5.2 (Replacing Λ with a Siegel model). *We have*

$$\begin{aligned}&\mathbb{E}_{n\leqslant x}\Lambda(n+h_1)\cdots\Lambda(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell')\\ &\quad\approx \mathbb{E}_{n\leqslant x}\Lambda_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda_{\mathrm{Siegel}}(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell').\end{aligned}$$

In view of (5.1), it suffices to show that

$$\begin{aligned}&\mathbb{E}_{n\leqslant x}\Lambda\nu(n+h_1)\cdots\Lambda\nu(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell')\\ &\quad\approx \mathbb{E}_{n\leqslant x}\Lambda_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda_{\mathrm{Siegel}}(n+h_k)\lambda_{\mathrm{Siegel}}(n+h_1')\cdots\lambda_{\mathrm{Siegel}}(n+h_\ell').\end{aligned}$$

By (4.2) and the triangle inequality, it suffices to show that

$$\mathbb{E}_{n\leqslant x}|\Lambda\nu(n+h_1)\cdots\Lambda\nu(n+h_k)-\Lambda_{\text{Siegel}}(n+h_1)\cdots\Lambda_{\text{Siegel}}(n+h_k)|\approx 0.$$

From (5.3), we have

$$\Lambda_{\text{Siegel}}(n+h_j)=\Lambda\nu(n+h_j)+O(E(n+h_j)+F(n+h_j)+G(n+h_j))$$

for $j=1,\dots,k$. Multiplying these estimates together, we conclude that

$$\begin{aligned}\Lambda_{\text{Siegel}}(n+h_1)\cdots\Lambda_{\text{Siegel}}(n+h_k)&=\Lambda\nu(n+h_1)\cdots\Lambda\nu(n+h_k)\\&\quad+O\Bigg(\sum_{j=1}^{k}E(n+h_j)\prod_{1\leqslant j'\leqslant k:\,j'\neq j}(\Lambda\nu+E+F+G)(n+h_{j'})\Bigg)\\&\quad+O\Bigg(\sum_{j=1}^{k}F(n+h_j)\prod_{1\leqslant j'\leqslant k:\,j'\neq j}(\Lambda\nu+F+G)(n+h_{j'})\Bigg)\\&\quad+O\Bigg(\sum_{j=1}^{k}G(n+h_j)\prod_{1\leqslant j'\leqslant k:\,j'\neq j}(\Lambda\nu+G)(n+h_{j'})\Bigg).\end{aligned}$$

By the triangle inequality and relabeling, it thus suffices to establish the bounds

$$\mathbb{E}_{n\leqslant x}E(n+h_1)\prod_{j=2}^{k}(\Lambda\nu+E+F+G)(n+h_j)\approx 0 \tag{5.7}$$

and

$$\mathbb{E}_{n\leqslant x}F(n+h_1)\prod_{j=2}^{k}(\Lambda\nu+F+G)(n+h_j)\approx 0 \tag{5.8}$$

and

$$\mathbb{E}_{n\leqslant x}G(n+h_1)\prod_{j=2}^{k}(\Lambda\nu+G)(n+h_j)\approx 0. \tag{5.9}$$

We begin with (5.9), which is a variant of (4.5). We can bound $(\Lambda\nu+G)(n+h_j)$ by $O(\log(2x)1_{(\geqslant R^{1/4})}(n+h_j))$, and we also have the bound

$$G(n+h_1)\ll\log(2x)\sum_{\sqrt{2x}<p^*\leqslant 2x/R^{1/2}}1_{p^*|n+h_1}1_{(\geqslant R^{1/4})}(n+h_1)$$

unless $n+h_j$ is of the form p^m for some $p<R^{1/4}$ and $m\geqslant 1$, or $p'p^m$ for some $p<R^{1/4}$, $m\geqslant 2$, and $\sqrt{2x}\leqslant p'\leqslant 2x/R^{1/2}$.

There are only $O(x \log x/R^{1/4})$ such exceptional values of n and their contribution is easily seen to be negligible using (3.1). Thus it will suffice to show that

$$(\log^k x) \sum_{\sqrt{2x}<p^*\leqslant 2x/R^{1/2}} \mathbb{E}_{n\leqslant x} 1_{p^*|n+h_1} \prod_{j=1}^{k} 1_{(\geqslant R^{1/4})}(n+h_j) \approx 0. \tag{5.10}$$

Making the change of variables $n = p^* n' - h_1$ and using Lemma 3.2 and Mertens' theorem (3.6), we see that

$$\mathbb{E}_{n\leqslant x} 1_{p^*|n+h_1} \prod_{j=1}^{k} 1_{(\geqslant R^{1/4})}(n+h_j) \ll \frac{1}{p^* \log^k R}.$$

The claim (5.10) now follows from Corollary 3.6 and (2.8).

Now we turn to (5.7). Observe using (3.10) that

$$(\Lambda\nu + E + F + G)(n+h_j) \ll \left(\sum_{d_j\leqslant D : d_j|n+h_j} \tau(d_j)^{O(1)} \right) \nu(n+h_j) \log x$$

and so we can bound the left-hand side of (5.7) by

$$\ll (\log^k x) \left(\sum_{R_0<p^*\leqslant\sqrt{2x}} a_{p^*} + \sum_{R_0<p_0\leqslant\sqrt{2x}} a_{p_0^2} \right) \tag{5.11}$$

where

$$a_d := \sum_{d_1,\dots,d_k\leqslant D} \tau(d_1\cdots d_k)^{O(1)} \mathbb{E}_{n\leqslant x} 1_{d|n+h_1} \prod_{j=1}^{k} 1_{d_j|n+h_j} \nu(n+h_j).$$

Let d be equal to p_0 or p_0^2 for some prime $p_0 \leqslant \sqrt{2x}$. If $d > \sqrt{2x}$, then $d = p_0^2$ for some $p_0 \gg x^{1/4}$. From (3.1), one has the crude bound $a_d \ll_\varepsilon \frac{x^\varepsilon}{d} = \frac{x^\varepsilon}{p_0^2}$ in this case, which certainly gives a negligible contribution. Hence we may assume that $d \leqslant \sqrt{2x}$.

Applying Lemma 3.4 and (3.1), we have

$$a_d \ll_A \frac{1}{\log^k R} \int_1^\infty f(\sigma) \frac{d\sigma}{\sigma^A} + \frac{D^{k+1}R^{2k}}{x}$$

for any $A > 0$, where

$$f(\sigma) := \sum_{d_1,\dots,d_k\leqslant\sqrt{2x}} \frac{\tau([d,d_1]d_2\cdots d_k)^{O(1)}}{[d,d_1]d_2\cdots d_k} \left(\prod_{p|dd_1\cdots d_k} \min(\sigma \log_R p, 1) \right).$$

Using Euler products (2.5), we can bound

$$f(\sigma) \leqslant \prod_{p\leqslant\sqrt{2x}} E_p(\sigma),$$

where

$$E_p(\sigma) := \sum_{d_1,\dots,d_k \in \mathbb{N}_{(p)}} \frac{\tau([d_{(p)}, d_1]d_2 \cdots d_k)^{O(1)}}{[d_{(p)}, d_1]d_2 \cdots d_k} \min(\sigma \log_R p, 1)^{1_{p|dd_1\cdots d_k}}.$$

If $p \neq p_0$, then $d_{(p)} = 1$, and we can calculate

$$E_p(\sigma) \leqslant 1 + O\left(\frac{\min(\sigma \log_R p, 1)}{p}\right).$$

From (3.4), we then have

$$\prod_{p \leqslant \sqrt{2x}} E_p(\sigma) \ll (1 + \sigma \log_R \sqrt{2x})^{O(1)} E_{p_0}(\sigma) \ll (\log_R^{O(1)} x)\sigma^{O(1)} E_{p_0}(\sigma).$$

Also, we have the crude bound

$$E_{p_0}(\sigma) \ll \frac{\tau(d)^{O(1)}}{d}.$$

Putting all these estimates together, and choosing A large enough, we conclude that

$$a_d \ll \tau(d)^{O(1)} \frac{\log_R^{O(1)} x}{d \log^k R} + \frac{D^{k+1}R^{2k}}{x}.$$

Inserting this into (5.11) and using Corollary 3.6, (2.17), (2.16), we obtain the claim (5.7).

Finally, we establish (5.8). Observe from (5.5) and (5.6) that

$$(\Lambda\nu + F + G)(n + h_{j'}) \ll \left(\sum_{d_j \leqslant D : d_j | n + h_j} \tau_{(\leqslant R_0)}(d_j)^{O(1)}\right) \nu(n + h_j) \log x$$

and so it suffices to show that

$$(\log^k x) \sum_{d_1,\dots,d_k \leqslant D : d_1 > 1} \left(\prod_{j=1}^k \tau_{(\leqslant R_0)}(d_j)^{O(1)}\right) \mathbb{E}_{n \leqslant x} \prod_{j=1}^k 1_{d_j | n + h_j} \nu(n + h_j) \approx 0.$$

Applying Lemma 3.4 and (3.1), we may estimate the left-hand side as

$$\ll_A (\log_R^k x) \sum_{d_1,\dots,d_k \leqslant D : d_1 > 1} \frac{\prod_{j=1}^k \tau_{(\leqslant R_0)}(d_j)^{O(1)}}{d_1 \cdots d_k} \int_1^\infty \prod_{p | d_1 \cdots d_k} \min(\sigma \log_R p, 1) \frac{d\sigma}{\sigma^A} + \frac{R^{2k} D^{k+1} (\log^k x)}{x}$$

for any $A > 0$. The second term is ≈ 0 by (2.16). Replacing the condition $d_1 > 1$ by $(d_1, \dots, d_k) \neq (1, \dots, 1)$, removing the constraints $d_1, \dots, d_k \leqslant D$, and factoring the Euler product using (2.5), the first term can be bounded by

$$(\log_R^k x) \int_1^\infty \left(\prod_{p \leqslant R_0} \tilde{E}_p(\sigma) - 1\right) \frac{d\sigma}{\sigma^A}, \tag{5.12}$$

where

$$\tilde{E}_p(\sigma) := \sum_{d_1,\dots,d_k \in \mathbb{N}_{(p)}\rangle} \frac{\prod_{j=1}^k \tau(d_j)^{O(1)}}{d_1 \cdots d_k} \min(\sigma \log_R p, 1)^{1_{p|d_1\cdots d_k}}.$$

Direct calculation gives

$$\tilde{E}_p(\sigma) = 1 + O\left(\frac{\min(\sigma \log_R p, 1)}{p}\right).$$

From (3.4), we have

$$\prod_{p \leqslant R_0} \tilde{E}_p(\sigma) \leqslant (1 + \sigma \log_R R_0)^{O(1)}$$

and hence

$$\left(\prod_{p \leqslant R_0} \tilde{E}_p(\sigma)\right) - 1 \ll \sigma^{O(1)} \log_R R_0.$$

We can thus bound (5.12) for A large enough by

$$\ll (\log_R^k x) \log_R R_0,$$

which is ≈ 0 by (2.14). This concludes the proof of (5.9) and hence of Proposition 5.2.

6 | THIRD STEP: REPLACING THE LIOUVILLE SIEGEL MODEL WITH A TYPE I APPROXIMANT

We now execute step (iii) of the strategy outlined in the introduction. From (4.1), (2.7), (2.6) and Möbius inversion, we have

$$\begin{aligned}\lambda_{\text{Siegel}} &= \lambda_{(\leqslant R)} * (\mu\chi)_{(\leqslant R)} * \chi_{(\leqslant R)} * \chi_{(>R)} \\ &= (\lambda * \mu\chi)_{(\leqslant R)} * \chi.\end{aligned}$$

We now split

$$\lambda_{\text{Siegel}} = \lambda^\sharp_{\text{Siegel}} + \lambda^\flat_{\text{Siegel}}, \tag{6.1}$$

where $\lambda^\sharp_{\text{Siegel}}$ is the Type I approximant

$$\lambda^\sharp_{\text{Siegel}} := (\lambda * \mu\chi)_{(\leqslant R)} \psi_{\leqslant D} * \chi \tag{6.2}$$

and $\lambda^\flat_{\text{Siegel}}$ is the error

$$\lambda^\flat_{\text{Siegel}} := (\lambda * \mu\chi)_{(\leqslant R)} \psi_{>D} * \chi. \tag{6.3}$$

Here $\psi_{\leqslant D}, \psi_{>D}$ are the smooth cutoffs defined in (2.21) and (2.22). In particular, we see that $\lambda_{\text{Siegel}}(n) = \lambda^{\sharp}_{\text{Siegel}}(n)$ whenever $n_{(\leqslant R)} \leqslant \sqrt{D}$. Since $\sqrt{D}$ is significantly larger than R, and R-smooth numbers become extremely sparse at scales much larger than R, we thus see that λ_{Siegel}, $\lambda^{\sharp}_{\text{Siegel}}$ agree with each other for 'typical' n, and would thus be heuristically expected to be close to each other; in other words, $\lambda^{\flat}_{\text{Siegel}}$ would be expected to be small on average.

Unfortunately, $\lambda^{\sharp}_{\text{Siegel}}, \lambda^{\flat}_{\text{Siegel}}$ are not bounded. However, we can still obtain a reasonable bound on the latter quantity:

Lemma 6.1. *For any $n \leqslant 2x$, we have*

$$\lambda^{\flat}_{\text{Siegel}}(n) \ll H(n), \tag{6.4}$$

where

$$H(n) := \sum_{d \leqslant D : d|n} \alpha(d) \tag{6.5}$$

and $\alpha(d)$ are non-negative quantities obeying the bounds

$$\sum_{d \leqslant D} \tau(d)^A \frac{\alpha(d)}{d} \ll_A \exp\left(-\frac{1}{8}\log_R D\right) \tag{6.6}$$

for any $A \geqslant 1$.

Remark 6.2. Note that by (2.19), we have

$$\exp(-\frac{1}{8}\log_R D) \ll_A \log^{-A} \eta \tag{6.7}$$

for any $A \geqslant 1$. We shall need (6.7) later, but we stated Lemma 6.1 in a stronger form to emphasize that it does not use any information on exceptional characters.

Proof. From (6.3) and the triangle inequality, we have

$$\lambda^{\flat}_{\text{Siegel}}(n) = \beta * \chi_{(>R)}(n) \ll |\beta| * 1_{(>R)}(n), \tag{6.8}$$

where

$$\beta := (\lambda * \mu\chi)_{(\leqslant R)} \psi_{>D} * \chi_{(\leqslant R)}. \tag{6.9}$$

To control β, we perform a Fourier expansion on $\psi_{>D}$, which is the only term on the right-hand side of (6.9) which is not multiplicative. Applying Fourier inversion (3.13) to the function $g(u) := e^{-u}(1 - \psi((\log_D R)u)$ and setting $u := \log_R n$, we conclude the identity

$$\psi_{>D}(n) = \int_{\mathbb{R}} n^{\frac{1+it}{\log R}} f(t)\, dt, \tag{6.10}$$

where

$$f(t) := \frac{1}{2\pi}\int_0^{\infty} e^{-(1+it)x}(1 - \psi((\log_D R)x))\, dx.$$

From the triangle inequality, we have

$$f(t) \ll \exp(-\frac{1}{4}\log_R D) \ll_A \log^{-A}\eta$$

for any $A > 0$, while from repeated integration by parts, we have

$$f(t) \ll_A \frac{1}{|t|^A}$$

for any positive integer A. Combining the two bounds, we conclude that

$$f(t) \ll_A \frac{\exp(-\frac{1}{8}\log_R D)}{(1+|t|)^A} \tag{6.11}$$

for any $A > 0$.

From (6.9) and (6.10), we have

$$\beta(n) = \int_{\mathbb{R}} \beta_t(n) f(t)\, dt,$$

where

$$\beta_t := (\lambda * \mu\chi)_{(\leqslant R)}(\cdot)^{\frac{1+it}{\log R}} * \chi_{(\leqslant R)}. \tag{6.12}$$

From (6.8) and the triangle inequality, we then have

$$\lambda^\flat_{\text{Siegel}}(n) \leqslant \int_{\mathbb{R}} |\beta_t| * 1_{(>R)}(n) |f(t)|\, dt.$$

The function $|\beta_t|$ is multiplicative and supported on $\mathbb{N}_{(\leqslant R)}$, thus

$$|\beta_t| * 1_{(>R)}(n) \ll \prod_{p \leqslant R} |\beta_t(n_{(p)})|.$$

From (6.12), we see that

$$|\beta_t(p^j)| = 1$$

when $p \leqslant R$ and $\chi(p) = -1$ (because $\lambda * \mu\chi$ agrees with $\lambda * \mu\lambda = 1_{\{1\}}$ on $\mathbb{N}_{(p)}$), and

$$|\beta_t(p^j)| = p^{\frac{j}{\log R}} = \exp(j \log_R p)$$

when $p \leqslant R$ and $\chi(p) = 0$. For $\chi(p) = +1$, the situation is more complicated: direct calculation gives

$$\beta_t(p^j) = P_j(p^{\frac{1+it}{\log R}}),$$

where P_j is the polynomial

$$P_j(z) := 1 - 2z + 2z^2 - \cdots + (-1)^j 2z^j.$$

Note that $|P_j(1)| \leqslant 1$ and $P_j'(z) \ll j^{O(1)}(1+|z|^{j-1})$ for any z, hence by the fundamental theorem of calculus

$$|P_j(z)| \leqslant 1 + O(|z-1| j^{O(1)})$$

whenever $|z| \leqslant 1 + \frac{1}{j}$. Also from the triangle inequality, we have $|P_j(z)| \ll j|z|^j$ for $|z| \geqslant 1$. We thus have

$$|P_j(z)| \leqslant \min(1 + j^{O(1)}|z-1|, \exp(O(j \log|z| + 1 + \log j)))$$

for $|z| \geqslant 1$. Thus regardless of the value of $\chi(p)$, we have the upper bound

$$|\beta_t(p^j)| \leqslant \exp(O(a_{t,p^j}))$$

for $p \leqslant R$, where a_{t,p^j} is the quantity

$$a_{t,p^j} := \min(j^C(1+|t|)\log_R p, j \log_R p + 1 + \log j)$$

for some large constant $C \geqslant 1$ and for all $j \geqslant 1$, with the convention $a_{t,1} = 0$. We conclude that

$$|\beta_t| * 1_{(>R)}(n) \ll \exp\left(O\left(\sum_{p \leqslant R} a_{t,n_{(p)}}\right)\right).$$

To convert the right-hand side into Type I sums, we apply Lemma 3.1(i) to split

$$n = n_{(>R)} n_1 \cdots n_m,$$

where $m = O(1)$ and $n_1, \dots, n_m \leqslant D$ lie in $\mathbb{N}_{(\leqslant R)}$. We then have

$$n_{(p)} = (n_1)_{(p)} \cdots (n_m)_{(p)}$$

for all $p \leqslant R$, and hence

$$a_{t,n_{(p)}} \ll \sum_{i=1}^m a_{t,(n_i)_{(p)}}.$$

Using the definition of a_{t,p^j} and the inequality $(j_1+j_2)^C \ll_C j_1^C + j_2^C$, we conclude that

$$|\beta_t| * 1_{(>R)}(n) \ll \exp\left(O\left(\sum_{i=1}^m \sum_{p \leqslant R} a_{t,(n_i)_{(p)}}\right)\right)$$

and hence (since $m = O(1)$), we have

$$\begin{aligned} |\beta_t| * 1_{(>R)}(n) &\ll \exp\left(O\left(\sum_{p \leqslant R} a_{t,(n_i)_{(p)}}\right)\right) \\ &= \prod_{p \leqslant R} \left(1 + (\exp(O(a_{t,(n_i)_{(p)}})) - 1)\right) \end{aligned}$$

for some $i = 1, \dots, m$. In particular, we see that

$$|\beta_t| * 1_{(>R)}(n) \ll \sum_{d \leqslant D : d|n} 1_{(\leqslant R)}(d) \prod_{p \leqslant R} \Big(\exp(O(a_{t,d_{(p)}})) - 1 \Big).$$

We therefore obtain the bound (6.4) with

$$\alpha(d) := 1_{(\leqslant R)}(d) \int_{\mathbb{R}} \prod_{p \leqslant R} \Big(\exp(O(a_{t,d_{(p)}})) - 1 \Big) |f(t)| \, dt.$$

It remains to establish the bound (6.6). We use Fubini's theorem and Euler product expansion (2.5) to bound

$$\sum_{d \leqslant D} \tau(d)^A \frac{\alpha(d)}{d} \ll \int_{\mathbb{R}} \prod_{p \leqslant R} \left(1 + \sum_{j=1}^{\infty} \frac{(1+j)^A (\exp(O(a_{t,p^j})) - 1)}{p^j} \right) |f(t)| \, dt.$$

For $j \geqslant 2$, we use the crude bound

$$\exp(O(a_{t,p^j})) \ll j^{O(1)} p^{O(j/\log R)} \ll e^{O(j)}$$

for $p \leqslant R$ to conclude that

$$\sum_{j=2}^{\infty} \frac{(1+j)^A (\exp(O(a_{t,p^j})) - 1)}{p^j} \ll_A \frac{1}{p^2}$$

for any $p \leqslant R$. For $j = 1$, we have

$$\exp(O(a_{t,p^j})) \leqslant 1 + O(\min((1+|t|) \log_R p, 1)),$$

and thus using $1 + x \leqslant e^x$, we obtain

$$\begin{aligned} 1 + \sum_{j=1}^{\infty} \frac{(1+j)^A (\exp(O(a_{t,p^j})) - 1)}{p^j} &\leqslant 1 + \frac{O_A(\min((1+|t|) \log_R p, 1))}{p} + O_A\left(\frac{1}{p^2}\right) \\ &\leqslant \exp\left(O_A\left(\frac{\min((1+|t|) \log_R p, 1)}{p} + \frac{1}{p^2} \right) \right). \end{aligned}$$

From Mertens' theorem (3.3), we have

$$\begin{aligned} \prod_{p \leqslant R} \exp\left(O_A\left(\frac{\min((1+|t|) \log_R p, 1))}{p} + \frac{1}{p^2} \right) \right) &= \exp\left(O_A\left(\sum_{p \leqslant R} \left(\frac{\min((1+|t|) \log_R p, 1)}{p} + \frac{1}{p^2} \right) \right) \right) \\ &\ll (2+|t|)^{O_A(1)}, \end{aligned}$$

and hence

$$\sum_{d \leqslant D} \tau(d)^A \frac{\alpha(d)}{d} \ll_A \int_{\mathbb{R}} (2+|t|)^{O_A(1)} |f(t)| \, dt.$$

Using (6.11), we obtain (6.6) as required. □

Now we can prove

Proposition 6.3 (Replacing λ_{Siegel} with a Type I approximant). *We have*

$$\begin{aligned}&\mathbb{E}_{n\leqslant x}\Lambda_{\text{Siegel}}(n+h_1)\cdots\Lambda_{\text{Siegel}}(n+h_k)\lambda_{\text{Siegel}}(n+h'_1)\cdots\lambda_{\text{Siegel}}(n+h'_{\ell})\\&\quad\approx \mathbb{E}_{n\leqslant x}\Lambda_{\text{Siegel}}(n+h_1)\cdots\Lambda_{\text{Siegel}}(n+h_k)\lambda^{\sharp}_{\text{Siegel}}(n+h'_1)\cdots\lambda^{\sharp}_{\text{Siegel}}(n+h'_{\ell}).\end{aligned}$$

From Lemma 6.1, we have

$$\lambda^{\sharp}_{\text{Siegel}}(n+h'_j)=\lambda_{\text{Siegel}}(n+h'_j)+O(H(n+h'_j))$$

for $j=1,\dots,\ell$. Multiplying these estimates using (4.2) and the triangle inequality, and relabeling, we reduce to showing that

$$\mathbb{E}_{n\leqslant x}|\Lambda_{\text{Siegel}}(n+h_1)\cdots\Lambda_{\text{Siegel}}(n+h_k)|H(n+h'_1)\cdots H(n+h'_{\ell'})\approx 0$$

for any $1\leqslant\ell'\leqslant\ell$. By (5.2), it suffices to show that

$$(\log^k x)\mathbb{E}_{n\leqslant x}\tau\nu(n+h_1)\cdots\tau\nu(n+h_k)H(n+h'_1)\cdots H(n+h'_{\ell'})\approx 0. \tag{6.13}$$

Expanding out (6.5), the left-hand side is

$$\sum_{d'_1,\dots,d'_{\ell'}\leqslant D}\alpha(d'_1)\cdots\alpha(d'_{\ell'})(\log^k x)\mathbb{E}_{n\leqslant x}\tau\nu(n+h_1)\cdots\tau\nu(n+h_k)1_{d'_1|n+h'_1,\dots,d'_{\ell'}|n+h'_{\ell'}};$$

using (3.10), one can bound this further by

$$\ll\sum_{d_1,\dots,d_{k'}\leqslant D}\tau(d_1)^{O(1)}\cdots\tau(d_k)^{O(1)}\alpha(d_{k+1})\cdots\alpha(d_{k'})(\log^k x)\mathbb{E}_{n\leqslant x}\prod_{j=1}^{k}\nu(n+h_j)\prod_{j=1}^{k'}1_{d_i|n+h_i}$$

where we use the notations $k':=k+\ell'$, $h_{k+j}:=h'_j$, and $d_{k+j}:=d'_j$ for $j=1,\dots,\ell'$. Applying Lemma 3.4, we can bound this by

$$\begin{aligned}&\ll_A(\log_R^k x)\int_1^{\infty}\sum_{d_1,\dots,d_{k'}\leqslant D}\frac{\tau(d_1)^{O(1)}\cdots\tau(d_{k'})^{O(1)}\alpha(d_{k+1})\cdots\alpha(d_{k'})}{d_1\cdots d_{k'}}\\&\qquad\Big(\prod_{p|d_1\cdots d_k}\min(\sigma\log_R p,1)\Big)\frac{d\sigma}{\sigma^A}+\frac{D^{k'}R^{2k}}{x}\end{aligned}$$

for any $A>0$. The contribution of the latter term $\frac{D^{k'}R^{2k}}{x}$ is ≈ 0 due to (2.16). By (6.6), the former term can be bounded by

$$\ll_A\frac{\log_R^k x}{\log^A\eta}\int_1^{\infty}\sum_{d_1,\dots,d_k\leqslant D}\frac{\tau(d_1)^{O(1)}\cdots\tau(d_k)^{O(1)}}{d_1\cdots d_k}\Big(\prod_{p|d_1\cdots d_k}\min(\sigma\log_R p,1)\Big)\frac{d\sigma}{\sigma^A}$$

for any $A > 0$, which by (2.5) can be bounded by

$$\ll_A \frac{\log_R^k x}{\log^A \eta} \int_1^\infty \prod_{p \leqslant D} E_p(\sigma) \frac{d\sigma}{\sigma^A}$$

where

$$E_p(\sigma) := \sum_{d_1,\dots,d_k \in \mathbb{N}_{(p)}} \frac{\tau(d_1)^{O(1)} \cdots \tau(d_k)^{O(1)}}{d_1 \cdots d_k} \min(\sigma \log_R p, 1)^{1_{p|d_1 \cdots d_k}}.$$

We can bound

$$E_p(\sigma) \leqslant 1 + O\left(\frac{\min(\sigma \log_R p, 1)}{p}\right)$$

so by (3.4), (2.15) and setting A large enough we conclude (6.13). This completes the proof of Proposition 6.3.

7 | FOURTH STEP: REPLACING THE VON MANGOLDT SIEGEL MODEL WITH A TYPE I APPROXIMANT

We now execute step (iv) of the strategy outlined in the introduction. In this step, we will achieve power savings in many of our error terms, and as a consequence we can often afford to lose factors such as $x^{O(\varepsilon)}$, in contrast to other sections where even a loss of $\log x$ is often unacceptable.

It is convenient to perform a smooth dyadic decomposition of the convolution $\chi * \log$ in order to run a smoothed version of the Dirichlet hyperbola method. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a smooth even function supported on $[-1, 1]$ of total mass one. For any $t > 0$, define the function

$$\Phi_t(n) := \phi\Big(\log \frac{n}{t}\Big),$$

which is a smooth cutoff to the interval $[t/e, et]$. Then for any natural number n, one has the identity

$$\begin{aligned} \int_0^\infty \Phi_t(n) \log t \frac{dt}{t} &= \int_0^\infty \phi\Big(\log \frac{n}{t}\Big) \log t \frac{dt}{t} \\ &= \int_{\mathbb{R}} \phi(u)(\log n - u)\, du \\ &= \log n, \end{aligned} \tag{7.1}$$

where we made the change of variables $u := \log n - \log t$. We conclude that

$$\chi * \log(n) = \int_0^\infty \chi * \Phi_t(n) \log t \frac{dt}{t}. \tag{7.2}$$

As it turns out, the Dirichlet convolution $\chi * \Phi_t$ is of an adequate 'Type I' form when $t \leqslant Dq_\chi^2$ or $x/t \leqslant (Dq_\chi^2)^2$. Accordingly, we split

$$\chi * \log = (\chi * \log)^\sharp + (\chi * \log)^\flat,$$

where $(\chi * \log)^\sharp$ is the Type I approximant

$$\begin{aligned}(\chi * \log)^\sharp(n) = &\int_0^{Dq_\chi^2} \chi * \Phi_t(n) \log t \frac{dt}{t} \\ &+ \int_{Dq_\chi^2}^\infty \psi_{\leqslant (Dq_\chi^2)^2}(x/t)\chi * \Phi_t(n) \log t \frac{dt}{t} \\ &+ \int_{Dq_\chi^2}^\infty \psi_{>(Dq_\chi^2)^2}(x/t)\chi * \Phi_{Dq_\chi^2}(n) \log t \frac{dt}{t},\end{aligned} \tag{7.3}$$

and $(\chi * \log)^\flat$ is the error

$$(\chi * \log)^\flat(n) := \int_{Dq_\chi^2}^\infty \psi_{>(Dq_\chi^2)^2}(x/t)\chi * (\Phi_t - \Phi_{Dq_\chi^2}) \log t \frac{dt}{t}. \tag{7.4}$$

Thus, $(\chi * \log)^\sharp$ is the modification of $\chi * \log$ formed by replacing the cutoff Φ_t with $\Phi_{Dq_\chi^2}$ in the intermediate range $Dq_\chi^2 \leqslant t \leqslant \frac{x}{(Dq_\chi^2)^2}$ of t (using a smoothed version of the upper cutoff $t \leqslant \frac{x}{(Dq_\chi^2)^2}$ in order to facilitate some technical computations in the next section). As it turns out, it will be the second term in the right-hand side of (7.3) (in which the Φ_t term is supported in values $\gg x/(Dq_\chi^2)^2$, so that the χ term is supported in values $\ll (Dq_\chi^2)^2$) that will give the main contributions, being a more complicated version of the (untwisted) Type I sum $(\chi\psi_{\leqslant (Dq_\chi^2)^2}) * \log$.

We then have a similar spliting

$$\Lambda_{\mathrm{Siegel}} = \Lambda^\sharp_{\mathrm{Siegel}} + \Lambda^\flat_{\mathrm{Siegel}},$$

where

$$\Lambda^\sharp_{\mathrm{Siegel}} := (\chi * \log)^\sharp \nu$$

and

$$\Lambda^\flat_{\mathrm{Siegel}} := (\chi * \log)^\flat \nu.$$

We have good bounds on the distribution of $(\chi * \log)^\flat$ or $\Lambda^\flat_{\mathrm{Siegel}}$ in residue classes a (q) with q almost as large as $x^{2/3}$, as long as (a, q) is not too large:

Proposition 7.1 (2/3 level of distribution). *Let $0 < \varepsilon < \frac{1}{2}$, $1 \leqslant q \leqslant x$, and a be an integer. Let I be a subinterval of $[0, 2x]$. Let $f : \mathbb{Z} \to [-1, 1]$ be a q_χ-periodic function.*

(i) *We have*

$$\sum_{n \in I : n = a\ (q)} (\chi * \log)^\flat(n) f\left(\frac{n-a}{q}\right) \ll_\varepsilon \frac{x}{q}\left(\frac{(a,q)^{3/2} q_\chi^{9/2} q^{3/2}}{x^{1-O(\varepsilon)}} + \frac{x^{O(\varepsilon)}(a,q)^2}{D^{1/2}} + \frac{1}{x^\varepsilon}\right).$$

(ii) *If ε is sufficiently small depending on k, ℓ, ε_0, then we have*

$$\sum_{n\in I: n=a\ (q)} \Lambda^\flat_{\text{Siegel}}(n) f\left(\frac{n-a}{q}\right) \ll_\varepsilon (a,q)^{O(1)} \frac{x}{q}\left(\frac{q_\chi^{9/2} q^{3/2}}{x^{1-\varepsilon_0}} + \frac{1}{x^\varepsilon}\right).$$

The powers of (a,q) and q_χ are of minor importance and these terms can be neglected on a first reading. The key point here is that we can have a power savings over the trivial bound of $O_\varepsilon(x^{1+\varepsilon}/q)$ even when q is somewhat above $x^{1/2}$ (indeed, the above bounds can remain non-trivial as q approaches $x^{2/3}$).

Proof. We first prove (i). Note that $0 \leqslant (\chi * \log)^\flat(n) \leqslant \chi * \log(n)$. From (3.1), we may bound the left-hand side of the claim by $O_\varepsilon(x^{O(\varepsilon)}(1 + x/q))$. From this, we see that we may assume without loss of generality that we may take ε is sufficiently small depending on k', and we may also assume that $q \leqslant x^{2/3}$, since otherwise the above crude bound is already dominated by $x^{O(\varepsilon)} q^{1/2}$ and hence by $\frac{x}{q}\frac{(a,q)^{3/2} q_\chi^{9/2} q^{3/2}}{x^{1-O(\varepsilon)}}$. By shrinking I slightly (and using (3.1) to treat the error), we may assume that $I \subset [x^{1-\varepsilon}, 2x]$.

The integrand in (7.4) is only non-zero in the range $Dq_\chi^2 \leqslant t \leqslant x/(Dq_\chi^2)^2$. By the fundamental theorem of calculus, one has

$$\Phi_t(n) - \Phi_{Dq_\chi^2}(n) = -\int_{Dq_\chi^2}^t \tilde{\Phi}_{t'}(n) \frac{dt'}{t'},$$

where

$$\tilde{\Phi}_t(n) := \phi'\left(\log \frac{n}{t}\right)$$

so by the triangle inequality (and increasing ε slightly), it will suffice to show that

$$\sum_{n=a\ (q)} f\left(\frac{n-a}{q}\right) 1_I(n) \chi * \tilde{\Phi}_t(n) \ll_\varepsilon x^{O(\varepsilon)} (a,q)^{3/2} q_\chi^{9/2} q^{1/2} + \frac{x^{1+O(\varepsilon)}(a,q)^2}{D^{1/2} q} + \frac{x^{1-\varepsilon}}{q}$$

for all $Dq_\chi^2 \leqslant t \leqslant x/(Dq_\chi^2)^2$.

We can approximate 1_I by a cutoff $\psi_I : \mathbb{R} \to \mathbb{R}$ supported on I obeying $\psi_I(y) = 1$ whenever $\operatorname{dist}(y, I) \geqslant x^{1-2\varepsilon}$, and additionally obeying the derivative estimates

$$\psi_I^{(j)}(y) \ll_j x^{-(1-2\varepsilon)j} \tag{7.5}$$

for all $j \geqslant 0$ and $y \in \mathbb{R}$, with the error being acceptable by (3.1). It thus remains to establish the bound

$$\sum_{n=a\ (q)} f\left(\frac{n-a}{q}\right) \psi_I(n) \chi * \tilde{\Phi}_t(n) \ll_\varepsilon x^{O(\varepsilon)} (a,q)^{3/2} q_\chi^{9/2} q^{1/2} + \frac{x^{1+O(\varepsilon)}(a,q)^2}{D^{1/2} q}.$$

The left-hand side can be rewritten as

$$\sum_{n_1,n_2} \chi(n_1) \tilde{\Phi}_t(n_2) \psi_I(n_1 n_2) 1_{n_1 n_2 = a\ (q)} f\left(\frac{n_1 n_2 - a}{q}\right).$$

By the triangle inequality, it suffices to show that

$$\sum_{n_1,n_2} \chi(n_1)\tilde{\Phi}_t(n_2)\psi_I(n_1n_2)1_{n_1n_2=a' \ (qq_\chi)} \ll_\varepsilon x^{O(\varepsilon)}(a,q)^{3/2}q_\chi^{7/2}q^{1/2} + \frac{x^{1+O(\varepsilon)}(a,q)^2}{D^{1/2}qq_\chi},$$

for any $a' \in \mathbb{Z}/(qq_\chi\mathbb{Z})$ with $a' = a \ (q)$. If we set $q_0 := (a,q)q_\chi$, then (a', qq_χ) divides q_0. Writing

$$q_0' := ((a', qq_\chi)q_0, qq_\chi)$$

and using Lemma 3.9 and the triangle inequality and (3.1), we obtain the bound

$$\sum_{n_1,n_2} \chi(n_1)\tilde{\Phi}_t(n_2)\psi_I(n_1n_2)1_{n_1n_2=a' \ (qq_\chi)} \ll_\varepsilon x^\varepsilon\left(\frac{q_0'}{qq_\chi}X + q_0^{3/2}q^{-3/2}Y\right), \tag{7.6}$$

where

$$X := \left|\sum_{n_1,n_2} \chi(n_1)\tilde{\Phi}_t(n_2)\psi_I(n_1n_2)1_{n_1n_2=a' \ (q_0')}1_{(n_1n_2,qq_\chi)=(a',qq_\chi)}\right|$$

and

$$Y := \sum_{u_1,u_2 \in \mathbb{Z}/(qq_\chi\mathbb{Z}):\, qq_\chi/q_0 \nmid u_1,u_2} (u_1,u_2,qq_\chi)^{1/2}\left|\sum_{n_1,n_2} \tilde{\Phi}_t(n_2)\psi_I(n_1n_2)e_{qq_\chi}(u_1n_1+u_2n_2)\right|.$$

We first estimate the quantity Y. From repeated summation by parts, we have

$$\begin{aligned}\sum_{n_1,n_2} \tilde{\Phi}_t(n_2)\psi_I(n_1n_2)e_q(u_1n_1+u_2n_2) &\ll_\varepsilon x^{-1+O(\varepsilon)}\frac{x/t}{\|u_1/(qq_\chi)\|_{\mathbb{R}/\mathbb{Z}}}\frac{t}{\|u_2/(qq_\chi)\|_{\mathbb{R}/\mathbb{Z}}}\\ &= \frac{x^{O(\varepsilon)}}{\|u_1/qq_\chi\|_{\mathbb{R}/\mathbb{Z}}\|u_2/qq_\chi\|_{\mathbb{R}/\mathbb{Z}}}.\end{aligned}$$

Writing $u_1 = du_1'$, $u_2 = du_2'$ with $d = (u_1,u_2,qq_\chi)^{1/2}$, we then have

$$\begin{aligned}Y &\ll_\varepsilon \sum_{d|qq_\chi} d^{1/2} \sum_{1\leqslant u_1',u_2' < \frac{qq_\chi}{d}} \frac{x^{O(\varepsilon)}}{\|u_1'/(qq_\chi/d)\|_{\mathbb{R}/\mathbb{Z}}\|u_2'/(qq_\chi/d)\|_{\mathbb{R}/\mathbb{Z}}}\\ &\ll_\varepsilon x^{O(\varepsilon)}\sum_{d|qq_\chi} d^{1/2}\left(\frac{qq_\chi}{d}\log\left(2+\frac{qq_\chi}{d}\right)\right)^2\\ &\ll_\varepsilon x^{O(\varepsilon)}q^2q_\chi^2.\end{aligned}$$

Thus we see that the contribution of Y is acceptable. Now we consider the contribution of X. From Möbius inversion, we have

$$1_{(n_1n_2,qq_\chi)=(a',qq_\chi)} = \sum_{d:(a',qq_\chi)|d|qq_\chi} \mu\left(\frac{d}{(a',qq_\chi)}\right)1_{d|n_1n_2}.$$

By (3.1) and the triangle inequality, one thus has

$$X \ll_\varepsilon x^{O(\varepsilon)} \sum_{n_1,n_2} \chi(n_1)\tilde{\Phi}_t(n_2)\psi_I(n_1n_2)1_{n_1n_2=a' \ (q_0')}1_{d|n_1n_2}$$

for some d with $(a', qq_\chi)|d|qq_\chi$. On the one hand, we see from (3.1) (noting that the constraints $n_1n_2 = a' \ (q_0')$, $d|n_1n_2$ constrain n_1n_2 to at most one residue class modulo $[d, q_0']$) that

$$X \ll_\varepsilon x^{O(\varepsilon)} \frac{x}{[d, q_0']}. \tag{7.7}$$

On the other hand, we can write

$$X \ll_\varepsilon x^{O(\varepsilon)} \left| \sum_{n_1,n_2} F(n_1, n_2)\phi'\Big(\log \frac{n_2}{t}\Big)\psi_I(n_1n_2)\right|,$$

where F is the $[d, q_0']$-periodic function

$$F(n_1, n_2) := \chi(n_1)1_{n_1n_2=a' \ (q_0')}1_{d|n_1n_2}.$$

By Fourier expansion and Poisson summation, we can then write

$$X \ll_\varepsilon x^{O(\varepsilon)} \left| \sum_{\xi_1,\xi_2 \in \frac{\mathbb{Z}}{dq_0'}} \hat{F}(\xi_1, \xi_2)\Psi(\xi_1, \xi_2)\right|,$$

where

$$\hat{F}(\xi_1, \xi_2) := \mathbb{E}_{n_1,n_2} F(n_1, n_2)e(-n_1\xi_1 - n_2\xi_2),$$

and

$$\Psi(\xi_1, \xi_2) := \int_0^\infty \int_0^\infty \tilde{\Phi}_t(x_2)\psi_I(x_1x_2)e(x_1\xi_1 + x_2\xi_2)\, dx_1dx_2.$$

From the area-preserving change of variables $(u_1, u_2) := (\log \frac{x_2}{t}, x_1x_2)$ and the fundamental theorem of calculus, we have

$$\Psi(0, 0) = \int_{\mathbb{R}} \int_0^\infty \phi'(u_1)\psi_I(u_2)\, du_1du_2 = 0,$$

and from integration by parts, one has the bounds

$$\Psi(\xi_1, \xi_2) \ll_m \frac{x^{1+O_m(\varepsilon)}}{(1 + t|\xi_1|)^m(1 + \frac{x}{t}|\xi_2|)^m}$$

for any $m \geqslant 0$. Meanwhile, using the trivial bound $|\hat{F}(\xi_1, \xi_2)| \leqslant 1$ and $t, \frac{x}{t} \geqslant Dq_\chi^2$, we have

$$X \ll_{m,\varepsilon} x^{O_m(\varepsilon)} \sum_{\xi \in (\frac{\mathbb{Z}}{[d,q_0']})^2 \setminus \{(0,0)\}} (1 + Dq_\chi^2|\xi|)^{-m}$$

for any $m > 0$, and thus

$$X \ll_{m,\varepsilon} x^{O_m(\varepsilon)} \left(\frac{Dq_\chi}{[d, q_0']} \right)^{-m}$$

for any $m > 0$. In particular, if $[d, q_0'] \leqslant D^{1/2} q_\chi^2$, we have $X \ll_\varepsilon 1$ (say) by choosing m large enough. Comparing this with (7.7), we conclude that

$$X \ll_\varepsilon x^{O(\varepsilon)} \frac{x}{D^{1/2} q_\chi^2}$$

in all cases. Inserting these bounds back into (7.6) and writing $q_0 = (a, q) q_\chi$ and bounding

$$q_0' \leqslant (a', q q_\chi) q_0 \leqslant q_0^2 = (a, q)^2 q_\chi^2,$$

we conclude that

$$\sum_{n_1, n_2} \chi(n_1) \tilde{\Phi}_t(n_2) \psi_I(n_1 n_2) 1_{n_1 n_2 = a' \ (q q_\chi)} \ll_\varepsilon x^{O(\varepsilon)} \left(\frac{x (a, q)^2}{D^{1/2} q q_\chi} + (a, q)^{3/2} q_\chi^{7/2} q^{1/2} \right),$$

and the claim (i) follows.

Now we prove (ii). Expanding out the Selberg sieve ν as

$$\nu(n) = \sum_{d \leqslant R^2} a_d 1_{d|n} \tag{7.8}$$

for some sieve weights a_d that can be crudely bounded using (3.1) as

$$a_d \ll \tau(d) \ll_\varepsilon x^\varepsilon, \tag{7.9}$$

we see from the triangle inequality and pigeonhole principle (noting that $\sum_{d \leqslant R^2} \frac{1}{d} \ll_\varepsilon x^\varepsilon$) that

$$\sum_{n \in I : n = a \ (q)} \Lambda^\flat_{\text{Siegel}}(n) f\left(\frac{n-a}{q} \right) \ll_\varepsilon x^{2\varepsilon} \left| d \sum_{n \in I : n = a \ (q); d|n} (\chi * \log)^\flat(n) f\left(\frac{n-a}{q} \right) \right|$$

for some $d \leqslant R^2$. We can restrict attention to those d with $(d, q) | (a, q)$, since otherwise the sum is empty. The conditions $n = a \ (q)$, $d|n$ can then be combined into a single congruence class $n = a' \ ([q, d])$, with $(a', [q, d]) \leqslant d(a, q)$; on this class, the quantity $f(\frac{n-a}{q})$ can be viewed as a q_χ-periodic function of $\frac{n-a'}{[q,d]}$. Applying (i) (with ε replaced by 3ε), we have

$$d \sum_{n \in I : n = a \ (q); d|n} (\chi * \log)^\flat(n) f\left(\frac{n-a}{q} \right)$$
$$\ll \frac{dx}{[q, d]} \left(\frac{d^{3/2} (a, q)^{3/2} q_\chi^{9/2} [q, d]^{3/2}}{x^{1 - O(\varepsilon)}} + \frac{x^{O(\varepsilon)} d^{3/2} (a, q)^2}{D^{1/2}} + \frac{1}{x^{3\varepsilon}} \right).$$

Writing $\frac{d}{[q,d]} = \frac{(d,q)}{q} \leqslant \frac{(a,q)}{q}$ and then bounding $[q,d] \leqslant qd$ and $d \leqslant R^2$, we conclude

$$\sum_{n\in I:\, n=a\ (q)} \Lambda^\flat_{\mathrm{Siegel}}(n) f\left(\frac{n-a}{q}\right) \ll_\varepsilon \frac{x}{q}\left(\frac{R^6(a,q)^{5/2} q_\chi^{9/2} q^{3/2}}{x^{1-O(\varepsilon)}} + \frac{x^{O(\varepsilon)} R^3 (a,q)^3}{D^{1/2}} + \frac{(a,q)}{x^\varepsilon}\right).$$

Using (2.10) and (2.9), we obtain the claim for ε small enough. □

Now we can prove

Proposition 7.2 (Replacing $\Lambda_{\mathrm{Siegel}}$ with a Type I approximant). *Assume $k \leqslant 2$. Then we have*

$$\begin{aligned}&\mathbb{E}_{n\leqslant x} \Lambda_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda_{\mathrm{Siegel}}(n+h_k)\lambda^\sharp_{\mathrm{Siegel}}(n+h'_1)\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h'_\ell)\\ &\quad\approx \mathbb{E}_{n\leqslant x} \Lambda^\sharp_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda^\sharp_{\mathrm{Siegel}}(n+h_k)\lambda^\sharp_{\mathrm{Siegel}}(n+h'_1)\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h'_\ell).\end{aligned}$$

Proof. The claim is trivial for $k = 0$, so we may assume that $k \in \{1,2\}$. By the triangle inequality and relabeling it suffices to show the bounds

$$\mathbb{E}_{n\leqslant x} \Lambda^\flat_{\mathrm{Siegel}}(n+h_1)\lambda^\sharp_{\mathrm{Siegel}}(n+h'_1)\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h'_\ell) \approx 0 \tag{7.10}$$

when $k = 1$, and the bounds

$$\mathbb{E}_{n\leqslant x} \Lambda^\flat_{\mathrm{Siegel}}(n+h_1)\Lambda_{\mathrm{Siegel}}(n+h_2)\lambda^\sharp_{\mathrm{Siegel}}(n+h'_1)\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h'_\ell) \approx 0 \tag{7.11}$$

and

$$\mathbb{E}_{n\leqslant x} \Lambda^\flat_{\mathrm{Siegel}}(n+h_1)\Lambda^\sharp_{\mathrm{Siegel}}(n+h_2)\lambda^\sharp_{\mathrm{Siegel}}(n+h'_1)\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h'_\ell) \approx 0 \tag{7.12}$$

when $k = 2$.

We begin with (7.10). Let $\varepsilon > 0$ be a sufficiently small quantity. From (6.2) and (3.1), we have

$$\lambda^\sharp_{\mathrm{Siegel}}(n) = \sum_{d\leqslant D} b_d 1_{d|n}\chi(n/d) \tag{7.13}$$

for some weights b_d of size

$$b_d \ll \tau(d)\log^{O(1)} x \ll_\varepsilon x^\varepsilon. \tag{7.14}$$

Since

$$\sum_{d'_1,\dots,d'_\ell\leqslant D} \frac{1}{[d'_1,\dots,d'_\ell]} \leqslant \sum_{d\leqslant D^{k'-1}} \frac{\tau(d)^{k'-1}}{d} \ll_\varepsilon x^\varepsilon$$

(due to (3.1)), we thus see from the pigeonhole principle that the left-hand side of (7.10) is bounded by

$$\ll_\varepsilon x^{O(\varepsilon)}[d'_1,\dots,d'_\ell]\left|\mathbb{E}_{n\leqslant x}\Lambda^\flat_{\mathrm{Siegel}}(n+h_1)\prod_{j=1}^{\ell} 1_{d'_j|n+h'_j}\chi\left(\frac{n+h'_j}{d'_j}\right)\right|$$

for some $d'_1, \dots, d'_\ell \leqslant D$. By translating (and removing negligible errors), we may assume that $h_1 = 0$. Setting $d := [d'_1, \dots, d'_\ell]$, we see that the constraints $d'_j | n + h'_j$ are either inconsistent, or restrict n to a single residue class a (d) with $(a, d) \ll 1$, and then $\prod_{j=1}^{\ell} \chi(\frac{n+h'_j}{d'_j})$ is a q_χ-periodic function of $\frac{n-a}{d}$. Applying Proposition 7.1(ii) (with a suitable multiple of ε), we bound the left-hand side of (7.10) by

$$\ll_\varepsilon \frac{x^{O(\varepsilon)} q_\chi^{9/2} d^{3/2}}{x^{1-\varepsilon_0}} + \frac{1}{x^\varepsilon}. \tag{7.15}$$

Bounding $d \leqslant D^{k'-1}$ and using (2.20), we see that the right-hand side is ≈ 0 for ε a sufficiently small constant, giving the claim.

Now we consider (7.11) and (7.12). Again let $\varepsilon > 0$ be sufficiently small. From (7.2), (7.3) and the pigeonhole principle, we can bound both left-hand sides (up to negligible errors) by

$$\begin{aligned} \ll_\varepsilon x^{O(\varepsilon)} \Bigg| \mathbb{E}_{n \leqslant x} \Lambda^\flat_{\mathrm{Siegel}}(n + h_1) \Bigg(\sum_{d_2 | n + h_2} \chi(d_2) \Phi_t\Big(\frac{n+h_2}{d_2}\Big) \Bigg) \nu(n + h_2) \\ \lambda^\sharp_{\mathrm{Siegel}}(n + h'_1) \cdots \lambda^\sharp_{\mathrm{Siegel}}(n + h'_\ell) \Bigg| \end{aligned} \tag{7.16}$$

for some $1 \ll t \ll x$ (note that the summation vanishes for t outside this range).

We now use a version of the Dirichlet hyperbola method. First suppose that $t \geqslant \sqrt{x}$, then the summand vanishes unless $d_2 \ll \sqrt{x}$. Expanding out using (7.13), (7.8) much as before and now using

$$\sum_{d_2 \ll \sqrt{x}; \tilde{d}_2 \leqslant R^2; d'_1, \dots, d'_\ell \leqslant D} \frac{1}{[d_2, \tilde{d}_2, d'_1, \dots, d'_\ell]} \ll_\varepsilon x^\varepsilon,$$

we can bound the contribution of the $d_2 \leqslant \sqrt{x}$ case by

$$\ll_\varepsilon x^{O(\varepsilon)} [d_2, \tilde{d}_2, d'_1, \dots, d'_\ell] \Bigg| \mathbb{E}_{n \leqslant x} \Lambda^\flat_{\mathrm{Siegel}}(n + h_1) 1_{d_2, \tilde{d}_2 | n + h_2} \Phi_t\Big(\frac{n+h_2}{d_2}\Big) \prod_{j=1}^{\ell} 1_{d'_j | n + h'_j} \chi\Big(\frac{n+h'_j}{d'_j}\Big) \Bigg|$$

for some $d_2 \leqslant \sqrt{x}$, $\tilde{d}_2 \leqslant R^2$, and $d'_1, \dots, d'_\ell \leqslant D$. Writing $d := [d_2, \tilde{d}_2, d'_1, \dots, d'_\ell]$ and arguing as before, using summation by parts to deal with the slowly varying function $\Phi_t(\frac{n+h_2}{d_2})$, we can again bound this expression by (7.15). Bounding $d \ll \sqrt{x} R^2 D^{k'-2}$ and using (2.20), we see that the right-hand side is ≈ 0 for ε small enough, giving the claim.

Finally, suppose that $t < \sqrt{x}$. Now we make the change of variables $d_2^* := \frac{n+h_2}{d_2}$ and rewrite the bound as

$$\begin{aligned} \ll_\varepsilon x^{O(\varepsilon)} \Bigg| \mathbb{E}_{n \leqslant x} \Lambda^\flat_{\mathrm{Siegel}}(n + h_1) \Bigg(\sum_{d_2^* | n + h_2} \Phi_t(d_2^*) \chi\Big(\frac{n+h_2}{d_2^*}\Big) \Bigg) \nu(n + h_2) \\ \lambda^\sharp_{\mathrm{Siegel}}(n + h'_1) \cdots \lambda^\sharp_{\mathrm{Siegel}}(n + h'_\ell) \Bigg|. \end{aligned}$$

Observe that the summand vanishes unless $d_2^* \ll \sqrt{x}$. Now we can repeat the previous arguments (using d_2^* in place of d_2, and the q_χ-periodic function χ in place of Φ_t, noting that (2.20) can handle several additional losses of q_χ) to conclude. □

8 | FIFTH STEP: COMPUTING THE TYPE I CORRELATIONS

We now execute step (v) of the strategy outlined in the introduction by establishing

Proposition 8.1 (Evaluating the Type I correlation). *We have*

$$\mathbb{E}_{n\leqslant x}\Lambda^\sharp_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda^\sharp_{\mathrm{Siegel}}(n+h_k)\lambda^\sharp_{\mathrm{Siegel}}(n+h_1')\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h_\ell')\approx \mathfrak{S}, \tag{8.1}$$

where $\mathfrak{S}$ is the quantity in Conjecture 1.3.

Clearly Theorem 1.6 follows immediately from concatenating together Propositions 4.2, 5.2, 6.3, 7.2, 8.1 using (1.7).

We first dispose of the easy case $\ell > 0$, in which $\mathfrak{S}$ vanishes. For $1 \leqslant j \leqslant k$, we see from (7.3) and replacing d by n/d in the first and third factors, and truncating the very small or very large values of t (where the summand vanishes) that

$$\begin{aligned}(\chi * \log)^\sharp(n) = &\int_{1/100}^{Dq_\chi^2}\sum_{d|n}\Phi_t(d)\chi(n/d)\log t\frac{dt}{t}\\ &+\int_{Dq_\chi^2}^{100x}\psi_{\leqslant (Dq_\chi^2)^2}(x/t)\sum_{d|n}\Phi_t(n/d)\chi(d)\log t\frac{dt}{t}\\ &+\int_{Dq_\chi^2}^{100x}\psi_{>(Dq_\chi^2)^2}(x/t)\sum_{d|n}\Phi_{Dq_\chi^2}(d)\chi(n/d)\log t\frac{dt}{t}.\end{aligned} \tag{8.2}$$

In all of these terms, the summands vanish unless $d \ll (Dq_\chi^2)^2$. One can then write

$$(\chi * \log)^\sharp(n) = \sum_{d\ll (Dq_\chi^2)^2 : d|n}(\Psi(n/d)\chi(d) + c_d\chi(\frac{n}{d})),$$

where $\Psi : \mathbb{R}^+ \to \mathbb{R}$ is the smooth function

$$\Psi(y) := \int_{Dq_\chi^2}^{100x}\psi_{\leqslant (Dq_\chi^2)^2}(x/t)\Phi_t(y)\log t\frac{dt}{t} \tag{8.3}$$

and c_d is the coefficient

$$\begin{aligned}c_d := &\int_{1/100}^{Dq_\chi^2}\Phi_t(d)\log t\frac{dt}{t}\\ &+\int_{Dq_\chi^2}^{100x}\psi_{>(Dq_\chi^2)^2}(x/t)\Phi_{Dq_\chi^2}(d)\log t\frac{dt}{t}.\end{aligned}$$

For the current analysis, we will need the crude bound

$$\|\Psi\|_{\mathrm{TV}} \ll \log^{O(1)} x, c_d \ll \log^{O(1)} x,$$

where we use the total variation norm

$$\|f\|_{\mathrm{TV}} := \sup_{y>0} |f(y)| + \int_{\mathbb{R}} |f'(y)|\, dy.$$

Combining this with the expansion (7.8), we see that

$$\Lambda^\sharp(n) = \sum_{d \ll R^2 (Dq_\chi^2)^2 : d|n} \Psi_d(n) + \sum_{d \ll R^2 (Dq_\chi^2)^2 : d'|d|n} g_{d,d'} \chi(\frac{n}{d'}), \tag{8.4}$$

where $\Psi_d : \mathbb{R}^+ \to \mathbb{R}$ is a smooth function and $g_{d,d'}$ is a coefficient obeying the bounds

$$\|\Psi_d\|_{\mathrm{TV}} \ll \tau(d)^{O(1)}, g_{d,d'} \ll \tau(d)^{O(1)} \log^{O(1)} x. \tag{8.5}$$

Using the decomposition (7.13) to expand $\lambda^\sharp_{\mathrm{Siegel}}(n+h'_j)$, we can thus write $\Lambda^\sharp_{\mathrm{Siegel}}(n+h_1)\cdots\Lambda^\sharp_{\mathrm{Siegel}}(n+h_k)\lambda^\sharp_{\mathrm{Siegel}}(n+h'_1)\cdots\lambda^\sharp_{\mathrm{Siegel}}(n+h'_\ell)$ as

$$\sum_{J\subset\{1,\dots,k\}} \sum_{d_1,\dots,d_k \ll R^2(Dq_\chi^2)^2; d_{k+1},\dots,d_{k+\ell} \leqslant D; d'_j | d_j \forall j\in J} h_{d_1,\dots,d_{k+\ell},d'_1,\dots,d'_k}(n)$$

$$\left(\prod_{j=1}^{k+\ell} 1_{d|n+h_j}\right) \prod_{j\in J\cup\{k+1,\dots,k+\ell\}} \chi\left(\frac{n+h_j}{d'_j}\right)$$

for some smooth functions $h_{d_1,\dots,d_{k+\ell},d'_1,\dots,d'_k} : \mathbb{R} \to \mathbb{R}$ with

$$\|h_{d_1,\dots,d_{k+\ell},d'_1,\dots,d'_k}\|_{\mathrm{TV}} \ll \tau(d_1)^{O(1)} \cdots \tau(d_k)^{O(1)} \log^{O(1)} x,$$

and with the convention that $h_{k+j} = h'_j$ and $d'_{k+j} = d_{k+j}$ for $j = 1, \dots, \ell$. From Lemma 3.7(ii) and summation by parts to deal with the $h_{d_1,\dots,d_{k+\ell},d'_1,\dots,d'_k}$ coefficients, we may thus bound the left-hand side of (8.1) by

$$\ll_\varepsilon q_\chi^{1/2+\varepsilon} \log^{O(1)} x \sum_{J\subset\{1,\dots,k\}} \sum_{d_1,\dots,d_k \ll R^2(Dq_\chi^2)^2; d_{k+1},\dots,d_{k+\ell} \leqslant D; d'_j | d_j \forall j\in J}$$

$$(d_1\cdots d_k, q_\chi)^{1/2} \tau(d_1)^{O(1)} \cdots \tau(d_k)^{O(1)} \left(\frac{1}{q_\chi d_1 \cdots d_k} + \frac{1}{x}\right)$$

which on evaluating the d'_j sums, and then writing $d := d_1 \cdots d_k$, can be bounded by

$$\ll_\varepsilon q_\chi^{1/2+\varepsilon} \log^{O(1)} x \sum_{d \ll D^{2(k+\ell)} (Rq_\chi^2)^{2k}} (d, q_\chi)^{1/2} \tau(d)^{O(1)} \left(\frac{1}{q_\chi d} + \frac{1}{x}\right). \tag{8.6}$$

From (2.16), we see that

$$d \ll D^{2k(k+\ell)}(Rq_\chi^2)^{2k} \ll \frac{x}{q_\chi^{1/2+\varepsilon_0/2}}$$

so that

$$\frac{1}{q_\chi d} + \frac{1}{x} \ll \frac{1}{q_\chi^{1/2+\varepsilon_0/2} d},$$

and then by (2.5), we can bound (8.6) by

$$\ll_\varepsilon q_\chi^{-\frac{\varepsilon_0}{2}+\varepsilon} \log^{O(1)} x \prod_{p \leqslant x} \sum_{d \in \mathbb{N}_{(p)}} \frac{(d, q_\chi)^{1/2} \tau(d)^{O(1)}}{d}. \tag{8.7}$$

One can calculate

$$\sum_{d \in \mathbb{N}_{(p)}} \frac{(d, q_\chi)^{1/2} \tau(d)^{O(1)}}{d} \leqslant 1 + O(\frac{1}{p})$$

when $p \nmid q_\chi$ and

$$\sum_{d \in \mathbb{N}_{(p)}} \frac{(d, q_\chi)^{1/2} \tau(d)^{O(1)}}{d} \ll 1$$

otherwise, thus by (3.7) the preceding expression (8.7) is

$$\ll_\varepsilon q_\chi^{-\frac{\varepsilon_0}{2}+\varepsilon} \tau(q_\chi)^{O(1)} \log^{O(1)} x$$

which by (3.1) is

$$\ll_\varepsilon q_\chi^{-\frac{\varepsilon_0}{4}} \log^{O(1)} x$$

if ε is small enough. Applying (2.18), we conclude that

$$\mathbb{E}_{n \leqslant x} \Lambda^\sharp_{\text{Siegel}}(n + h_1) \cdots \Lambda^\sharp_{\text{Siegel}}(n + h_k) \lambda^\sharp_{\text{Siegel}}(n + h'_1) \cdots \lambda^\sharp_{\text{Siegel}}(n + h'_\ell) \approx 0.$$

This concludes the treatment of the $\ell > 0$ case.

Now suppose that $\ell = 0$. The above arguments allow us to dispose of the $g_{d,d'}$ contributions in (8.4), leaving us with the task of showing that

$$\mathbb{E}_{n \leqslant x} \prod_{j=1}^{k} \sum_{d_j \ll R^2 (Dq_\chi^2)^2 : d_j | n + h_j} \Psi_{d_j}(n + h_j) \approx \mathfrak{S}.$$

This is a correlation of Goldston–Yıldırım type and can be calculated by a lengthy but straightforward calculation, basically a more careful variant of Lemma 3.4. We follow the Fourier-analytic method laid out in [9, Appendix D], as follows. Using Lemma 3.3, (8.5), and summation by parts,

we can write the left-hand side here as

$$\sum_{d_1,\dots,d_k \ll R^2(Dq_\chi^2)^2} \left(\frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{[d_1,\dots,d_k]} \frac{1}{x} \int_0^x \prod_{j=1}^k \Psi_{d_j}(y+h_j)\,dy \right.$$
$$\left. + O\left(\frac{\tau(d_1)^{O(1)} \cdots \tau(d_k)^{O(1)} \log^{O(1)} x}{x} \right) \right).$$

Using (3.1), the contribution of the error term is at most

$$\ll_\varepsilon \frac{(RDq_\chi^2)^{2k} x^\varepsilon}{x}$$

for any $\varepsilon > 0$, which is ≈ 0 for ε small enough due to (2.16). Thus it remains to show that

$$\sum_{d_1,\dots,d_k \ll (RDq_\chi^2)^{2k}} \frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{[d_1,\dots,d_k]} \frac{1}{x} \int_0^x \prod_{j=1}^k \Psi_{d_j}(y+h_j)\,dy \approx \mathfrak{S}.$$

The contribution of those y with $y \leqslant x^{1-\varepsilon_0^2}$ is bounded by

$$x^{-\varepsilon_0^2} \log^{O(1)} x \sum_{d_1,\dots,d_k \ll R^2(Dq_\chi^2)^2} \frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{[d_1,\dots,d_k]} \tau(d_1)^{O(1)} \cdots \tau(d_k)^{O(1)}.$$

Bounding $1_{(d_i,d_j)|h_i-h_j}[d_1,\dots,d_k] \ll \frac{1}{d_1\cdots d_k}$ and using (3.1), we see that this contribution is ≈ 0. Thus it will suffice to establish the pointwise bound

$$\sum_{d_1,\dots,d_k \ll R^2(Dq_\chi^2)^2} \frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{[d_1,\dots,d_k]} \prod_{j=1}^k \Psi_{d_j}(y+h_j) \approx \mathfrak{S} \tag{8.8}$$

for all $x^{1-\varepsilon_0^2} \leqslant y \leqslant x$. Note that the restrictions on $d_1,\dots,d_k$ can be dropped due to the support of the Ψ_{d_j}.

By construction, we have

$$\Psi_d(y+h_j) = \sum_{d_0,d_1,d_2 : [d,d_1,d_2]=d} \chi(d_0)\mu(d_1)\mu(d_2)\Phi\left(\frac{y+h_j}{d_0}\right)\psi_{\leqslant R}(d_1)\psi_{\leqslant R}(d_2) \tag{8.9}$$

This function is not multiplicative in d, but it can be Fourier expanded as a linear combination of multiplicative functions:

Lemma 8.2 (Fourier expansion). *We have*

$$\psi_{\leqslant R}(d) = \int_{\mathbb{R}} \frac{1}{d^{\frac{1+it}{\log R}}} f(t)\,dt, \tag{8.10}$$

and

$$\Phi\left(\frac{y+h_j}{d}\right) = \log x \int_{\mathbb{R}} \frac{1}{d^{\frac{1+it}{\log x}}} F_j(t)\,dt \tag{8.11}$$

for all real $d \geqslant 1$, and some measurable functions $f, F_j : \mathbb{R} \to \mathbb{C}$ obeying the decay estimates

$$f(t) \ll_m (1+|t|)^{-m} \tag{8.12}$$

and

$$F_j(t) \ll_m (1+|t|)^{-m} \tag{8.13}$$

for all $t \in \mathbb{R}$ and $m \geqslant 0$, as well as the identities

$$\int_{\mathbb{R}} f(t)\,dt = 1 \tag{8.14}$$

and

$$\int_{\mathbb{R}} F_j(t)(1+it)\,dt = 1. \tag{8.15}$$

Proof. From (2.21) and (3.14), we obtain (8.10) with

$$f(t) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{(1+it)u} \psi(u)\,du$$

the Fourier transform of $u \mapsto e^u \psi(u)$. From repeated integration by parts, we have the rapid decrease (8.12), while from (3.15) we have

$$\int_{\mathbb{R}} f(t)\,dt = e^0 \psi(0) = 1$$

giving (8.14).

Next, from (8.3), we have

$$\Phi\left(\frac{y+h_j}{d}\right) = \int_0^\infty \psi\left(\frac{\log(x/t)}{2\log(Dq_\chi^2)}\right) \Phi_{dt}(y+h_j) \log t \frac{dt}{t},$$

for any natural number d. Writing $s = \frac{\log(x/t)}{\log x}$, we can rewrite this as

$$\Phi\left(\frac{y+h_j}{d}\right) = \log^2 x \int_{\mathbb{R}} \psi\left(\frac{\log x}{2\log(Dq_\chi^2)} s\right) \phi\left(\log\frac{y+h_j}{x} + s\log x - \log d\right)(1-s)\,ds.$$

By Fourier inversion (3.14), we then have (8.11) where

$$F_j(t) := \frac{\log x}{2\pi} \int_{\mathbb{R}} e^{(1+it)u} \int_{\mathbb{R}} \psi\left(\frac{\log x}{2\log(Dq_\chi^2)} s\right) \phi\left(\log\frac{y+h_j}{x} + (s-u)\log x\right)(1-s)\,ds du,$$

which on making the change of variables $v := u - s$ factors as

$$F_j(t) = \frac{\log x}{2\pi}\left(\int_{\mathbb{R}} e^{(1+it)v}\phi\left(\log\frac{y+h_j}{x} - v\log x\right)dv\right)\left(\int_{\mathbb{R}} e^{(1+it)s}\psi\left(\frac{\log x}{2\log(Dq_\chi^2)}s\right)(1-s)\,ds\right).$$

From the triangle inequality, one has

$$\int_{\mathbb{R}} e^{(1+it)v}\phi\left(\log\frac{y+h_j}{x} - v\log x\right)dv \ll \frac{1}{\log x}$$

while from integration by parts (and (2.9)), one has

$$\int_{\mathbb{R}} e^{(1+it)s}\psi\left(\frac{\log x}{2\log(Dq_\chi^2)}s\right)(1-s)\,ds \ll_m (1+|t|)^{-m}$$

for any $m \geqslant 0$, thus yielding (8.13). Also, from (3.16), and integration by parts one has

$$\begin{aligned}
\int_{\mathbb{R}} F_j(t)(1+it)\,dt &= -\frac{d}{dx}\Phi\left(\frac{y+h_j}{x}\right)|_{x=0}\\
&= (y+h_j)\Phi'(y+h_j)\\
&= \int_0^\infty \psi\left(\frac{\log(x/t)}{2\log(Dq_\chi^2)}\right)\phi'\left(\log\frac{y+h_j}{t}\right)\log t\,\frac{dt}{t}\\
&= \int_0^\infty \phi'\left(\log\frac{y+h_j}{t}\right)\log t\,\frac{dt}{t}\\
&= \int_{\mathbb{R}} \phi'(u)(\log(y+h_j)-u)\,du\\
&= \int_{\mathbb{R}} \phi(u)\,du\\
&= 1
\end{aligned}$$

where we have used the observation that $\psi(\frac{\log(x/t)}{2\log(Dq_\chi^2)})$ equals to 1 on the support of $\phi'(\log\frac{y+h_j}{t})$ (since one then has $x/t \asymp x/y \ll x^{\varepsilon_0^2}$). This gives (8.15). □

Inserting the expansions (8.10), (8.11) back into (8.9), we see that

$$\Psi_d(y+h_j) = \log^k x \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} \sum_{d_0,d_1,d_2:[d,d_1,d_2]=d} \frac{\chi(d_0)\mu(d_1)\mu(d_2)}{d_0^{\frac{1+it_0}{\log x}} d_1^{\frac{1+it_1}{\log R}} d_2^{\frac{1+it_2}{\log R}}} F_j(t_0)f(t_1)f(t_2)dt_0dt_1dt_2.$$

Inserting this back into the left-hand side of (8.8) and factoring the Euler product using (2.3), we can thus write that left-hand side as

$$\log^k x \int_{\mathbb{R}^{3k}} \prod_p E_{p,t_{0,1},\dots,t_{2,k}} \prod_{j=1}^k F_j(t_{0,j})f(t_{1,j})f(t_{2,j})dt_{0,j}dt_{1,j}dt_{2,j}, \tag{8.16}$$

where

$$E_{p,t_{0,1},\dots,t_{2,k}} := \sum_{d_1,\dots,d_k \in \mathbb{N}_{(p)}} \frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{[d_1,\dots,d_k]} \prod_{j=1}^{k} c_{d_j,t_{0,j},t_{1,j},t_{2,j}}, \tag{8.17}$$

and

$$c_{d,t_0,t_1,t_2} := \sum_{d_0,d_1,d_2 : [d_0,d_1,d_2]=d} \frac{\chi(d_0)\mu(d_1)\mu(d_2)}{d_0^{\frac{1+it_0}{\log x}} d_1^{\frac{1+it_1}{\log R}} d_2^{\frac{1+it_2}{\log R}}}. \tag{8.18}$$

From the triangle inequality, one has the crude bound

$$E_{p,t_{0,1},\dots,t_{2,k}} = 1 + O\left(\frac{1}{p^{1+1/\log R}}\right) \tag{8.19}$$

and thus by Mertens' theorem (3.7)

$$\prod_p E_{p,t_{0,1},\dots,t_{2,k}} \ll \log^{O(1)} R.$$

Using (8.12) and (8.13), we see that the contribution of the integral in which the quantity

$$|t| := \sup_{0\leqslant i\leqslant 2; 1\leqslant j\leqslant k} |t_{i,j}|$$

exceeds (say) $\log^{1/10} R$ is negligible. Thus we may restrict attention to the regime

$$|t| \leqslant \log^{1/10} R.$$

We can improve the above analysis to restrict the region of t further. From Taylor expansion, one has the more precise bound

$$E_{p,t_{0,1},\dots,t_{2,k}} = 1 - \frac{k}{p} + O\left(\frac{(1+|t|)^3 \log_R p}{p}\right) + O(\frac{1}{p^2})$$

when $p \leqslant R$. Using this bound in place of (8.19) when $\log p \leqslant (1+|t|)^{-3} \log R$ and using Mertens' theorem (3.3) and (3.6), we obtain the refined estimate

$$\prod_{p\geqslant C} E_{p,t_{0,1},\dots,t_{2,k}} \ll_C (1+|t|)^{O(1)} \log^{-k} R \tag{8.20}$$

for any $C \geqslant 1$. Using (8.12) and (8.13), we see that the contribution of the integral in which $|t| \geqslant \log^{1/(100k)} \eta$ (say) is negligible. Thus we may restrict attention to the regime

$$|t| \leqslant \log^{1/(100k)} \eta. \tag{8.21}$$

We now perform an even more precise analysis of the Euler factors $E_{p,t_{0,1},\dots,t_{2,k}}$. Let us first suppose that p is larger than C_0 for some sufficiently large C_0 (depending on $h_1,\dots,h_k,k$). Then p does not divide $\prod_{1\leqslant i<j\leqslant k}(h_i-h_j)$. Thus in order for the sum in (8.17) to be non-zero, at most one

of the d_j can be greater than 1, and hence

$$E_{p,t_{0,1},\dots,t_{2,k}} = 1 + \sum_{j=1}^{k} \sum_{l=1}^{\infty} \frac{c_{p^l,t_{0,j},t_{1,j},t_{2,j}}}{p^l}.$$

We expand $c_{p^l,t_{0,j},t_{1,j},t_{2,j}}$ using (8.18). For $l > 1$, the sum in (8.18) only consists of those terms with $d_0 = p^l$, and thus

$$c_{p^l,t_{0,j},t_{1,j},t_{2,j}} = \frac{\chi(p^l)}{p^{l(1+\frac{1+it_{0,j}}{\log x})}} \left(1 - \frac{1}{p^{\frac{1+it_{1,j}}{\log R}}}\right)\left(1 - \frac{1}{p^{\frac{1+it_{2,j}}{\log R}}}\right).$$

In particular, from Taylor expansion, we have

$$c_{p^l,t_{0,j},t_{1,j},t_{2,j}} \ll \min\left((1+|t|)\log_R p, 1\right)^2). \tag{8.22}$$

For $l = 1$, the sum in (8.18) consists of those terms with $d_0, d_1, d_2 \in \{1, p\}$, excluding the triple $d_0 = d_1 = d_2 = 0$, thus

$$c_{p,t_0,t_1,t_2} = \left(1 + \frac{\chi(p)}{p^{\frac{1+it_0}{\log x}}}\right)\left(1 - \frac{1}{p^{\frac{1+it_1}{\log R}}}\right)\left(1 - \frac{1}{p^{\frac{1+it_2}{\log R}}}\right) - 1.$$

We thus have

$$E_{p,t_{0,1},\dots,t_{2,k}} = 1 + \sum_{j=1}^{k} \frac{c_{p,t_{0,j},t_{1,j},t_{2,j}}}{p}\left(1 - \frac{1}{p^{\frac{1+it_{1,j}}{\log R}}}\right)\left(1 - \frac{1}{p^{\frac{1+it_{2,j}}{\log R}}}\right) - \frac{1}{p}$$
$$+ O(\frac{\min((1+|t|)\log_R p, 1)^2}{p^2}).$$

For $p \geqslant C_0$, we may use the trivial bound $c_{p,t_{0,j},t_{1,j},t_{2,j}} \ll 1$ to factor

$$E_{p,t_{0,1},\dots,t_{2,k}} = \left(1 + \sum_{j=1}^{k} \frac{c_{p,t_{0,j},t_{1,j},t_{2,j}}}{p}\right) \exp\left(O\left(\frac{\min((1+|t|)\log_R p, 1)^2}{p^2}\right)\right).$$

Since

$$\sum_p \frac{\min((1+|t|)\log_R p, 1)^2}{p^2} \ll \frac{(1+|t|)^2}{\log^2 R} \sum_p \frac{1}{p^{3/2}} \ll \frac{1}{\log R},$$

we thus have

$$\prod_{p \geqslant C_0} E_{p,t_{0,1},\dots,t_{2,k}} = \exp\left(O\left(\frac{1}{\log R}\right)\right) \prod_{p \geqslant C_0} \left(1 + \sum_{j=1}^{k} \frac{c_{p,t_{0,j},t_{1,j},t_{2,j}}}{p}\right).$$

Let us compare $c_{p,t_{0,j},t_{1,j},t_{2,j}}$ against the quantity

$$c'_{p,t_{0,j},t_{1,j},t_{2,j}} := \left(1 - \frac{1}{p^{\frac{1+it_{0,j}}{\log x}}}\right)\left(1 - \frac{1}{p^{\frac{1+it_{1,j}}{\log R}}}\right)\left(1 - \frac{1}{p^{\frac{1+it_{2,j}}{\log R}}}\right) - 1.$$

The two quantities agree unless p is exceptional. From the triangle inequality, we have the crude bound

$$c_{p,t_{0,j},t_{1,j},t_{2,j}} - c'_{p,t_{0,j},t_{1,j},t_{2,j}} \ll \frac{1}{p^{\frac{1}{\log x}}},$$

and when $p \leqslant x$ we can use Taylor expansion and (8.21) to also obtain the bound

$$c_{p,t_{0,j},t_{1,j},t_{2,j}} - c'_{p,t_{0,j},t_{1,j},t_{2,j}} \ll (\log^{1/(100k)} \eta \log_R p)^2.$$

Thus in all cases one has the bound

$$c_{p,t_{0,j},t_{1,j},t_{2,j}} - c'_{p,t_{0,j},t_{1,j},t_{2,j}} \ll 1_{p \text{ exceptional}} \frac{(\log^{1/(100k)} \eta \log_R p)^2}{p^{\frac{1}{\log x}}}. \tag{8.23}$$

Applying Corollary 3.6, we conclude that

$$\prod_{p \geqslant C_0} E_{p,t_{0,1},\dots,t_{2,k}} = \exp\left(O\left(\frac{1}{\log^{1/(7k)} \eta}\right)\right) \prod_{p \geqslant C_0} \left(1 + \sum_{j=1}^{k} \sum_{p \geqslant C_0} \frac{c'_{p,t_{0,j},t_{1,j},t_{2,j}}}{p}\right).$$

Bounding $\frac{1}{p^{\frac{1+it_{1,j}}{\log R}}}, \frac{1}{p^{\frac{1+it_{2,j}}{\log R}}} = O(\exp(-\log_R p))$, we have

$$c'_{p,t_{0,j},t_{1,j},t_{2,j}} = -\frac{1}{p^{\frac{1+it_{0,j}}{\log x}}} + O(\exp(-2\log_R p)),$$

while from Taylor expansion we see for $p \leqslant x$ that

$$\begin{aligned} c'_{p,t_{0,j},t_{1,j},t_{2,j}} &= -1 + O((1+|t|)\log_x p) \\ &= -\frac{1}{p^{\frac{1+it_{0,j}}{\log x}}} + O((1+|t|)\log_x p) \\ &= -\frac{1}{p^{\frac{1+it_{0,j}}{\log x}}} + O\left(\frac{1}{\log^{1/(6k)} \eta} \log_R p\right) \end{aligned} \tag{8.24}$$

due to (8.21), (2.8). Combining the bounds, we see that

$$c'_{p,t_{0,j},t_{1,j},t_{2,j}} = -\frac{1}{p^{\frac{1+it_{0,j}}{\log x}}} + O\left(\min\left(\frac{\log_R p}{\log^{1/(6k)} \eta}, \exp(-2\log_R p)\right)\right)$$

for all $p \geqslant C_0$. From Mertens' theorem ((3.3) for $\log_R p \leqslant \log^{1/(100k)} \eta$ and (3.5) for $\log_R p > \log^{1/(100k)} \eta$), we have

$$\sum_p \frac{\min\left(\frac{1}{\log^{1/(6k)} \eta} \log_R p, \exp(-2\log_R p)\right)}{p} \ll \frac{1}{\log^{1/(7k)} \eta}$$

(say). We conclude that

$$\prod_{p \geqslant C_0} E_{p,t_{0,1},\dots,t_{2,k}} = \exp\left(O\left(\frac{1}{\log^{1/(7k)} \eta}\right)\right) \prod_{p \geqslant C_0} \left(1 - \sum_{j=1}^k \frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}\right).$$

The function

$$\prod_{p \geqslant C_0} \frac{1 - \sum_{j=1}^k \frac{1}{p^{s_j}}}{\prod_{j=1}^k (1 - \frac{1}{p^{s_j}})}$$

converges to a holomorphic function of $s_1, \dots, s_k$ in the polydisk $\prod_{j=1}^k \{s_j : |s_j - 1| \leqslant 1/2\}$ which is bounded in magnitude by $O(1)$ (since each factor here is $1 + O(1/p^2)$). From the Cauchy integral formula, we conclude that

$$\prod_{p \geqslant C_0} \frac{1 - \sum_{j=1}^k \frac{1}{p^{s_j}}}{\prod_{j=1}^k (1 - \frac{1}{p^{s_j}})} = \prod_{p \geqslant C_0} \frac{1 - \sum_{j=1}^k \frac{1}{p}}{\prod_{j=1}^k (1 - \frac{1}{p})} (1 + \max(|s_1 - 1|, \dots, |s_k - 1|))$$

when $|s_1 - 1|, \dots, |s_k - 1| \leqslant \frac{1}{4}$. Observing from (1.2) that

$$\frac{1 - \sum_{j=1}^k \frac{1}{p}}{\prod_{j=1}^k (1 - \frac{1}{p})} = \beta_p$$

for $p \geqslant C_0$, we conclude (using (8.21)) that

$$\prod_{p \geqslant C_0} \frac{1 - \sum_{j=1}^k \frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}}{\prod_{j=1}^k \left(1 - \frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}\right)} = \exp\left(O\left(\frac{1+|t|}{\log x}\right)\right) \prod_{p \geqslant C_0} \beta_p$$

and thus (by (8.21), (2.8))

$$\prod_{p \geqslant C_0} E_{p,t_{0,1},\dots,t_{2,k}} = \exp\left(O\left(\frac{1}{\log^{1/(7k)} \eta}\right)\right) \prod_{p \geqslant C_0} \beta_p \prod_{j=1}^k \left(1 - \frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}\right). \tag{8.25}$$

Now we turn attention to the small primes $p < C_0$. Using (8.17) and (8.22), we have

$$E_{p,t_{0,1},\dots,t_{2,k}} = \sum_{d_1,\dots,d_k \in \{1,p\}} \frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{[d_1,\dots,d_k]} \prod_{j=1}^{k} c_{d_j,t_{0,j},t_{1,j},t_{2,j}} + O\left(\left((1+|t|)\frac{\log C_0}{\log R}\right)^2\right)$$

for $p \leqslant C_0$, which we rewrite as

$$E_{p,t_{0,1},\dots,t_{2,k}} = 1 + \sum_{d_1,\dots,d_k \in \{1,p\}:[d_1,\dots,d_k]=p} \frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{p} \prod_{j=1}^{k} c_{d_j,t_{0,j},t_{1,j},t_{2,j}} + O\left(\left((1+|t|)\frac{\log C_0}{\log R}\right)^2\right)$$

From (8.21) and (2.8), the error term is certainly $O(\frac{1}{\log^{1/(7k)}\eta})$. From (8.23), (8.24), (8.21) and (2.8), we similarly have

$$c_{p,t_{0,j},t_{1,j},t_{2,j}} = -1 + O\left(\frac{1}{\log^{1/(7k)}\eta}\right)$$

for $p \leqslant C_0$, and thus

$$c_{d_j,t_{0,j},t_{1,j},t_{2,j}} = \mu(d_j) + O\left(\frac{1}{\log^{1/(7k)}\eta}\right)$$

for $j = 0,\dots,k$. This gives

$$E_{p,t_{0,1},\dots,t_{2,k}} = 1 + \sum_{d_1,\dots,d_k \in \{1,p\}:[d_1,\dots,d_k]=p} \mu(d_1)\cdots\mu(d_k) \frac{\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}}{p} + O\left(\frac{1}{\log^{1/(7k)}\eta}\right).$$

If the h_i occupy m distinct residue classes $b_1,\dots,b_m$ modulo p, then the constraint $\prod_{1\leqslant i<j\leqslant k} 1_{(d_i,d_j)|h_i-h_j}$ constrains the index set $\{i : d_i = p\}$ to be a subset of one of the sets $\{i : h_i = b_j\ (p)\}$ for $j = 1,\dots,m$, which must be non-empty if $[d_1,\dots,d_k]$ is to equal p. From the alternating sign of the Möbius function, each j has a net contribution of $-\frac{1}{p}$ to the above sum, thus

$$E_{p,t_{0,1},\dots,t_{2,k}} = 1 - \frac{m}{p} + O\left(\frac{1}{\log^{1/(7k)}\eta}\right).$$

From (1.2), we have

$$\beta_p = \left(1-\frac{m}{p}\right)\left(1-\frac{1}{p}\right)^{-k}$$

and thus

$$E_{p,t_{0,1},\dots,t_{2,k}} = \left(1-\frac{1}{p}\right)^{k}\beta_p + O\left(\frac{1}{\log^{1/(7k)}\eta}\right);$$

by Taylor expansion and (8.21), we then have

$$E_{p,t_{0,1},\dots,t_{2,k}} = \beta_p \prod_{j=1}^{k}\left(1-\frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}\right)+O\left(\frac{1}{\log^{1/(7k)}\eta}\right)$$

for $p < C_0$. If we now fix C_0 so that all the previous estimates are justified, we have

$$\prod_{p<C_0} E_{p,t_{0,1},\dots,t_{2,k}} = \prod_{p<C_0}\beta_p \prod_{j=1}^{k}\left(1-\frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}\right)+O\left(\frac{1}{\log^{1/(7k)}\eta}\right)$$

and hence by (8.25), (3.5) and (1.1)

$$\begin{aligned}\prod_{p} E_{p,t_{0,1},\dots,t_{2,k}} &= \mathfrak{S}\prod_{p}\prod_{j=1}^{k}\left(1-\frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}\right)+O\left(\frac{1}{\log^{1/(7k)}\eta}\prod_{p}\left(1-\frac{1}{p^{1+\frac{1}{\log x}}}\right)^{k}\right)\\ &= \mathfrak{S}\prod_{p}\prod_{j=1}^{k}\left(1-\frac{1}{p^{1+\frac{1+it_{0,j}}{\log x}}}\right)+O\left(\frac{\log^{-k}x}{\log^{1/(7k)}\eta}\right).\end{aligned}$$

From the Euler product formula (3.2) as well as (8.21), we conclude that

$$\prod_{p} E_{p,t_{0,1},\dots,t_{2,k}} = \mathfrak{S}\log^{-k}x\prod_{j=1}^{k}(1+t_{0,j})+O\left(\frac{\log^{-k}x}{\log^{1/(7k)}\eta}\right).$$

Inserting this bound into (8.16), and using (8.12) and (8.13) to remove the restriction (8.21), we can thus write the left-hand side of (8.8) as

$$\approx \mathfrak{S}\int_{\mathbb{R}^{3k}}\prod_{j=1}^{k}(1+t_{0,j})F_j(t_{0,j})f(t_{1,j})f(t_{2,j})\,dt_{0,j}dt_{1,j}dt_{2,j}.$$

Applying (8.14) and (8.15), this is $\approx \mathfrak{S}$, giving the claim. This (finally!) concludes the proof of Theorem 1.6.

ACKNOWLEDGEMENTS

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-1764034. JT was supported by a Titchmarsh Fellowship and Academy of Finland grant no. 340098.

The authors thank Kaisa Matomäki and Jori Merikoski for pointing out a slight correction to the proof of Proposition 3.5 in an earlier version of this paper. The authors would also like to thank the referee for helpful comments and suggestions.

Open access publishing facilitated by *The University of Newcastle*, as part of the Wiley - The University of Newcastle agreement via the Council of Australian University Librarians.

REFERENCES

1. D. A. Burgess, *On character sums and primitive roots*, Proc. London Math. Soc. (3) **12** (1962), 179–192.
2. J. Chinis, *Siegel zeros and Sarnak's conjecture*. Preprint, arXiv:2105.14653 (2021).
3. S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*, Math. Appl., vol. 4, Gordon and Breach Science Publishers, New York–London–Paris, 1965.
4. T. Estermann, *Über die darstellungen einer zahl als differenz von zwei produkten*, J. reine angew. Math. **164** (1931), 173–182.
5. T. Estermann, *On Kloosterman's sum*, Mathematika **8** (1961), 83–86.
6. L. Germán and I. Kátai, *On multiplicative functions on consecutive integers*, Lith. Math. J. **50** (2010), no. 1, 43–53.
7. D. A. Goldston and C. Y. Yıldırım, *Higher correlations of divisor sums related to primes. I. Triple correlations*, Integers **3** (2003), A5, 66.
8. A. Granville and K. Soundararajan, *Large character sums: pretentious characters and the Pólya-Vinogradov theorem*, J. Amer. Math. Soc. **20** (2007), no. 2, 357–384.
9. B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850.
10. 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** (1923), no. 1, 1–70.
11. D. R. Heath-Brown, *Prime twins and Siegel zeros*, Proc. Lond. Math. Soc. (3) **47** (1983), no. 2, 193–224.
12. H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc. Colloquium Publ., vol. 53, Amer. Math. Soc., Providence, R.I., 2004.
13. H. D. Kloosterman, *On the representation of numbers in the form* $ax^2 + by^2 + cz^2 + dt^2$, Acta Math. **49** (1927), no. 3-4, 407–464.
14. B. Landreau, *A new proof of a theorem of van der Corput*, Bull. Lond. Math. Soc. **21** (1989), no. 4, 366–368.
15. J. D. Lichtman, *Averages of the Möbius function on shifted primes*, Quart. J. Math. (2020), no. 2, 729–757.
16. J. D. Lichtman and J. Teräväinen, *On the Hardy-Littlewood-Chowla conjecture on average*. Forum Math. Sigma 2022.
17. K. Matomäki and J. Merikoski, *Siegel zeros, twin primes, Goldbach's conjecture, and primes in short intervals*. Preprint, arXiv:2112.11412, 2021.
18. K. Matomäki, M. Radziwiłł, X. Shao, T. Tao, and J. Teräväinen, *Higher uniformity of arithmetic functions in short intervals II. Almost all intervals*. Under preparation.
19. K. Matomäki, X. Shao, T. Tao, and J. Teräväinen, *Higher uniformity of arithmetic functions in short intervals I. All intervals*. Preprint, arXiv:2204.03754, 2022.
20. H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Stud. Adv. Math., vol. 97, Cambridge Univ. Press, Cambridge, 2007.
21. G. I. Perelmuter, *Estimate of a sum along an algebraic curve*, Mat. Z. **5** (1969), 373–380.
22. J. Pintz, *Are there arbitrarily long arithmetic progressions in the sequence of twin primes?* In An irregular mind, Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., Budapest, 2010, pp. 525–559.
23. J. Pintz, *An approximation to the twin prime conjecture and the parity phenomenon*, Indag. Math. (N.S.) **26** (2015), no. 5, 883–896.
24. D. H. J. Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes*, Res. Math. Sci. **1** (2014), 12.
25. M. Ram Murty and A. Vatwani, *Twin primes and the parity problem*, J. Number Theory **180** (2017), 643–659.
26. W. Sawin and M. Shusterman, *On the Chowla and twin primes conjectures over* $\mathbb{F}_q[T]$, Ann. of Math. **196** (2022), no. 2, 457–506.
27. T. Tao, *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations*, Forum Math. Pi **4** (2016), e8, 36.
28. T. Tao and J. Teräväinen, *Odd order cases of the logarithmically averaged Chowla conjecture*, J. Théor. Nombres Bordeaux **30** (2018), no. 3, 997–1015.
29. T. Tao and J. Teräväinen, *The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures*, Duke Math. J. **168** (2019), no. 11, 1977–2027.
30. A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U.S.A. **34** (1948), 204–207.