Correlations of almost primes

Evans, Natalie

Awarding institution:
King's College London

Correlations of almost primes

Natalie Rebecca Evans

A thesis presented for the degree of

Doctor of Philosophy in Pure Mathematics Research

King's College London

December 2022

Abstract

We prove that analogues of the Hardy-Littlewood generalised twin prime conjecture for almost primes hold on average. Our main theorem establishes an asymptotic formula for the number of integers $n = p_1 p_2 \leq X$ such that $n + h$ is a product of exactly two primes which holds for almost all $|h| \leq H$ with $(\log X)^{19+\varepsilon} \leq H \leq X^{1-\varepsilon}$, under a restriction on the size of one of the prime factors of n and $n + h$. Additionally, we consider correlations $n, n + h$ where n is a prime and $n + h$ has exactly two prime factors, establishing an asymptotic formula which holds for almost all $|h| \leq H$ with $X^{1/6+\varepsilon} \leq H \leq X^{1-\varepsilon}$.

Declaration

This thesis incorporates the following publication:

[7] N. Evans. Correlations of almost primes. Mathematical Proceedings of the Cambridge Philosophical Society, page 1–44, 2022.

DOI: 10.1017/S0305004122000251

Acknowledgements

I would first like to thank my supervisor Dr. Stephen Lester for his constant support and encouragement. I am very grateful for his guidance, for suggesting interesting problems and for countless helpful discussions and comments.

I would like to thank my secondary supervisors, Prof. Igor Wigman and Prof. Abhishek Saha, for their support. I am grateful to Prof. Andrew Granville for his encouragement and advice, and for introducing me to research.

I would like to thank Dr. Julio Andrade and Dr. Ian Petrow for their careful reading of this thesis and their helpful comments. I am also grateful to Dr. Aled Walker for acting as the facilitator during my viva, and for his support.

I would like to thank Dr. Joni Teräväinen for helpful discussions and useful comments on an earlier version of this work, which included Theorem 1.0.4 and its proof, and an improvement on the log exponent in Theorem 1.0.2. I would also like to thank Prof. Kaisa Matomäki for helpful comments on an earlier draft, and the anonymous referee for a careful reading and useful comments on the draft of the published version of this work.

This work was supported by the Engineering and Physical Sciences Research Council [EP/R513106/1]. I would also like to thank Katie Hale for her kindness and help.

I would like to thank Scott, Rick, Anna, Kandy, Lexi, Josh, Amy, Kobe, Ava, Chloe and Cassie for their love and for always being there for me. I am grateful to my friends for their support, and to Jasna, Sophie, Antonino, Niki and Aryan for going above and beyond.

Lastly, I would like to thank my Mum, as this would not have been possible without her unconditional love, support and belief in me. I wish she could have made it to the end of this journey, and I hope she would be proud.

Contents

1 Introduction **8**

2 Correlations of almost primes **17**

2.1 Proof Sketch 19

2.2 Preliminaries 28

2.2.1 Notation 28

2.2.2 Preliminary Lemmas 29

2.3 Proof of Theorem 1.0.2 36

2.3.1 Applying the Circle Method 36

2.3.2 The Minor Arcs 44

2.3.3 The Major Arcs 50

2.3.3.1 Expanding the Exponential Sum 50

2.3.3.2 Evaluating the Main Term 54

2.3.4 The Error Term of the Major Arcs 61

2.3.4.1 Reduction of the problem 62

2.3.4.2 Bounding $B_2(X)$ 66

2.3.4.3 Bounding $B_1(X)$ 70

2.3.5 Preliminaries on Dirichlet Polynomials 76

2.3.5.1 Definitions 76

2.3.5.2 Decomposing Dirichlet Polynomials 77

2.3.5.3 Mean Value Theorems for Dirichlet Polynomials 79

2.3.5.4 Large Value Theorems 80

2.3.5.5 Moments of Dirichlet Polynomials 82

2.3.6 Bounding the Mean Value of a Dirichlet Polynomial . . 87

2.3.6.1 The contribution of $\mathcal{S}_1$ 93

2.3.6.2 The contribution of the complement of $\mathcal{S}_1$ . . 97

2.3.6.2.1 Type II Sums. 101

2.3.6.2.2 Type I Sums. 111

2.3.6.3 Completing the proof of Proposition 2.3.21 . . 118

2.4 Proof of Theorem 1.0.4 122

2.5 Proof of Theorem 1.0.5 131

3 Future Outlook **137**

A Primes in short arithmetic progressions **141**

A.1 Preliminary Results 142

A.2 Proof of the Lemma 143

Chapter 1

Introduction

We have known since Euclid that there are infinitely many primes. Naturally, we then ask how the primes are distributed and whether there exist patterns among them. The prime number theorem was proved independently by de la Vallée Poussin [40] and Hadamard [13] in 1896 and describes the distribution of the primes, establishing the asymptotic formula

$$\pi(X) := \#\{p \leq X : p \text{ prime}\} \sim \frac{X}{\log X}. \tag{1.0.1}$$

When looking for patterns among the primes, we see that pairs with difference two often appear, and the twin prime conjecture famously states that there are infinitely many primes p such that $p+2$ is also prime. More generally, it is conjectured that there are infinitely many primes p such that $p+h$ is prime, where h is an even integer. These questions remain open today.

Hardy and Littlewood [14] conjectured the following asymptotic formula for the number of primes $p \le X$ such that $p + h$ is also prime

$$\#\{p \le X : p, p + h \text{ both prime}\} \sim \mathfrak{S}(h) \frac{X}{\log^2 X} \tag{1.0.2}$$

as $X \to \infty$, where $\mathfrak{S}(h)$ is the singular series defined by

$$\mathfrak{S}(h) := 2\Pi_2 \prod_{\substack{p|h \\ p>2}} \frac{p-1}{p-2} \tag{1.0.3}$$

if h is an even integer and zero if h is odd. Here $\Pi_2 := \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right)$ is the twin prime constant. The Hardy-Littlewood conjecture (1.0.2) is equivalent to showing for any fixed non-zero integer h that

$$\frac{1}{X} \sum_{X<n\le 2X} \mathbb{1}_{\mathbb{P}}(n)\mathbb{1}_{\mathbb{P}}(n+h) \sim \mathfrak{S}(h) \left(\frac{1}{X} \sum_{X<n\le 2X} \mathbb{1}_{\mathbb{P}}(n) \right)^2, \tag{1.0.4}$$

where $\mathbb{1}_{\mathbb{P}}$ is the indicator function of the primes, as $X \to \infty$.

The Hardy-Littlewood conjecture remains wide open, and is not known for any fixed even h. However, there are several results showing it holds on average, that is, the asymptotic formula (1.0.4) holds for almost all shifts $|h| \le H = H(X)$, where H grows with X. We would like to take H as small as possible, with the aim to establish that we can take an average of bounded length.

Van der Corput [41] and Lavrik [22] proved that the Hardy-Littlewood conjecture (1.0.4) holds for almost all $|h| \leq X$. In 1989, Wolke [43] improved on this, proving that if $X^{5/8+\varepsilon} \leq H \leq X^{1-\varepsilon}$, then for all but at most $O_{\varepsilon,A}(H \log^{-A} X)$ values of $|h| \leq H$ we have that (1.0.4) holds for any fixed $A > 0$. Under the assumption of the density hypothesis, Wolke was able to obtain the range $X^{1/2+\varepsilon} \leq H \leq X$. Mikawa [32] was able to go beyond this unconditionally in 1991, improving the range to $X^{1/3+\varepsilon} \leq H \leq X$. The shortest known average is due to Matomäki, Radziwiłł and Tao [27][1], who showed that if $0 \leq h_0 \leq X^{1-\varepsilon}$ and $X^{8/33+\varepsilon} \leq H \leq X^{1-\varepsilon}$, then (1.0.4) holds for all but $O_{\varepsilon,A}(H \log^{-A} X)$ values of h such that $|h - h_0| \leq H$.

Our first result establishes an analogue of the Hardy-Littlewood conjecture for integers which have exactly two prime factors (called E_2 numbers) which holds on average, provided we restrict the size of one of the prime factors.

Definition 1.0.1. Given $P > 0$ and fixed $\delta > 0$ we define $E_2' := E_2'(P)$ to be the set of integers $n = p_1 p_2$ with exactly two prime factors such that $p_1 \in (P, P^{1+\delta}]$.

The presence of the two prime factors gives the problem a bilinear structure which enables us to go further and we show an asymptotic formula for

[1]Matomäki, Radziwiłł and Tao note that their result can also be proved in the range $X^{1-\varepsilon} \leq H \leq X$ by their methods. Theirs and the preceding results are proved with a better error term.

the correlation

$$\frac{1}{X}\sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n)\mathbb{1}_{E_2'}(n+h),$$

where $\mathbb{1}_{E_2'}$ is the indicator function of the set E_2', which holds for almost all $|h| \leq H$ with $(\log X)^{19+\varepsilon} \leq H \leq X\log^{-A} X$ and $A > 3$.

Theorem 1.0.2. *Let $\varepsilon > 0$, $A > 3$ be fixed and let $(\log X)^{19+\varepsilon} \leq H \leq X\log^{-A} X$. Then, there exists some $\eta = \eta(\varepsilon) > 0$ such that*

$$\frac{1}{X}\sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n)\mathbb{1}_{E_2'}(n+h) \sim \mathfrak{S}(h)\left(\frac{1}{X}\sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n)\right)^2$$

holds for all but at most $O(H\log^{-\eta} X)$ values of $0 < |h| \leq H$. Here we define E_2' as in Definition 1.0.1 with

$$P := \begin{cases} (\log X)^{17+\varepsilon}, & \textit{if } (\log X)^{19+\varepsilon} \leq H \leq \exp((\log X)^{\varepsilon^3}), \\ \exp\left((\log\log X)^2\right), & \textit{if } \exp((\log X)^{\varepsilon^3}) < H \leq X\log^{-A} X. \end{cases}$$

Remark 1.0.3. Here, it is crucial that the integers n and $n+h$ have exactly two prime factors, not just at most two (such integers are called P_2 numbers). As we will discuss later in this chapter, there are previous results considering P_2 numbers which are proved using sieve theory.

Also, the range $X\log^{-A} X \leq H \leq X$ can be dealt with by the same methods, see for example [32], [27].

We can prove a similar asymptotic formula for correlations of general E_2

numbers which holds on average using the same methods. Here, E_2 is the set of integers with exactly two prime factors, with no restriction on the sizes of the prime factors as in Definition 1.0.1 of E_2'. Making some adjustments to the proof of Theorem 1.0.2, we obtain an asymptotic formula for correlations $n, n+h \in E_2$ which holds for almost all $|h| \leq H$. The cost of considering the set of E_2 numbers is taking H larger than in the previous theorem, although we still go beyond what is known for primes.

Theorem 1.0.4. *Let* $\varepsilon > 0$, $B > 0$, $A > 3$ *be fixed and let* $\exp\left((\log X)^{1-\varepsilon}\right) \leq H \leq X \log^{-A} X$. *Then, we have that*

$$\frac{1}{X} \sum_{X < n \leq 2X} \mathbb{1}_{E_2}(n) \mathbb{1}_{E_2}(n+h) \sim \mathfrak{S}(h) \left(\frac{1}{X} \sum_{X < n \leq 2X} \mathbb{1}_{E_2}(n) \right)^2$$

for all but at most $O(H \log^{-B} X)$ *values of* $0 < |h| \leq H$.

We can also combine our argument with the work of Mikawa [32] on correlations of primes to study correlations $n, n+h$ where n is a prime and $n+h$ is an E_2 number on average. We are still able to take advantage of the bilinear structure provided by the almost prime to go further than what is known for primes and prove an asymptotic formula which holds for almost all $|h| \leq H$ with H as small as $X^{1/6+\varepsilon}$.

Theorem 1.0.5. *Let* $\varepsilon > 0$ *be fixed sufficiently small,* $B > 0$, $A > 5$ *be fixed* *and let* $X^{1/6+\varepsilon} \leq H \leq X \log^{-A} X$*. Then, we have that*

$$\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{\mathbb{P}}(n)\mathbb{1}_{E_2}(n+h) \sim \mathfrak{S}(h) \Bigg(\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{\mathbb{P}}(n) \Bigg) \Bigg(\frac{1}{X} \sum_{X<m\leq 2X} \mathbb{1}_{E_2}(m) \Bigg)$$

for all but at most $O(H \log^{-B} X)$ *values of* $0 < |h| \leq H$*.*

There are a number of previous works which use sieve theory to obtain results on gaps between primes and almost primes. In the last twenty years, there have been several breakthroughs on bounded gaps between primes. Goldston, Pintz and Yıldırım [12] proved in 2005 that there exist consecutive primes closer than any arbitrarily small multiple of the average spacing. In particular, if p_n is the n-th prime, then by the prime number theorem (1.0.1) the average spacing is $\log p_n$, and Goldston, Pintz and Yıldırım's result states

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

Assuming the Elliott-Halberstam conjecture, the authors prove that there are infinitely many pairs of consecutive primes differing by at most 16, that is

$$\liminf_{n\to\infty} (p_{n+1} - p_n) \leq 16. \tag{1.0.5}$$

The Elliott-Halberstam conjecture [6] (see also [4, 8]) concerns the distribution of primes in arithmetic progressions.

Conjecture 1.0.6 (Elliott-Halberstam Conjecture)**.** *For every* $A > 0$ *and*

$0 < \theta < 1$ *we have that*

$$\sum_{q \le x^\theta} \max_{(a,q)=1} \left| \sum_{\substack{n \le x \\ n \equiv a(q)}} \Lambda(n) - \frac{x}{\varphi(q)} \right| \ll \frac{x}{\log^A x}.$$

In 2013, Zhang [44] in breakthrough work established unconditionally that there exist infinitely many bounded gaps between the primes. In particular, Zhang showed that

$$\liminf_{n \to \infty} (p_{n+1} - p_n) < 7 \times 10^7.$$

Later in 2013, Maynard [30] introduced a new idea which substantially simplified the proof of this result and established the improved bound 600 for (1.0.5) unconditionally. Under the Elliott-Halberstam conjecture, Maynard obtains the bound 12. The Polymath8b project [36] subsequently improved the unconditional bound to 246, and under the assumption of the generalised Elliott-Halberstam conjecture obtained (1.0.5) with 6. The twin prime conjecture would amount to proving (1.0.5) with the bound 2.

Goldston, Graham, Pintz and Yıldırım [11] proved an almost prime analogue of (1.0.5) which holds unconditionally; if $q_1 < q_2 < \cdots$ denotes the sequence of products of exactly two distinct primes, then

$$\liminf_{n \to \infty} (q_{n+1} - q_n) \le 6.$$

These results on bounded gaps between primes and almost primes are proved using sieve theory and the arguments do not establish an asymptotic formula. However, sieve methods can be used to obtain an upper bound for the correlation (1.0.4), namely

$$\frac{1}{X}\sum_{X<n\leq 2X}\mathbb{1}_{\mathbb{P}}(n)\mathbb{1}_{\mathbb{P}}(n+h)\ll\frac{\mathfrak{S}(h)}{(\log X)^2}. \tag{1.0.6}$$

Chen's theorem gives that $p+2=q$ such that p is prime and q is either a prime or a product of two primes holds for infinitely many primes p. Debouzy [5] proved under the Elliott-Halberstam conjecture that given any $0\leq\beta<\gamma$ there exists X_0 such that for all $X\geq X_0$ we have that

$$\sum_{n\leq X}\Lambda(n)\Lambda(n+2)+\frac{1}{\gamma-\beta}\sum_{n\leq X}\Lambda(n+2)\sum_{\substack{d_1d_2=n\\ n^\beta\leq d_1\leq n^\gamma}}\frac{\Lambda(d_1)\Lambda(d_2)}{\log n}=2\Pi_2X(1+o(1)).$$

This result is proved using an improvement of the Bombieri asymptotic sieve.

More generally, Bombieri [2] had previously considered pairs P_k and $P_k+2=p$ with p prime and P_k an almost prime with at most k factors. More precisely, defining $\Lambda_k(n):=(\mu*\log^k)(n)$ to be the generalised von Mangoldt function where $*$ denotes Dirichlet convolution, Bombieri proved that if $k\geq 1$ is an integer and $x\geq x_0(k)$ we have

$$\sum_{n\leq X}\Lambda_k(n)\Lambda(n+2)=2\Pi_2X(\log X)^{k-1}(k+O(k^{4/3}2^{-k/3}))$$

and, assuming the Elliott-Halberstam conjecture, for $k\geq 2$ we have the asymptotic

$$\sum_{n\leq X} \Lambda_k(n)\Lambda(n+2) \sim 2\Pi_2 kX(\log X)^{k-1}. \tag{1.0.7}$$

Unlike in these results on P_2 numbers with at most two prime factors, integers with exactly two prime factors cannot be counted using sieve methods due to the parity problem, even assuming the Elliott-Halberstam conjecture. In general, sieve methods have difficulty distinguishing between integers with an even and an odd number of prime factors (see, for example, [15, Chapter 2]). So, while these methods can be used to obtain asymptotics such as (1.0.7) for problems involving P_2 numbers which have either one or two prime factors, we would need additional input to separate the contributions coming from each of these sets of integers. In particular, we cannot currently expect to obtain asymptotics for our questions on integers with exactly two prime factors using sieve methods, and the best we can hope for is to establish an upper bound, similar to the prime case (1.0.6). To prove our results we will instead apply the circle method as in the previously discussed works on correlations of primes [27], [32].

Chapter 2

Correlations of almost primes

In this chapter we prove our results on correlations of almost primes and primes. First, in Section 2.3 we will prove that an analogue of the Hardy - Littlewood conjecture for almost primes which have exactly two prime factors holds on average under a restriction on the size of one of the prime factors. We recall the definition of the set E_2':

Definition 1.0.1. Given $P > 0$ and fixed $\delta > 0$ we define $E_2' := E_2'(P)$ to be the set of integers $n = p_1 p_2$ with exactly two prime factors such that $p_1 \in (P, P^{1+\delta}]$.

Then, for correlations of integers $n, n+h \in E_2'$, we will prove that the expected asymptotic formula holds for almost all $|h| \leq H$ with $H \geq (\log X)^{19+\varepsilon}$:

Theorem 1.0.2. *Let* $\varepsilon > 0$*,* $A > 3$ *be fixed and let* $(\log X)^{19+\varepsilon} \leq H \leq$

$X \log^{-A} X$. *Then, there exists some* $\eta = \eta(\varepsilon) > 0$ *such that*

$$\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n)\mathbb{1}_{E_2'}(n+h) \sim \mathfrak{S}(h) \left(\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n) \right)^2$$

holds for all but at most $O(H \log^{-\eta} X)$ *values of* $0 < |h| \leq H$. *Here we define* E_2' *as in Definition 1.0.1 with*

$$P := \begin{cases} (\log X)^{17+\varepsilon}, & \textit{if } (\log X)^{19+\varepsilon} \leq H \leq \exp((\log X)^{\varepsilon^3}), \\ \exp\left((\log\log X)^2\right), & \textit{if } \exp((\log X)^{\varepsilon^3}) < H \leq X \log^{-A} X. \end{cases}$$

Remark 2.0.1. Here, the main term is of size $\sim \mathfrak{S}(h) \frac{\log^2(1+\delta)}{\log^2 X}$ by the prime number theorem and Mertens' theorem (Lemma 2.2.2), where $\delta > 0$ is fixed as in Definition 1.0.1. The choice $A > 3$ ensures that the second term of (2.2.2), which arises in the application of Gallagher's Lemma (Lemma 2.2.4), gives sufficient cancellation.

In Section 2.4, we will adapt the proof of Theorem 1.0.2 to prove that the expected asymptotic formula for correlations of $n, n+h \in E_2$ holds for almost all $|h| \leq H$ with a longer average H than the previous result.

Theorem 1.0.4. *Let* $\varepsilon > 0$, $B > 0$, $A > 3$ *be fixed and let* $\exp\left((\log X)^{1-\varepsilon}\right) \leq H \leq X \log^{-A} X$. *Then, we have that*

$$\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{E_2}(n)\mathbb{1}_{E_2}(n+h) \sim \mathfrak{S}(h) \left(\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{E_2}(n) \right)^2$$

for all but at most $O(H\log^{-B}X)$ *values of* $0<|h|\le H$.

Remark 2.0.2. Here, the main term is of size $\sim \mathfrak{S}(h)\frac{(\log\log X)^2}{\log^2 X}$ by the prime number theorem and Mertens' theorem (Lemma 2.2.2).

Lastly, in Section 2.5 we adapt the argument to establish that the conjectured asymptotic formula for the number of primes p such that $p+h$ has exactly two prime factors holds for almost all $|h|\le H$ with $X^{1/6+\varepsilon}\le H$.

Theorem 1.0.5. *Let* $\varepsilon>0$ *be fixed sufficiently small,* $B>0$, $A>5$ *be fixed and let* $X^{1/6+\varepsilon}\le H\le X\log^{-A}X$. *Then, we have that*

$$\frac{1}{X}\sum_{X<n\le 2X}\mathbb{1}_{\mathbb{P}}(n)\mathbb{1}_{E_2}(n+h)\sim\mathfrak{S}(h)\left(\frac{1}{X}\sum_{X<n\le 2X}\mathbb{1}_{\mathbb{P}}(n)\right)\left(\frac{1}{X}\sum_{X<m\le 2X}\mathbb{1}_{E_2}(m)\right)$$

for all but at most $O(H\log^{-B}X)$ *values of* $0<|h|\le H$.

Remark 2.0.3. Here, the main term is of size $\sim \mathfrak{S}(h)\frac{\log\log X}{\log^2 X}$ by the prime number theorem and Mertens' theorem (Lemma 2.2.2). The choice $A>5$ ensures that the second term of (2.2.2), which arises in the application of Gallagher's Lemma (Lemma 2.2.4), gives sufficient cancellation.

2.1 Proof Sketch

We now discuss the main ideas of the proof of Theorem 1.0.2. We apply the Hardy-Littlewood circle method (see, for example, [42]), first expressing the correlation

$$\sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n)\mathbb{1}_{E_2'}(n+h)$$

in terms of the integral

$$\int_0^1 \left| \sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n)e(n\alpha)\right|^2 e(-h\alpha)d\alpha. \tag{2.1.1}$$

We need to understand which points on the unit circle contribute the main term. Dirichlet's approximation theorem states that for each $Q \geq 1$ there exists $a/q \in \mathbb{Q}$ with $(a,q) = 1$, $1 \leq q \leq Q$ and $|\alpha - a/q| \leq 1/(qQ)$. So, we first aim to understand the behaviour of the exponential sum appearing in (2.1.1) at a rational point a/q with $(a,q) = 1$ on the unit circle. We have that

$$\begin{aligned}\sum_{X<n\leq 2X} \mathbb{1}_{E_2'}(n)e\left(\frac{an}{q}\right) &= \sum_{b=1}^{q} e\left(\frac{ab}{q}\right) \sum_{\substack{X<n\leq 2X \\ n\equiv b \bmod q}} \mathbb{1}_{E_2'}(n) \\ &= \sum_{b=1}^{q} e\left(\frac{ab}{q}\right) \sum_{P<p_1\leq P^{1+\delta}} \sum_{\substack{\frac{X}{p_1}<p_2\leq \frac{2X}{p_1} \\ p_2\equiv b\overline{p_1} \bmod q}} 1.\end{aligned}$$

Heuristically, applying results on primes in arithmetic progressions (e.g. the Siegel-Walfisz Theorem) on the inner sum followed by Mertens' Theorem (Lemma 2.2.2) on the sum over p_1, we would expect that this is

$$\approx \frac{\mu(q)c_\delta X}{\varphi(q)\log X},$$

where $c_\delta > 0$ is some constant depending on δ. This suggests that the larger contributions arise when $\alpha \in (0, 1)$ is well approximated by a rational a/q with a small denominator q.

We therefore split the integral (2.1.1) over the unit circle into integrals over the major arcs, the set of points in $(0, 1)$ which are well approximated by a rational with a small denominator, i.e. the set of $\alpha \in (0, 1)$ such that $|\alpha - a/q| \le 1/(qQ)$ for some integers $(a, q) = 1$ with $1 \le q \le \log^{A'} X$ for some bounded $A' > 0$, and the minor arcs consisting of the rest of the circle. Here, Q is slightly larger than the size of the smaller prime factor P, and depends on the size of $H(X)$ (in particular, it is a power of $\log X$, or larger when $H \ge \exp((\log X)^\varepsilon)$). In order to achieve the smallest possible H, we want to take A', P and Q as small as possible.

In many problems of this type (see e.g. [27], [32]) where the Hardy-Littlewood circle method is applied, it is usual that the major arcs are treated in a standard way to provide the main term and an error term which is not too difficult to control, while the contribution from the minor arcs is more difficult to bound suitably. Since the correlation

$$\sum_{X<n\le 2X} \mathbb{1}_{E_2'}(n)\mathbb{1}_{E_2'}(n+h) = \sum_{P<p_1,p_3\le P^{1+\delta}} \sum_{\substack{X<p_1p_2,p_3p_4\le 2X \\ p_3p_4=p_1p_2+h}} 1$$

has a bilinear structure, we are in fact able to bound the integral over the minor arcs with relative ease using standard results on bilinear exponential sums. For the major arcs, while we are still able to evaluate the main term in the usual way, the difficulty now lies in estimating the error term.

We will first treat the integral over the minor arcs in Proposition 2.3.6. We find cancellation in the contribution on average over the shift h:

$$\sum_{0<|h|\leq H}\left|\int_{\mathfrak{m}}\left|\sum_{X<n\leq 2X}\mathbb{1}_{E_2'}(n)e(n\alpha)\right|^2 e(-h\alpha)d\alpha\right|^2.$$

We next expand the square, apply Poisson summation (Lemma 2.2.1) and Gallagher's Lemma [10, Lemma 1]:

Lemma 2.2.4 (Gallagher's Lemma)**.** *Let* $2<y<X/2$*. For arbitrary complex numbers* a_n*, we have*

$$\int_{|\beta|\leq\frac{1}{2y}}\left|\sum_{X<n\leq 2X}a_ne(\beta n)\right|^2 d\beta \ll \frac{1}{y^2}\int_X^{2X}\left|\sum_{x<n\leq x+y}a_n\right|^2 dx + y\left(\max_{X<n\leq 2X}|a_n|\right)^2. \tag{2.2.2}$$

This reduces the problem to bounding an integral of the form

$$\sup_{\alpha\in\mathfrak{m}}\int_X^{2X}\left|\sum_{x<n\leq x+H}\mathbb{1}_{E_2'}(n)e(n\alpha)\right|^2 dx = \sup_{\alpha\in\mathfrak{m}}\int_X^{2X}\left|\sum_{\substack{x<p_1p_2\leq x+H\\ P<p_1\leq P^{1+\delta}}}e(\alpha p_1p_2)\right|^2 dx.$$

The bilinear structure of these sums means we get the required cancellation, as seen in the work of Mikawa [32]. We apply the Cauchy-Schwarz inequality before separating the contributions of the diagonal and off-diagonal terms. The diagonal terms are bounded trivially and a standard argument for bounding bilinear exponential sums is used to bound the off-diagonal terms.

The major arcs, treated in Proposition 2.3.7, contribute the main term, which is evaluated in a standard way, and an error term. On the major arcs, we can write $\alpha = a/q + \beta$ with $q \leq Q_0$, $(a, q) = 1$ and $|\beta| \leq \frac{1}{qQ}$. We then need to evaluate

$$\sum_{q \leq Q_0} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \int_{|\beta| \leq \frac{1}{qQ}} \left| \sum_{X < n \leq 2X} \mathbb{1}_{E_2'}(n) e\left(\frac{an}{q}\right) e(n\beta) \right|^2 e\left(-\frac{ah}{q}\right) e(-h\beta) d\beta.$$

We expand the additive character $e(an/q)$ in terms of Dirichlet characters χ mod q, and after applying character orthogonality the problem is transformed into understanding

$$\sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{|\beta| \leq \frac{1}{qQ}} \left| \sum_{X < n \leq 2X} \mathbb{1}_{E_2'}(n) \chi(n) e(\beta n) \right|^2 d\beta.$$

A suitable approximation to the principal character then provides the main term.

To the remaining terms, we again apply Gallagher's Lemma (Lemma 2.2.4) to reduce the problem to understanding almost primes in almost all short intervals. We add and subtract a sum over a longer interval, so that we aim to estimate an expression of the form

$$\sum_{q\le \log^{A'} X} \frac{q}{\varphi(q)} \sum_{\substack{\chi(q)\\ \chi\neq\chi_0}} \Bigg(\int_X^{2X} \Bigg| \Bigg(\frac{2}{qQ} \sum_{x<n\le x+qQ/2} - \frac{2}{q\Delta} \sum_{x<n\le x+q\Delta/2} \Bigg) \mathbb{1}_{E_2'}(n)\chi(n) \Bigg|^2 dx + \int_X^{2X} \Bigg| \frac{2}{q\Delta} \sum_{x<n\le x+q\Delta/2} \mathbb{1}_{E_2'}(n)\chi(n) \Bigg|^2 dx \Bigg), \tag{2.1.2}$$

with Δ slightly smaller than X. To the second term, we first apply Cauchy-Schwarz to separate the two prime factors. Since the length of the interval is close to X and we need an estimate for almost all intervals, we only need a result which is slightly stronger than the prime number theorem. For the estimation of the first term, we adapt the work of Teräväinen [38] on almost primes in almost all short intervals (which in turn adapts the work of Matomäki and Radziwiłł [25] on multiplicative functions in short intervals). In particular, we first use a Parseval-type bound (Lemma 2.2.7) in order to bound the integral in terms of the mean square of the associated Dirichlet polynomial

$$\sum_{q\le Q_0} \sum_{\substack{\chi(q)\\ \chi\neq\chi_0}} \int_{-T}^{T} \Bigg| \sum_{X<n\le 2X} \frac{\mathbb{1}_{E_2'}(n)\chi(n)}{n^{1+it}} \Bigg|^2 dt. \tag{2.1.3}$$

Finding cancellation in this mean value is the crux of the argument, and is covered in detail in Sections 2.3.5 and 2.3.6.

For now, if we were only interested in achieving the range $H \ge X^{\varepsilon}$, we would first choose the smaller prime factor of the almost primes $n, n+h \in E_2'$ to instead have size $P = \exp((\log X)^{3/4})$. We choose the parameters of the circle method to be $Q_0 = \log^{A'} X$ with $A' > 4$ fixed and $Q = \exp((\log X)^{4/5})$. To ensure cancellation in the contribution of the minor arcs, we will need $H > Q$, so we take $H \geq X^{\varepsilon}$.

Heuristically, we first factorise the Dirichlet polynomial appearing in (2.1.3) into two Dirichlet polynomials, each associated with one of the prime factors of $n = p_1 p_2$. Ignoring remainder terms, we would need to find cancellation in

$$\sum_{q \leq Q_0} \sum_{\substack{\chi(q) \\ \chi \neq \chi_0}} \int_{-T}^{T} \left| \sum_{P < p_1 \leq 2P} \frac{\chi(p_1)}{p_1^{1+it}} \sum_{X/(2P) < p_2 \leq 2X/P} \frac{\chi(p_2)}{p_2^{1+it}} \right|^2 dt.$$

As $P = \exp((\log X)^{3/4})$, we can find cancellation pointwise in the shorter polynomial using the Vinogradov-Korobov zero-free region for Dirichlet L-functions. We then apply the mean value theorem (Lemma 2.3.27) to the mean value of the Dirichlet polynomial over p_2, so that the above is bounded by

$$\begin{aligned} &\ll \exp(-c(\log X)^{\varepsilon}) \sum_{q \leq Q_0} \sum_{\substack{\chi(q) \\ \chi \neq \chi_0}} \int_{-T}^{T} \left| \sum_{X/(2P) < p_2 \leq 2X/P} \frac{\chi(p_2)}{p_2^{1+it}} \right|^2 dt \\ &\ll \sum_{q \leq Q_0} \left(\frac{\varphi(q) T P}{X} + \frac{\varphi(q)}{q} \right) \exp(-c(\log X)^{\varepsilon}), \end{aligned}$$

for some constant $c > 0$. As $T \approx X/(qQ)$, $P = \exp((\log X)^{3/4})$, $Q =$

$\exp((\log X)^{4/5})$ and $Q_0 = \log^{A'} X$, this is

$$\ll \left(\frac{P}{Q} + 1\right) \exp(-c(\log X)^{\varepsilon}) \sum_{q \leq Q_0} \frac{\varphi(q)}{q} \ll \exp(-c'(\log X)^{\varepsilon}),$$

for some constant $c' > 0$, which is sufficient.

To improve the range of H further and achieve a power of $\log X$, we need to take a smaller prime factor P. In particular, we want to choose P to be a power of $\log X$, which means we can no longer apply the Vinogradov-Korobov zero-free region to obtain cancellation. Instead, we need to use a more involved argument with additional tools and ideas. First, we factorise this Dirichlet polynomial into a short Dirichlet polynomial corresponding to the smaller prime factor p_1 and a longer polynomial corresponding to the larger prime factor p_2. The domain of integration is split according to whether the short polynomial is pointwise small. When the shorter polynomial is small, we apply the pointwise bound followed by a mean value theorem. When this shorter polynomial is large, to get sufficient cancellation we further decompose the long Dirichlet polynomial into products of shorter polynomials using Heath-Brown's identity [17, Eq. (8)], reducing the problem to estimating type I and type II sums. The type I sums occur when these polynomials are sufficiently long and are in fact partial sums related to Dirichlet L-functions. In this case we are able to apply the Cauchy-Schwarz inequality followed by a result on the twisted fourth moment of partial sums of Dirichlet L-functions. Otherwise, for the type II sums, we then further split the domain according to whether one of these polynomials is small, in which case it is bounded pointwise before we use a mean value theorem. When the polynomial is large, we apply the Halász-Montgomery inequality [34, Theorem 7.8] followed by large value theorems.

The proof of Theorem 1.0.4 also follows the argument given above, but we need to make appropriate adjustments to the parameters when applying the circle method and take more care when using the Cauchy-Schwarz inequality. On both the major and minor arcs the application of Cauchy-Schwarz to sums over the smaller prime factor is now too inefficient, but we can overcome this by splitting these sums into dyadic intervals and then combining the contributions. For the proof of Theorem 1.0.5, we combine these ideas for the almost primes with the work of Mikawa [32] on the primes.

We lastly remark that recently, the methods of Matomäki and Radziwiłł [25] have been combined with the Hardy-Littlewood circle method to make progress on other problems in analytic number theory. Matomäki, Radziwiłł and Tao [28] obtained short averages (of length $\log^B X$ for some large $B > 0$) for correlations of divisor functions and the von Mangoldt function, at the cost of weaker error terms. Matomäki, Radziwiłł and Tao [26] use these ideas to establish that Chowla's conjecture [3] holds on average as soon as the length of the average grows with X. Recent work of Lichtman and Teräväinen [24] shows that a hybrid of Chowla's conjecture and the Hardy-Littlewood conjecture holds on average (see also [23]), with average of length a power of $\log X$.

2.2 Preliminaries

2.2.1 Notation

Throughout p, p_i, are used to denote prime numbers, while k, l, m, n, q, r, v (with or without subscripts) are positive integers.

As usual, $\mu(\cdot)$ is the Möbius function and $\varphi(\cdot)$ is the Euler totient function. We let $d_r(n)$ denote the number of solutions to $n = a_1 \cdots a_r$ in positive integers. We let $c_q(\cdot)$ be the Ramanujan sum, defined by

$$c_q(n) := \sum_{\substack{a=1 \\ (a,q)=1}}^{q} e\left(\frac{an}{q}\right).$$

We write $\tau(\cdot)$ for the Gauss sum defined on Dirichlet characters χ modulo q by

$$\tau(\chi) := \sum_{n=1}^{q} e\left(\frac{n}{q}\right) \chi(n), \tag{2.2.1}$$

which satisfies $\tau(\chi_0) = \mu(q)$.

We use $e : \mathbb{T} \to \mathbb{R}$ to denote $e(x) := e^{2\pi i x}$, where $\mathbb{T}$ is the unit circle. The notation $\mathbb{1}_S(\cdot)$ is the indicator function of the set S; in particular, we write $\mathbb{1}_S(n) = 1$ if $n \in S$ and $\mathbb{1}_S(n) = 0$ otherwise. Let $\|x\| := \min_{n\in\mathbb{Z}} |x - n|$ denote distance to the nearest integer. For a function $f \in L^1(\mathbb{R})$, we define its Fourier transform to be $\hat{f}(\xi) = \int_{\mathbb{R}} f(x)e(x\xi)dx$ for all $\xi \in \mathbb{R}$.

We will use (a, b) to denote the greatest common divisor of natural numbers a and b, while we write $a \mid b$ if a divides b. The shorthand $a \equiv b(q)$ is used to denote that a and b are congruent modulo q.

We use the shorthand $\chi(q)$ to denote that the summation is taken over all Dirichlet characters modulo q. For complex functions g_1, g_2 we use the usual asymptotic notation $g_1(x) = O(g_2(x))$ or $g_1(x) \ll g_2(x)$ to denote that there exist real x_0 and $C > 0$ such that for every $x \geq x_0$ we have that $|g_1(x)| \leq C|g_2(x)|$. We write $g_1(x) = o(g_2(x))$ if for every $\varepsilon > 0$ there exists x_0 such that $|g_1(x)| \leq \varepsilon |g_2(x)|$ for all $x \geq x_0$. We use the convention that $\varepsilon > 0$ may be different from line to line.

2.2.2 Preliminary Lemmas

We now state several results we will need throughout the argument. We will need to apply a version of the Poisson summation formula.

Lemma 2.2.1. *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a Schwartz function and suppose that $u \in \mathbb{R}$ and $v \in \mathbb{R}^+$. Then*

$$\sum_{m \in \mathbb{Z}} f(vm + u) = \frac{1}{v} \sum_{n \in \mathbb{Z}} \hat{f}\left(\frac{n}{v}\right) e\left(\frac{un}{v}\right).$$

Proof. See [18, Eq. (4.24)] or [35, Theorem D.3]. □

We will frequently make use of Mertens' Theorem:

Lemma 2.2.2 (Mertens' Theorem)**.** *For $x \geq 2$, we have that*

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + b + O\left(\frac{1}{\log x}\right),$$

where $b = \gamma - \sum_p \sum_{k=2}^{\infty} \frac{1}{kp^k}$ *and* γ *is Euler's constant.*

Proof. See Mertens' [31] or [35, Theorem 2.7]. □

We will need the following bound on primes p such that $p + h$ is a prime, and also a bound for the singular series:

Lemma 2.2.3. *Let* $h \leq x$ *be an even non-zero integer and suppose that* $y \geq 4$*. The number of primes* $p \in (x, x + y]$ *such that* $p + h$ *is also prime is*

$$\ll \frac{\mathfrak{S}(h)y}{(\log y)^2}.$$

Furthermore, we have that

$$\sum_{h \leq x} \mathfrak{S}(h) \ll x$$

and

$$\mathfrak{S}(h) \ll \log \log h.$$

Proof. See [35, Corollary 3.14] and the subsequent exercises. The final bound follows from Mertens' theorem (Lemma 2.2.2). □

We will also need Gallagher's Lemma, which will reduce bounding integrals over the major and minor arcs to studying almost primes in short intervals.

Lemma 2.2.4 (Gallagher's Lemma)**.** *Let* $2 < y < X/2$*. For arbitrary com-*

plex numbers a_n, we have

$$\int_{|\beta|\le\frac{1}{2y}} \left|\sum_{X<n\le 2X} a_n e(\beta n)\right|^2 d\beta \ll \frac{1}{y^2}\int_X^{2X} \left|\sum_{x<n\le x+y} a_n\right|^2 dx + y\left(\max_{X<n\le 2X}|a_n|\right)^2. \tag{2.2.2}$$

Proof. This lemma is a modification of [10, Lemma 1] (see also [32, Lemma 1]). We use a similar argument to [33, Lemma 1.10]. We suppose that the sequence of complex numbers a_n is supported on $(X, 2X]$. We begin by considering the integral

$$\begin{aligned}
&\frac{1}{y^2}\int_X^{2X-y} \left|\sum_{x<n\le x+y} a_n\right|^2 dx \\
&= \frac{1}{y^2}\int_X^{2X} \left|\sum_{x<n\le x+y} a_n\right|^2 dx - \frac{1}{y^2}\int_{2X-y}^{2X} \left|\sum_{x<n\le x+y} a_n\right|^2 dx \\
&\ll \frac{1}{y^2}\int_X^{2X} \left|\sum_{x<n\le x+y} a_n\right|^2 dx + y\left(\max_{X<n\le 2X}|a_n|\right)^2
\end{aligned}$$

giving us the right hand side of (2.2.2). By substitution, we have that

$$\frac{1}{y^2}\int_X^{2X-y} \left|\sum_{x<n\le x+y} a_n\right|^2 dx = \frac{1}{y^2}\int_{X+y/2}^{2X-y/2} \left|\sum_{x-y/2<n\le x+y/2} a_n\right|^2 dx.$$

We define

$$F_y(x) := \begin{cases} \frac{1}{y}, & \text{if } |x| \leq \frac{y}{2}, \\ 0, & \text{if } |x| > \frac{y}{2}, \end{cases}$$

$$C_y(x) := \frac{1}{y} \sum_{|n-x| \leq y/2} a_n,$$

$$S(\beta) := \sum_n a_n e(\beta n).$$

Then $C_y(x) = \sum_n a_n F_y(x-n)$ and, taking Fourier transforms,

$$\begin{aligned} \hat{C}_y(\xi) &= \int_{-\infty}^{\infty} \sum_n a_n F_y(x-n) e(x\xi) dx \\ &= \sum_n a_n e(n\xi) \int_{-\infty}^{\infty} F_y(x-n) e((x-n)\xi) dx \\ &= (S \cdot \hat{F}_y)(\xi). \end{aligned}$$

Note that since only finitely many terms in the sum over n do not vanish, we can interchange the order of summation and integration. The series S is absolutely convergent, so C_y is square-integrable. Therefore, by Plancherel's theorem, we have that

$$\begin{aligned} \frac{1}{y^2} \int_X^{2X-y} \left| \sum_{x<n \leq x+y} a_n \right|^2 dx = \int_{-\infty}^{\infty} |C_y(x)|^2 \, dx &= \int_{-\infty}^{\infty} \left| \hat{C}_y(\beta) \right|^2 d\beta \\ &= \int_{-\infty}^{\infty} \left| S(\beta) \hat{F}_y(\beta) \right|^2 d\beta. \end{aligned}$$

Now, we have that

$$\hat{F}_y(\beta) = \frac{\sin(\pi\beta y)}{\pi\beta y} \gg 1 \text{ for } |\beta| \leq \frac{1}{2y}$$

and therefore

$$\frac{1}{y^2}\int_X^{2X-y}\left|\sum_{x<n\leq x+y} a_n\right|^2 dx \gg \int_{-\frac{1}{2y}}^{\frac{1}{2y}} |S(\beta)|^2 \, d\beta,$$

as claimed. □

We will need a result on primes in short intervals:

Lemma 2.2.5. *Let* $\varepsilon > 0$ *and define* $\psi(x) := \sum_{n\leq x}\Lambda(n)$*. For all* y *with* $x^{7/12+\varepsilon} \leq y \leq x$ *we have*

$$\psi(x+y) - \psi(x) = y + O\left(y\exp(-c(\log x)^{1/3-\varepsilon})\right)$$

for some constant $c > 0$*.*

Proof. This can be proved following the argument of [18, Theorem 10.5]. □

Once we have applied Gallagher's Lemma in the treatment of the major arcs, part of the error term is reduced to a Dirichlet character analogue of a problem on primes in almost all short intervals. We will use the following result adapted from the work of Koukoulopoulos [21] to bound the second term arising in (2.1.2):

Lemma 2.2.6. *Let $A \geq 1$ and $\varepsilon \in (0, \frac{1}{3}]$ be fixed. Let $X \geq 1$, $1 \leq Q \leq \frac{\Delta}{X^{1/6+\varepsilon}}$ and $\Delta = X^\theta$ with $\frac{1}{6} + 2\varepsilon \leq \theta \leq 1$. Then we have that*

$$\sum_{q \leq Q} \sum_{\chi(q)} \int_X^{2X} \left| \sum_{x < n \leq x + q\Delta} (\Lambda(n)\chi(n) - \delta_\chi) \right|^2 dx \ll \frac{Q^3 \Delta^2 X}{\log^A X},$$

where we define $\delta_\chi = 1$ if $\chi = \chi_0$ and $\delta_\chi = 0$ otherwise.

Proof. The proof can be adapted from the proof given in [21, Section 4], as described in the Appendix. □

We use the following Parseval-type result to reduce the problem of finding almost primes in short intervals (cf. the first term of (2.1.2)) to finding cancellation in the mean square of the associated Dirichlet polynomial:

Lemma 2.2.7 (Parseval Bound)**.** *Let a_n be arbitrary complex numbers, and let $2 \leq h_1 \leq h_2 \leq \frac{X}{(T')^3}$ with $T' \geq 1$. Define $F(s) := \sum_{X < n \leq 2X} \frac{a_n}{n^s}$. Then*

$$\frac{1}{X} \int_X^{2X} \left| \frac{1}{h_1} \sum_{x < n \leq x + h_1} a_n - \frac{1}{h_2} \sum_{x < n \leq x + h_2} a_n \right|^2 dx$$
$$\ll \frac{1}{(T')^2} \max_{X < n \leq 2X} |a_n|^2 + \int_{T'}^{\frac{X}{h_1}} |F(1+it)|^2 dt + \max_{T \geq \frac{X}{h_1}} \frac{X}{T h_1} \int_T^{2T} |F(1+it)|^2 dt.$$

Proof. This is [38, Lemma 1], which is a variant of [25, Lemma 14]. □

We record an exponential sum bound and a related bound on the sum of the reciprocal of the distance to the nearest integer function which provide the necessary cancellation in the estimation of the minor arcs.

Lemma 2.2.8. *Let* $\beta \in \mathbb{R}$*, then*

$$\sum_{n \leq x} e(\beta n) \ll \min\left(x, \frac{1}{\|\beta\|}\right).$$

Proof. This is a standard result, see for example [18, Chapter 8, Eq. (8.6)].

□

Lemma 2.2.9. *If* $1 < X \leq Y$ *and* $\alpha \in \mathbb{R}$ *satisfies* $\alpha = a/q + O(q^{-2})$ *with* $(a,q) = 1$*, then we have*

$$\sum_{n \leq X} \min\left(\frac{Y}{n}, \frac{1}{\|\alpha n\|}\right) \ll \left(\frac{Y}{q} + X + q\right) \log(qX).$$

Proof. This is a standard result, see for example [18, Chapter 13, Page 346].

□

We will also need to apply the Brun-Titchmarsh inequality:

Lemma 2.2.10. *If* $(a,q) = 1$*, then for any* $\varepsilon > 0$ *and* $q < x^{1-\varepsilon}$ *we have the bound*

$$\pi(x; q, a) \ll \frac{x}{\varphi(q) \log(x/q)}.$$

Proof. See [39, Theorem 2]. □

2.3 Proof of Theorem 1.0.2

2.3.1 Applying the Circle Method

To prove Theorem 1.0.2, we will apply the Hardy-Littlewood circle method. First, we will set the size of the smaller prime factor:

Definition 2.3.1. Let $\varepsilon > 0$ be small and fixed. Define $P > 0$ according to the size of H as follows:

$$P := \begin{cases} (\log X)^{17+\varepsilon}, & \text{if } (\log X)^{19+\varepsilon} \leq H \leq \exp((\log X)^{\varepsilon^3}), \\ \exp\left((\log\log X)^2\right), & \text{if } \exp((\log X)^{\varepsilon^3}) < H \leq X \log^{-A} X. \end{cases}$$

It will be more convenient throughout the argument to have a log weight attached to the indicator function of E_2' as follows:

Definition 2.3.2. Let P be defined as in Definition 2.3.1. We define the arithmetic function $\varpi_2 : \mathbb{N} \to \mathbb{R}$ to be

$$\varpi_2(n) = \begin{cases} \log p_2, & \text{if } n = p_1 p_2 \text{ with } P < p_1 \leq P^{1+\delta}, \\ 0, & \text{otherwise.} \end{cases}$$

From now on we fix $\delta > 0$ sufficiently small. We will prove the following asymptotic formula, from which Theorem 1.0.2 follows immediately after applying dyadic decomposition:

Theorem 2.3.3. *Let $\varepsilon > 0$, $A > 3$ be fixed and let $(\log X)^{19+\varepsilon} \leq H \leq X \log^{-A} X$. Then, there exists some $\eta = \eta(\varepsilon) > 0$ such that for all but at most $O(H \log^{-\eta} X)$ values of $0 < |h| \leq H$ we have that*

$$\sum_{X<n\leq 2X} \varpi_2(n)\varpi_2(n+h) = \mathfrak{S}(h)X \left(\sum_{P<p\leq P^{1+\delta}} \frac{1}{p} \right)^2 + O\left(\frac{X}{\log^{\eta} X} \right),$$

where $\mathfrak{S}(h)$ is the singular series defined in (1.0.3).

Remark 2.3.4. As H becomes an arbitrarily large power of $\log X$, or is larger than any power of $\log X$, we are able to improve the bound on the error terms to $O(X \log^{-A} X)$ for $A > 0$ once we have suitably modified the dependencies between H, P and the parameters of the circle method. We also note that, after appropriately modifying the main term, using this result we can in fact prove Theorem 1.0.2 with a better error term.

We consider the integral

$$\int_0^1 |S(\alpha)|^2 e(-h\alpha) d\alpha = \sum_{X<m,n\leq 2X} \varpi_2(m)\varpi_2(n) \int_0^1 e(\alpha(m-n-h)) d\alpha, \tag{2.3.1}$$

where for $\alpha \in (0,1)$ we define the exponential sum

$$S(\alpha) := \sum_{X<n\leq 2X} \varpi_2(n) e(n\alpha).$$

Then, by the integral identity

$$\int_0^1 e(nx)dx = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{2.3.2}$$

we have that the integral in (2.3.1) vanishes unless $m = n + h$. Thus (2.3.1) becomes

$$\begin{aligned} \int_0^1 |S(\alpha)|^2 e(-h\alpha)d\alpha &= \sum_{X<n\leq 2X-h} \varpi_2(n)\varpi_2(n+h) \\ &= \sum_{X<n\leq 2X} \varpi_2(n)\varpi_2(n+h) + O(h\log^2 X). \end{aligned} \tag{2.3.3}$$

This error term will be negligible by our choice of H. Thus, except for an acceptable error, we can represent the correlation by an integral over the unit circle.

We split the domain of integration into the major and minor arcs. We define the major arcs $\mathfrak{M}$ to be the set of real $\alpha \in (0, 1)$ such that

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{qQ} \text{ for some } 1 \leq q \leq Q_0, a < q, (a, q) = 1 \tag{2.3.4}$$

with $Q_0 := \log^{A'} X$ and $Q := P \log X$. Here we define $A' > 0$ according to the size of H as follows

$$A' := \begin{cases} 1+\varepsilon^2, & \text{if } (\log X)^{19+\varepsilon} \leq H \leq \exp((\log X)^{\varepsilon^3}), \\ 3+\varepsilon^2, & \text{if } \exp((\log X)^{\varepsilon^3}) < H \leq X \log^{-A} X. \end{cases} \tag{2.3.5}$$

We define the minor arcs $\mathfrak{m}$ to be the rest of the circle, that is, the set of real $\alpha \in (0,1)$ such that

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{qQ} \text{ for some } Q_0 < q \leq Q, a < q, (a,q) = 1. \tag{2.3.6}$$

Remark 2.3.5. The parameters satisfy $Q_0 < P < Q < H$. Decreasing the size we can take for P would directly reduce how small we are able to take H.

In Section 2.3.2, we will prove the following estimate for the integral over the minor arcs:

Proposition 2.3.6 (Minor Arc Estimate)**.** *Let $A > 3$ be fixed and let $\varepsilon > 0$ be fixed sufficiently small. Let $Q(\log X)^{1+\varepsilon} \leq H \leq X \log^{-A} X$. With $\mathfrak{m}$ defined as in* (2.3.6)*, for $\alpha \in \mathfrak{m}$ there exists some $\eta = \eta(\varepsilon) > 0$ such that*

$$\int_{\mathfrak{m} \cap [\alpha - \frac{1}{2H}, \alpha + \frac{1}{2H}]} |S(\theta)|^2 d\theta \ll \frac{X}{(\log X)^{1+\eta}}. \tag{2.3.7}$$

Sections 2.3.3 to 2.3.6 will be dedicated to proving the following expression for the integral over the major arcs:

Proposition 2.3.7 (Major Arc Integral)**.** *Let $A > 3$ be fixed and let $\varepsilon > 0$ be fixed sufficiently small. Let $(\log X)^{19+\varepsilon} \leq H \leq X \log^{-A} X$. With $\mathfrak{M}$ defined as in (2.3.4) and $\delta > 0$ sufficiently small, there exists some $\eta = \eta(\varepsilon) > 0$ such that for all but at most $O(HQ_0^{-1/3})$ values of $0 < |h| \leq H$ we have that*

$$\int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha) d\alpha = \mathfrak{S}(h) X \left(\sum_{P < p \leq P^{1+\delta}} \frac{1}{p} \right)^2 + O\left(\frac{X}{\log^\eta X} \right),$$

where $\mathfrak{S}(h)$ is the singular series given in (1.0.3).

Assuming Proposition 2.3.6 and Proposition 2.3.7, we can now prove Theorem 2.3.3.

Proof of Theorem 2.3.3. We follow the arguments in [27, Pages 32-34]. By (2.3.3), we have that

$$\sum_{0<|h|\leq H} \left| \sum_{X<n\leq 2X-h} \varpi_2(n)\varpi_2(n+h) - \int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha) d\alpha \right|^2$$
$$\ll \sum_{0<|h|\leq H} \left| \int_{\mathfrak{m}} |S(\alpha)|^2 e(-h\alpha) d\alpha \right|^2.$$

We now apply a smoothing; we multiply the above by an even non-negative Schwartz function $\Phi : \mathbb{R} \to \mathbb{R}^+$ such that $\Phi(x) \geq 1$ for $x \in [-1, 1]$ and its Fourier transform $\hat{\Phi}$ is supported in $[-1/2, 1/2]$. Thus, the above is bounded by

$$\sum_h \Phi\left(\frac{h}{H}\right) \int_{\mathfrak{m}} \int_{\mathfrak{m}} |S(\alpha_1)|^2 |S(\alpha_2)|^2 e(-h(\alpha_1 - \alpha_2)) d\alpha_1 d\alpha_2.$$

By substitution we have that

$$\int_{\mathbb{R}} \Phi\left(\frac{x}{H}\right) e(x(\alpha_2 - \alpha_1 - \xi))dx = H \int_{\mathbb{R}} \Phi(t) e(Ht(\alpha_2 - \alpha_1 - \xi))dt$$
$$= H\hat{\Phi}(H(\alpha_2 - \alpha_1 - \xi)).$$

Therefore, applying the Poisson summation formula (Lemma 2.2.1), we have that

$$\sum_h \Phi\left(\frac{h}{H}\right) e(-h(\alpha_1 - \alpha_2)) = H \sum_n \hat{\Phi}(H(\alpha_2 - \alpha_1 + n)).$$

Note that due to the support of $\hat{\Phi}$, this expression vanishes unless $n = 0, -1, 1$. Using a change of variables and periodicity, we can reduce this to needing to treat $\alpha_1 \in [\alpha_2 - \frac{1}{2H}, \alpha_2 + \frac{1}{2H}]$. Therefore, we have that

$$\begin{aligned} &\sum_{0<|h|\leq H} \left| \int_{\mathfrak{m}} |S(\alpha)|^2 e(-h\alpha) d\alpha \right|^2 \\ &\ll H \int_{\mathfrak{m}} |S(\alpha_2)|^2 \int_{\mathfrak{m} \cap [\alpha_2 - \frac{1}{2H}, \alpha_2 + \frac{1}{2H}]} |S(\alpha_1)|^2 d\alpha_1 d\alpha_2 \\ &\ll H \int_0^1 |S(\alpha_2)|^2 \int_{\mathfrak{m} \cap [\alpha_2 - \frac{1}{2H}, \alpha_2 + \frac{1}{2H}]} |S(\alpha_1)|^2 d\alpha_1 d\alpha_2. \end{aligned}$$

By Proposition 2.3.6, there exists some $\eta = \eta(\varepsilon) > 0$ such that

$$\sup_{\alpha \in \mathfrak{m}} \int_{\mathfrak{m} \cap [\alpha - \frac{1}{2H}, \alpha + \frac{1}{2H}]} |S(\beta)|^2 d\beta \ll \frac{X}{(\log X)^{1+\eta}}.$$

We have that

$$\int_0^1 |S(\alpha)|^2 d\alpha = \sum_{X<n\leq 2X} \varpi_2^2(n) = \sum_{P<p_1\leq P^{1+\delta}} \sum_{\frac{X}{p_1}<p_2\leq \frac{2X}{p_1}} \log^2 p_2.$$

By partial summation and the prime number theorem, we have that

$$\begin{aligned}\sum_{\frac{X}{p_1}<p_2\leq \frac{2X}{p_1}} \log^2 p_2 &= \left(\log \frac{X}{p_1}\right) \sum_{\frac{X}{p_1}<p_2\leq \frac{2X}{p_1}} \log p_2 + \int_{X/p_1}^{2X/p_1} \frac{\sum_{p_2\leq t} \log p_2}{t} dt \\ &= \left(\log \frac{X}{p_1}\right) \sum_{\frac{X}{p_1}<p_2\leq \frac{2X}{p_1}} \log p_2 + O\left(\frac{X}{p_1}\right).\end{aligned}$$

By Mertens' theorem (Lemma 2.2.2), we have the bound

$$\begin{aligned}\int_0^1 |S(\alpha)|^2 d\alpha &\ll \sum_{P<p_1\leq P^{1+\delta}} \log \frac{X}{p_1} \sum_{\frac{X}{p_1}<p_2\leq \frac{2X}{p_1}} \log p_2 \\ &\ll X \log X \sum_{P<p\leq P^{1+\delta}} \frac{1}{p} \\ &\ll X \log X.\end{aligned}$$

Therefore, we have that

$$\sum_{0<|h|\leq H} \left| \sum_{X<n\leq 2X-h} \varpi_2(n)\varpi_2(n+h) - \int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha) d\alpha \right|^2 \ll \frac{HX^2}{\log^\eta X}.$$

Thus, by Chebyshev's inequality (see, for example, [37, Page 185]), we have that the size of the set of integers $0 < |h| \leq H$ such that

$$\left| \sum_{X < n \leq 2X-h} \varpi_2(n)\varpi_2(n+h) - \int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha)d\alpha \right| \gg \frac{X}{(\log X)^{\eta/3}}$$

is bounded by

$$\begin{aligned}
&\ll \frac{(\log X)^{2\eta/3}}{X^2} \cdot \sum_{0<|h|\leq H} \left| \sum_{X<n\leq 2X-h} \varpi_2(n)\varpi_2(n+h) - \int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha)d\alpha \right|^2 \\
&\ll \frac{(\log X)^{2\eta/3}}{X^2} \cdot \frac{HX^2}{\log^\eta X} \\
&\ll \frac{H}{(\log X)^{\eta/3}}.
\end{aligned}$$

In particular, we have that

$$\sum_{X<n\leq 2X-h} \varpi_2(n)\varpi_2(n+h) - \int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha)d\alpha = O\left(\frac{X}{(\log X)^{\eta/3}}\right)$$

for all but $O(H(\log X)^{-\eta/3})$ integers $0 < |h| \leq H$. Finally, applying Proposition 2.3.7, we have that

$$\sum_{X<n\leq 2X} \varpi_2(n)\varpi_2(n+h) = \mathfrak{S}(h)X\left(\sum_{P<p\leq P^{1+\delta}} \frac{1}{p}\right)^2 + O\left(\frac{X}{(\log X)^{\eta/3}}\right),$$

for all but $O(H(\log X)^{-\eta/3})$ integers $0 < |h| \leq H$, as claimed. □

2.3.2 The Minor Arcs

We first treat the integral over the minor arcs, proving Proposition 2.3.6 by following the proof of [32, Lemma 8].

Proof of Proposition 2.3.6. Starting with the minor arc integral (2.3.7), we make the substitution $\theta = \alpha + \beta$ to see that

$$I := \int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]} |S(\theta)|^2 d\theta = \int_{\substack{\alpha+\beta\in\mathfrak{m} \\ |\beta|\le\frac{1}{2H}}} |S(\alpha+\beta)|^2 d\beta.$$

We apply Lemma 2.2.4 to the integral to get

$$I \ll \frac{1}{H^2} \int_X^{2X} \left| \sum_{x<n\le x+H} \varpi_2(n) e(n\alpha) \right|^2 dx + H \log^2 X.$$

The second term is $\ll X/\log^{1+\eta} X$ by our choice of H, so it remains to bound the first term.

Case 1. $H \le \exp((\log X)^{\varepsilon^3})$.

We apply the Cauchy-Schwarz inequality to the integrand to get

$$\begin{aligned}
&\left|\sum_{\substack{x<p_1p_2\le x+H\\ P<p_1\le P^{1+\delta}}}(\log p_2)e(\alpha p_1p_2)\right|^2\\
&=\left|\sum_{P<m\le P^{1+\delta}}\mathbb{1}_{\mathbb{P}}(m)\left(\sum_{x<mp\le x+H}(\log p)e(\alpha mp)\right)\right|^2\\
&\le\left(\sum_{P<m_1\le P^{1+\delta}}|\mathbb{1}_{\mathbb{P}}(m_1)|^2\right)\left(\sum_{P<m_2\le P^{1+\delta}}\left|\sum_{x<m_2p\le x+H}(\log p)e(\alpha m_2p)\right|^2\right).
\end{aligned} \tag{2.3.8}$$

By the prime number theorem the first term is $\ll \frac{P^{1+\delta}}{\log P}$, while the second term is equal to

$$\sum_{\substack{x<mp_1,mp_2\le x+H\\ P<m\le P^{1+\delta}}}(\log p_1)(\log p_2)e(\alpha m(p_1-p_2)).$$

Next, we perform the integration on this sum. Note that $X<mp_i\le x+H\le 3X$, so we include this condition on the summation. We now trivially extend the domain of integration to $x\in[0,3X]$ as the integrand is positive and define the set $\Omega:=\{x:0\le x\le 3X, mp_i-H\le x<mp_i, i=1,2\}$. Exchanging the order of integration and summation, we have that

$$I\ll\frac{P^{1+\delta}}{H^2\log P}\sum_{P<m\le P^{1+\delta}}\left|\sum_{X<mp_1,mp_2\le 3X}(\log p_1)(\log p_2)e(\alpha m(p_1-p_2))\cdot|\Omega|\right|.$$

If $m|p_1-p_2|>H$, then $|\Omega|=0$. Since we have that $mp_i-H>X-H>0$ and

$mp_i \leq 3X$ for $i = 1, 2$, the condition $0 \leq x \leq 3X$ is weaker than the condition $\max(mp_1, mp_2) - H \leq x < \min(mp_1, mp_2)$. Therefore, if $m|p_1 - p_2| \leq H$ we have that $|\Omega| = H - m|p_1 - p_2|$.

We now split the sum into the diagonal terms, $p_1 = p_2$, and the off-diagonal terms, $p_1 \neq p_2$, denoted by S_1 and S_2 respectively. The diagonal terms contribute

$$S_1 \ll \frac{P^{1+\delta}}{H \log P} \sum_{P < m \leq P^{1+\delta}} \sum_{\frac{X}{m} < p \leq \frac{3X}{m}} \log^2 p \ll \frac{X P^{1+\delta} \log(X/P)}{H}. \tag{2.3.9}$$

Now we bound the off-diagonal terms S_2. Let $r = |p_1 - p_2|$. Noting that $0 < mr \leq H$, we have that S_2 is

$$\ll \frac{P^{1+\delta}}{H^2 \log P} \sum_{0 < r \leq H} \sum_{\substack{\frac{X}{P^{1+\delta}} < p_1, p_2 \leq \frac{3X}{P} \\ r = |p_1 - p_2|}} (\log p_1)(\log p_2) \left| \sum_{\substack{P < m \leq P^{1+\delta} \\ 0 < m \leq H/r}} e(\alpha m r)(H - mr) \right|.$$

Noting that $0 < m \leq H/r$ and $P < m \leq P^{1+\delta}$, we have that $0 < r \leq H/P$. We apply partial summation and Lemma 2.2.8 to the sum over m to see that

$$S_2 \ll \frac{P^{1+\delta}}{H \log P} \sum_{0 < r \leq \frac{H}{P}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right) \sum_{\substack{\frac{X}{P^{1+\delta}} < p_1, p_2 \leq \frac{3X}{P} \\ r = |p_1 - p_2|}} (\log p_1)(\log p_2).$$

By partial summation followed by Lemma 2.2.3, we have that the sum over

p_1, p_2 is bounded by

$$\sum_{\substack{\frac{X}{P^{1+\delta}} < p_1, p_2 \le \frac{3X}{P} \\ r = |p_1 - p_2|}} (\log p_1)(\log p_2) \ll (\log X)^2 \sum_{\substack{\frac{X}{P^{1+\delta}} < p_1, p_2 \le \frac{3X}{P} \\ r = |p_1 - p_2|}} 1 \ll \frac{\mathfrak{S}(r) X}{P}.$$

Therefore the contribution of the off-diagonal terms can be bounded by

$$S_2 \ll \frac{X P^{1+\delta}}{H P \log P} \sum_{0 < r \le \frac{H}{P}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right) \mathfrak{S}(r).$$

We have that $\mathfrak{S}(r) \ll \log\log r$ by Lemma 2.2.3, so applying partial summation we have that

$$S_2 \ll \frac{X P^{\delta}}{H \log P} \log\log X \sum_{0 < r \le \frac{H}{P}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right).$$

Next, we apply Lemma 2.2.9 to the sum over r to get

$$S_2 \ll \frac{X P^{\delta}}{H} \left(\frac{H}{Q_0} + \frac{H}{P} + Q\right) \log \frac{QH}{P}, \tag{2.3.10}$$

recalling that since $\alpha \in \mathfrak{m}$ we have that $Q_0 \le q \le Q$.

Since we are in the case $H \le \exp((\log X)^{\varepsilon^3})$, we have that $\log \frac{QH}{P} \ll (\log X)^{\varepsilon^3}$. Therefore, combining the contributions of the diagonal terms (2.3.9) and the off-diagonal terms (2.3.10), we find

$$I \ll X P^{\delta} \left((\log X)^{\varepsilon^3} \left(\frac{1}{Q_0} + \frac{1}{P} + \frac{Q}{H}\right) + \frac{P \log(X/P)}{H}\right).$$

By our choices of $Q_0 = (\log X)^{1+\varepsilon^2}$, $Q(\log X)^{1+\varepsilon} = P(\log X)^{2+\varepsilon} \ll H$, we have that

$$I \ll \frac{X}{(\log X)^{1+\eta}}$$

for some $\eta = \eta(\varepsilon) > 0$.

Case 2: $H > \exp((\log X)^{\varepsilon^3})$.

We split the sum over $P \leq p_1 \leq P^{1+\delta}$ in (2.3.8) into $O(\log P)$ dyadic intervals $[P_1, 2P_1]$ before applying the triangle inequality and Cauchy-Schwarz to obtain

$$\begin{aligned}
&\left| \sum_{\substack{x < p_1 p_2 \leq x+H \\ P_1 < p_1 \leq 2P_1}} (\log p_2) e(\alpha p_1 p_2) \right|^2 \\
&\leq \left(\sum_{P_1 < m_1 \leq 2P_1} |\mathbb{1}_{\mathbb{P}}(m_1)|^2 \right) \left(\sum_{P_1 < m_2 \leq 2P_1} \left| \sum_{x < m_2 p \leq x+H} (\log p) e(\alpha m_2 p) \right|^2 \right) \\
&\leq \frac{P_1}{\log P_1} \sum_{\substack{x < mp_1, mp_2 \leq x+H \\ P_1 < m \leq 2P_1}} (\log p_1)(\log p_2) e(\alpha m (p_1 - p_2)).
\end{aligned}$$

As in the previous case, we perform the integration and then split into considering the diagonal ($p_1 = p_2$) and off-diagonal ($p_1 \neq p_2$) terms, which we denote by S_1' and S_2' respectively. The diagonal terms contribute

$$S_1' \ll \frac{P_1}{H \log P_1} \sum_{P_1 < m \leq 2P_1} \sum_{\frac{X}{m} < p \leq \frac{3X}{m}} \log^2 p \ll \frac{X P_1 \log(X/P_1)}{H \log P_1}.$$

The off-diagonal terms S_2' contribute

$$\ll \frac{P_1}{H^2 \log P_1} \sum_{0<r\le H} \sum_{\substack{\frac{X}{2P_1}<p_1,p_2\le \frac{3X}{P_1} \\ r=|p_1-p_2|}} (\log p_1)(\log p_2) \left| \sum_{\substack{P_1<m\le 2P_1 \\ 0<m\le H/r}} e(\alpha m r)(H-mr) \right|.$$

We once again apply partial summation and Lemma 2.2.8 to the sum over m, followed by applying partial summation and Lemma 2.2.3 to the sum over p_1, p_2, so that

$$\begin{aligned} S_2' &\ll \frac{P_1}{H \log P_1} \sum_{0<r\le \frac{H}{P_1}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right) \sum_{\substack{\frac{X}{2P_1}<p_1,p_2\le \frac{3X}{P_1} \\ r=|p_1-p_2|}} (\log p_1)(\log p_2) \\ &\ll \frac{X P_1}{H P_1 \log P_1} \sum_{0<r\le \frac{H}{P_1}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right) \mathfrak{S}(r) \\ &\ll \frac{X \log\log X}{H \log P_1} \sum_{0<r\le \frac{H}{P_1}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right). \end{aligned}$$

Again, we apply Lemma 2.2.9 to the sum over r to see this off-diagonal contribution is bounded by

$$S_2' \ll \frac{X}{H \log\log X} \left(\frac{H}{Q_0} + \frac{H}{P_1} + Q\right) \log \frac{QH}{P_1}.$$

We have that $\log \frac{QH}{P_1} \ll \log X$ and $Q_0 = (\log X)^{3+\varepsilon^2}$. Combining all of the dyadic sums contributes $O(\log^2 P) = O((\log\log X)^4)$, so that the total contribution is

$$
\begin{aligned}
I &\ll X(\log\log X)^3 \left(\log X \left(\frac{1}{Q_0} + \frac{1}{P} + \frac{Q}{H}\right)\right) + \frac{XP^{1+\delta}\log(X/P)\log P}{H} \\
&\ll \frac{X}{(\log X)^{2+\eta}},
\end{aligned}
$$

for some $\eta = \eta(\varepsilon) > 0$, which is acceptable. □

2.3.3 The Major Arcs

We now shift our attention to evaluating the contribution of the integral over the major arcs. We will first expand the exponential sum $S(\alpha)$ in terms of Dirichlet characters and suitably approximate the contribution of the principal character, which will provide the main term. We will then evaluate this main term and the sequel will be dedicated to bounding the error terms that arise from this expansion.

2.3.3.1 Expanding the Exponential Sum

First, we rewrite the integral over the major arcs by expanding the exponential sum $S(\alpha)$ in terms of Dirichlet characters. We first define the following.

Definition 2.3.8. Let $\alpha = a/q + \beta$ satisfy (2.3.4), P be defined as in Definition 2.3.1 and $\mathfrak{M}$ as in (2.3.4). We define

$$\begin{aligned} a(\alpha) &:= \frac{\mu(q)}{\varphi(q)} \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \sum_{X<n\le 2X} e(\beta n), \\ b(\alpha) &:= \frac{1}{\varphi(q)} \sum_{\chi(q)} \tau(\overline{\chi})\chi(a) \sum_{X<n\le 2X} \left(\varpi_2(n)\chi(n) - \delta_\chi \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right) e(\beta n), \\ A^2(X) &:= \int_{\mathfrak{M}} |a(\alpha)|^2 d\alpha, \\ B^2(X) &:= \int_{\mathfrak{M}} |b(\alpha)|^2 d\alpha, \end{aligned}$$

where $\tau(\chi)$ denotes the Gauss sum as defined in (2.2.1), and $\delta_\chi = 1$ when $\chi = \chi_0$ and is zero otherwise.

We will now find the following expression for the integral over the major arcs, once we have expanded the exponential sum:

Lemma 2.3.9. *Let $\mathfrak{M}$ be defined as in* (2.3.4) *and* $a(\alpha), A^2(X), B^2(X)$ *be as in Definition 2.3.8. We have that*

$$\int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha) d\alpha = \int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha) d\alpha + O\left(A(X)B(X) + B^2(X)\right).$$

Proof. Let $\alpha \in \mathfrak{M}$, so that $\alpha = \frac{a}{q} + \beta$ with $q \le Q_0$, $(a,q) = 1$ and $|\beta| \le \frac{1}{qQ}$. Then

$$S(\alpha) = \sum_{X<n\le 2X} \varpi_2(n) e\left(\frac{an}{q}\right) e(\beta n).$$

By Definition 2.3.2, we have that $n = p_1 p_2$ with $P < p_1 \le P^{1+\delta}$. As we have $P > Q_0$, we must have that $(p_1, q) = (p_2, q) = 1$ and therefore that $(n, q) = 1$.

We can now rewrite our expression for $S(\alpha)$ by applying the identity

$$e\left(\frac{a}{q}\right) = \frac{1}{\varphi(q)} \sum_{\chi(q)} \chi(a)\tau(\overline{\chi})$$

which holds for $(a, q) = 1$. This gives

$$\begin{aligned} S(\alpha) &= \frac{1}{\varphi(q)} \sum_{\chi(q)} \tau(\overline{\chi})\chi(a) \sum_{X<n\leq 2X} \varpi_2(n)\chi(n)e(\beta n) \\ &= \frac{1}{\varphi(q)} \sum_{\chi(q)} \tau(\overline{\chi})\chi(a) \sum_{\substack{X<p_1p_2\leq 2X \\ P<p_1\leq P^{1+\delta}}} \chi(p_1)\chi(p_2)(\log p_2)e(\beta p_1 p_2), \end{aligned} \tag{2.3.11}$$

where we have applied the definition of ϖ_2 in the last line. Now we approximate the contribution of the principal character, which will become the main term. First, note that since we have $q \leq Q_0 < P < p_1$ we must have that $(p_1p_2, q) = 1$ for $X < p_1p_2 \leq 2X$, so we must have $(\log p_2)\chi_0(p_1)\chi_0(p_2) = \log p_2$ in these ranges. By the prime number theorem, we have that

$$\sum_{X<n\leq 2X} \varpi_2(n) = \sum_{P<p_1\leq P^{1+\delta}} \sum_{\frac{X}{p_1}<p_2\leq \frac{2X}{p_1}} \log p_2 \sim X \sum_{P<p\leq P^{1+\delta}} \frac{1}{p}.$$

Therefore we choose to approximate $\sum_{X<n\leq 2X} \varpi_2(n)$ by

$$\sum_{P<p\leq P^{1+\delta}} \frac{1}{p} \sum_{X<n\leq 2X} 1 = X \sum_{P<p\leq P^{1+\delta}} \frac{1}{p} + O(1).$$

Using this and the fact that $\tau(\chi_0) = \mu(q)$, we approximate the contribution of the principal character to the exponential sum $S(\alpha)$ by

$$\frac{\mu(q)}{\varphi(q)}\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\sum_{X<n\le 2X}e(\beta n).$$

Adding and subtracting this approximation in our expression (2.3.11) for $S(\alpha)$, we have that

$$\begin{aligned}
S(\alpha) =&\frac{\mu(q)}{\varphi(q)}\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\sum_{X<n\le 2X}e(\beta n)\\
&+\frac{\mu(q)}{\varphi(q)}\sum_{X<n\le 2X}\left(\varpi_2(n)\chi_0(n)-\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)e(\beta n)\\
&+\frac{1}{\varphi(q)}\sum_{\substack{\chi(q)\\ \chi\ne\chi_0}}\tau(\overline{\chi})\chi(a)\sum_{X<n\le 2X}\varpi_2(n)\chi(n)e(\beta n)\\
=&\frac{\mu(q)}{\varphi(q)}\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\sum_{X<n\le 2X}e(\beta n)\\
&+\frac{1}{\varphi(q)}\sum_{\chi(q)}\tau(\overline{\chi})\chi(a)\sum_{X<n\le 2X}\left(\varpi_2(n)\chi(n)-\delta_\chi\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)e(\beta n)\\
=&a(\alpha)+b(\alpha).
\end{aligned}$$

Finally, expanding the square and applying the Cauchy-Schwarz inequality, we have that

$$\begin{aligned}\int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha)d\alpha &= \int_{\mathfrak{M}} |a(\alpha)+b(\alpha)|^2 e(-h\alpha)d\alpha \\ &= \int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha)d\alpha + \int_{\mathfrak{M}} |b(\alpha)|^2 e(-h\alpha)d\alpha \\ &\quad + \int_{\mathfrak{M}} (a(\alpha)\overline{b}(\alpha)+\overline{a}(\alpha)b(\alpha))e(-h\alpha)d\alpha \\ &= \int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha)d\alpha + O\left(A(X)B(X)+B^2(X)\right),\end{aligned}$$

as required. □

Thus, in order to prove Proposition 2.3.7 we need to evaluate

$$\int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha)d\alpha$$

(which will also provide a bound for $A^2(X)$), and suitably bound $B^2(X)$.

2.3.3.2 Evaluating the Main Term

In this section we evaluate the integral $\int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha)d\alpha$, giving the main term of the asymptotic (and a bound for $A^2(X)$):

Proposition 2.3.10. *Let $\varepsilon > 0$ be fixed sufficiently small, let $A > 3$ be fixed. Let Q_0 be defined as in (2.3.5) and let $(\log X)^{19+\varepsilon} \le H \le X \log^{-A} X$. Then for all but at most $O(HQ_0^{-1/3})$ values of $0 < |h| \le H$ we have that*

$$\int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha)d\alpha = \mathfrak{S}(h)X\left(\sum_{P<p\le P^{1+\delta}} \frac{1}{p}\right)^2 + O\left(\frac{X}{\log^{\eta} X}\right),$$

for some $\eta = \eta(\varepsilon) > 0$*, where we define the singular series* $\mathfrak{S}(h)$ *as in* (1.0.3)*.*

Remark 2.3.11. More generally, if the smaller prime factor lies in the range $p \in [P_1, P_2]$ for some $P_1 = P_1(X)$ and $P_2 = P_2(X)$, the error term here is in fact

$$\ll XQ_0^{-1/3}(\log H)\left(\sum_{P_1<p\le P_2}\frac{1}{p}\right)^2.$$

In the proofs of Theorems 1.0.4 and 1.0.5, we will make larger choices of Q_0, P_1 and P_2, which will lead to the improved error term. In particular, we will choose $Q_0 = \log^{A'} X$ for some suitable $A' > 0$, $P_1 = \exp((\log X)^{o(1)})$ and $P_2 = \exp((\log X)^{1-o(1)})$.

Before we can prove Proposition 2.3.10, we need an expression involving the singular series $\mathfrak{S}(h)$.

Lemma 2.3.12 (The Singular Series)**.** *Let* h *be a non-zero even integer and* Q_0 *be defined as in* (2.3.5)*. Let* $A > 3$ *be fixed and let* $(\log X)^{19+\varepsilon} \le H \le X\log^{-A} X$*. Then, for all but at most* $O(HQ_0^{-1/3})$ *values of* $0 < |h| \le H$ *we have that*

$$\sum_{q\le Q_0}\frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} = \mathfrak{S}(h) + O(Q_0^{-1/3}\log H).$$

Proof. For similar results, see [27, Page 39] and [42, Page 35]. Let h be a non-zero even integer. Rewriting the sum over q, we have that

$$\sum_{q\le Q_0}\frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} = \left(\sum_{q=1}^{\infty} - \sum_{q>Q_0}\right)\frac{\mu^2(q)c_q(-h)}{\varphi^2(q)},$$

noting that this is valid as the series is absolutely convergent. We first evaluate the singular series. Since each of the functions in the summand is multiplicative, we can calculate the Euler product expansion

$$\sum_{q=1}^{\infty} \frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} = \prod_p \sum_{k=0}^{\infty} \frac{\mu^2(p^k)c_{p^k}(-h)}{\varphi^2(p^k)} = \prod_p \left(1 + \frac{c_p(-h)}{\varphi^2(p)}\right).$$

We know that for Ramanujan's sum we have

$$c_p(-h) = \begin{cases} \varphi(p), & \text{if } p \mid h, \\ -1, & \text{if } p \nmid h, \end{cases}$$

so we have

$$\begin{aligned}
\sum_{q=1}^{\infty} \frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} &= \prod_{p \nmid h} \left(1 - \frac{1}{\varphi^2(p)}\right) \prod_{p \mid h} \left(1 + \frac{1}{\varphi(p)}\right) \\
&= \prod_{p \nmid h} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{p \mid h} \left(1 + \frac{1}{p-1}\right) \\
&= 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{\substack{p \mid h \\ p>2}} \left(\frac{p}{p-1} \cdot \frac{(p-1)^2}{p(p-2)}\right) \\
&= 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{\substack{p \mid h \\ p>2}} \left(\frac{p-1}{p-2}\right) \\
&= \mathfrak{S}(h).
\end{aligned}$$

It remains to bound the tail of the sum. Following [42, Page 35], we have

that

$$c_q(-h) = \frac{\mu(q/(q,h))\varphi(q)}{\varphi(q/(q,h))},$$

and therefore, letting $(q,h) = d$, that

$$\sum_{q>Q_0} \frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} = \sum_{\substack{q'>Q_0/d \\ d|h \\ (q',h)=1}} \frac{\mu^2(q'd)}{\varphi(q'd)} \frac{\mu(q')}{\varphi(q')}.$$

Note that $\mu^2(q'd)$ is zero unless $(q',d) = 1$, so we have

$$\sum_{q>Q_0} \frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} = \sum_{d|h} \frac{\mu^2(d)}{\varphi(d)} \sum_{\substack{q'>Q_0/d \\ (q',h)=1}} \frac{\mu^3(q')}{\varphi^2(q')} = \sum_{d|h} \frac{\mu^2(d)}{\varphi(d)} \sum_{\substack{q'>Q_0/d \\ (q',h)=1}} \frac{\mu(q')}{\varphi^2(q')}.$$

To bound $\sum_{q>Q_0} 1/\varphi^2(q)$, we use the result

$$\sum_{n\le x} \left(\frac{n}{\varphi(n)}\right)^2 \ll x$$

(see [35, Corollary 2.15 and Eq. (2.32)]) and apply partial summation. This gives $\sum_{q>Q_0} 1/\varphi^2(q) \ll Q_0^{-1}$, so we have that the above is bounded by

$$\ll \sum_{d|h} \frac{\mu^2(d)}{\varphi(d)} \min\left(\frac{d}{Q_0}, 1\right).$$

Therefore the tail is bounded by $\ll \log h$. We will need more cancellation in this bound, therefore we consider the average

$$\begin{aligned}\sum_{h\leq H}\left|\sum_{q>Q_0}\frac{\mu^2(q)c_q(-h)}{\varphi^2(q)}\right|^2 &\ll (\log H)\sum_{h\leq H}\sum_{d|h}\frac{\mu^2(d)}{\varphi(d)}\min\left(\frac{d}{Q_0},1\right)\\ &\ll (\log H)\sum_{d\leq H}\frac{H\mu^2(d)}{d\varphi(d)}\min\left(\frac{d}{Q_0},1\right)\\ &\ll \frac{H\log H}{Q_0}\sum_{d\leq H}\frac{\mu^2(d)}{\varphi(d)}\\ &\ll \frac{H\log^2 H}{Q_0}.\end{aligned}$$

By Chebyshev's inequality, we have for all but at most $O(HQ_0^{-1/3})$ values of h the bound

$$\sum_{q>Q_0}\frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} \ll Q_0^{-1/3}\log H.$$

Overall, we have that

$$\sum_{q\leq Q_0}\frac{\mu^2(q)c_q(-h)}{\varphi^2(q)} = \mathfrak{S}(h) + O(Q_0^{-1/3}\log H),$$

for all but at most $O(HQ_0^{-1/3})$ values of h, as claimed. □

We are now able to complete the proof of Proposition 2.3.10.

Proof of Proposition 2.3.10. Applying the definition of the major arcs (2.3.4) and expanding the square, we have that

$$\begin{aligned}
&\int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha) d\alpha \\
&= \sum_{q \le Q_0} \sum_{\substack{1 \le a \le q \\ (a,q)=1}} \int_{|\beta| \le \frac{1}{qQ}} \left| \frac{\mu(q)}{\varphi(q)} \sum_{P < p \le P^{1+\delta}} \frac{1}{p} \sum_{X < n \le 2X} e(\beta n) \right|^2 e\left(\frac{-ha}{q} - h\beta \right) d\beta \\
&= \left(\sum_{P < p \le P^{1+\delta}} \frac{1}{p} \right)^2 \sum_{q \le Q_0} \frac{\mu^2(q) c_q(-h)}{\varphi^2(q)} \int_{|\beta| \le \frac{1}{qQ}} \sum_{X < m,n \le 2X} e(\beta(m-n-h)) d\beta \\
&= \left(\sum_{P < p \le P^{1+\delta}} \frac{1}{p} \right)^2 \sum_{q \le Q_0} \frac{\mu^2(q) c_q(-h)}{\varphi^2(q)} I_1, \qquad (2.3.12)
\end{aligned}$$

say. We rewrite the integral I_1 as

$$\begin{aligned}
I_1 &= \left\{ \int_0^1 - \int_{\frac{1}{qQ}}^{1-\frac{1}{qQ}} \right\} \sum_{X < m,n \le 2X} e(\beta(m-n-h)) d\beta \\
&=: I_2 - I_3,
\end{aligned}$$

say. To the first term I_2, we apply the identity (2.3.2) to get

$$I_2 = \sum_{\substack{X < m,n \le 2X \\ m = n+h}} 1 = X + O(H). \qquad (2.3.13)$$

Returning to (2.3.12), the error term contributes $\ll X \log^{-\eta} X$, since $H \le X \log^{-A} X$, which is acceptable. Now we bound the integral I_3. Note that β is never an integer in the domain of integration, so applying Lemma 2.2.8 to the sums over m and n we have that

$$
\begin{aligned}
I_3 &= \int_{\frac{1}{qQ}}^{1-\frac{1}{qQ}} \sum_{X<m,n\le 2X} e(\beta(m-n-h))d\beta \ll \int_{\frac{1}{qQ}}^{1-\frac{1}{qQ}} \frac{1}{\|\beta\|^2} d\beta \\
&\ll \int_{\frac{1}{qQ}}^{\frac{1}{2}} \frac{1}{\beta^2} d\beta + \int_{\frac{1}{2}}^{1-\frac{1}{qQ}} \frac{d\beta}{(1-\beta)^2} \\
&\ll \int_{\frac{1}{qQ}}^{\frac{1}{2}} \frac{1}{\beta^2} d\beta \\
&\ll qQ.
\end{aligned}
$$

Therefore, combining this with (2.3.13), we have that

$$
I_1 = X + O\left(qQ + H\right).
$$

We now substitute this expression for I_1 into (2.3.12) to get

$$
\begin{aligned}
&\int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha) d\alpha \\
&= \sum_{q\le Q_0} \frac{\mu^2(q) c_q(-h)}{\varphi^2(q)} \left(X \left(\sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right)^2 + O\left(qQ + H\right) \right).
\end{aligned}
$$

To complete the proof, it remains to treat the sum over q. By Lemma 2.3.12 and our definitions of H and Q_0, we find immediately that for all but at most $O(HQ_0^{-1/3})$ values of $0 < |h| \le H$ we have that

$$
\int_{\mathfrak{M}} |a(\alpha)|^2 e(-h\alpha) d\alpha = \mathfrak{S}(h) X \left(\sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right)^2 + O\left(\frac{X}{\log^\eta X} \right),
$$

for some $\eta = \eta(\varepsilon) > 0$, as claimed. □

2.3.4 The Error Term of the Major Arcs

In order to complete the proof of Proposition 2.3.7, and therefore the proof of Theorem 1.0.2, we need to find sufficient cancellation in the error term $B^2(X)$ arising on the major arcs. Recall that $B^2(X)$ is defined as in Definition 2.3.8 to be

$$B^2(X) := \int_{\mathfrak{M}} |b(\alpha)|^2 d\alpha,$$
$$b(\alpha) := \frac{1}{\varphi(q)} \sum_{\chi(q)} \tau(\overline{\chi})\chi(a) \sum_{X<n\leq 2X} \left(\varpi_2(n)\chi(n) - \delta_\chi \sum_{P<p\leq P^{1+\delta}} \frac{1}{p} \right) e(\beta n).$$

In this section we prove the following bound for $B^2(X)$, which immediately completes the proof of Proposition 2.3.7 when combined with Proposition 2.3.10:

Proposition 2.3.13. *Let $\varepsilon > 0$ be fixed sufficiently small, then there exists some $\eta = \eta(\varepsilon) > 0$ such that*

$$B^2(X) \ll \frac{X}{\log^\eta X}.$$

Remark 2.3.14. If the smaller prime factor lies in the range $p \in [P_1, P_2]$ for some $P_1 = P_1(X)$ and $P_2 = P_2(X)$ instead of $p \in (P, P^{1+\delta}]$, we have the bound

$$B^2(X) \ll \frac{X}{\log^C X} + \frac{XQ_0}{U} + XU^2 P_2^{-2\alpha_1} \log^2 P_2 + Q_0^3 Q \log^2 X + \frac{X}{Q^2} + X^{49/50} Q_0^2 \log^2 X$$

for any $C > 0$. Here, U is a parameter chosen later in terms of Q_0. For the bound given in Proposition 2.3.13, we will choose $P_1 = P$, $P_2 = P^{1+\delta}$, $U = Q_0^{1+\varepsilon^2}$, and $\alpha_1 := \frac{3}{34} - \varepsilon'$ with ε' sufficiently small in terms of ε. For the proofs of Theorems 1.0.4 and 1.0.5, we will make different choices for the parameters, namely $P_1 = \exp((\log X)^{o(1)})$, $P_2 = \exp((\log X)^{1-o(1)})$, $Q_0 = \log^{A'} X$, $U = Q_0^E$ for some suitable $A', E > 0$, and $\alpha_1 = \varepsilon'$, which will give a better error term in these results.

2.3.4.1 Reduction of the problem

First, using Gallagher's Lemma (Lemma 2.2.4), we will reduce the problem of estimating $B^2(X)$ to understanding almost primes in almost all short intervals. We first define the following:

Definition 2.3.15. Define $T_0 := X^{1/100}$ and $\Delta := \frac{2X}{qT_0^3}$, where $q \leq Q_0$.

Let Q_0 be as defined in Definition 2.3.5, P as in Definition 2.3.1 and

$Q := P \log X$. Let ϖ_2 be as in Definition 2.3.2. Then we define $B_1(X)$ to be

$$\sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_X^{2X} \left| \frac{2}{qQ} \sum_{x<n\le x+qQ/2} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right) \right. \\ \left. - \frac{2}{q\Delta} \sum_{x<n\le x+q\Delta/2} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right) \right|^2 dx \tag{2.3.14}$$

and $B_2(X)$ to be

$$\sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_X^{2X} \left| \frac{2}{q\Delta} \sum_{x<n\le x+q\Delta/2} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right) \right|^2 dx. \tag{2.3.15}$$

Now we are able to state a bound for $B^2(X)$ in terms of $B_1(X)$ and $B_2(X)$:

Proposition 2.3.16. *We have that*

$$B^2(X) \ll B_1(X) + B_2(X) + \exp(2(\log\log X)^2).$$

Remark 2.3.17. The final error term bounds $Q_0^3 Q \log^2 X$. In the proofs of Theorems 1.0.4 and 1.0.5, we will make larger choices of Q_0 and Q. However, this error term will still be at most $\ll X^{1/6+\varepsilon} \ll X \log^{-A} X$, which will be acceptable.

Then, if we can prove that $B_i(X) \ll X \log^{-\eta} X$ for $i = 1, 2$, we will immediately be able to conclude Proposition 2.3.13.

Proof. By definition, we have that $B^2(X)$ is equal to

$$\sum_{q\le Q_0}\sum_{\substack{1\le a\le q\\(a,q)=1}}\int_{|\beta|\le\frac{1}{qQ}}\left|\frac{1}{\varphi(q)}\sum_{\chi(q)}\tau(\overline{\chi})\chi(a)\right.$$
$$\left.\times\sum_{X<n\le 2X}\left(\chi(n)\varpi_2(n)-\delta_\chi\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)e(\beta n)\right|^2 d\beta.$$

Expanding the square, we have that $B^2(X)$ is equal to

$$\sum_{q\le Q_0}\frac{1}{\varphi^2(q)}\sum_{\chi,\chi'(q)}\tau(\overline{\chi})\tau(\chi')\sum_{\substack{1\le a\le q\\(a,q)=1}}\chi(a)\overline{\chi'}(a)$$
$$\times\int_{|\beta|\le\frac{1}{qQ}}\sum_{X<m\le 2X}\left(\chi(m)\varpi_2(m)-\delta_\chi\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)$$
$$\times\sum_{X<n\le 2X}\left(\overline{\chi'}(n)\varpi_2(n)-\delta_{\chi'}\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)e(\beta(m-n))d\beta.$$

Now, using the definition of Dirichlet characters to trivially extend the sum over a to all $1\le a\le q$, we may apply the character orthogonality relation

$$\sum_{a=1}^{q}\chi(a)\overline{\chi'}(a)=\begin{cases}\varphi(q), & \text{if } \chi=\chi',\\ 0, & \text{if } \chi\ne\chi',\end{cases}$$

to see that

$$B^2(X)$$
$$= \sum_{q\le Q_0} \frac{1}{\varphi(q)} \sum_{\chi(q)} |\tau(\overline{\chi})|^2 \int_{|\beta|\le\frac{1}{qQ}} \left| \sum_{X<n\le 2X} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right) e(\beta n) \right|^2 d\beta$$
$$\ll \sum_{q\le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{|\beta|\le\frac{1}{qQ}} \left| \sum_{X<n\le 2X} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right) e(\beta n) \right|^2 d\beta,$$

where we have used that $\tau(\overline{\chi}) \ll q^{1/2}$ in the last line. Now we apply Lemma 2.2.4 to the integral term to get that $B^2(X)$ is bounded by

$$\sum_{q\le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \left(\int_X^{2X} \left| \frac{2}{qQ} \sum_{x<n\le x+qQ/2} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P<p\le P^{1+\delta}} \frac{1}{p} \right) \right|^2 dx \right.$$
$$\left. + qQ\log^2 X \right).$$

The second term contributes

$$Q\log^2 X \sum_{q\le Q_0} \sum_{\chi(q)} \frac{q^2}{\varphi(q)} \ll QQ_0^3 \log^2 X \ll \exp(2(\log\log X)^2),$$

to $B^2(X)$. As in Definition 2.3.15, let $\Delta = \frac{2X}{qT_0^3}$ with $T_0 = X^{1/100}$. Then we have that $B^2(X)$ is bounded by

$$\begin{aligned}
&\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_X^{2X}\Bigg|\frac{2}{qQ}\sum_{x<n\le x+qQ/2}\left(\chi(n)\varpi_2(n)-\delta_\chi\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)\\
&\qquad -\frac{2}{q\Delta}\sum_{x<n\le x+q\Delta/2}\left(\chi(n)\varpi_2(n)-\delta_\chi\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)\Bigg|^2 dx\\
&+\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_X^{2X}\left|\frac{2}{q\Delta}\sum_{x<n\le x+q\Delta/2}\left(\chi(n)\varpi_2(n)-\delta_\chi\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)\right|^2 dx\\
=&B_1(X)+B_2(X),
\end{aligned}$$

as claimed. □

2.3.4.2 Bounding $B_2(X)$

First, we prove the the following estimate for $B_2(X)$, which will be reduced to a Dirichlet character analogue of a problem on primes in almost all short intervals. We recall that $B_2(X)$ is defined as in Definition 2.3.15, (2.3.15) to be

$$\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_X^{2X}\left|\frac{2}{q\Delta}\sum_{x<n\le x+q\Delta/2}\left(\chi(n)\varpi_2(n)-\delta_\chi\sum_{P<p\le P^{1+\delta}}\frac{1}{p}\right)\right|^2 dx.$$

Proposition 2.3.18. *Let $C>0$ be fixed, then with $B_2(X)$ as defined in* (2.3.15) *we have*

$$B_2(X)\ll\frac{X}{\log^C X}.$$

Proof. We separate the cases $\chi = \chi_0$ and $\chi \neq \chi_0$. If $\chi = \chi_0$, we have that

$$\frac{2}{q\Delta} \sum_{x<n\leq x+q\Delta/2} \left(\varpi_2(n) - \sum_{P<p\leq P^{1+\delta}} \frac{1}{p} \right)$$
$$= \frac{2}{q\Delta} \sum_{P<p_1\leq P^{1+\delta}} \sum_{\frac{x}{p_1}<p_2\leq \frac{x+q\Delta/2}{p_1}} \log p_2 - \sum_{P<p\leq P^{1+\delta}} \frac{1}{p} + O\left(\frac{1}{q\Delta}\right),$$

where Δ is as defined in Definition 2.3.15. We now apply the prime number theorem in short intervals (Lemma 2.2.5), finding that

$$\frac{2}{q\Delta} \sum_{P<p_1\leq P^{1+\delta}} \sum_{\frac{x}{p_1}<p_2\leq \frac{x+q\Delta/2}{p_1}} \log p_2 = \sum_{P<p\leq P^{1+\delta}} \frac{1}{p} + O\left(\exp(-c(\log x)^{1/3-\varepsilon})\right).$$

Substituting this back into the above, we have that

$$\frac{2}{q\Delta} \sum_{x<n\leq x+q\Delta/2} \left(\varpi_2(n) - \sum_{P<p\leq P^{1+\delta}} \frac{1}{p} \right) = O\left(\exp(-c(\log x)^{1/3-\varepsilon})\right).$$

Returning to the integral and summing over q, we find that the contribution of the principal character to $B_2(X)$ is

$$\ll X \exp(-c'(\log X)^{1/3-\varepsilon}) \sum_{q\leq Q_0} \frac{q}{\varphi(q)} \ll XQ_0 \exp(-c'(\log X)^{1/3-\varepsilon})$$
$$\ll X \exp(-c''(\log X)^{1/3-\varepsilon}),$$

which is $\ll X \log^{-C} X$, so this contribution is acceptable.

We now consider the case $\chi \neq \chi_0$. By the definition of ϖ_2, we have that

$B_2(X)$ is

$$= \sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\substack{\chi(q) \\ \chi \ne \chi_0}} \int_X^{2X} \left| \frac{2}{q\Delta} \sum_{x < n \le x + q\Delta/2} \chi(n) \varpi_2(n) \right|^2 dx$$

$$= \sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\substack{\chi(q) \\ \chi \ne \chi_0}} \int_X^{2X} \left| \frac{2}{q\Delta} \sum_{P < p_1 \le P^{1+\delta}} \chi(p_1) \sum_{\frac{x}{p_1} < p_2 \le \frac{x+q\Delta/2}{p_1}} \chi(p_2) \log p_2 \right|^2 dx. \tag{2.3.16}$$

Next, we apply the Cauchy-Schwarz inequality to the sum over p_1 to obtain

$$\left| \sum_{P < p_1 \le P^{1+\delta}} \chi(p_1) \sum_{\frac{x}{p_1} < p_2 \le \frac{x+q\Delta/2}{p_1}} \chi(p_2) \log p_2 \right|^2$$

$$\le \sum_{P < p_1' \le P^{1+\delta}} |\chi(p_1')|^2 \sum_{P < p_1 \le P^{1+\delta}} \left| \sum_{\frac{x}{p_1} < p_2 \le \frac{x+q\Delta/2}{p_1}} \chi(p_2) \log p_2 \right|^2.$$

By the prime number theorem, the sum over p_1' is $\ll P^{1+\delta}/\log P$. Therefore, (2.3.16) is bounded by

$$\ll \frac{P^{1+\delta}}{\log P} \sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{P < p_1 \le P^{1+\delta}} \frac{4}{(q\Delta)^2} \sum_{\substack{\chi(q) \\ \chi \ne \chi_0}} \int_X^{2X} \left| \sum_{\frac{x}{p_1} < p_2 \le \frac{x+q\Delta/2}{p_1}} \chi(p_2) \log p_2 \right|^2 dx. \tag{2.3.17}$$

We make the change of variables $u = x/p_1$ to the integral, so that

$$\int_X^{2X} \left| \sum_{\frac{x}{p_1} < p_2 \leq \frac{x+q\Delta/2}{p_1}} \chi(p_2) \log p_2 \right|^2 dx = p_1 \int_{X/p_1}^{2X/p_1} \left| \sum_{u < p_2 \leq u + \frac{q\Delta}{2p_1}} \chi(p_2) \log p_2 \right|^2 du.$$

First, in the case $H \leq \exp((\log X)^{\varepsilon^3})$, we now apply Lemma 2.2.6 to get that

$$\begin{aligned} B_2(X) &\ll \frac{P^{1+\delta}}{\log P} \frac{\log\log Q_0}{Q_0^2 \Delta^2} \sum_{P < p \leq P^{1+\delta}} p \left(\frac{Q_0^3 \Delta^2 X}{p^3 \log^B X} \right) \\ &\ll \frac{P^{1+\delta}}{\log P} \frac{X Q_0 \log\log Q_0}{\log^D X} \sum_{P < p \leq P^{1+\delta}} \frac{1}{p^2} \\ &\ll \frac{X}{\log^C X} \end{aligned}$$

for $C > 0$, as required. In the case $H > \exp((\log X)^{\varepsilon^3})$, we first split the sum over $P < p_1 \leq P^{1+\delta}$ in (2.3.16) into $O(\log P)$ dyadic intervals $[P_1, 2P_1]$. We then need to bound

$$\sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\substack{\chi(q) \\ \chi \neq \chi_0}} \int_X^{2X} \left| \frac{2}{q\Delta} \sum_{P_1 < p_1 \leq 2P_1} \chi(p_1) \sum_{\frac{x}{p_1} < p_2 \leq \frac{x+q\Delta/2}{p_1}} \chi(p_2) \log p_2 \right|^2 dx.$$

Applying Cauchy-Schwarz to the sum over p_1 and then the prime number theorem as in the previous case, this is bounded by

$$\ll \frac{P_1}{\log P_1} \sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{P_1 < p_1 \leq 2P_1} \frac{4}{(q\Delta)^2} \sum_{\substack{\chi(q) \\ \chi \neq \chi_0}} \int_X^{2X} \left| \sum_{\frac{x}{p_1} < p_2 \leq \frac{x+q\Delta/2}{p_1}} \chi(p_2) \log p_2 \right|^2 dx.$$

We again make the substitution $u = x/p_1$ in the integral and then apply Lemma 2.2.6 to obtain

$$\begin{aligned}
&\ll \frac{P_1}{\log P_1} \frac{\log\log Q_0}{Q_0^2\Delta^2} \sum_{P_1 < p \leq 2P_1} p \left(\frac{Q_0^3 \Delta^2 X}{p^3 \log^B X} \right) \\
&\ll \frac{P_1}{\log P_1} \frac{X Q_0 \log\log Q_0}{\log^D X} \sum_{P_1 < p \leq 2P_1} \frac{1}{p^2} \\
&\ll \frac{X}{\log^{C'} X},
\end{aligned}$$

for $C' > 0$. Combining all of the dyadic intervals introduces a factor of $\log^2 P = (\log\log X)^4$, which gives the bound $\ll X \log^{-C} X$ for some $C > 0$, as required. □

2.3.4.3 Bounding $B_1(X)$

It now remains to prove the required bound for $B_1(X)$. Recall that $B_1(X)$ is defined in Definition 2.3.15, (2.3.14) to be

$$\sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_X^{2X} \left| \frac{2}{qQ} \sum_{x < n \leq x + qQ/2} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P < p \leq P^{1+\delta}} \frac{1}{p} \right) \right. \\ \left. - \frac{2}{q\Delta} \sum_{x < n \leq x + q\Delta/2} \left(\chi(n)\varpi_2(n) - \delta_\chi \sum_{P < p \leq P^{1+\delta}} \frac{1}{p} \right) \right|^2 dx.$$

This problem can be reduced to finding cancellation in the mean square of a Dirichlet polynomial.

Proposition 2.3.19. *Let $\varepsilon > 0$ be fixed sufficiently small. With $B_1(X)$ as defined in Definition 2.3.15,* (2.3.14)*, there exists some $\eta = \eta(\varepsilon) > 0$ such* *that*

$$B_1(X) \ll \frac{X}{\log^\eta X}.$$

Remark 2.3.20. If the smaller prime factor lies in the range $p \in [P_1, P_2]$ for some $P_1 = P_1(X)$ and $P_2 = P_2(X)$ instead of $p_1 \in (P, P^{1+\delta}]$, we have the bound

$$B_1(X) \ll \frac{XQ_0}{U} + XU^2 Q_0 P_2^{-2\alpha_1} \log^2 P_2 + \frac{X}{Q^2} + X^{49/50} Q_0^2 \log^2 X.$$

Here, U is a parameter chosen later in terms of Q_0. For the bound given in Proposition 2.3.19, we will choose $P_1 = P$, $P_2 = P^{1+\delta}$ and $U = Q_0^{1+\varepsilon^2}$, and $\alpha_1 := \frac{3}{34} - \varepsilon'$ with ε' sufficiently small in terms of ε. For the proofs of Theorems 1.0.4 and 1.0.5, we will make different choices for the parameters, namely $P_1 = \exp((\log X)^{o(1)})$, $P_2 = \exp((\log X)^{1-o(1)})$, $Q_0 = \log^{A'} X$, $U = Q_0^E$ for some suitable $A', E > 0$, and $\alpha_1 = \varepsilon'$, which will give a better error term.

To prove this result, we will need the following variant of a result of Teräväinen [38] on the mean square of the Dirichlet polynomial

$$F(s, \chi) := \sum_{\substack{X < p_1 p_2 \le 2X \\ P < p_1 \le P^{1+\delta}}} \frac{\chi(p_1)\chi(p_2)}{(p_1 p_2)^s}, \tag{2.3.18}$$

where $s \in \mathbb{C}$ and χ is a Dirichlet character modulo q, to be proved in Section 2.3.6:

Proposition 2.3.21. *Let $\varepsilon > 0$ be fixed sufficiently small. Define $T_0 = X^{1/100}$ as in Definition 2.3.15 and $F(s,\chi)$ to be the Dirichlet polynomial defined in* (2.3.18)*, with P and $\delta > 0$ as in Definition 2.3.1. Then, for $T \geq T_0$, there exists some $\eta = \eta(\varepsilon) > 0$ such that*

$$\begin{aligned} B_3(X) &:= \sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{T_0}^{T} |F(1+it,\chi)|^2 dt \\ &\ll \frac{1}{Q_0 (\log X)^{2+\eta}} \sum_{q \leq Q_0} \left(\frac{qTP\log X}{X} + \frac{q}{\varphi(q)} \right). \end{aligned}$$

Remark 2.3.22. More generally, if the smaller prime factor lies in the range $p_1 \in [P_1, P_2]$ for some $P_1 = P_1(X)$ and $P_2 = P_2(X)$, we have the bound

$$B_3(X) \ll \left(\frac{1}{U \log^2 X} + \frac{U^2 P_2^{-2\alpha_1} \log^2 P_2}{\log^2 X} \right) \sum_{q \leq Q_0} \left(\frac{qTP_2 \log X}{X} + \frac{q}{\varphi(q)} \right).$$

Here, U is a parameter chosen later in terms of Q_0. For the statement given in Proposition 2.3.21, we choose $P_1 = P$, $P_2 = P^{1+\delta}$, $U = Q_0^{1+\varepsilon^2}$, and $\alpha_1 := \frac{3}{34} - \varepsilon'$ with ε' chosen sufficiently small in terms of ε.

For the proofs of Theorems 1.0.4 and 1.0.5, we will make different choices of these parameters, which will lead to an improved error term. In particular, we will choose $P_1 = \exp((\log X)^{o(1)})$, $P_2 = \exp((\log X)^{1-o(1)})$, $Q_0 = \log^{A'} X$, $U = Q_0^E$ for some suitable $A', E > 0$, and $\alpha_1 = \varepsilon'$.

Proof of Proposition 2.3.19 assuming Proposition 2.3.21. First we consider when $\chi = \chi_0$ as we have a different summand in this case. We have

$$\frac{2}{qQ}\sum_{x<n\leq x+qQ/2}\left(\varpi_2(n)-\sum_{P<p\leq P^{1+\delta}}\frac{1}{p}\right)$$
$$-\frac{2}{q\Delta}\sum_{x<n\leq x+q\Delta/2}\left(\varpi_2(n)-\sum_{P<p\leq P^{1+\delta}}\frac{1}{p}\right),$$

where $\Delta = \frac{2X}{qT_0^3}$ as in Definition 2.3.15 and $Q = P\log X$. We first consider the contribution of the second and fourth terms, namely

$$\sum_{P<p\leq P^{1+\delta}}\frac{1}{p}\left(\frac{2}{q\Delta}\sum_{x<n\leq x+q\Delta/2}1-\frac{2}{qQ}\sum_{x<n\leq x+qQ/2}1\right)\ll\frac{1}{qQ}+\frac{1}{q\Delta}. \quad (2.3.19)$$

Returning to our expression for $B_1(X)$, by our choice of Q_0, Q and Δ we have that (2.3.19) contributes

$$\ll\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\int_X^{2X}\frac{1}{(qQ)^2}dx\ll\frac{X}{Q^2}\sum_{q\leq Q_0}\frac{1}{q\varphi(q)}\ll\frac{X}{Q^2}\ll\frac{X}{(\log X)^{36+2\varepsilon}},$$

which is $\ll X\log^{-\eta}X$, so is acceptable. Therefore, when considering the principal character χ_0, we need only to bound

$$\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\int_X^{2X}\left|\frac{2}{qQ}\sum_{x<n\leq x+qQ/2}\varpi_2(n)-\frac{2}{q\Delta}\sum_{x<n\leq x+q\Delta/2}\varpi_2(n)\right|^2dx$$
$$=\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\int_X^{2X}\left|\left(\frac{2}{qQ}\sum_{x<n\leq x+qQ/2}-\frac{2}{q\Delta}\sum_{x<n\leq x+q\Delta/2}\right)\varpi_2(n)\chi_0(n)\right|^2dx,$$

noting that in the range of summation we must have $(n, q) = 1$, in particular $\varpi_2\chi_0(n) = \varpi_2(n)$ for each $X < n \leq 2X$. Thus, from now on we are able to unify the treatment of the principal character χ_0 with the rest of the characters modulo q at the cost of a negligible error.

We now apply Lemma 2.2.7 with $T' = T_0$, $h_1 = qQ/2$ and $h_2 = q\Delta/2$ to the integral with respect to x to get

$$B_1(X) \ll X \sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \left(\frac{\log^2 X}{T_0^2} + \int_{T_0}^{\frac{2X}{qQ}} |F_1(1+it,\chi)|^2 dt \right.$$
$$\left. + \max_{T \geq \frac{2X}{qQ}} \frac{X}{TqQ} \int_T^{2T} |F_1(1+it,\chi)|^2 dt \right),$$

with $T_0 = X^{1/100}$ and

$$F_1(s,\chi) := \sum_{X < n \leq 2X} \frac{\varpi_2(n)\chi(n)}{n^s} = \sum_{\substack{X < p_1p_2 \leq 2X \\ P < p_1 \leq P^{1+\delta}}} \frac{\chi(p_1)\chi(p_2)\log p_2}{(p_1p_2)^s}.$$

By the definition of $T_0 := X^{1/100}$, the first term is

$$X^{49/50} \log^2 X \sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} 1 \ll X^{49/50} Q_0^2 \log^2 X \ll \frac{X}{\log^\eta X},$$

so is acceptable. Applying partial summation, we have that

$$B_1(X) \ll X(\log X)^2 \sum_{q\leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \left(\int_{T_0}^{\frac{2X}{qQ}} |F(1+it,\chi)|^2 dt \right.$$
$$\left. + \max_{T\geq \frac{2X}{qQ}} \frac{X}{TqQ} \int_{T}^{2T} |F(1+it,\chi)|^2 dt \right).$$

We now apply Proposition 2.3.21. Note that we have $P\log X = Q$, so that the first term in our bound for $B_1(X)$ is bounded by

$$\ll \frac{X}{Q_0 \log^\eta X} \sum_{q\leq Q_0} \left(\frac{P\log X}{Q} + \frac{q}{\varphi(q)} \right) \ll \frac{X}{\log^\eta X},$$

as needed. For the second term, we want to bound

$$\frac{X^2\log^2 X}{Q} \sum_{q\leq Q_0} \frac{1}{\varphi(q)} \max_{T\geq \frac{2X}{qQ}} \frac{1}{T} \sum_{\chi(q)} \int_{T}^{2T} |F(1+it,\chi)|^2 dt.$$

Applying Proposition 2.3.21, we have the bound

$$\ll \frac{X^2}{Q_0 Q\log^\eta X} \sum_{q\leq Q_0} \max_{T\geq \frac{2X}{qQ}} \left(\frac{P\log X}{X} + \frac{1}{T\varphi(q)} \right) \ll \frac{X}{\log^\eta X},$$

again using that $P\log X = Q$. Overall we have that

$$B_1(X) \ll \frac{X}{\log^\eta X},$$

for some $\eta = \eta(\varepsilon) > 0$, as required. □

2.3.5 Preliminaries on Dirichlet Polynomials

Before we can prove Proposition 2.3.21, we first need the following preliminary lemmas on Dirichlet polynomials.

2.3.5.1 Definitions

We will in some instances take the mean square over a sparse set of points, which can be split into "well-spaced" subsets:

Definition 2.3.23. [Well-Spaced Set] We say a set $\mathcal{T}$ is *well-spaced* if for any $t, u \in \mathcal{T}$ with $t \neq u$ we have that $|t-u| \geq 1$.

After decomposing our Dirichlet polynomial, we will have factors which are "prime-factored", that is, polynomials which satisfy certain pointwise bounds:

Definition 2.3.24 (Prime-factored polynomial, [38])**.** Let $s \in \mathbb{C}$, $M \geq 1$ and

$$M(s,\chi) = \sum_{M<m\leq 2M} \frac{a_m \chi(m)}{m^s}$$

be a Dirichlet polynomial with $|a_m| \ll d_r(m)$ for some fixed r. Let $T \geq 1$, $q \geq 2$, $\mathcal{T} \subset [-T,T]$ be a well-spaced set, and $\mathcal{S} = \mathcal{T} \times \{\chi \mod q\}$. Suppose that $\min\{|t| : (t,\chi) \in \mathcal{S}\} \gg \log^A N$ for all $A > 0$ if $\chi = \chi_0$. We say that $M(s,\chi)$ is *prime-factored* if for each $C > 0$ we have

$$\sup_{(t,\chi)\in\mathcal{S}} |M(1+it,\chi)| \ll \frac{1}{\log^C M}$$

when $\exp((\log M)^{9/10}) \leq t \leq M^{C \log\log M}$.

The use of the term "prime-factored" is in reference to Dirichlet polynomials which satisfy

$$\sum_{R<r\leq 2R} \frac{c_r}{r^{1/2+it}} \ll \frac{R^{1/2}}{\log^A R}$$

for any $A > 0$, where the coefficient c_r is the characteristic function of a set of numbers with a bounded number of prime factors restricted to certain ranges, or the characteristic function of the primes. Such polynomials arise for example when applying the fundamental lemma of the sieve to problems on primes in short intervals (see [15, Chapter 7.2]).

2.3.5.2 Decomposing Dirichlet Polynomials

As in the work of Teräväinen [38] and Matomäki, Radziwiłł [25], we take advantage of the bilinear structure to factorise our Dirichlet polynomial.

Lemma 2.3.25 (Factorisation of Dirichlet Polynomials)**.** *Let* $s \in \mathbb{C}$ *and* $v \in \mathbb{Z}$*. Define*

$$F(s) := \sum_{\substack{X<mn\leq 2X \\ M\leq m\leq M'}} \frac{a_m b_n}{(mn)^s}$$

for some $M' > M \geq 2$ *and arbitrary complex numbers* a_m, b_n*. Let* $U \geq 1$ *and define*

$$A_v(s) := \sum_{e^{\frac{v}{U}} \leq m < e^{\frac{v+1}{U}}} \frac{a_m}{m^s}, \quad B_v(s) := \sum_{Xe^{-\frac{v}{U}} < n \leq 2Xe^{-\frac{v}{U}}} \frac{b_n}{n^s}.$$

Then

$$F(s) = \sum_{v \in I \cap \mathbb{Z}} A_v(s) B_v(s) + \sum_{\substack{k \in [Xe^{-1/U}, Xe^{1/U}] \\ or\ k \in [2X, 2Xe^{1/U}]}} \frac{d_k}{k^s} \tag{2.3.20}$$

where $I = [U \log M, U \log M']$ *and*

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

Proof. This is [38, Lemma 2] (see also [25, Lemma 12]). □

In some cases we will use the Heath-Brown identity to decompose a long polynomial into products of shorter polynomials.

Lemma 2.3.26 (Heath-Brown decomposition)**.** *Let* $k \geq 1$ *be a fixed integer,* $T \geq 2$ *and fix* $\varepsilon > 0$*. For* $s \in \mathbb{C}$ *and* χ *a Dirichlet character modulo* q*, define the Dirichlet polynomial* $P(s, \chi) := \sum_{P' \leq p < P_1} \chi(p) p^{-s}$ *with* $P' \gg T^{\varepsilon}$*,* $P_1 \in \left[P' + \frac{P'}{\log T}, 2P'\right]$*. Then, there exist Dirichlet polynomials* $Q_1(s, \chi), \ldots, Q_L(s, \chi)$ *and a constant* $C > 0$ *such that* $L \leq \log^C X$ *and*

$$|P(1 + it, \chi)| \ll (\log^C X)(|Q_1(1 + it, \chi)| + \cdots + |Q_L(1 + it, \chi)|)$$

for all $t \in [-T, T]$*. Here, each* $Q_j(s, \chi)$ *is of the form*

$$Q_j(s, \chi) = \prod_{i \leq J_j} M_i(s, \chi), \quad J_j \leq 2k,$$

where each $M_i(s, \chi)$ *is a prime-factored Dirichlet polynomial (depending on* *j) of the form*

$$\sum_{M_i<n\leq 2M_i} \frac{\chi(n)\log n}{n^s}, \quad \sum_{M_i<n\leq 2M_i} \frac{\chi(n)}{n^s}, \quad \text{or} \quad \sum_{M_i<n\leq 2M_i} \frac{\mu(n)\chi(n)}{n^s},$$

whose lengths satisfy $M_1 \cdots M_{J_j} = X^{1+o(1)}, M_i \gg \exp\left(\frac{\log P}{\log\log P}\right)$. *Furthermore, if in fact* $M_i > X^{1/k}$*, then* $M_i(s,\chi)$ *is of the form*

$$\sum_{M_i<n\leq 2M_i} \frac{\chi(n)\log n}{n^s} \quad \text{or} \quad \sum_{M_i<n\leq 2M_i} \frac{\chi(n)}{n^s}.$$

Proof. This is [38, Lemma 10] with the coefficient twisted by a Dirichlet character, and follows from the same argument. □

2.3.5.3 Mean Value Theorems for Dirichlet Polynomials

Now we state two mean value theorems, the first being the classical result:

Lemma 2.3.27 (Mean Value Theorem)**.** *Let* $q, X \geq 1$ *and let* a_n *be arbitrary complex numbers. Let* $s \in \mathbb{C}$ *and* χ *be a Dirichlet character mod* q*, and define* $F(s,\chi) := \sum_{X<n\leq 2X} \frac{a_n\chi(n)}{n^s}$. *Then*

$$\sum_{\chi(q)} \int_{-T}^{T} |F(it,\chi)|^2 dt \ll \left(\varphi(q)T + \frac{\varphi(q)}{q}X\right) \sum_{\substack{X<n\leq 2X \\ (n,q)=1}} |a_n|^2.$$

Proof. See, for example, [33, Chapter 6, Eq. (6.14)]. □

Next we state a variant of the mean value theorem which will allow us to save a $\log X$ in certain parts of the proof.

Lemma 2.3.28. *With the same assumptions as Lemma 2.3.27, we have that*

$$\sum_{\chi(q)} \int_{-T}^{T} |F(it,\chi)|^2 dt \ll T\varphi(q) \Bigg(\sum_{\substack{X<n\leq 2X \\ (n,q)=1}} |a_n|^2 + \sum_{\substack{1\leq h\leq \frac{X}{T} \\ q|h}} \sum_{\substack{X<n\leq 2X \\ (n(n+h),q)=1}} |a_{n+h}||a_n| \Bigg).$$

Proof. This is the Dirichlet character analogue of [38, Lemma 4], which follows from [18, Lemma 7.1]. The proof is contained in the proof of [28, Lemma 5.2]. □

After factorising the Dirichlet polynomial F and splitting the domain of integration according to the size of the factors, there will be cases where the mean value is taken over a well-spaced set. In this case, we will apply the Halász-Montgomery inequality:

Lemma 2.3.29 (Halász-Montgomery Inequality)**.** *Let* $T \geq 1$, $q \geq 2$. *Let* $\mathcal{T} \subset [-T,T]$ *be a well-spaced set, and* $\mathcal{S} = \mathcal{T} \times \{\chi \mod q\}$. *With the same assumptions as Lemma 2.3.27, we have that*

$$\sum_{(t,\chi)\in\mathcal{S}} |F(it,\chi)|^2 \ll \left(\frac{\varphi(q)}{q} X + |\mathcal{S}|(qT)^{1/2} \right) (\log(2qT)) \sum_{\substack{X<n\leq 2X \\ (n,q)=1}} |a_n|^2.$$

Proof. This is [20, Lemma 7.4]. □

2.3.5.4 Large Value Theorems

There will be subsets of the domain of integration where a short Dirichlet polynomial factor is large, in which case we apply the following large value theorem.

Lemma 2.3.30 (Large Value Theorem)**.** *Let $P' \geq 1, V > 0$ and $T \geq 10$. For $s \in \mathbb{C}$ and χ mod q, define $F(s,\chi) = \sum_{P'<p\leq 2P'} \frac{a_p\chi(p)}{p^s}$ with $|a_p| \leq 1$. Let $\mathcal{T} \subset [-T,T]$ be a well-spaced set and $\mathcal{S} = \mathcal{T} \times \{\chi \mod q\}$ such that $|F(1+it,\chi)| \geq V$ for all $(t,\chi) \in \mathcal{S}$. Then*

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

Proof. See [20, Lemma 7.5]. This is the Dirichlet character analogue of [38, Lemma 6] and [25, Lemma 8]. □

Remark 2.3.31. As remarked in [38, Remark 6], this Lemma can still be applied to polynomials with coefficients $|a_n| \leq 1$ not only supported on the primes in the case we will have in our application, $P' \gg \exp\left(\frac{\log X}{\log\log X}\right)$. The coefficient of the exponent $\frac{\log(qT)\log\log(qT)}{\log P'}$ is replaced with $\frac{\log^2(qT)\log\log(P')}{\log^2 P'}$.

Alternatively, in the case that we have a longer Dirichlet polynomial factor which is large, we will apply a result of Jutila on large values.

Lemma 2.3.32 (Jutila's Large Value Theorem)**.** *Let $\varepsilon > 0$ be fixed. For $s \in \mathbb{C}$ and χ mod q define $F(s,\chi) = \sum_{X<n\leq 2X} \frac{a_n\chi(n)}{n^s}$ with $|a_n| \leq d_r(n)$ for some fixed r. Let k be a fixed positive integer. Let $\mathcal{T} \subset [-T,T]$ be a well-spaced set and $\mathcal{S} = \mathcal{T} \times \{\chi \mod q\}$ such that $|F(1+it,\chi)| \geq V$ for all*

$(t,\chi) \in \mathcal{S}$. *Then,*

$$|\mathcal{S}| \ll \left(V^{-2} + \left(\frac{qTV^{-4}}{X^2} \right)^k + \frac{qTV^{-8k}}{X^{2k}} \right) (qTX)^{\varepsilon}. \tag{2.3.21}$$

Proof. This is the first bound of the main theorem in [19]. □

2.3.5.5 Moments of Dirichlet Polynomials

After decomposing the Dirichlet polynomial using Heath-Brown's decomposition (Lemma 2.3.26), we can have a long polynomial which is the partial sum of a Dirichlet L-function (or its derivative). In this case, we will apply the Cauchy-Schwarz inequality to enable us to use the following bound on the twisted fourth moment of such sums:

Lemma 2.3.33 (Twisted Fourth Moment Estimate)**.** *Let* $Q' \le T^{\varepsilon}$, $T^{\varepsilon} \le T' \le T$, $1 \le M, N \le T^{1+o(1)}$ *and for* $s \in \mathbb{C}$ *and* χ *mod* q *define the Dirichlet polynomials*

$$N(s,\chi) = \sum_{N<n\le 2N} \frac{\chi(n)}{n^s} \quad or \quad \sum_{N<n\le 2N} \frac{\chi(n)\log n}{n^s},$$
$$M(s,\chi) = \sum_{M<m\le 2M} \frac{a_m \chi(m)}{m^s},$$

with a_m any complex numbers. Then we have that

$$\sum_{q\leq Q'}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T'}^{T}|N(1+it,\chi)|^4|M(1+it,\chi)|^2dt$$
$$\ll\left(\frac{Q'T}{MN^2}(1+M^2(Q'T)^{-1/2})+\frac{1}{T'}\right)(Q'T)^{\varepsilon}\max_{M<m\leq 2M}|a_m|^2. \tag{2.3.22}$$

Proof. This is the Dirichlet character analogue of [38, Lemma 9] and we follow the same argument. We split the domain of integration $t\in[T_0,T]$. In the case $Q't\leq N$, we use the hybrid result of Fujii, Gallagher and Montgomery [9, Theorem 1]

$$\sum_{n\leq N}\chi(n)n^{it}=\frac{\delta_\chi\varphi(q)N^{1+it}}{q(1+it)}+O((q\tau)^{1/2}\log(q\tau)),$$

with $\tau:=|t|+2$ (noting we apply partial summation to deal with the $1/n$ factor) in place of the zeta sum bound to get that

$$\sum_{q\leq Q'}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T'}^{N/Q'}|N(1+it,\chi)|^4|M(1+it,\chi)|^2dt$$
$$\ll T^{\varepsilon}\max_{M<m\leq 2M}|a_m|^2\sum_{q\leq Q'}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T'}^{N/Q'}\left(\frac{\varphi(q)}{q(1+|t|)}\right)^4+\frac{\log^4(q\tau)}{\tau^2}dt$$
$$\ll\frac{T^{\varepsilon}}{T'}\max_{M<m\leq 2M}|a_m|^2,$$

providing the third term of (2.3.22).

In the case $N\leq Q't\leq Q'T$, we apply in place of Watt's twisted moment result its Dirichlet character analogue [16, Theorem 2]. This analogue, due to Harman, Watt and Wong, states that for $\varepsilon > 0$ given and Q'' a positive integer, we have for all $M \geq 1$, $T \geq (Q'')^{3/5}$ that

$$\begin{aligned} &\sum_{q \leq Q''} \frac{1}{\varphi(q)} \sum_{\chi(q)} \int_{-T}^{T} |L(\tfrac{1}{2} + it, \chi)|^4 \Big| \sum_{m \leq M} a_m \chi(m) m^{-it} \Big|^2 dt \\ &\ll (TMQ'')^{1+\varepsilon} \big(1 + M^2 (TQ'')^{-1/2}\big) \max_{m \leq M} |a_m|^2. \end{aligned} \tag{2.3.23}$$

Then, as in [38, Lemma 9], we obtain the first two terms of (2.3.22) using the argument of [1, Lemma 2]. We first split the integral

$$\sum_{q \leq Q'} \frac{1}{\varphi(q)} \sum_{\chi(q)} \int_{N/Q'}^{T} |N(1 + it, \chi)|^4 |M(1 + it, \chi)|^2 dt$$

into dyadic ranges $[T_1, 2T_1]$ with $N/Q' \leq T_1 \leq T$.

Case 1. $N, M \leq (Q'T_1)^{1/2}$.

The proof is included in the proof of this main result (2.3.23) of Harman, Watt and Wong [16, Theorem 2].

Case 2. $N \leq (Q'T_1)^{1/2}$ and $M > (Q'T_1)^{1/2}$.

We trivially bound $|M(1 + it, \chi)|^2$ and apply the mean value theorem

(Lemma 2.3.27) to obtain

$$\begin{aligned}
&\sum_{q\le Q'}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T_1}^{2T_1}|N(\tfrac{1}{2}+it,\chi)|^4|M(1+it,\chi)|^2dt\\
&\qquad\ll T^{\varepsilon}\Big(\max_{M<m\le 2M}|a_m|^2\Big)\sum_{q\le Q'}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T_1}^{2T_1}|N(\tfrac{1}{2}+it,\chi)|^4dt\\
&\qquad\ll T^{\varepsilon}\Big(\max_{M<m\le 2M}|a_m|^2\Big)\sum_{q\le Q'}\left(T_1+\frac{N^2}{q}\right)\\
&\qquad\ll T^{\varepsilon}\left(Q'T_1+N^2\log Q'\right)\max_{M<m\le 2M}|a_m|^2\\
&\qquad\ll T^{\varepsilon}M(Q'T)^{1/2}\max_{M<m\le 2M}|a_m|^2.
\end{aligned}$$

This contributes to the second term of (2.3.22) following an application of partial summation.

Case 3. $N>(Q'T_1)^{1/2}$.

We apply the approximate functional equation to replace $N(\frac{1}{2}+it,\chi)$ with another sum which is the partial sum of a Dirichlet L-function or its derivative and has length $N'\le(Q'T_1)^{1/2}$. This is treated as in the previous cases. The error terms can be treated as in [1, Lemma 2]. As in [16, Lemma 1], the error term is of size $\ll 1+R(\frac{1}{2}+it,\chi)$, where $R(\frac{1}{2}+it,\chi)\ge 0$ and

$$\sum_{\chi(q)}{}^{*}R(\tfrac{1}{2}+it,\chi)\ll\varphi(q)q^{\varepsilon}T_1^{\varepsilon-1/2}.$$

This contributes

$$\ll \sum_{q\le Q'} \frac{1}{\varphi(q)} \sum_{\chi(q)} \int_{T_1}^{2T_1} |1 + R(\tfrac{1}{2} + it, \chi)||M(1 + it, \chi)|^2 dt. \qquad (2.3.24)$$

We treat the first term of (2.3.24) using the mean value theorem (Lemma 2.3.27) to give

$$\ll T^{\varepsilon} \sum_{q\le Q'} \left(\frac{T_1}{M} + \frac{1}{q}\right) \max_{M<m\le 2M} |a_m|^2 \ll T^{\varepsilon} \left(\frac{Q'T}{M} + \log Q'\right) \max_{M<m\le 2M} |a_m|^2,$$

which contributes to the first term of (2.3.22). Applying the trivial bound $R(\frac{1}{2} + it, \chi) \ll \varphi(q) q^{\varepsilon} T_1^{\varepsilon-1/2}$ (handling the imprimitive characters as in [16]) and the mean value theorem (Lemma 2.3.27) $M(1 + it, \chi)$, the second term of (2.3.24) contributes

$$\begin{aligned} &\ll T^{\varepsilon} T_1^{-1/2} \sum_{q\le Q'} q^{\varepsilon} \left(\frac{\varphi(q)T_1}{M} + \frac{\varphi(q)}{q}\right) \max_{M<m\le 2M} |a_m|^2 \\ &\ll \left(\frac{(Q')^{2+\varepsilon} T^{1/2+\varepsilon}}{M} + \frac{(Q')^{1+\varepsilon} T^{\varepsilon}}{(T')^{1/2}}\right) \max_{M<m\le 2M} |a_m|^2. \end{aligned}$$

Following an application of partial summation (which introduces a factor of N^{-2}), this contribution is acceptable in comparison with the terms of (2.3.22). □

2.3.6 Bounding the Mean Value of a Dirichlet Polynomial

We will now prove Proposition 2.3.21, establishing the bound

$$\begin{aligned} B_3(X) &:= \sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{T_0}^{T} |F(1+it, \chi)|^2 dt \\ &\ll \frac{1}{Q_0 (\log X)^{2+\eta}} \sum_{q \leq Q_0} \left(\frac{qTP \log X}{X} + \frac{q}{\varphi(q)} \right). \end{aligned}$$

This will complete the proof of Proposition 2.3.13, which treats error terms on the major arcs, and consequently the main result Theorem 1.0.2. We will adapt the argument appearing in [38, Sections 2-4]. For $s \in \mathbb{C}$ and $\chi \mod q$, We first recall the definition (2.3.18) of the Dirichlet polynomial

$$F(s, \chi) := \sum_{\substack{X < p_1 p_2 \leq 2X \\ P < p_1 \leq P^{1+\delta}}} \frac{\chi(p_1)\chi(p_2)}{(p_1 p_2)^s}.$$

We will factorise this polynomial before bounding the contribution of the remainder terms, that is, the second term of (2.3.20).

Lemma 2.3.34. *Let $\varepsilon > 0$ be fixed sufficiently small and $T_0 = X^{1/100}$. Let $s \in \mathbb{C}$ and $v \in \mathbb{Z}$. Denote*

$$G_v(s, \chi) := \sum_{e^{\frac{v}{U}} < p \leq e^{\frac{v+1}{U}}} \frac{\chi(p)}{p^s}, \quad H_v(s, \chi) := \sum_{Xe^{-\frac{v}{U}} < p \leq 2Xe^{-\frac{v}{U}}} \frac{\chi(p)}{p^s},$$

then we have the bound

$$B_3(X) \ll \sum_{q\le Q_0}\left(\frac{qU^2\log^2 P}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{T}|G_{v_0}(1+it,\chi)|^2|H_{v_0}(1+it,\chi)|^2dt + \frac{1}{Q_0(\log X)^{2+\eta}}\left(\frac{qT\log X}{X}+\frac{q}{\varphi(q)}\right)\right),$$

for some $\eta=\eta(\varepsilon)>0$*, where* $v_0\in I$ *a suitable integer with* $I=[U\log P,(1+\delta)U\log P]$ *and* $U:=Q_0^{1+\varepsilon^2}$*.*

Remark 2.3.35. We note that the factor

$$\frac{1}{Q_0(\log X)^{2+\eta}}$$

appearing in the second and third terms of the sum over q arises from

$$\frac{1}{U\log^2 X}.$$

In the proofs of Theorems 1.0.4 and 1.0.5 we will make larger choices of Q_0 and U. In particular, we will choose $Q_0=\log^{A'}X$ and $U=Q_0^E$ for some suitable $A',E>0$.

Proof. We factorise $F(s,\chi)$ using Lemma 2.3.25 with $M=P$, $M'=P^{1+\delta}$, $a_m=\mathbb{1}_{\mathbb{P}}(m)\chi(m)$, $b_n=\mathbb{1}_{\mathbb{P}}(n)\chi(n)$. This gives

$$F(s,\chi)=\sum_{v\in I\cap\mathbb{Z}}G_v(s,\chi)H_v(s,\chi)+\sum_{\substack{k\in[Xe^{-1/U},Xe^{1/U}]\\ \text{or } k\in[2X,2Xe^{1/U}]}}\frac{d_k\chi(k)}{k^s},$$

where $I = [U \log P, (1+\delta)U \log P]$, $U := Q_0^{1+\varepsilon^2}$ and

$$|d_k| \leq \sum_{\substack{k=p_1p_2 \\ P<p_1\leq P^{1+\delta}}} 1,$$

which is bounded.

Therefore, taking the maximum in the sum over I, the mean square of the Dirichlet polynomial is bounded by

$$\begin{aligned}
\int_{T_0}^{T} |F(1+it,\chi)|^2 dt \ll & \int_{T_0}^{T} \left| \sum_{v\in I\cap\mathbb{Z}} G_v(1+it,\chi)H_v(1+it,\chi) \right|^2 dt \\
& + \int_{T_0}^{T} \left| \sum_{\substack{k\in[Xe^{-1/U}, Xe^{1/U}] \\ \text{or } k\in[2X, 2Xe^{1/U}]}} \frac{d_k\chi(k)}{k^{1+it}} \right|^2 dt \\
\ll & |I|^2 \int_{T_0}^{T} |G_{v_0}(1+it,\chi)|^2 |H_{v_0}(1+it,\chi)|^2 dt \\
& + \int_{T_0}^{T} \left| \sum_{\substack{k\in[Xe^{-1/U}, Xe^{1/U}] \\ \text{or } k\in[2X, 2Xe^{1/U}]}} \frac{d_k\chi(k)}{k^{1+it}} \right|^2 dt,
\end{aligned}$$

where $v_0 \in I$ is the integer maximising the right hand side. Applying

Lemma 2.3.28 to the second integral, we have that

$$
\begin{aligned}
&\sum_{\chi(q)} \int_{T_0}^{T} |F(1+it,\chi)|^2 dt \qquad (2.3.25)\\
\ll & U^2 (\log P)^2 \sum_{\chi(q)} \int_{T_0}^{T} |G_{v_0}(1+it,\chi)|^2 |H_{v_0}(1+it,\chi)|^2 dt\\
&+ T\varphi(q) \sum_{\substack{k\in[Xe^{-1/U},Xe^{1/U}] \\ \text{or } k\in[2X,2Xe^{1/U}] \\ (k,q)=1}} \frac{|d_k|^2}{k^2} \qquad (2.3.26)\\
&+ T\varphi(q) \sum_{\substack{1\le h\le \frac{2Xe^{1/U}}{T} \\ q|h}} \sum_{\substack{m-n=h \\ m,n\in[Xe^{-1/U},Xe^{1/U}] \\ \text{or } m,n\in[2X,2Xe^{1/U}] \\ (mn,q)=1}} \frac{|d_m||d_n|}{mn}. \qquad (2.3.27)
\end{aligned}
$$

We now bound the last two terms. We consider only the sums where $k \in [Xe^{-1/U}, Xe^{1/U}]$, with the sums over $k \in [2X, 2Xe^{1/U}]$ being treated analogously. For the first sum (2.3.26), we have

$$
\sum_{\substack{k=p_1p_2 \\ Xe^{-1/U}\le k\le Xe^{1/U} \\ P<p_1\le P^{1+\delta}}} \frac{1}{k^2} \ll \frac{e^{2/U}}{X^2} \sum_{P<p_1\le P^{1+\delta}} \sum_{\frac{Xe^{-1/U}}{p_1}\le p_2\le \frac{Xe^{1/U}}{p_1}} 1. \qquad (2.3.28)
$$

By the Brun-Titchmarsh inequality (Lemma 2.2.10), we have the bound

$$
\sum_{\frac{Xe^{-1/U}}{p_1}\le p_2\le \frac{Xe^{1/U}}{p_1}} 1 \ll \frac{X(e^{1/U}-e^{-1/U})}{p_1 \log X}.
$$

Returning to (2.3.28), by Mertens' theorem (Lemma 2.2.2) we have that

$$\sum_{\substack{Xe^{-1/U}\leq p_1p_2\leq Xe^{1/U}\\ P<p_1\leq P^{1+\delta}}} \frac{1}{(p_1p_2)^2} \ll \frac{e^{3/U}-e^{1/U}}{X\log X}\sum_{P<p\leq P^{1+\delta}}\frac{1}{p} \ll \frac{1}{XU\log X}. \quad (2.3.29)$$

We will use Brun's sieve to bound the second of these sums (2.3.27). We may trivially bound

$$\begin{aligned}&|\{n\leq 2X : n=p_1p_2, p_1\in(P,P^{1+\delta}]\}|\\ &\ll |\{n\leq 2X : n=p_1m, p_1\in(P,P^{1+\delta}], (m,P(z))=1\}|\end{aligned}$$

where we define $P(z)=\prod_{p<z}p$ with $z=X^{1/\beta}$ and $\beta>1$ suitably large. Let Π be the product of all primes in $\tilde{I}:=(P,P^{1+\delta}]\cap[1,z)$ and $P'(z)=\prod_{p<z,p\nmid h}p$. Therefore, we have that

$$\begin{aligned}&\sum_{\substack{1\leq h\leq\frac{2Xe^{1/U}}{T}\\ q|h}}\ \sum_{Xe^{-1/U}\leq m\leq Xe^{1/U}}\frac{|d_m||d_{m+h}|}{m(m+h)}\\ &\ll \frac{e^{2/U}}{X^2}\sum_{\substack{1\leq h\leq\frac{2Xe^{1/U}}{T}\\ q|h}}\left|\left\{m\in[Xe^{-1/U},Xe^{1/U}] : \left(m(m+h),\frac{P'(z)}{\Pi}\right)=1\right\}\right|.\end{aligned}$$

Brun's sieve then gives the bound

$$\begin{aligned}
&\left|\left\{m \in [Xe^{-1/U}, Xe^{1/U}] : \left(m(m+h), \frac{P'(z)}{\Pi}\right) = 1\right\}\right| \\
&\ll X(e^{1/U} - e^{-1/U})\frac{h}{\varphi(h)} \prod_{\substack{p<z \\ p \notin \tilde{I}}} \left(1 - \frac{2}{p}\right) \\
&\ll \frac{X}{U \log^2 z} \frac{h}{\varphi(h)}.
\end{aligned}$$

Therefore we have that

$$\begin{aligned}
&\sum_{\substack{1 \le h \le \frac{2Xe^{1/U}}{T} \\ q|h}} \sum_{Xe^{-1/U} \le m \le Xe^{1/U}} \frac{|d_m||d_{m+h}|}{m(m+h)} \\
&\ll \frac{e^{2/U}}{XU \log^2 z} \sum_{\substack{1 \le h \le \frac{2Xe^{1/U}}{T} \\ q|h}} \frac{h}{\varphi(h)} \qquad (2.3.30) \\
&\ll \frac{1}{\varphi(q) TU \log^2 X}.
\end{aligned}$$

Combining the diagonal (2.3.29) and off-diagonal (2.3.30) estimates and applying the definition $U := Q_0^{1+\varepsilon^2}$, the two terms (2.3.26) and (2.3.27) contribute

$$\ll \frac{q}{\varphi(q)} \left(\frac{T\varphi(q)}{XU \log X} + \frac{1}{U \log^2 X}\right) \ll \left(\frac{\varphi(q) T \log X}{X} + 1\right) \frac{1}{Q_0 (\log X)^{2+\eta}}$$

to (2.3.25) for some $\eta = \eta(\varepsilon) > 0$, as needed. □

It remains to estimate the integral appearing in Lemma 2.3.34. We split the domain of integration $[T_0, T]$ according to the size of the polynomial

$$G_{v_0}(1+it,\chi) := \sum_{e^{\frac{v_0}{U}} < p \leq e^{\frac{v_0+1}{U}}} \frac{\chi(p)}{p^{1+it}}.$$

We will first treat the set where G_{v_0} is pointwise small:

$$\mathcal{S}_1 := \{(t,\chi) \in [T_0,T] \times \{\chi \mod q\} : |G_{v_0}(1+it,\chi)| \leq e^{-\frac{\alpha_1 v_0}{U}}\},$$

where $\alpha_1 := \frac{3}{34} - \varepsilon'$ and $\varepsilon' > 0$ is sufficiently small in terms of $\varepsilon > 0$. We may write

$$\mathcal{S}_1 = \bigcup_{\chi \mod q} \{\chi\} \times \mathcal{T}_{1,\chi} \tag{2.3.31}$$

for some $\mathcal{T}_{1,\chi} \subset [T_0, T]$.

2.3.6.1 The contribution of $\mathcal{S}_1$

We first treat the contribution of the integral over $\mathcal{T}_{1,\chi}$, where the polynomial $G_{v_0}(1+it,\chi)$ is pointwise small. In this case, we apply the pointwise bound for G_{v_0} to find the required cancellation, before applying the mean value theorem to the longer Dirichlet polynomial

$$H_{v_0}(1+it,\chi) := \sum_{Xe^{-\frac{v_0}{U}} < p \leq 2Xe^{-\frac{v_0}{U}}} \frac{\chi(p)}{p^{1+it}},$$

which corresponds to the larger prime factor.

Lemma 2.3.36. *Let $\varepsilon > 0$ be fixed sufficiently small and $\mathcal{T}_{1,\chi}$ be defined as* *in* (2.3.31)*. Let* $v_0 \in I = [U \log P, (1+\delta)U \log P]$ *be a suitable integer as in Lemma 2.3.34, with* $U := Q_0^{1+\varepsilon^2}$*. Then, there exists some* $\eta = \eta(\varepsilon) > 0$ *such that*

$$U^2(\log P)^2 \sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{\mathcal{T}_{1,\chi}} |G_{v_0}(1+it)|^2 |H_{v_0}(1+it)|^2 dt \ll \frac{1}{Q_0(\log X)^{2+\eta}} \sum_{q \le Q_0} \left(\frac{qPT \log X}{X} + 1 \right).$$

Remark 2.3.37. We note that if the smaller prime factor satisfies $p_1 \in [P_1, P_2]$ for some $P_1 = P_1(X)$ and $P_2 = P_2(X)$, we have the bound

$$\frac{U^2 P_2^{-2\alpha_1} \log^2 P_2}{\log^2 X} \sum_{q \le Q_0} \left(\frac{qP_2T \log X}{X} + \frac{q}{\varphi(q)} \right).$$

In the proofs of Theorems 1.0.4 and 1.0.5, we will make larger choices of Q_0, U, P_1 and P_2, which leads to the improved error term. In particular, we will choose $Q_0 = \log^{A'} X$, $U = Q_0^E$ for some suitable $A', E > 0$, $P_1 = \exp((\log X)^{o(1)})$ and $P_2 = \exp((\log X)^{1-o(1)})$.

Proof. First we apply the definition of $\mathcal{T}_{1,\chi}$, bounding pointwise

$$|G_{v_0}(1+it,\chi)| \le e^{-\frac{\alpha_1 v_0}{U}} \le P^{-\alpha_1},$$

where $\alpha_1 := \frac{3}{34} - \varepsilon'$ and $\varepsilon' > 0$ sufficiently small in terms of $\varepsilon > 0$. We can then bound the integral over $\mathcal{T}_{1,\chi}$ by

$$\int_{\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)|^2 |H_{v_0}(1+it,\chi)|^2 dt \ll P^{-2\alpha_1} \int_{\mathcal{T}_{1,\chi}} |H_{v_0}(1+it,\chi)|^2 dt.$$

Applying Lemma 2.3.28, we have that

$$\begin{aligned}
&\sum_{\chi(q)} \int_{\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)|^2 |H_{v_0}(1+it,\chi)|^2 dt \\
&\quad \ll P^{-2\alpha_1} T\varphi(q) \left(\sum_{\frac{X}{e^{v_0/U}} < p \le \frac{2X}{e^{v_0/U}}} \frac{1}{p^2} + \sum_{\substack{1 \le h \le \frac{X}{Te^{v_0/U}} \\ q|h}} \sum_{\substack{\frac{X}{e^{v_0/U}} < p,p' \le \frac{2X}{e^{v_0/U}} \\ p-p'=h}} \frac{1}{pp'} \right) \\
&\quad \ll \frac{P^{-2\alpha_1} T\varphi(q) e^{2v_0/U}}{X^2} \left(\sum_{\frac{X}{e^{v_0/U}} < p \le \frac{2X}{e^{v_0/U}}} 1 + \sum_{\substack{1 \le h \le \frac{X}{Te^{v_0/U}} \\ q|h}} \sum_{\substack{\frac{X}{e^{v_0/U}} < p_1,p_2 \le \frac{2X}{e^{v_0/U}} \\ p_1-p_2=h}} 1 \right).
\end{aligned}$$

By Chebyshev's inequality, we have that the diagonal terms are bounded by

$$\sum_{\frac{X}{e^{v_0/U}} < p \le \frac{2X}{e^{v_0/U}}} 1 \ll \frac{X}{e^{v_0/U} \log X}. \tag{2.3.32}$$

For the off-diagonal terms, we have by Lemma 2.2.3 that

$$\begin{aligned}
\sum_{\substack{1 \le h \le \frac{X}{Te^{v_0/U}} \\ q|h}} \sum_{\substack{\frac{X}{e^{v_0/U}} < p_1,p_2 \le \frac{2X}{e^{v_0/U}} \\ p_1-p_2=h}} 1 &\ll \frac{X}{e^{v_0/U} \log^2 X} \sum_{\substack{1 \le h \le \frac{X}{Te^{v_0/U}} \\ q|h}} \mathfrak{S}(h) \\
&\ll \frac{X^2}{e^{2v_0/U} qT \log^2 X}.
\end{aligned} \tag{2.3.33}$$

Combining the diagonal (2.3.32) and off-diagonal (2.3.33) estimates, we have

that

$$\begin{aligned}\sum_{\chi(q)}\int_{\mathcal{T}_{1,\chi}} &|G_{v_0}(1+it,\chi)|^2|H_{v_0}(1+it,\chi)|^2dt\\ &\ll \frac{P^{-2\alpha_1}T\varphi(q)e^{2v_0/U}}{X^2}\left(\frac{X}{e^{v_0/U}\log X}+\frac{X^2}{e^{2v_0/U}qT\log^2 X}\right)\\ &\ll \frac{\varphi(q)P^{\delta-2\alpha_1}}{q\log^2 X}\left(\frac{qPT\log X}{X}+1\right).\end{aligned}$$

Thus the overall contribution to the sum $B_3(X)$ is

$$\begin{aligned}&U^2(\log P)^2\sum_{q\le Q_0}\frac{q}{\varphi(q)}\frac{\varphi(q)P^{\delta-2\alpha_1}}{q\log^2 X}\left(\frac{qPT\log X}{X}+1\right)\\ =&\frac{U^2P^{\delta-2\alpha_1}\log^2 P}{\log^2 X}\sum_{q\le Q_0}\left(\frac{qPT\log X}{X}+1\right).\end{aligned}$$

Now, to estimate the first term we split into cases depending on the size of H, as this determines the size of Q_0 and P.

In the case $H\le\exp((\log X)^{\varepsilon^3})$, we have that $Q_0:=(\log X)^{1+\varepsilon^2}$ and $P:=(\log X)^{17+\varepsilon}$. By the choice of $U:=Q_0^{1+\varepsilon^2}$ and the definition of $\alpha_1:=\frac{3}{34}-\varepsilon'$ with ε' sufficiently small in terms of ε, we have that

$$\begin{aligned}\frac{U^2P^{\delta-2\alpha_1}\log^2 P}{\log^2 X} &\ll \frac{(\log X)^{(2+2\varepsilon^2)(1+\varepsilon^2)+(\delta-3/17+2\varepsilon')(17+\varepsilon)}(\log\log X)^2}{\log^2 X}\\ &\ll \frac{1}{Q_0(\log X)^{2+\eta}},\end{aligned}$$

for some $\eta=\eta(\varepsilon)>0$ as $\delta>0$ is sufficiently small.

In the case $H>\exp((\log X)^{\varepsilon^3})$, we have that $Q_0:=(\log X)^{3+\varepsilon^2}$ and

$P := \exp((\log\log X)^2)$, so that

$$\begin{aligned}&\frac{U^2 P^{\delta-2\alpha_1}\log^2 P}{\log^2 X}\\ &\ll \frac{(\log X)^{(2+2\varepsilon^2)(3+\varepsilon^2)}\exp((\delta-2\alpha_1)(\log\log X)^2)(\log\log X)^4}{\log^2 X}\\ &\ll \frac{1}{Q_0(\log X)^{2+\eta}},\end{aligned}$$

for some $\eta = \eta(\varepsilon) > 0$ as $\delta > 0$ is sufficiently small. □

2.3.6.2 The contribution of the complement of $\mathcal{S}_1$

It remains to consider the contribution of the complement of $\mathcal{S}_1$, where the polynomial

$$G_{v_0}(1+it,\chi) := \sum_{e^{\frac{v_0}{U}} < p \le e^{\frac{v_0+1}{U}}} \frac{\chi(p)}{p^{1+it}}$$

is pointwise large. The polynomial

$$H_{v_0}(1+it,\chi) := \sum_{Xe^{-\frac{v_0}{U}} < p \le 2Xe^{-\frac{v_0}{U}}} \frac{\chi(p)}{p^{1+it}}$$

is too long to find the cancellation we need, so we introduce further decomposition into shorter polynomials using Heath-Brown's identity (Lemma 2.3.26) with $k = 3$. This decomposes the polynomial H_{v_0} into

$$|H_{v_0}(1+it,\chi)| \ll (\log^C X)\left(|Q_1(1+it,\chi)| + \cdots + |Q_L(1+it,\chi)|\right),$$

where $L \leq \log^C X$ for some $C > 0$. Each $Q_j(s, \chi)$ is of the form $Q_j(s, \chi) = \prod_{i \leq J_j} R_i(s, \chi)$ with $J_j \leq 6$ for each $1 \leq j \leq L$, where $R_i(s, \chi)$ are prime-factored Dirichlet polynomials of the form

$$\sum_{R_i < n \leq 2R_i} \frac{\chi(n) \log n}{n^s}, \quad \sum_{R_i < n \leq 2R_i} \frac{\chi(n)}{n^s}, \text{ or } \sum_{R_i < n \leq 2R_i} \frac{\mu(n)\chi(n)}{n^s},$$

whose lengths satisfy $R_1 \cdots R_{J_j} = X^{1+o(1)}$, $R_i \gg \exp\left(\frac{\log X}{\log \log X}\right)$ for each i. These polynomials $Q_j(s, \chi)$ are either a product of many shorter prime-factored polynomials, or a product of two longer polynomials which are partial sums of a Dirichlet L-function (or its derivative).

We will treat each type of $Q_j(s, \chi)$ with different methods, so we now split into two cases according to the the lengths R_i of the factors as follows:

Case 1: Type II Sums. When $Q_j(s, \chi)$ is a product of many prime-factored polynomials, this has arisen from having many localised summation variables, so we will describe this case as Type II.

Suppose we have $Q_j(s, \chi) = \prod_{i \leq J_j} R_i(s, \chi)$ for some $1 \leq j \leq L$ with $R_i \leq X^{1/3+\varepsilon'}$ for some $i \leq J_j \leq 6$. Then, we rewrite $Q_j(s, \chi) = M_1(s, \chi) M_2(s, \chi)$ with $\exp\left(\frac{\log X}{\log \log X}\right) \ll M_1 \leq X^{1/3+\varepsilon'}$ and $M_2 = X^{1+o(1)}/M_1$. Where the coefficient $\log n$ appears, we apply partial summation. The polynomial $M_2(s, \chi)$ is a product of polynomials, and the coefficients are given by convolving coefficients which are one of the sequences $(\mu(n)), (1)$. Thus the coefficients of the polynomial $M_2(s, \chi)$ are bounded in absolute value by $\ll d_r(n)$ with

$r \leq 5$. We will therefore need to find sufficient cancellation in

$$\sum_{q \leq Q_0} \frac{q}{\varphi(q)} \int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)M_1(1+it,\chi)M_2(1+it,\chi)|^2 dt.$$

In this case, we once again have a short polynomial $M_1(s,\chi)$ to work with. So, as before, we split the domain of integration according to whether $M_1(1+it,\chi)$ is pointwise small. When $M_1(1+it,\chi)$ is pointwise small, we will apply this pointwise bound. Previously, for the set $\mathcal{S}_1$ where the shorter polynomial $G_{v_0}(1+it,\chi)$ was pointwise small, we were able to find the required cancellation by applying the mean value theorem to the remaining longer polynomial $H_{v_0}(1+it,\chi)$. However, $M_1(s,\chi)$ can be much longer than $G_{v_0}(s,\chi)$, so following the same strategy as for the set $\mathcal{S}_1$ and applying the mean value theorem to $G_{v_0}M_2(1+it,\chi)$ will not be enough.

To overcome this issue and sufficiently increase the length of this polynomial, we introduce a suitable $2(\ell-1)$th moment of $G_{v_0}(1+it,\chi)$. In this domain we have $|G_{v_0}(1+it,\chi)P^{\alpha_1}|^{2(\ell-1)} \geq 1$, as this polynomial is large. Now, the polynomial $G_{v_0}^{\ell-1}M_2(1+it,\chi)$ has length which is comparable to the length of integration, and the mean value theorem can be applied effectively.

Otherwise, where the polynomial $M_1(1+it,\chi)$ is pointwise large we will apply the prime-factored bound. This set will be sparse, so we will apply the Halász-Montgomery inequality followed by large value theorems.

Case 2: Type I Sums. Otherwise, $Q_j(s,\chi)$ is a product of two longer polynomials which are partial sums of Dirichlet L-functions (or derivatives).

These polynomials arise from larger variables in the Heath-Brown identity, so we will describe this case as Type I.

We may write $Q_j(s,\chi) = N_1(s,\chi)N_2(s,\chi)$, where each $N_i(s,\chi)$ is of the form

$$\sum_{N_i<n\leq 2N_i} \frac{\chi(n)\log n}{n^s}, \text{ or } \sum_{N_i<n\leq 2N_i} \frac{\chi(n)}{n^s},$$

with lengths satisfying $N_1N_2 = X^{1+o(1)}$. Note that if in fact only one of the lengths N_i satisfies $N_i > X^{1/3+\varepsilon'}$, then one of $N_1(s,\chi), N_2(s,\chi)$ can be the constant polynomial 1^{-s}. Since we have that $N_1N_2 = X^{1+o(1)}$, without loss of generality we may take that $N_1 > X^{1/2-\varepsilon'}$, so that $X^{1/3+\varepsilon'} < N_2 \leq X^{1/2+\varepsilon'}$.

In this case, we will therefore need to find sufficient cancellation in

$$\sum_{q\leq Q_0} \frac{q}{\varphi(q)} \int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)N_1(1+it,\chi)N_2(1+it,\chi)|^2 dt.$$

We will again introduce a suitable $2(l-1)$th moment of $G_{v_0}(1+it,\chi)$ using $|G_{v_0}(1+it,\chi)P^{\alpha_1}|^{2(l-1)} \geq 1$ to ensure that the polynomial $G_{v_0}^{l-1}N_1(1+it,\chi)$ is sufficiently long. We will then introduce fourth moments and separate the polynomials $G_{v_0}^{l-1}N_1(1+it,\chi)$ and $N_2(1+it,\chi)$ using the Cauchy-Schwarz inequality. To find the required cancellation we apply the mean value theorem to $N_2^2(1+it,\chi)$ and the twisted fourth moment result for partial sums of Dirichlet L-functions (Lemma 2.3.33).

2.3.6.2.1 Type II Sums. We first handle the contribution of the Type II sums, that is, the prime-factored polynomials

$$M_1(1+it,\chi) = \sum_{M_1<m\le 2M_1} \frac{a_m\chi(m)}{m^{1+it}}, \qquad M_2(1+it,\chi) = \sum_{M_2<n\le 2M_2} \frac{b_n\chi(n)}{n^{1+it}},$$

with a_m either the coefficient 1 or the Möbius function, $|b_n| \le d_5(n)$, and lengths satisfying $\exp\left(\frac{\log X}{\log\log X}\right) \ll M_1 \le X^{1/3+\varepsilon'}$, $M_1M_2 = X^{1+o(1)}$.

We need to find cancellation in

$$\sum_{q\le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)M_1(1+it,\chi)M_2(1+it,\chi)|^2 dt.$$

To treat this contribution, we split the complement of $\mathcal{S}_1$ according to the size of the shorter polynomial $M_1(1+it,\chi)$:

$$\mathcal{S}_2 := \{(t,\chi) \in [T_0,T] \times \{\chi \mod q\} : |M_1(1+it,\chi)| \le M_1^{-\alpha_2}\} \setminus \mathcal{S}_1,$$
$$\mathcal{S} := ([T_0,T] \times \{\chi \mod q\}) \setminus (\mathcal{S}_1 \cup \mathcal{S}_2),$$

with $\alpha_2 := \frac{2}{17} - \varepsilon' > \alpha_1$. As before, we may write

$$\mathcal{S}_2 = \bigcup_{\chi \mod q} \{\chi\} \times \mathcal{T}_{2,\chi}, \qquad \mathcal{S} = \bigcup_{\chi \mod q} \{\chi\} \times \mathcal{T}_{\chi}, \tag{2.3.34}$$

for some $\mathcal{T}_{2,\chi}, \mathcal{T}_\chi \subset [T_0,T]$.

We first consider the contribution of the integral over $\mathcal{T}_{2,\chi}$. In this set, the polynomial $G_{v_0}(1+it,\chi)$ is large while the polynomial $M_1(1+it,\chi)$ is small. We will pointwise bound the polynomial $M_1(1+it,\chi)$ to find some cancellation. We will then introduce a suitable $2(\ell-1)$-th moment $|G_{v_0}(1+it,\chi)P^{\alpha_1}|^{2(\ell-1)} \geq 1$, which will ensure the polynomial $G_{v_0}^{2(\ell-1)}M_1(1+it,\chi)$ is of a length similar to the length of integration and we can effectively apply the mean value theorem.

Lemma 2.3.38. *Let $\varepsilon > 0$ be fixed sufficiently small. Let $\mathcal{T}_{2,\chi}$ be defined as in* (2.3.34). *For $s \in \mathbb{C}$ and $\chi \mod q$, define $M_1(s,\chi), M_2(s,\chi)$ to be prime-factored polynomials*

$$M_1(s,\chi) = \sum_{M_1 < m \leq 2M_1} \frac{a_m \chi(m)}{m^s}, \qquad M_2(s,\chi) = \sum_{M_2 < n \leq 2M_2} \frac{b_n \chi(n)}{n^s},$$

with a_m either the coefficient 1 or the Möbius function, $|b_n| \leq d_5(n)$, and lengths satisfying $\exp\left(\frac{\log X}{\log\log X}\right) \ll M_1 \leq X^{1/3+\varepsilon'}$, $M_1 M_2 = X^{1+o(1)}$. *Then*

$$\sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |G_{v_0}(1+it,\chi)|^2 |M_1(1+it,\chi) M_2(1+it,\chi)|^2 dt \ll \log^{-F} X,$$

for some suitably large $F > 0$.

Proof. This proof is similar to [25, Lemma 13] and [38, Proposition 2]. By definition of $\mathcal{T}_{2,\chi}$, we have that

$$(|G_{v_0}(1+it,\chi)|P^{\alpha_1})^{2(\ell-1)} \geq 1$$

for integers $\ell > 0$. We choose $\ell = \lceil \log M_1 / \log P \rceil$. Therefore, we have

$$\sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |G_{v_0}(1+it,\chi)|^2 |M_1(1+it,\chi) M_2(1+it,\chi)|^2 dt$$
$$\ll M_1^{-2\alpha_2} P^{2\alpha_1(\ell-1)} \sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |G_{v_0}^{\ell}(1+it,\chi) M_2(1+it,\chi)|^2 dt. \quad (2.3.35)$$

By the choice of ℓ, we have that

$$P^{2\alpha_1 \ell} \ll \exp\left(\frac{2\alpha_1 \log P \log M_1}{\log P}\right) = M_1^{2\alpha_1}.$$

Therefore (2.3.35) is bounded by

$$\ll M_1^{2\alpha_1 - 2\alpha_2} P^{-2\alpha_1} \sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |G_{v_0}^{\ell}(1+it,\chi) M_2(1+it,\chi)|^2 dt$$
$$\ll M_1^{2\alpha_1 - 2\alpha_2} P^{-2\alpha_1} \sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |A(1+it,\chi)|^2 dt,$$

where we define

$$A(s,\chi) := \sum_{n \in J} \frac{A_n \chi(n)}{n^s},$$

with $J := (M_2 e^{\ell v_0/U}, 2M_2 e^{\ell(v_0+1)/U}]$ and the coefficients A_n satisfying

$$|A_n| \leq \sum_{\substack{n = p_1 \cdots p_\ell m \\ e^{v_0/U} < p_i \leq e^{(v_0+1)/U} \\ i = 1, \ldots, \ell \\ M_2 < m \leq 2M_2}} d_r(m),$$

where $r \leq 5$, as before. Note that the primes $p_1, \ldots, p_\ell$ are not necessarily distinct and m may also have prime factors in the range $(e^{v_0/U}, e^{(v_0+1)/U}]$. Applying Lemma 2.3.27 to the integral, we have that

$$\sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |A(1+it,\chi)|^2 dt \ll \left(\varphi(q)T + \frac{\varphi(q)}{q} M_2 e^{\ell v_0/U}(2e^{\ell/U}-1)\right) \sum_{\substack{n\in J \\ (n,q)=1}} \frac{|A_n|^2}{n^2}.$$

The number of ways we can write $d = p_1 \cdots p_\ell$ with p_i not necessarily distinct is at most $\ell!$. Then we have the bound

$$|A_n| \ll \ell! \sum_{\substack{n=md \\ p|d \Rightarrow e^{v_0/U}<p\le e^{(v_0+1)/U} \\ \Omega(d)>0}} d_r(m) \ll \ell! d_{r+1}(n).$$

trivially extending the range of summation for m. Therefore, we have that

$$\begin{aligned}
\sum_{\substack{n\in J \\ (n,q)=1}} \frac{|A_n|^2}{n^2} &\ll \frac{\ell!}{e^{kv_0/U}M_2} \sum_{\substack{n\in J \\ (n,q)=1}} \frac{|A_n| d_{r+1}(n)}{n} \\
&\ll \frac{\ell!}{e^{\ell v_0/U}M_2} \sum_{M_2<m\le 2M_2} \frac{d_5(m)d_6(m)}{m} \sum_{\substack{e^{v_0/U}<p_i\le e^{(v_0+1)/U} \\ i=1,\dots,\ell}} \frac{d_6(p_1\cdots p_\ell)}{p_1\cdots p_\ell} \\
&\ll \frac{\ell!}{e^{\ell v_0/U}M_2} \sum_{M_2<m\le 2M_2} \frac{d_6^2(m)}{m} \left(\sum_{e^{v_0/U}<p_i\le e^{(v_0+1)/U}} \frac{6}{p} \right)^{\ell} \\
&\ll \frac{\ell!}{e^{\ell v_0/U}M_2} (\log M_2)^{35} \left(6 \log\left(1+\frac{1}{v_0}\right)\right)^{\ell} \\
&\ll \frac{M_1^{\frac{\log 6}{\log P}} (\log X)^{35} \ell!}{e^{\ell v_0/U}M_2} \frac{\ell}{v_0},
\end{aligned}$$

noting that by the definition of ℓ we have $6^\ell \ll \exp(\frac{(\log 6)(\log M_1)}{\log P}) \ll M_1^{\frac{\log 6}{\log P}}$. By the definition of ℓ and since $v_0 \in I$, we have that

$$\frac{\ell}{v_0} \ll \frac{\log M_1}{U \log^2 P} \ll 1.$$

By the definition of ℓ, we also have that

$$\ell! \ll (\log M_1)^{\frac{\log M_1}{\log P}} \ll \exp\left(\frac{\log\log M_1 \log M_1}{(17+\varepsilon)\log\log X}\right) \ll M_1^{1/(17+\varepsilon)}.$$

Therefore, we can bound the integral over $\mathcal{T}_{2,\chi}$ by

$$\begin{aligned}&\sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |A(1+it,\chi)|^2 dt \\ &\ll \left(\varphi(q)\frac{T}{e^{\ell v_0/U} M_2} + \frac{\varphi(q)}{q}(2e^{\ell/U}-1)\right) M_1^{\frac{1}{17+\varepsilon}+\frac{\log 6}{\log P}} (\log X)^{35}.\end{aligned}$$

Since $v_0 \in I$, we have that $e^{\ell v_0/U} \gg P^\ell \gg M_1$ by the definition of ℓ. We also have that $2e^{\ell/U} - 1 \ll 1$ and therefore we can bound the above integral by

$$\begin{aligned}\sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |A(1+it,\chi)|^2 dt &\ll \left(\varphi(q)\frac{T}{M_1 M_2} + \frac{\varphi(q)}{q}\right) M_1^{\frac{1}{17+\varepsilon}+\frac{\log 6}{\log P}} (\log X)^{35} \\ &\ll \varphi(q) M_1^{\frac{1}{17+\varepsilon}+\frac{\log 6}{\log P}} (\log X)^{35},\end{aligned}$$

as we have $M_1 M_2 = X^{1+o(1)}$ and $T \leq X^{1+o(1)}$. Returning to (2.3.35), we have the bound

$$\sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |G_{v_0}(1+it,\chi)|^2 |M_1(1+it,\chi)M_2(1+it,\chi)|^2 dt$$
$$\ll \varphi(q) M_1^{\frac{1}{17+\varepsilon}+\frac{\log 6}{\log P}+2\alpha_1-2\alpha_2} P^{-2\alpha_1} (\log X)^{35}.$$

With our choices of α_1, α_2, we have that $2\alpha_1 - 2\alpha_2 = -1/17$. Summing over q introduces a factor of Q_0^2. Recalling that $\exp\left(\frac{\log X}{\log\log X}\right) \ll M_1$, we find that

$$\sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{\mathcal{T}_{2,\chi}} |G_{v_0}(1+it,\chi)|^2 |M_1(1+it,\chi)M_2(1+it,\chi)|^2 dt$$
$$\ll Q_0^2 P^{-2\alpha_1} X^{-\varepsilon/(300\log\log X)} (\log X)^{35},$$

and choosing $\varepsilon' > 0$ sufficiently small in terms of $\varepsilon > 0$ ensures the above is bounded by $\log^{-F} X$ for some suitable $F > 0$, as needed. $\square$

For the Type II case, it remains to treat the contribution of the integral over $\mathcal{T}_\chi$, where both polynomials

$$G_{v_0}(1+it,\chi) := \sum_{e^{\frac{v_0}{U}} < p \le e^{\frac{v_0+1}{U}}} \frac{\chi(p)}{p^{1+it}}$$
$$M_1(1+it,\chi) = \sum_{M_1 < n \le 2M_1} \frac{\chi(n)}{n^{1+it}}, \text{ or } \sum_{M_1 < n \le 2M_1} \frac{\mu(n)\chi(n)}{n^{1+it}},$$

are large. To handle this contribution, we will first apply the Halász - Montgomery inequality as we are in a sparse set, followed by large value theorems.

Lemma 2.3.39. *Let $\mathcal{T}_\chi$ be defined as in* (2.3.34). *Let $E > 0$ be fixed suf-*

ficiently large. For $s \in \mathbb{C}$ and $\chi \mod q$, define $M_1(s,\chi), M_2(s,\chi)$ to be prime-factored polynomials

$$M_1(s,\chi) = \sum_{M_1 < m \le 2M_1} \frac{a_m \chi(m)}{m^s}, \qquad M_2(s,\chi) = \sum_{M_2 < n \le 2M_2} \frac{b_n \chi(n)}{n^s},$$

with a_m either the coefficient 1 or the Möbius function, $|b_n| \le d_5(n)$, and lengths satisfying $\exp\left(\frac{\log X}{\log\log X}\right) \ll M_1 \le X^{1/3+\varepsilon'}$, $M_1 M_2 = X^{1+o(1)}$. Then, we have that

$$\sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{\mathcal{T}_\chi} |G_{v_0}(1+it,\chi) M_1(1+it,\chi) M_2(1+it,\chi)|^2 dt \ll \frac{1}{\log^E X},$$

where $v_0 \in I$ is a suitable integer, with $I = [U \log P, (1+\delta) U \log P]$ and $U := Q_0^{1+\varepsilon^2}$.

Proof. This is similar to [38, Section 4.1]. We first replace the integral over $\mathcal{T}_\chi$ with a sum over a well-spaced set. For each character χ mod q, cover $\mathcal{T}_\chi$ with intervals of unit length and from each interval take the point which maximises the integral over that interval. This set is not yet necessarily well-spaced, but we can split it into $O(1)$ well-spaced subsets. Therefore we may write

$$\sum_{\chi(q)} \int_{\mathcal{T}_\chi} |G_{v_0}(1+it,\chi) M_1(1+it,\chi) M_2(1+it,\chi)|^2 dt$$
$$\ll \sum_{(t,\chi) \in \mathcal{T}'} |G_{v_0}(1+it,\chi) M_1(1+it,\chi) M_2(1+it,\chi)|^2,$$

where $\mathcal{T}'$ is the well-spaced subset which maximises the right hand side. We now apply the prime-factored property $|M_1(1+it,\chi)|^2 \ll \log^{-F'} X$ with $F' > 0$ sufficiently large and then Lemma 2.3.29 to get that

$$\begin{aligned}
&\sum_{(t,\chi)\in\mathcal{T}'} |G_{v_0}(1+it,\chi)M_1(1+it,\chi)M_2(1+it,\chi)|^2 \\
&\ll (\log X)^{-F'} \sum_{(t,\chi)\in\mathcal{T}'} |G_{v_0}(1+it,\chi)M_2(1+it,\chi)|^2 \\
&\ll (\log X)^{-F'} \left(\frac{\varphi(q)}{q} M_2 e^{(v_0+1)/U} + |\mathcal{T}'|(qT)^{1/2}\right) \sum_{\substack{n=pm \\ e^{v_0/U}<p\le e^{(v_0+1)/U} \\ M_2<m\le 2M_2}} \frac{d_r^2(m)}{n^2} \\
&\ll (\log X)^{-F} \left(\frac{\varphi(q)}{q} e^{1/U} + \frac{|\mathcal{T}'|(qT)^{1/2}}{M_2 e^{v_0/U}}\right), \qquad (2.3.36)
\end{aligned}$$

where $F > 0$ is suitably large and $r \le 5$.

Case 1: $\exp\left(\frac{\log X}{\log\log X}\right) \ll M_1 \ll X^{\nu}$ for all $\nu > 0$.

In this case we apply Lemma 2.3.30 with $V = M_1^{-\alpha_2}$ to see that

$$|\mathcal{T}'| \ll (qT)^{2\alpha_2} M_1^{2\alpha_2} \exp\left((1+o(1))(\log\log X)^3\right) \ll X^{4/17-2\varepsilon'+\varepsilon},$$

since $\exp\left(\frac{\log X}{\log\log X}\right) \ll M_1 \ll X^{\nu}$ for all $\nu > 0$ and $\alpha_2 := \frac{2}{17} - \varepsilon'$. Substituting this bound into (2.3.36) and summing over q, in this case we have that

$$\begin{aligned}&\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{(t,\chi)\in\mathcal{T}'}|G_{v_0}(1+it,\chi)M_1(1+it,\chi)M_2(1+it,\chi)|^2\\ &\ll(\log X)^{-F}\sum_{q\le Q_0}\left(1+\frac{q^{1/2}X^{4/17+1/2+\varepsilon}}{M_2e^{v_0/U}}\right)\\ &\ll(\log X)^{-E'}\left(1+\frac{X^{4/17+1/2+\varepsilon+\nu}}{X^{1+o(1)}P}\right),\end{aligned}$$

noting that $M_2=X^{1+o(1)}/M_1\gg X^{1-\nu+o(1)}$ for any $\nu>0$ in this case. The above is then $\ll(\log X)^{-E'}$ for some suitable $E'>0$, as required.

Otherwise, we may write $M_1=X^{\nu+\varepsilon'}$ for some $0<\nu\le 1/3$. If we can show that $|\mathcal{T}'|\ll X^{1/2-\nu-\varepsilon^2}$, then we will have that

$$\sum_{\chi(q)}\int_{\mathcal{T}_\chi}|G_{v_0}(1+it,\chi)M_1(1+it,\chi)M_2(1+it,\chi)|^2dt\ll\frac{1}{\log^{E'}X}$$

for some suitable $E'>0$. Summing over q, we have that

$$\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{\mathcal{T}_\chi}|G_{v_0}(1+it,\chi)M_1(1+it,\chi)M_2(1+it,\chi)|^2dt\ll\frac{1}{\log^{E}X},$$

where $E>0$ is sufficiently large.

Thus, it remains to prove that $|\mathcal{T}'|\ll X^{1/2-\nu-\varepsilon^2}$. The large value theorem we apply to obtain the required bound will depend on the size of M_1.

Case 2: $\frac{17}{58}\le\nu\le\frac{1}{3}$.

We apply Lemma 2.3.32 with $M_1(1+it,\chi)^2, V=M_1^{-2\alpha_2}$ and $k=2$, we have that

$$\begin{aligned}|\mathcal{T}'| &\ll X^{\varepsilon^3}\left(M_1^{4\alpha_2} + X^2 M_1^{8(2\alpha_2-1)} + X M_1^{8(4\alpha_2-1)}\right)\\ &\ll X^{\max(\frac{8}{17}\nu, 2-\frac{104}{17}\nu, 1-\frac{72}{17}\nu)-2\varepsilon^2}.\end{aligned}$$

We have that $\nu \leq \frac{1}{3}+\varepsilon'$. The inequality $\frac{8}{17}\nu \leq \frac{1}{2}-\nu$ holds when $\nu \leq \frac{17}{50}$ and we have that $2-\frac{104}{17}\nu \geq \frac{8}{17}\nu$ when $\nu \leq \frac{17}{56}$. Note that $2-\frac{104}{17}\nu \leq \frac{1}{2}-\nu$ fails if $\nu < \frac{17}{58}$. Therefore, in this case we have the bound $|\mathcal{T}'| \ll X^{1/2-\nu-\varepsilon^2}$.

Case 3: $\frac{51}{278} \leq \nu \leq \frac{17}{58}$**.**

We again apply Lemma 2.3.32 with $M_1(1+it,\chi)^3, V = M_1^{-3\alpha_2}$ and $k=2$, to obtain

$$\begin{aligned}|\mathcal{T}'| &\ll X^{\varepsilon^3}\left(M_1^{6\alpha_2} + X^2 M_1^{12(2\alpha_2-1)} + X M_1^{12(4\alpha_2-1)}\right)\\ &\ll X^{\max(\frac{12}{17}\nu, 2-\frac{156}{17}\nu, 1-\frac{108}{17}\nu)-2\varepsilon^2}.\end{aligned}$$

Similar to the previous case, we have that $\frac{12}{17}\nu \leq \frac{1}{2}-\nu$ holds for $\nu \leq \frac{17}{58}$ and $2-\frac{156}{17}\nu \geq \frac{12}{17}\nu$ when $\nu \leq \frac{17}{84}$. We have that $2-\frac{156}{17}\nu \leq 1/2-\nu$ fails when $\nu < \frac{51}{278}$, so we again have the required bound for $|\mathcal{T}'|$ in this range.

Case 4: $\nu < \frac{51}{278}$**.**

Further iterations of Lemma 2.3.32 with higher powers of $M_1(1+it,\chi)$ are not enough to cover the full range of ν. In this case, ν is now small enough for us to apply Lemma 2.3.30 effectively with $V = M_1^{-\alpha_2}$, which gives

$$|\mathcal{T}'| \ll (qT)^{2\alpha_2} X^{2\nu\alpha_2+\varepsilon} \ll X^{\frac{4}{17}(1+\nu)+100\varepsilon} \ll X^{1/2-\nu-\varepsilon^2},$$

as required. $\square$

2.3.6.2.2 Type I Sums. Now that we have handled the Type II contribution, it remains to treat the contribution of polynomials which are partial sums of a Dirichlet L-function or its derivative. These polynomials are of the form

$$N_i(1+it,\chi) = \sum_{N_i<n\leq 2N_i} \frac{\chi(n)\log n}{n^{1+it}}, \text{ or } \sum_{N_i<n\leq 2N_i} \frac{\chi(n)}{n^{1+it}},$$

for $i = 1, 2$, with lengths satisfying $N_1N_2 = X^{1+o(1)}$ with $N_1 > X^{1/2-\varepsilon'}$ and $X^{1/3+\varepsilon'} < N_2 \leq X^{1/2+\varepsilon'}$.

We need to find cancellation in

$$\sum_{q\leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)N_1(1+it,\chi)N_2(1+it,\chi)|^2 dt.$$

To do this we will first introduce a suitable $2(l-1)$th moment of $G_{v_0}(1+it,\chi)$, using that $|G_{v_0}(1+it,\chi)P^{\alpha_1}|^{2(l-1)} \geq 1$ as we are in the set where $G_{v_0}(1+it,\chi)$ is pointwise large. This will ensure that the polynomial $G_{v_0}^{l-1}N_1$ is sufficiently long when we apply the twisted fourth moment of partial sums of Dirichlet L-functions (Lemma 2.3.33). We will apply the Cauchy-Schwarz inequality to separate the polynomials and introduce a fourth moment, so that we can then use this twisted fourth moment result.

Lemma 2.3.40. *Let $\varepsilon > 0$ be fixed sufficiently small. For $s \in \mathbb{C}$, $\chi \mod q$ and $i = 1, 2$, define*

$$N_i(s, \chi) = \sum_{N_i < n \leq 2N_i} \frac{\chi(n) \log n}{n^s}, \text{ or } \sum_{N_i < n \leq 2N_i} \frac{\chi(n)}{n^s},$$

with lengths satisfying $N_1 N_2 = X^{1+o(1)}$ and $N_1 > X^{1/2-\varepsilon'}, X^{1/3+\varepsilon'} < N_2 \leq X^{1/2+\varepsilon'}$. Let $\mathcal{T}_{1,\chi}$ be defined as in (2.3.31). Then, we have that

$$\sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{[T_0,T] \setminus \mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi) N_1(1+it,\chi) N_2(1+it,\chi)|^2 dt \ll X^{-\varepsilon/2},$$

where $v_0 \in I$ is a suitable integer with $I := [U \log P, (1+\delta) U \log P]$ and $U := Q_0^{1+\varepsilon^2}$.

Proof. This is similar to [38, Proposition 3]. As we are in the complement of $\mathcal{S}_1$, we have that

$$|G_{v_0}(1+it,\chi) P^{\alpha_1}|^{2(l-1)} \geq 1,$$

where we choose $l = \lfloor \varepsilon \log X / \log P \rfloor$. Therefore, we have that

$$\begin{aligned} &\sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{[T_0,T] \setminus \mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi) N_1(1+it,\chi) N_2(1+it,\chi)|^2 dt \\ &\ll P^{2\alpha_1(l-1)} \sum_{q \leq Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{T_0}^{T} |G_{v_0}^l(1+it,\chi) N_1(1+it,\chi) N_2(1+it,\chi)|^2 dt. \end{aligned}$$

We will split the domain of integration according to whether $t \leq N_1$. Applying the Cauchy-Schwarz inequality three times in total (to the integral and the sums over χ and q), we have that

$$\begin{aligned}
&P^{2\alpha_1(l-1)}\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{N_1}|G_{v_0}^l(1+it,\chi)N_1(1+it,\chi)N_2(1+it,\chi)|^2dt\\
\ll&P^{2\alpha_1(l-1)}\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\left(\sum_{\chi(q)}\int_{T_0}^{N_1}|G_{v_0}(1+it,\chi)|^{4l}|N_1(1+it,\chi)|^4dt\right)^{\frac{1}{2}}\\
&\times\left(\sum_{\chi(q)}\int_{T_0}^{N_1}|N_2(1+it,\chi)|^4dt\right)^{\frac{1}{2}}\\
\ll&P^{2\alpha_1(l-1)}\left(\sum_{q\leq Q_0}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{N_1}|G_{v_0}(1+it,\chi)|^{4l}|N_1(1+it,\chi)|^4dt\right)^{\frac{1}{2}}\\
&\times\left(\sum_{q\leq Q_0}\frac{q^2}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{N_1}|N_2(1+it,\chi)|^4dt\right)^{\frac{1}{2}}.
\end{aligned}\tag{2.3.37}$$

We have the analogous inequality

$$\begin{aligned}
&P^{2\alpha_1(l-1)}\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{N_1}^{T}|G_{v_0}^l(1+it,\chi)N_1(1+it,\chi)N_2(1+it,\chi)|^2dt\\
\ll&P^{2\alpha_1(l-1)}\left(\sum_{q\leq Q_0}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{N_1}^{T}|G_{v_0}(1+it,\chi)|^{4l}|N_1(1+it,\chi)|^4dt\right)^{\frac{1}{2}}\\
&\times\left(\sum_{q\leq Q_0}\frac{q^2}{\varphi(q)}\sum_{\chi(q)}\int_{N_1}^{T}|N_2(1+it,\chi)|^4dt\right)^{\frac{1}{2}}
\end{aligned}\tag{2.3.38}$$

for the integral over $[N_1,T]$.

We will first treat the integral over $[T_0, N_1]$. In this case, to treat the first integral appearing in (2.3.37), we will apply the same argument as in the proof of Lemma 2.3.33, with $N(1+it,\chi) = N_1(1+it,\chi)$ and $M(1+it,\chi) = G_{v_0}^{2l}(1+it,\chi)$. We will again use the hybrid result of Fujii, Gallagher and Montgomery

$$\sum_{n\le N}\chi(n)n^{it} = \frac{\delta_\chi \varphi(q)N^{1+it}}{q(1+it)} + O((q\tau)^{1/2}\log(q\tau)),$$

where $\tau = |t|+2$, to bound $N_1(1+it,\chi)$. We have that

$$|G_{v_0}(1+it,\chi)|^{4l} = \left|\sum_{e^{v_0/U}<p\le e^{(v_0+1)/U}} \frac{\chi(p)}{p^{1+it}}\right|^{4l} = \left|\sum_{e^{2lv_0/U}<n\le e^{2l(v_0+1)/U}} \frac{\chi(n)a(n)}{n^{1+it}}\right|^{2},$$

where $a(n)=0$ unless n is a product of $2l$ primes, not necessarily distinct, each lying in the interval $(e^{v_0/U}, e^{(v_0+1)/U}]$. Writing n in terms of its prime factorisation $n = p_1^{a_1}\cdots p_b^{a_b}$ with $b\le 2l$, we have that $a(n) = \binom{2l}{a_1,\dots,a_b}$ when it is non-zero and therefore that $a(n)\ll (2l)!$. Therefore, we have that

$$\begin{aligned}
&\sum_{q\le Q_0}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{N_1} |N_1(1+it,\chi)|^4 |G_{v_0}^l(1+it,\chi)|^2 dt \\
&\quad\ll X^{\varepsilon/10}(2l)!^2 \sum_{q\le Q_0}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{N_1}\left(\frac{\varphi(q)}{q(1+|t|)}\right)^4 + \frac{q^2\log^4(q\tau)}{\tau^2}dt \\
&\quad\ll \frac{X^{\varepsilon/5}}{T_0}(l!)^{4+\varepsilon}.
\end{aligned} \tag{2.3.39}$$

We apply Lemma 2.3.27 to the second integral in (2.3.37). Noting that

$N_2(1+it,\chi)$ either has coefficients 1 or $\log n$, we find that

$$\begin{aligned}
&\sum_{q\le Q_0}\frac{q^2}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{N_1}|N_2(1+it,\chi)|^4dt\\
&\quad\ll \sum_{q\le Q_0} q^2\left(N_1+\frac{N_2^2}{q}\right)\sum_{N_2^2\le n\le 4N_2^2}\frac{d^2(n)\log^4 n}{n^2}\\
&\quad\ll (\log X)^7\sum_{q\le Q_0} q^2\left(\frac{N_1+N_2^2/q}{N_2^2}\right)\\
&\quad\ll (\log X)^{16+\varepsilon}\left(\frac{N_1+N_2^2}{N_2^2}\right). \qquad (2.3.40)
\end{aligned}$$

Since $N_1N_2 = X^{1+o(1)}$ with $N_1 > X^{1/2-\varepsilon'}$ and $X^{1/3+\varepsilon'} < N_2 \le X^{1/2+\varepsilon'}$, we have that

$$\frac{N_1+N_2^2}{N_2^2}\ll 1.$$

Returning to (2.3.37) and combining the estimates (2.3.39) and (2.3.40), we have that

$$\begin{aligned}
&P^{2\alpha_1(l-1)}\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{N_1}|G_{v_0}^l(1+it,\chi)N_1(1+it,\chi)N_2(1+it,\chi)|^2dt\\
\ll&\frac{P^{2\alpha_1(l-1)}X^{\varepsilon/10}(l!)^{2+\varepsilon}}{T_0^{1/2}}.
\end{aligned}$$

By the definition of l, we have that $(l!)^{2+\varepsilon}\ll(\log^2 X)^{l(1+\varepsilon)}$, and therefore

$$\begin{aligned}
(P^{2\alpha_1-1}\log^2 X)^{l(1+\varepsilon)} &\ll \exp\left((1+\varepsilon)\varepsilon\left(2\alpha_1-1+\frac{2}{17+\varepsilon}\right)\log X\right)\\
&\ll X^{-2\varepsilon/3}. \qquad (2.3.41)
\end{aligned}$$

Therefore, the overall contribution of the integral over $[T_0, N_1]$ is

$$P^{2\alpha_1(l-1)} \sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{T_0}^{N_1} |G_{v_0}^l(1+it,\chi) N_1(1+it,\chi) N_2(1+it,\chi)|^2 dt$$
$$\ll X^{\varepsilon/10 - 2\varepsilon/3 - 1/200}$$
$$\ll X^{-\varepsilon/2},$$

as required.

Since the contribution of the integral over $[T_0, N_1]$ is acceptable, it remains to bound (2.3.38), the contribution of the integral over $[N_1, T]$. To the first integral, we again apply Lemma 2.3.27, obtaining

$$\sum_{q \le Q_0} \frac{q^2}{\varphi(q)} \sum_{\chi(q)} \int_{N_1}^{T} |N_2(1+it,\chi)|^4 dt$$
$$\ll \sum_{q \le Q_0} q^2 \left(T + \frac{N_2^2}{q} \right) \sum_{N_2^2 \le n \le 4N_2^2} \frac{d^2(n) \log^4 n}{n^2}$$
$$\ll (\log X)^{16+\varepsilon} \left(\frac{T + N_2^2}{N_2^2} \right). \tag{2.3.42}$$

For the second integral of (2.3.38), since we have $N_1 \le t$ for all $t \in [N_1, T]$, we can apply Lemma 2.3.33. We choose $M(1+it,\chi) = G_{v_0}^{2l}(1+it,\chi)$, $M = P^{2l}$

and N corresponding to N_1 to find that

$$\begin{aligned}
&\sum_{q\le Q_0}\frac{1}{\varphi(q)}\sum_{\chi(q)}\int_{N_1}^{T}|G_{v_0}(1+it,\chi)|^{4l}|N_1(1+it,\chi)|^4dt\\
&\ll X^{\varepsilon/10}(2l)!^2\left(\frac{Q_0T}{N_1^2P^{2l}}\left(1+P^{4l}(Q_0T)^{-1/2}\right)+\frac{1}{N_1}\right)\\
&\ll X^{\varepsilon/10}(l!)^{4+\varepsilon}\left(\frac{Q_0T}{N_1^2P^{2l}}+\frac{1}{N_1}\right), \qquad (2.3.43)
\end{aligned}$$

as the definition of l ensures that $P^{4l}(Q_0T_1)^{-1/2}\ll 1$. Returning to (2.3.38) and substituting the estimates (2.3.42) and (2.3.43), we have that

$$\begin{aligned}
&P^{2\alpha_1(l-1)}\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{N_1}^{T}|G_{v_0}^{l}(1+it,\chi)N_1(1+it,\chi)N_2(1+it,\chi)|^2dt\\
&\ll P^{2\alpha_1(l-1)}X^{\varepsilon/10}(l!)^{2+\varepsilon}\left(\frac{Q_0T}{N_1^2P^{2l}}+\frac{1}{N_1}\right)^{1/2}\left(\frac{T+N_2^2}{N_2^2}\right)^{1/2}\\
&\ll P^{2\alpha_1(l-1)}X^{\varepsilon/10}(l!)^{2+\varepsilon}\left(\frac{Q_0T}{N_1^2N_2^2P^{2l}}(T+N_2^2)+\frac{T}{N_1N_2^2}+\frac{1}{N_1}\right)^{1/2}.
\end{aligned}$$

We have that $N_1N_2 = X^{1+o(1)}$ with $N_1 \ge X^{1/2-\varepsilon'}$ and $X^{1/3+\varepsilon'} \le N_2 \le X^{1/2+\varepsilon'}$. As we also have that $T \le X^{1+o(1)}$, the above is bounded by

$$\ll P^{2\alpha_1(l-1)}X^{\varepsilon/10}(l!)^{2+\varepsilon}\left(\frac{1}{P^l}+\frac{1}{N_2}+\frac{1}{N_1^{1/2}}\right).$$

Once again we apply (2.3.41) to see that $P^{2\alpha_1(l-1)}(l!)^{2+\varepsilon}\ll X^{-2\varepsilon/3}$. Overall we have the bound

$$\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}}|G_{v_0}(1+it,\chi)|^2|N_1(1+it,\chi)N_2(1+it,\chi)|^2dt \ll X^{-\varepsilon/2},$$

as required. Since both the integrals over $[T_0, N_1]$ and $[N_1, T]$ contribute $\ll X^{-\varepsilon/2}$, this completes the proof. □

2.3.6.3 Completing the proof of Proposition 2.3.21

Now that we have handled both the Type I and Type II contributions, we have found the cancellation we need over the complement of $\mathcal{S}_1$ (where the polynomial $G_{v_0}(1+it,\chi)$ is pointwise large). Having previously treated the set $\mathcal{S}_1$ where $G_{v_0}(1+it,\chi)$ is pointwise small, we may now combine these estimates to complete the proof of Proposition 2.3.21.

Proof of Proposition 2.3.21. By Lemma 2.3.34, we have that

$$\begin{aligned}
&\sum_{q\le Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{T}|F(1+it,\chi)|^2dt\\
&\ll \sum_{q\le Q_0}\Bigg(\frac{qU^2\log^2 P}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{T}|G_{v_0}(1+it,\chi)|^2|H_{v_0}(1+it,\chi)|^2dt\\
&\qquad\qquad+\frac{1}{Q_0(\log X)^{2+\eta}}\left(\frac{qT\log X}{X}+\frac{q}{\varphi(q)}\right)\Bigg)\\
&\ll \sum_{q\le Q_0}\Bigg(\frac{qU^2\log^2 P}{\varphi(q)}\sum_{\chi(q)}\Bigg(\int_{\mathcal{T}_{1,\chi}}|G_{v_0}(1+it,\chi)|^2|H_{v_0}(1+it,\chi)|^2dt\\
&\qquad\qquad+\int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}}|G_{v_0}(1+it,\chi)|^2|H_{v_0}(1+it,\chi)|^2dt\Bigg)\\
&\qquad\qquad+\frac{1}{Q_0(\log X)^{2+\eta}}\left(\frac{qT\log X}{X}+\frac{q}{\varphi(q)}\right)\Bigg)
\end{aligned}$$

for some $\eta=\eta(\varepsilon)>0$ and some suitable integer $v_0\in I$. We apply Lemma 2.3.36 to bound the contribution of the integral over $\mathcal{T}_{1,\chi}$, where the polynomial G_{v_0} is small, finding that the above is bounded by

$$\begin{aligned}
\ll \sum_{q\le Q_0}\Bigg(&\frac{qU^2\log^2 P}{\varphi(q)}\sum_{\chi(q)}\int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}}|G_{v_0}(1+it,\chi)|^2|H_{v_0}(1+it,\chi)|^2dt\\
&+\frac{1}{Q_0(\log X)^{2+\eta}}\left(\frac{qTP\log X}{X}+\frac{q}{\varphi(q)}\right)\Bigg).
\end{aligned}$$

We will combine Lemmas 2.3.38 to 2.3.40 to bound the contribution of the complement of $\mathcal{S}_1$, where the polynomial G_{v_0} is large. We have decomposed the polynomial H_{v_0} using Heath-Brown's identity (Lemma 2.3.26), which introduces a factor of $(\log X)^{2C}$. We have

$$\begin{aligned} &U^2(\log P)^2 \sum_{q\le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)|^2 |H_{v_0}(1+it,\chi)|^2 dt \\ \ll &U^2(\log P)^2(\log X)^{2C} \\ &\times \sum_{q\le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}} |G_{v_0}(1+it,\chi)|^2 \sum_{j\le L} |Q_j(1+it,\chi)|^2, \end{aligned} \tag{2.3.44}$$

where each $Q_j(s,\chi) = \prod_{i\le J_j} R_i(s,\chi)$ with $J_j \le 6$ for each $1\le j\le L$ and $L \le \log^C X$ and $R_i(s,\chi)$ are prime-factored Dirichlet polynomials of the form

$$\sum_{R_i<n\le 2R_i} \frac{\chi(n)\log n}{n^s}, \quad \sum_{R_i<n\le 2R_i} \frac{\chi(n)}{n^s}, \text{ or } \sum_{R_i<n\le 2R_i} \frac{\mu(n)\chi(n)}{n^s},$$

whose lengths satisfy $R_1\cdots R_{J_j} = X^{1+o(1)}$, $R_i \gg \exp\left(\frac{\log X}{\log\log X}\right)$ for each i. We treat these integrals according to whether Q_j is a Type I or Type II sum. For the Type II sums, we further split the domain of integration into the sets $\mathcal{T}_{2,\chi}$ and $\mathcal{T}_\chi$ according to the size of one of the factors of Q_j, as in (2.3.34). To the set $\mathcal{T}_{2,\chi}$, where this factor is small, we apply Lemma 2.3.38. To the set $\mathcal{T}_\chi$, where this factor is large, we apply Lemma 2.3.39. Overall, the Type

II sums contribute

$$\begin{aligned}
&U^2(\log P)^2(\log X)^{2C}\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}}|G_{v_0}(1+it,\chi)|^2|Q_j(1+it,\chi)|^2\\
&\ll\frac{U^2(\log X)^{2C}(\log P)^2}{\log^E X}\\
&\ll\frac{1}{\log^F X}
\end{aligned}$$

to (2.3.44) for some sufficiently large $F>0$. There are $\leq\log^C X$ Type II sums to consider, so as $F>0$ is sufficiently large this is negligible.

When Q_j is a Type I sum, we apply Lemma 2.3.40 to obtain the bound

$$\begin{aligned}
&U^2(\log P)^2(\log X)^{2C}\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{[T_0,T]\setminus\mathcal{T}_{1,\chi}}|G_{v_0}(1+it,\chi)|^2|Q_j(1+it,\chi)|^2\\
&\ll U^2(\log X)^{2C}(\log P)^2X^{-\varepsilon/2}\\
&\ll X^{-\varepsilon/4}.
\end{aligned}$$

There are $\leq\log^C X$ Type I sums to consider, so this contribution is negligible.

Thus, we have that

$$\begin{aligned}
&\sum_{q\leq Q_0}\frac{q}{\varphi(q)}\sum_{\chi(q)}\int_{T_0}^{T}|F(1+it,\chi)|^2dt\\
&\qquad\ll\frac{1}{Q_0(\log X)^{2+\eta}}\sum_{q\leq Q_0}\left(\frac{qTP\log X}{X}+\frac{q}{\varphi(q)}\right),
\end{aligned}$$

as required. □

2.4 Proof of Theorem 1.0.4

We now briefly outline how to adjust the argument to prove Theorem 1.0.4. The problem can be reduced to the set of E_2 numbers which factorise in the "typical" way. By Mertens' theorem, almost all products of exactly two primes $p_1p_2 \leq X$ with $p_1 \leq p_2$ satisfy

$$p_1 \in \left[\exp\left((\log X)^{\varepsilon(X)}\right), \exp\left((\log X)^{1-\varepsilon(X)}\right)\right] =: [P_1, P_2],$$

where $\varepsilon(X) = o(1)$. We define $E_2'' := E_2''(X)$ to be the set of E_2 numbers $n = p_1p_2 \in (X, 2X]$ which factorise in the typical way.

Lemma 2.4.1. *Let h be a fixed non-zero integer, $P_1 := \exp\left((\log X)^{\varepsilon(X)}\right)$ and $P_2 := \exp\left((\log X)^{1-\varepsilon(X)}\right)$. We have that*

$$\begin{aligned}
&\frac{1}{X}\sum_{X<n\leq 2X} \mathbb{1}_{E_2}(n)\mathbb{1}_{E_2}(n+h) - o\left(\frac{\mathfrak{S}(h)(\log\log X)^2}{(\log X)^2}\right) \\
&\leq \frac{1}{X}\sum_{X<n\leq 2X} \mathbb{1}_{E_2''}(n)\mathbb{1}_{E_2''}(n+h) \\
&\leq \frac{1}{X}\sum_{X<n\leq 2X} \mathbb{1}_{E_2}(n)\mathbb{1}_{E_2}(n+h).
\end{aligned}$$

Proof. The second inequality is trivial as E_2'' is a subset of E_2 and the summand is positive, so it remains to prove the first inequality. Since

$\mathbb{1}_{E_2} = \mathbb{1}_{E_2''} + \mathbb{1}_{E_2 \setminus E_2''}$, we have that

$$\begin{aligned}\sum_{X<n\leq 2X} \mathbb{1}_{E_2}(n)\mathbb{1}_{E_2}(n+h) = &\sum_{X<n\leq 2X} \mathbb{1}_{E_2''}(n)\mathbb{1}_{E_2''}(n+h)\\ &+ \sum_{X<n\leq 2X} \mathbb{1}_{E_2\setminus E_2''}(n)\mathbb{1}_{E_2\setminus E_2''}(n+h)\\ &+ \sum_{X<n\leq 2X} \mathbb{1}_{E_2\setminus E_2''}(n)\mathbb{1}_{E_2''}(n+h)\\ &+ \sum_{X<n\leq 2X} \mathbb{1}_{E_2''}(n)\mathbb{1}_{E_2\setminus E_2''}(n+h).\end{aligned}$$

We will bound the last three terms using sieve theory. We treat the second term, with the third and fourth being handled similarly. We have, by definition

$$\sum_{X<n\leq 2X} \mathbb{1}_{E_2\setminus E_2''}(n)\mathbb{1}_{E_2\setminus E_2''}(n+h) = \sum_{\substack{p_1,p_3\leq\sqrt{X}\\ p_1,p_3\notin[P_1,P_2]}} \sum_{\substack{X<p_1p_2\leq 2X\\ X-h<p_3p_4\leq 2X-h\\ p_1p_2-p_3p_4=h}} 1.$$

To find an upper bound, we will attach sieve weights λ_d to the inner sum (for example, we can use Brun's weights as in Lemma 2.3.34). Set $z = X^{1/\beta}$ with $\beta > 0$ suitably large, and $P(z) = \prod_{p<z, p\nmid h} p$. We then need to bound

$$\sum_{\substack{p_1,p_3\leq\sqrt{X}\\ p_1,p_3\notin[P_1,P_2]}} \sum_{\substack{X/p_1<n\leq 2X/p_1\\ \left(\frac{n}{p_3}(p_1n+h),P(z)\right)=1}} \left(\sum_{d|n} \lambda_d\right).$$

Note that we can suppose $(d,p_3) = 1$, as removing the terms $(d,p_3) > 1$

makes a negligible difference. Switching the order of summation, the inner sum is

$$\sum_{d|P(z)} \lambda_d \sum_{\substack{X/p_1<n\leq 2X/p_1 \\ d|\frac{n}{p_3}(p_1 n+h) \\ p_3|p_1 n-h}} 1. \tag{2.4.1}$$

In the inner sum we have congruence conditions mod d and p_3, which can be combined using the Chinese Remainder Theorem into a condition mod dp_3 since we assume $(d, p_3) = 1$. Therefore (2.4.1) is bounded by

$$\ll \frac{X}{p_1 p_3} \frac{h}{\varphi(h)} \prod_{p\leq z}\left(1-\frac{2}{p}\right) \ll \frac{X}{p_1 p_3 \log^2 z} \frac{h}{\varphi(h)} \ll \frac{X}{p_1 p_3 \log^2 X} \frac{h}{\varphi(h)}.$$

By Mertens' Theorem (Lemma 2.2.2), we have that

$$\begin{aligned}
\sum_{\substack{p\leq\sqrt{X} \\ p\notin[P_1,P_2]}} \frac{1}{p} &= \sum_{p<P_1} \frac{1}{p} + \sum_{P_2<p\leq\sqrt{X}} \frac{1}{p} \\
&= \log\log P_1 + \log\log\sqrt{X} - \log\log P_2 + O(1) \\
&= \log\log X - \log\log X + o(\log\log X) \\
&= o(\log\log X).
\end{aligned}$$

Therefore, summing over p_1 and p_3, we have that

$$\begin{aligned}
\sum_{X<n\leq 2X} \mathbb{1}_{E_2\setminus E_2''}(n)\mathbb{1}_{E_2\setminus E_2''}(n+h) &\ll \frac{X}{\log^2 X}\frac{h}{\varphi(h)} \sum_{\substack{p_1,p_3\leq\sqrt{X} \\ p_1,p_3\notin[P_1,P_2]}} \frac{1}{p_1 p_3} \\
&= o\left(\frac{\mathfrak{S}(h)X(\log\log X)^2}{(\log X)^2}\right),
\end{aligned}$$

as required. $\square$

Therefore, we can reduce the problem to considering the correlations of $n, n+h \in E_2''$. We modify every definition featuring $(P, P^{1+\delta}]$, replacing this interval with $[P_1, P_2]$. We will once again apply the Hardy-Littlewood circle method and in (2.3.4) and (2.3.6) we take

$$Q_0 := \log^{A'} X, A' > 4, \qquad Q := P_2 \log^C X, \qquad H \geq Q \log^D X, \tag{2.4.2}$$

where C is chosen sufficiently large in terms of A' and D is chosen sufficiently large in terms of A' and C. In Lemma 2.3.34 we instead define $I := [U \log P_1, U \log P_2]$ where $U := Q_0^E$, $E > 0$ and we define $\alpha_1 := \varepsilon' > 0$ sufficiently small in terms of $\varepsilon > 0$.

The error terms in Propositions 2.3.6, 2.3.10 and 2.3.18 need to be treated with more care. The error term in Proposition 2.3.10 is acceptable with this new choice of Q_0 (see Remark 2.3.11). For Propositions 2.3.6 and 2.3.18, we again need to split the sums over $p_1 \in [P_1, P_2]$ into dyadic intervals whenever we apply the Cauchy-Schwarz inequality, as we did previously for the range $H > \exp((\log X)^{\varepsilon^3})$. If we do not make this modification, the bounds will not provide the necessary cancellation.

For example, the minor arc integral would be bounded by

$$\begin{aligned}
&\int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]}|S(\theta)|^2 d\theta \\
&\qquad \ll \frac{1}{H^2}\int_X^{2X}\left|\sum_{x<n\leq x+H}\varpi_2(n)e(n\alpha)\right|^2 dx + H\log^2 X \\
&\qquad \ll \frac{XP_2}{P_1}\left((\log X)\left(\frac{1}{Q_0}+\frac{1}{P_1}+\frac{Q}{H}\right)+\frac{P_2\log(X/P_2)}{H}\right).
\end{aligned}$$

The term P_2/P_1 prevents us from finding the necessary cancellation. In particular, the first term is

$$\ll X\exp((\log X)^{1-\varepsilon(X)} - (\log X)^{\varepsilon(X)})\log^{-A'} X,$$

which is too large. We now outline how to modify the proof of Proposition 2.3.6.

Proposition 2.4.2. *Let $A > 3$, $B > 1$ be fixed and $\mathfrak{m}$ be defined as in (2.3.6) with Q_0, Q as in (2.4.2). Let $Q\log^D X \leq H \leq X\log^{-A} X$ with $D > 0$ sufficiently large. For $\alpha \in \mathfrak{m}$ we have that*

$$\int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]}|S(\theta)|^2 d\theta \ll \frac{X}{\log^B X}.$$

Proof. As before, we apply Lemma 2.2.4 to the minor arc integral so that we need to bound

$$\begin{aligned} I &:= \int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]} |S(\theta)|^2 d\theta \\ &\ll \frac{1}{H^2}\int_X^{2X} \left|\sum_{x<n\le x+H} \varpi_2(n)e(n\alpha)\right|^2 dx + H\log^2 X. \end{aligned}$$

The second term is $\ll X(\log X)^{-A+2}$ by our choice of H, so it remains to bound the first term. Now before applying Cauchy-Schwarz to the integrand we split the sum over p_1 into dyadic intervals $[P, 2P]$ with $P_1 \le P \le P_2$ before applying the triangle inequality, so that we instead need to integrate

$$\begin{aligned} &\left|\sum_{\substack{x<p_1p_2\le x+H \\ P<p_1\le 2P}} (\log p_2)e(\alpha p_1 p_2)\right|^2 \\ &\le \left(\sum_{P<m_1\le 2P} |\mathbb{1}_{\mathbb{P}}(m_1)|^2\right)\left(\sum_{P<m_2\le 2P}\left|\sum_{x<m_2p\le x+H}(\log p)e(\alpha m_2 p)\right|^2\right) \\ &\le \frac{P}{\log P}\sum_{\substack{x<mp_1,mp_2\le x+H \\ P<m\le 2P}} (\log p_1)(\log p_2)e(\alpha m(p_1-p_2)). \end{aligned}$$

Next, we perform the integration on this sum and split into the diagonal ($p_1 = p_2$) and off-diagonal terms ($p_1 \ne p_2$), which will be denoted by S_1 and S_2 respectively. The diagonal terms now contribute

$$S_1 \ll \frac{P}{H\log P}\sum_{P<m\le 2P}\sum_{\frac{X}{m}<p\le\frac{3X}{m}} \log^2 p \ll \frac{XP\log X}{H\log P} \ll \frac{X}{(\log X)^{C+D-1}}. \tag{2.4.3}$$

Once again applying Lemma 2.2.8 followed by Lemma 2.2.3, the off-diagonal terms S_2 contribute

$$\begin{aligned}
&\ll \frac{P}{H^2 \log P} \sum_{0<r\leq H} \sum_{\substack{\frac{X}{2P}<p_1,p_2\leq \frac{3X}{P} \\ r=|p_1-p_2|}} (\log p_1)(\log p_2) \left| \sum_{\substack{P<m\leq 2P \\ 0<m\leq H/r}} e(\alpha m r)(H-mr) \right| \\
&\ll \frac{P}{H \log P} \sum_{0<r\leq \frac{H}{P}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right) \sum_{\substack{\frac{X}{2P}<p_1,p_2\leq \frac{3X}{P} \\ r=|p_1-p_2|}} (\log p_1)(\log p_2) \\
&\ll \frac{XP}{HP \log P} \sum_{0<r\leq \frac{H}{P}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right) \mathfrak{S}(r) \\
&\ll \frac{X \log\log X}{H \log P} \sum_{0<r\leq \frac{H}{P}} \min\left(\frac{H}{r}, \frac{1}{\|\alpha r\|}\right).
\end{aligned}$$

Applying Lemma 2.2.9, the off-diagonal terms contribute

$$S_2 \ll \frac{X \log\log X \log X}{\log P} \left(\frac{1}{Q_0} + \frac{1}{P} + \frac{Q}{H}\right).$$

Since we have chosen $Q_0 := \log^{A'} X$ with $A' > 4$, $P_1 \leq P \leq P_2$ and $Q := P_2 \log^C X$, we have that

$$S_2 \ll \frac{X}{\log^{B'} X} \tag{2.4.4}$$

for $B' > 3$. Combining the contributions of diagonal (2.4.3) and off-diagonal terms (2.4.4), before combining the dyadic intervals $[P, 2P]$ introduces a factor of $O(\log^2 P_2) = O(\log^2 X)$, so that

$$I \ll \frac{X}{\log^B X}$$

for $B > 1$, as claimed. $\square$

Proposition 2.4.3. *Let $A > 3, B > 0$ be fixed. Let $\varepsilon > 0$ be fixed and $\exp((\log X)^{1-\varepsilon}) \leq H \leq X \log^{-A} X$. Let $\mathfrak{M}$ be defined as in (2.3.4) with Q_0, Q as in (2.4.2). Then, for all but at most $O(HQ_0^{-1/3})$ values of $0 < |h| \leq H$ we have that*

$$\int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha) d\alpha = \mathfrak{S}(h) X \left(\sum_{P_1 \leq p \leq P_2} \frac{1}{p} \right)^2 + O\left(\frac{X}{\log^B X} \right).$$

Proof. Recalling Lemma 2.3.9, we have the expansion

$$\begin{aligned} S(\alpha) =& \frac{\mu(q)}{\varphi(q)} \sum_{P_1 < p \leq P_2} \frac{1}{p} \sum_{X < n \leq 2X} e(\beta n) \\ &+ \frac{1}{\varphi(q)} \sum_{\chi(q)} \tau(\overline{\chi}) \chi(a) \sum_{X < n \leq 2X} \left(\varpi_2(n) \chi(n) - \delta_\chi \sum_{P_1 \leq p \leq P_2} \frac{1}{p} \right) e(\beta n) \\ =& a(\alpha) + b(\alpha) \end{aligned} \tag{2.4.5}$$

and following the argument of Section 2.3.3 we have that

$$\begin{aligned} \int_{\mathfrak{M}} |S(\alpha)|^2 e(-h\alpha) d\alpha =& \mathfrak{S}(h) X \left(\sum_{P_1 \leq p \leq P_2} \frac{1}{p} \right)^2 \\ &+ O\left(\frac{X}{\log^B X} + A(X)B(X) + B^2(X) \right). \end{aligned}$$

Note that $A^2(X) \ll X(\log\log X)^3$, so it remains to bound $B^2(X)$. Following Proposition 2.3.16 and Remark 2.3.17, we have that

$$B^2(X) \ll B_1(X) + B_2(X) + \exp((\log X)^{1-\varepsilon(X)}),$$

where $B_1(X)$ and $B_2(X)$ is defined in Definition 2.3.15 with $(P, P^{1+\delta}]$ replaced with $[P_1, P_2]$. With our choices of (2.4.2), P_1 and P_2, by Remarks 2.3.20 and 2.3.22 we now have that

$$B_1(X) \ll \frac{X}{\log^B X}.$$

The proof of Proposition 2.3.18 requires modifying in a similar way to Proposition 2.3.6. As in the case $H > \exp((\log X)^{\varepsilon^3})$, we split the sum over $P_1 \leq p_1 \leq P_2$ into dyadic intervals $P < p_1 \leq 2P$ before applying Cauchy-Schwarz, Lemma 2.2.6 and then combining the contributions of the dyadic sums. □

We are now able to complete the proof of Theorem 1.0.4.

Proof of Theorem 1.0.4. By partial summation and Lemma 2.2.2 we have the bound

$$\int_0^1 |S(\alpha)|^2 d\alpha = \sum_{X<n\leq 2X} \varpi_2^2(n) \ll \log X \sum_{X<n\leq 2X} \varpi_2(n) \ll X\log X \sum_{P_1\leq p\leq P_2} \frac{1}{p}$$
$$\ll X(\log X)\log\log X.$$

Therefore, following the proof of Theorem 2.3.3, the result may be deduced from combining this bound with Proposition 2.4.2, an application of Chebyshev's inequality and Proposition 2.4.3 followed by an application of partial summation. □

2.5 Proof of Theorem 1.0.5

We outline the modifications needed to prove Theorem 1.0.5. When applying the Hardy-Littlewood circle method, in (2.3.4) and (2.3.6) we now choose

$$Q_0 := \log^{A'} X, A' > 6, \qquad Q := X^{1/6+\varepsilon/2}, \qquad H \geq QX^{\varepsilon/2}. \tag{2.5.1}$$

As in Section 2.4, in Lemma 2.3.34 we instead define $I := [U \log P_1, U \log P_2]$ where $U := Q_0^E$, $E > 0$ and we define $\alpha_1 := \varepsilon' > 0$ sufficiently small in terms of $\varepsilon > 0$. Analogously to the almost prime case, we may write

$$\sum_{X<n\leq 2X} \Lambda(n)\varpi_2(n+h) = \int_0^1 S(\alpha)\overline{S'(\alpha)}e(-h\alpha)d\alpha + O(h\log^2 X),$$

where for $\alpha \in (0, 1)$ we define the exponential sum

$$S'(\alpha) := \sum_{X<n\leq 2X} \Lambda(n)e(n\alpha).$$

The error term is $\ll X(\log X)^{-A+2}$ by our choice of H, which will be acceptable. We have the following result for the major arcs.

Proposition 2.5.1. *Let $A > 5, B > 0$ be fixed and let $\varepsilon > 0$ be fixed sufficiently small. Let $X^{1/6+\varepsilon} \leq H \leq X \log^{-A} X$. Let $\mathfrak{M}$ be defined as in* (2.3.4) *with Q_0, Q as in* (2.5.1)*. Then, for all but at most $O(HQ_0^{-1/3})$ values of $0 < |h| \leq H$ we have that*

$$\int_{\mathfrak{M}} S(\alpha)\overline{S'(\alpha)}e(-h\alpha)d\alpha = \mathfrak{S}(h)X\left(\sum_{P_1 \leq p \leq P_2} \frac{1}{p}\right) + O\left(\frac{X}{\log^B X}\right).$$

Proof. We can expand S' in terms of Dirichlet characters (see for example [32]). We have that $\alpha = a/q + \beta$ with $q \leq Q_0$, $(a,q) = 1$ and $|\beta| \leq \frac{1}{qQ}$. Then

$$S'(\alpha) = \sum_{X < n \leq 2X} \Lambda(n) e\left(\frac{an}{q}\right) e(n\beta).$$

We again apply the identity

$$e\left(\frac{a}{q}\right) = \frac{1}{\varphi(q)} \sum_{\chi(q)} \chi(a)\tau(\overline{\chi}),$$

which holds for $(n,q) = 1$, where $\tau(\chi)$ is the Gauss sum (2.2.1). Then

$$\begin{aligned} S'(\alpha) =& \frac{1}{\varphi(q)} \sum_{\chi(q)} \tau(\overline{\chi})\chi(a) \sum_{X < n \leq 2X} \Lambda(n)\chi(n)e(n\beta) \\ &+ \sum_{\substack{X < n \leq 2X \\ (n,q) > 1}} \Lambda(n) e\left(\frac{an}{q}\right) e(n\beta). \end{aligned}$$

The second term is bounded by

$$\ll \sum_{\substack{X<p^a\leq 2X \\ (p,q)>1}} \log p \ll \log^2 X.$$

We now introduce the approximation to the principal character term, so that

$$\begin{aligned} S'(\alpha) =& \frac{\mu(q)}{\varphi(q)} \sum_{X<n\leq 2X} e(n\beta) \\ &+ \frac{1}{\varphi(q)} \sum_{\chi(q)} \tau(\bar{\chi})\chi(a) \sum_{X<n\leq 2X} (\Lambda(n)\chi(n) - \delta_\chi) e(n\beta) \\ &+ O(\log^2 X) \\ =& c(\alpha) + d(\alpha) + O(\log^2 X), \end{aligned}$$

say. Therefore, using the expansion (2.4.5) and Cauchy-Schwarz, we may write the integral over the major arcs as

$$\begin{aligned} &\int_{\mathfrak{M}} S(\alpha)\overline{S'(\alpha)} e(-h\alpha) d\alpha \\ =& \int_{\mathfrak{M}} a(\alpha)\overline{c(\alpha)} e(-h\alpha) d\alpha \\ &+ O\left(A(X)D(X) + B(X)(C(X) + D(X)) + (A(X) + B(X))\log^2 X\right), \end{aligned} \tag{2.5.2}$$

where we define $C^2(X) = \int_{\mathfrak{M}} |c(\alpha)|^2 d\alpha$ with $D(X)$ defined analogously. Evaluating $\int_{\mathfrak{M}} a(\alpha)\overline{c(\alpha)} e(-h\alpha) d\alpha$ as in Section 2.3.3.2 gives the required main term and an acceptable error. Mikawa [32, Section 3] proves that

$C^2(X) \ll X \log\log X$ and that

$$D^2(X) \ll \sum_{q \le Q_0} \frac{q}{\varphi(q)} \sum_{\chi(q)} \left(\int_X^{2X} \left| \frac{1}{qQ} \sum_{x < n \le x + qQ/2} (\Lambda(n)\chi(n) - \delta_\chi) \right|^2 dx + qQ \log^2 X \right).$$

The second term is $\ll X^{1/6+\varepsilon} \ll X \log^{-B} X$ by the definition of Q, which is acceptable. Noting that we have chosen $Q = X^{1/6+\varepsilon/2}$, we apply Lemma 2.2.6 to the first term to get

$$D^2(X) \ll \frac{X}{\log^B X}$$

for $B > 0$, as required. Combining this with our estimates for $A(X), B(X)$ (from Proposition 2.4.3) and $C(X)$ we have that the error term in (2.5.2) is $O(X \log^{-B} X)$, as required. □

Proof of Theorem 1.0.5. Analogously to the proof of Theorem 2.3.3, by [27, Proposition 3.1] we have that

$$\sum_{0 < |h| \le H} \left| \sum_{X < n \le 2X - h} \Lambda(n)\varpi_2(n+h) - \int_{\mathfrak{M}} S(\alpha)\overline{S'(\alpha)} e(-h\alpha) d\alpha \right|^2$$
$$\ll H \int_{\mathfrak{m}} |S(\alpha)||S'(\alpha)| \int_{\mathfrak{m} \cap [\alpha - \frac{1}{2H}, \alpha + \frac{1}{2H}]} |S(\beta)||S'(\beta)| d\beta d\alpha.$$

By Cauchy-Schwarz on the integral over β, we have that the above is bounded by

$$\ll H\int_{\mathfrak{m}}|S(\alpha)||S'(\alpha)|\left(\int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]}|S(\beta)|^2d\beta\right)^{1/2}$$
$$\times\left(\int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]}|S'(\beta)|^2d\beta\right)^{1/2}d\alpha$$
$$\ll H\left(\int_{\mathfrak{m}}|S(\alpha)||S'(\alpha)|d\alpha\right)\left(\int_0^1|S'(\beta)|^2d\beta\right)^{1/2}$$
$$\times\left(\sup_{\alpha\in\mathfrak{m}}\int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]}|S(\beta)|^2d\beta\right)^{1/2}.$$

In the last step we have trivially bounded one of the integrals over the minor arcs by the integral over the unit circle and taken the supremum over $\alpha\in\mathfrak{m}$. Applying Cauchy-Schwarz again to the integral over α, we have the bound

$$\ll H\left(\int_{\mathfrak{m}}|S(\alpha)|^2d\alpha\right)^{1/2}\left(\int_{\mathfrak{m}}|S'(\alpha)|^2d\alpha\right)^{1/2}\left(\int_0^1|S'(\beta)|^2d\beta\right)^{1/2}$$
$$\times\left(\sup_{\alpha\in\mathfrak{m}}\int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]}|S(\beta)|^2d\beta\right)^{1/2}$$
$$\ll H\left(\int_0^1|S'(\alpha)|^2d\alpha\right)\left(\int_0^1|S(\alpha)|^2d\alpha\right)^{1/2}$$
$$\times\left(\sup_{\alpha\in\mathfrak{m}}\int_{\mathfrak{m}\cap[\alpha-\frac{1}{2H},\alpha+\frac{1}{2H}]}|S(\beta)|^2d\beta\right)^{1/2}.$$

Again, in the last line we have trivially bounded the integral over the minor arcs by the integral over the unit circle. Trivially, we have that

$$\int_0^1|S(\alpha)|^2d\alpha\ll X(\log X)\log\log X,\qquad\int_0^1|S'(\alpha)|^2d\alpha\ll X\log X,$$

so, combining these estimates with Proposition 2.4.2 (suitably adjusting for the choices of Q_0, Q, H), we have that

$$\sum_{0<|h|\leq H} \left| \sum_{X<n\leq 2X-h} \Lambda(n)\varpi_2(n+h) - \int_{\mathfrak{M}} S(\alpha)\overline{S'(\alpha)}e(-h\alpha)d\alpha \right|^2 \ll \frac{HX^2}{\log^B X}.$$

Therefore, applying Chebyshev's inequality and Proposition 2.5.1 followed by partial summation gives the result. □

Chapter 3

Future Outlook

For our main result Theorem 1.0.2 on correlations of almost primes, the smallest possible choice of H is $(\log X)^{19+\varepsilon}$, however it may be possible to lower this exponent. In the proof of Theorem 1.0.2 we apply the argument of Teräväinen [38, Sections 2-4] showing that almost all intervals $[x, x + (\log x)^{5+\varepsilon}]$ contain an integer which has exactly two prime factors. The second half of Teräväinen's paper is dedicated to lowering the exponent $5+\varepsilon$ to 3.51 through an argument additionally using some sieve theory and the theory of exponent pairs. This result has recently been further improved by Matomäki and Teräväinen [29], who prove that for almost all x the interval $(x, x+(\log x)^{2.1}]$ contains E_2 numbers. We do not apply ideas from these arguments here, but it is possible that adapting some aspects to our proof could lower the exponent of H in Theorem 1.0.2.

A natural next question is whether we can establish an asymptotic formula for the number of integers $n = p_1p_2p_3 \le X$ such that $n+h$ has exactly three (or, more generally, $k \ge 3$) prime factors which holds for almost all $|h| \le H = H(X)$.

Question 3.0.1. Let $k \ge 3$ and define E_k be the set of integers $n = p_1 \cdots p_k$ with exactly k prime factors. Can we show an asymptotic formula of the form

$$\frac{1}{X}\sum_{X<n\le 2X} \mathbb{1}_{E_k}(n)\mathbb{1}_{E_k}(n+h) \sim \mathfrak{S}(h)\left(\frac{1}{X}\sum_{X<n\le 2X}\mathbb{1}_{E_k}(n)\right)^2$$

which holds for almost all $|h| \le H = H(k, X)$? How short an average H can we take?

In his work on almost primes in almost all short intervals, Teräväinen [38] proved results for E_k numbers with $k \ge 2$. In particular, Teräväinen showed that almost all intervals $[x, x + (\log x)^{3.51}x]$ with $x \le X$ contain an E_2 number, and almost all intervals $[x, x + (\log\log x)^{6+\varepsilon}\log x]$ contain an E_3 number. For $k \ge 4$, the author shows that there exists a constant $C_k > 0$ such that almost all intervals $[x, x+(\log_{k-1} x)^{C_k}\log x]$ contain an E_k number. It is therefore reasonable to expect that we could adapt the arguments used to prove these results along the lines of Sections 2.3.5 - 2.3.6 to decrease the size we can take the prime factor P in Proposition 2.3.21. Provided we can suitably adapt the argument treating the contribution of the minor arcs, it would be expected that we would then obtain a shorter average H over which the expected asymptotic formula for correlations of E_k numbers

($k \geq 3$) holds than for correlations of E_2 numbers.

If we could obtain such results, we could also ask about other correlations of primes and almost primes, for example

$$\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{\mathbb{P}}(n)\mathbb{1}_{E_k}(n+h),$$

with $k \geq 3$. We would not expect to be able to go beyond the range $H \geq X^{1/6+\varepsilon}$ established in Theorem 1.0.5 using our arguments, as the limitation there is due to the presence of the prime in the correlation.

Similarly, if asymptotic formulas for correlations of E_k numbers and correlations of E_ℓ numbers can be established on average for some $\ell \geq k \geq 2$, then for correlations of almost primes of the form

$$\frac{1}{X} \sum_{X<n\leq 2X} \mathbb{1}_{E_k}(n)\mathbb{1}_{E_\ell}(n+h),$$

we would expect the range of H to match the range known for correlations of E_k numbers.

Another future direction is to investigate whether we can extend these methods of Matomäki and Radziwiłł [25] and Teräväinen [38] to another setting, for example the Gaussian integers. In particular, we could ask whether we can establish an analogue of the work of Teräväinen [38] and show that almost all narrow sectors contain a Gaussian 'almost prime'.

Question 3.0.2. Let $0 \leq \phi \leq \frac{\pi}{2}$ and $0 < \delta = \delta(X) \leq \frac{\pi}{2}$ shrink as $X \to \infty$.

How small can we take δ such that almost all sectors $X < N(\mathfrak{a}) \leq 2X$, $\theta_{\mathfrak{a}} \in (\phi, \phi + \delta]$ contain a Gaussian almost prime $\mathfrak{a}$ with exactly two prime factors? Or exactly $k \geq 2$ factors?

If such results can be proved, it would then be natural to ask whether we can establish results on correlations of Gaussian 'almost primes' and primes.

Appendix A

Primes in short arithmetic progressions

In this appendix we prove the following variant of a theorem on primes in short arithmetic progressions, due to Koukoulopoulos [21].

Lemma 2.2.6. *Let* $A \geq 1$ *and* $\varepsilon \in (0, \frac{1}{3}]$ *be fixed. Let* $X \geq 1$, $1 \leq Q \leq \frac{\Delta}{X^{1/6+\varepsilon}}$ *and* $\Delta = X^{\theta}$ *with* $\frac{1}{6} + 2\varepsilon \leq \theta \leq 1$. *Then we have that*

$$\sum_{q \leq Q} \sum_{\chi(q)} \int_X^{2X} \left| \sum_{x < n \leq x + q\Delta} (\Lambda(n)\chi(n) - \delta_\chi) \right|^2 dx \ll \frac{Q^3 \Delta^2 X}{\log^A X},$$

where we define $\delta_\chi = 1$ *if* $\chi = \chi_0$ *and* $\delta_\chi = 0$ *otherwise.*

A.1 Preliminary Results

We will again reduce the problem to finding cancellation in Dirichlet polynomials

$$F(s,\chi) := \sum_{X<n\leq 2X} \frac{\Lambda(n)\chi(n)}{n^s},$$

which will be handled using mean and large value theorems. The large values result of Jutila (Lemma 2.3.32) is not sharp enough for this argument, as we would lose in the power of $(qTX)^{\varepsilon}$ which appears in (2.3.21), where $\Im(s) \in \mathcal{T} \subset [-T,T]$ and $\mathcal{T}$ is well-spaced. We will instead use a result derived from Huxley. First, we recall the definition of a well-spaced set:

Definition 2.3.23. [Well-Spaced Set] We say a set $\mathcal{T}$ is *well-spaced* if for any $t,u \in \mathcal{T}$ with $t \neq u$ we have that $|t-u| \geq 1$.

Lemma A.1.1. *Fix $m \in \mathbb{N}$, $r \geq 0$ and let $\{a_n\}_{n=1}^N$ be a sequence of complex numbers such that $|a_n| \leq d_m(n)(\log n)^r$ for all $n \leq N$. For each Dirichlet character χ, we let $A(s,\chi) = \sum_{n=1}^N \frac{a_n\chi(n)}{n^s}$ and consider a well-spaced set*

$$\mathcal{R} \subset \bigcup_{q\leq Q} \bigcup_{\substack{\chi(q) \\ \chi \text{ primitive}}} \{(t,\chi) : t\in\mathbb{R}, |A(\tfrac{1}{2}+it,\chi)| \geq V\},$$

where $V,Q,T \geq 1$ are some parameters. Then

$$|\mathcal{R}| \ll_{m,r} \min\left\{\frac{N+Q^2T}{V^2}, \frac{N}{V^2} + \frac{NQ^2T}{V^6}\right\} (\log 2N)^{3m^2+6r+18}.$$

Proof. See [21, Lemma 3.2]. This follows from Huxley's result [18, Theorem

9.18] □

We will also need to be able to bound a Dirichlet polynomial with a Dirichlet character by a shorter polynomial:

Lemma A.1.2. *Let χ be a primitive Dirichlet character modulo $q \in (1, Q]$ and let $g : [0, +\infty) \to [0, +\infty)$ be a smooth function supported on $[1, 4]$. Let $t \in \mathbb{R}$, $N \geq 1$ and r be a non-negative integer. If $|t| \leq T$ for some $T \geq 2$ and $M = \max\{1, (QT/N)^{1+\delta}\}$ for some fixed $\delta > 0$, then*

$$\sum_{n=1}^{\infty} \frac{g(n/N)\chi(n)(\log n)^r}{n^{1/2+it}} \ll_{r,\delta} (\log 2N)^r \int_{-\infty}^{\infty} \left| \sum_{n \leq M} \frac{\chi(n)}{n^{1/2+i(t+u)}} \right| \frac{du}{1+u^2}.$$

Proof. This is [21, Lemma 3.3]. □

A.2 Proof of the Lemma

We outline how to adapt the proof appearing in [21, Section 4]. First, we will show that the contribution of the imprimitive characters is acceptable.

Lemma A.2.1. *Let $A \geq 1$ and $\varepsilon \in (0, \frac{1}{3}]$ be fixed. Let $X \geq 1$, $1 \leq Q \leq \frac{\Delta}{X^{1/6+\varepsilon}}$ and $\Delta = X^\theta$ with $\frac{1}{6} + 2\varepsilon \leq \theta \leq 1$. Then we have that*

$$\sum_{q \leq Q} \sum_{\substack{\chi(q) \\ \chi \text{ imprimitive}}} \int_X^{2X} \left| \sum_{x < n \leq x+q\Delta} (\Lambda(n)\chi(n) - \delta_\chi) \right|^2 dx \ll \frac{Q^3\Delta^2 X}{\log^A X} + XQ^2 \log^4 X.$$

Proof. First, we treat the contribution of the principal character. We have

that

$$\begin{aligned}\left|\sum_{x<n\le x+q\Delta}(\Lambda(n)\chi_0(n)-1)-\sum_{x<n\le x+q\Delta}(\Lambda(n)-1)\right| &\le \sum_{\substack{x<n\le x+q\Delta\\(n,q)>1}}\Lambda(n)\\ &\le \omega(q)\log X\\ &\ll \log^2 X.\end{aligned}$$

This error contributes $\ll XQ^2\log^4 X$, which is acceptable. Then, since we have $\Delta \ge X^{1/6+2\varepsilon}$, we apply what is known about primes in almost all short intervals (Lemma 2.2.5) to get that

$$\sum_{q\le Q}\int_X^{2X}\left|\sum_{x<n\le x+q\Delta}(\Lambda(n)-1)\right|^2 dx \ll \sum_{q\le Q}\frac{(q\Delta)^2 X}{\log^A X} \ll \frac{Q^3\Delta^2 X}{\log^A X},$$

which contributes the first term of the bound.

Now we deal with the contribution of the remaining imprimitive characters. Suppose $\chi \mod q$ is non-principal and induced by $\chi_1 \mod q_1$, then we have that

$$\left|\sum_{x<n\le x+q\Delta}\Lambda(n)\chi(n)-\sum_{x<n\le x+q\Delta}\Lambda(n)\chi_1(n)\right| \le \sum_{\substack{x<n\le x+q\Delta\\(n,q)>1}}\Lambda(n) \ll \log^2 X.$$

Overall, this contributes $\ll XQ^2\log^4 X$, which contributes the second term.

□

Therefore we have reduced the problem to showing that

$$\begin{aligned}\sum_{2\le q\le Q}\sum_{q_1|q}\sideset{}{^*}\sum_{\chi(q_1)}\int_X^{2X}\left|\sum_{x<n\le x+q\Delta}\Lambda(n)\chi(n)\right|^2 dx\\ =\sum_{2\le q_2\le Q}\sum_{2/q_2\le q_1\le Q_0/q_2}\sideset{}{^*}\sum_{\chi(q_1)}\int_X^{2X}\left|\sum_{x<n\le x+q_1q_2\Delta}\Lambda(n)\chi(n)\right|^2 dx \qquad \text{(A.2.1)}\\ \ll\frac{Q^3\Delta^2X}{\log^A X},\end{aligned}$$

where $\sum^*$ indicates that we restrict the sum to primitive characters. We now split the sum over the modulus q into $O(\log X)$ dyadic intervals $[Q_1, 2Q_1]$ with $2/q_2 \le Q_1 \le Q/q_2$. Then (A.2.1) is reduced to showing that

$$\sum_{Q_1\le q_1\le 2Q_1}\sideset{}{^*}\sum_{\chi(q_1)}\int_X^{2X}\left|\sum_{x<n\le x+q_1q_2\Delta}\Lambda(n)\chi(n)\right|^2 dx \ll \frac{q_2^2Q_1^2\Delta^2X}{\log^{A+1}X}.$$

Let $Q_1 = \Delta X^{-\beta}(\log X)^{-A-1} = X^{\theta-\beta}(\log X)^{-A-1}$, so that $\beta \in [1/6+\varepsilon/2, \theta]$.

We will next apply Perron's formula to reduce the problem to finding cancellation in Dirichlet polynomials, then use the Heath-Brown identity to decompose the long polynomial which arises into shorter polynomials.

Lemma A.2.2. *Let $A \ge 1$ and $\varepsilon \in (0, \frac{1}{3}]$ be fixed. Let $X \ge 1$ and $\Delta = X^\theta$ with $\frac{1}{6}+2\varepsilon \le \theta \le 1$. Let $Q_1 = \Delta X^{-\beta}\log^{-A-1} X$ with $\beta \in [1/6+\varepsilon/2, \theta]$. Let $k_0 \ge 3$ be an integer and for $s \in \mathbb{C}$ and χ mod q_1 define $G_j(s, \chi) =$*

$\prod_{i\le J_j} F_i(s,\chi)$ *with* $J_j \le 2k_0$*, where each* $F_i(s,\chi)$ *is of the form*

$$\sum_{N_i<n\le 2N_i} \frac{\chi(n)\log n}{n^s}, \sum_{N_i<n\le 2N_i} \frac{\chi(n)}{n^s}, \ \ or \ \ \sum_{N_i<n\le 2N_i} \frac{\chi(n)\mu(n)}{n^s}$$

and the lengths satisfy $N_1\cdots N_{J_j} = X^{1+o(1)}$*,* $N_i \gg \exp\left(\frac{\log X}{\log\log X}\right)$*. Then, we have that*

$$\begin{aligned}&\sum_{Q_1<q_1\le 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_X^{2X} \left| \sum_{x<n\le x+q_1q_2\Delta} \Lambda(n)\chi(n) \right|^2 dx \\ &\ll X^2(\log X)^{2D+3} \min\left\{\frac{q_2Q_1\Delta}{X}, \frac{1}{T}\right\}^2 \sum_{j=1}^{L} \int_{T\le|t|+1\le 2T} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt.\end{aligned} \tag{A.2.2}$$

for some $D>0$ *and* $L \le (\log X)^D$*.*

Proof. We apply Perron's formula to the sum over n so that

$$\begin{aligned}&\sum_{Q_1<q_1\le 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_X^{2X} \left| \sum_{x<n\le x+q_1q_2\Delta} \Lambda(n)\chi(n) \right|^2 dx \\ =&\frac{1}{4\pi^2} \sum_{Q_1<q_1\le 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_X^{2X} \left| \int_{\substack{\Re(s)=1/2 \\ |\Im(s)|\le T_0}} F(s,\chi)\frac{(1+q_1q_2\Delta/x)^s-1}{s} x^s ds \right|^2 dx \\ &+ O(Q_1^2 X^{1+\varepsilon/5}),\end{aligned}$$

where for $s\in\mathbb{C}$ and χ mod q we define

$$F(s,\chi) := \sum_{n=1}^{\infty} \frac{\Lambda(n)\chi(n)}{n^s}.$$

We choose T_0 to be the unique integer of the form $2^m - 1 \in (\frac{X}{2}, X]$ and split the integral with respect to t into $O(\log X)$ dyadic intervals, so that

$$\begin{aligned}
&\sum_{Q_1<q_1\leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_X^{2X} \left| \sum_{x<n\leq x+q_1q_2\Delta} \Lambda(n)\chi(n) \right|^2 dx \\
\ll &(\log X)^2 \sum_{Q_1<q_1\leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_X^{2X} \left| \int_{\substack{\Re(s)=1/2 \\ T\leq|\Im(s)|+1\leq 2T}} F(s,\chi) \frac{(1+q_1q_2\Delta/x)^s - 1}{s} x^s ds \right|^2 dx \\
&+ Q_1^2 X^{1+\varepsilon/5}.
\end{aligned} \tag{A.2.3}$$

Next, we expand the square, apply the bound

$$\frac{(1+q_1q_2\Delta/x)^{1/2+it} - 1}{1/2+it} \ll \min\left\{ \frac{q_2 Q\Delta}{X}, \frac{1}{1+|t|} \right\}$$

and integrate with respect to x to get that the integral in (A.2.3) is bounded by

$$\begin{aligned}
&\int_X^{2X} \left| \int_{\substack{\Re(s)=1/2 \\ T\leq|\Im(s)|+1\leq 2T}} F(s,\chi) \frac{(1+q_1q_2\Delta/x)^s - 1}{s} x^s ds \right|^2 dx \\
&\ll X^2 \min\left\{ \frac{q_2Q_1\Delta}{X}, \frac{1}{T} \right\}^2 \int_{T\leq|t|+1\leq 2T} |F(\tfrac{1}{2}+it_1,\chi)|^2 \int_{-2T}^{2T} \frac{1}{1+|t_1-t_2|} dt_2 dt_1 \\
&\ll X^2 (\log X) \min\left\{ \frac{q_2Q_1\Delta}{X}, \frac{1}{T} \right\}^2 \int_{T\leq|t|+1\leq 2T} |F(\tfrac{1}{2}+it,\chi)|^2 dt.
\end{aligned} \tag{A.2.4}$$

We now apply Heath-Brown's decomposition (Lemma 2.3.26) with $k_0 \geq 3$

followed by the triangle inequality to get that

$$\int_{T\le|t|+1\le 2T} |F(\tfrac{1}{2}+it,\chi)|^2 dt \ll (\log X)^{2D} \sum_{j=1}^{L} \int_{T\le|t|+1\le 2T} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt,$$

such that $L \le \log^D X$ for some $D > 0$, and $G_j(s,\chi) = \prod_{i\le J_j} F_i(s,\chi)$ with $J_j \le 2k_0$ and each $F_i(s,\chi)$ is of the form

$$\sum_{N_i<n\le 2N_i} \frac{\chi(n)\log n}{n^s}, \sum_{N_i<n\le 2N_i} \frac{\chi(n)}{n^s}, \text{ or } \sum_{N_i<n\le 2N_i} \frac{\chi(n)\mu(n)}{n^s}.$$

The lengths satisfy $N_1\cdots N_{J_j} = X^{1+o(1)}$, $N_i \gg \exp\left(\frac{\log X}{\log\log X}\right)$. □

We will again use mean and large value theorems to bound this mean value of the Dirichlet polynomial $G_j(\frac{1}{2}+it,\chi)$.

Lemma A.2.3. *Let $A \ge 1$ and $\varepsilon \in (0, \frac{1}{3}]$ be fixed. Let $X \ge 1$ and $\Delta = X^\theta$ with $\frac{1}{6} + 2\varepsilon \le \theta \le 1$. Let $Q_1 = \Delta X^{-\beta}(\log X)^{-A-1}$ with $\beta \in [1/6+\varepsilon/2, \theta]$. Let $k_0 \ge 3$ be an integer and for $s \in \mathbb{C}$ and χ mod q_1 define $G_j(s,\chi) = \prod_{i\le J_j} F_i(s,\chi)$ with $J_j \le 2k_0$, where each $F_i(s,\chi)$ is of the form*

$$\sum_{N_i<n\le 2N_i} \frac{\chi(n)\log n}{n^s}, \sum_{N_i<n\le 2N_i} \frac{\chi(n)}{n^s}, \text{ or } \sum_{N_i<n\le 2N_i} \frac{\chi(n)\mu(n)}{n^s}$$

and the lengths satisfy $N_1\cdots N_{J_j} = X^{1+o(1)}$, $N_i \gg \exp\left(\frac{\log X}{\log\log X}\right)$. Let $1 \le$

$j \le L$ be a fixed integer. Then, we have that

$$\sum_{Q_1 < q_1 \le 2Q_1} \sum_{\chi(q_1)}^{*} \int_{T \le |t|+1 \le 2T} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt \\ \ll \max\left\{1, \frac{q_2 Q_1 \Delta T}{X}\right\}^2 \frac{X}{(\log X)^{A+3D+4}}.$$

Assuming this Lemma, we are able to prove Lemma 2.2.6.

Proof of Lemma 2.2.6 assuming Lemma A.2.2. First, by Lemma A.2.1, we are able to reduce the problem to handling the contribution of the primitive characters since

$$\sum_{q \le Q} \sum_{\chi(q)} \int_X^{2X} \left| \sum_{x < n \le x+q\Delta} (\Lambda(n)\chi(n) - \delta_\chi) \right|^2 dx \\ \ll \sum_{q \le Q} \sum_{\chi(q_1)}^{*} \int_X^{2X} \left| \sum_{x < n \le x+q\Delta} \Lambda(n)\chi(n) \right|^2 dx + \frac{Q^3 \Delta^2 X}{\log^A X} + XQ^2 \log^4 X.$$

We split the sum over $q \le Q$ into $O(\log X)$ dyadic intervals $[Q_1, 2Q_1]$ with $2/q_2 \le Q_1 \le Q/q_2$. Then, we apply Lemma A.2.2 to obtain

$$\sum_{q \le Q} \sum_{\chi(q)}^{*} \int_X^{2X} \left| \sum_{x < n \le x+q\Delta} \Lambda(n)\chi(n) \right|^2 dx \\ \ll X^2 (\log X)^{2D+3} \min\left\{\frac{q_2 Q_1 \Delta}{X}, \frac{1}{T}\right\}^2 \\ \times \sum_{j=1}^{L} \sum_{q \le Q} \sum_{\chi(q_1)}^{*} \int_{T \le |t|+1 \le 2T} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt.$$

For each $1 \le j \le L$, we apply Lemma A.2.3. Since there are $L \le \log^D X$

terms, this is bounded by

$$\begin{aligned}&\ll \min\left\{\frac{q_2Q_1\Delta}{X},\frac{1}{T}\right\}^2\times\max\left\{1,\frac{q_2Q_1\Delta T}{X}\right\}^2\frac{X^3}{\log^{A+1}X}\\&\ll\frac{q_2^2Q_1^2\Delta^2X}{(\log X)^{A+1}}.\end{aligned}$$

Then, combining the contribution of the $O(\log X)$ dyadic intervals $Q_1\leq q\leq 2Q_1$, we have that

$$\sum_{q\leq Q}\sum_{\chi(q)}\int_X^{2X}\left|\sum_{x<n\leq x+q\Delta}(\Lambda(n)\chi(n)-\delta_\chi)\right|^2dx\ll\frac{Q^3\Delta^2X}{\log^A X},$$

as required. □

As in Chapter 2, to prove Lemma A.2.3 we split the domain of integration according to the size of the Dirichlet polynomials $F_i(s,\chi)$. We fix a $1\leq j\leq L$. Fix some integers $U_1,\ldots,U_{J_j}$ such that $1\leq U_1\ll\sqrt{N_1}\log N_1$ (supposing w.l.o.g. that the first polynomial has the $\log n$ coefficient) and $1\leq U_j\ll\sqrt{N_i}$ for $j=2,\ldots,J_j$ and set $U:=U_1\cdots U_{J_j}$. Define $\mathcal{P}(\chi,T,\mathbf{U})$ to be

$$\left\{t\in\mathbb{R}:T\leq|t|+1\leq 2T,U_j\leq|F_i(\tfrac{1}{2}+it,\chi)|\leq 2U_i,1\leq i\leq J_j\right\},\quad\text{(A.2.5)}$$

so there are up to $O(\log^{2k_0}X)$ such subsets to consider. Therefore, in order to prove Lemma A.2.3, we need to show that

$$\sum_{Q_1<q_1\le 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{\mathcal{P}(\chi,T,\mathbf{U})} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt \\ \ll \max\left\{1, \frac{q_2 Q\Delta T}{X}\right\}^2 \frac{X}{(\log X)^{A+4k_0+3D+4}}.$$

There are assumptions we can make to simplify the problem about the size of the Dirichlet polynomials U, the length of integration T, and the lengths of the polynomials N_i.

Assumption 1. We have

$$U \le \min\left\{\frac{\sqrt{X}\log^B X}{\sqrt{Q_1}}, \frac{\sqrt{X}}{\log^C X}\right\}.$$

We first show that we may restrict to $U \le (\log X)^B\sqrt{X/Q_1}$; if instead $U > (\log X)^B\sqrt{X/Q_1}$, then by (A.2.4) and $|G_j(\frac{1}{2}+it,\chi)|^2/U^2 \ge 1$, we have

$$\sum_{Q_1<q_1\le 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{\mathcal{P}(\chi,T,\mathbf{U})} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt \\ \ll \frac{1}{U^2} \sum_{Q_1<q_1\le 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{-2T}^{2T} |G_j(\tfrac{1}{2}+it,\chi)|^4 dt \\ \ll \frac{Q_1}{X\log^{2B} X} \sum_{Q_1<q_1\le 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{-2T}^{2T} |G_j(\tfrac{1}{2}+it,\chi)|^4 dt$$

and, therefore, applying the mean value theorem (Lemma 2.3.27) we have

that

$$\sum_{Q_1 < q_1 \leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{\mathcal{P}(\chi,T,\mathbf{U})} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt \ll \frac{Q_1(Q_1^2 T + X^2)}{X(\log X)^{2B-16k_0^2-5}}.$$

Overall, since $Q_1 \leq Q/q_2$ the contribution to (A.2.2) of the above is

$$\min\left\{\frac{q_2 Q_1 \Delta}{X}, \frac{1}{T}\right\}^2 \frac{Q_1 X(Q_1^2 T + X^2)}{(\log X)^{2B-16k_0^2-4k_0-3D-9}} \ll \frac{Q^3 \Delta^2 X}{(\log X)^{2B-16k_0^2-4k_0-3D-9}},$$

which is acceptable as we assume that $B > 0$ is sufficiently large in terms of A and k_0.

The second claim can be proved in the same way as in [21]. If we have that $Q_1 > (\log X)^{2(B+C)}$, then we already have that $U \leq (\log X)^B \sqrt{X/Q_1} \leq \sqrt{X}/(\log X)^C$, so we now assume that $Q_1 \leq (\log X)^{2(B+C)}$. We define

$$\mathcal{I} := \{1 \leq i \leq J_j : N_i > X^{\delta^2}\}$$

where $\delta > 0$ is small and fixed. We may suppose that δ satisfies $\delta^2 < \frac{1}{J_j}$ so that $|\mathcal{I}| \geq 1$. If we let $i \in \mathcal{I}$ and χ be a primitive character modulo some $q_1 \in (Q_1, 2Q_1]$, then we have for all $t \in [-X, X]$ that

$$|F_i(\tfrac{1}{2}+it,\chi)| \ll \frac{\sqrt{N_i}}{(\log X)^{(C+1)/J_j}},$$

since F_i is prime-factored (by Lemma 2.3.26). Thus we may assume $U_i \ll$

$\sqrt{N_i}/(\log X)^{(C+1)/J_j}$ for all $i \in \mathcal{I}$ and since $|\mathcal{I}| \geq 1$ we have that

$$U = \prod_{j=1}^{J_j} U_i \ll \frac{1}{(\log X)^{C+1}} \prod_{j=1}^{J_j} \sqrt{N_i} \ll \frac{\sqrt{X}}{\log^C X},$$

as required.

Assumption 2. We have

$$T \leq \frac{X(\log X)^{A/2+8k_0^2+3D/2+3}}{q_2 Q_1 \Delta}. \tag{A.2.6}$$

Suppose we have that $T > \frac{X(\log X)^{A/2+8k_0^2+3D/2+3}}{q_2 Q_1 \Delta}$, then an application of the mean value theorem (Lemma 2.3.27) shows that

$$\begin{aligned}
&\sum_{Q_1 < q_1 \leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{\mathcal{P}(\chi,T,\mathbf{U})} |G_j(\tfrac{1}{2}+it,\chi)|^2 dt \\
&\ll (Q_1^2 T + X)(\log X)^{4k_0^2+2} \\
&\ll \max\left\{1, \frac{q_2 Q_1 \Delta T}{X}\right\}^2 \frac{X}{(\log X)^{A+4k_0+3D+4}},
\end{aligned}$$

which is acceptable. Note that since $Q_1 = \Delta X^{-\beta}(\log X)^{-A-1}$, we therefore may assume that

$$Q_1^2 T \leq \frac{X^{1-\beta}}{q_2}(\log X)^{-A/2+8k_0^2+3D/2+2}. \tag{A.2.7}$$

Assumption 3. When $F_i(s,\chi)$ has the coefficient $\log n$ or 1, we have

$$N_i \leq (Q_1 T)^{1/2+\delta} \log^B X. \tag{A.2.8}$$

Suppose that $N_i > (Q_1 T)^{1/2+\delta} \log^B X$ for some such F_i. For simplicity, assume that the coefficient is 1, with the $\log n$ case being handled analogously. Fix $\delta_1 > 0$ such that $\frac{1+\delta_1}{2+\delta_1} = 1/2+\delta$ and define $M := \max\{1, (2Q_1 T/N_i)^{1+\delta_1}\}$. By Lemma A.1.2 we have that

$$F_i(\tfrac{1}{2}+it,\chi) \ll \int_{-\infty}^{\infty} \left| \sum_{n \leq M} \frac{\chi(n)}{n^{1/2+i(t+u)}} \right| \frac{du}{1+u^2}.$$

Note that if we were in the log coefficient case the only difference here would be an additional factor of $\log X$. Substituting the above and applying

Cauchy-Schwarz to the integral with respect to u we have that

$$\sum_{Q_1<q_1\le 2Q_1}\sideset{}{^*}\sum_{\chi(q_1)}\int_{\mathcal{P}(\chi,T,\mathbf{U})}|G_j(\tfrac{1}{2}+it,\chi)|^2dt$$

$$\ll \sum_{Q_1<q_1\le 2Q_1}\sideset{}{^*}\sum_{\chi(q_1)}\int_{-2T}^{2T}\left(\int_{\mathbb{R}}\left|\sum_{n\le M}\frac{\chi(n)}{n^{1/2+i(t+u)}}\right|\frac{du}{1+u^2}\right)^2\prod_{\substack{1\le \ell\le J_j\\ \ell\ne j}}|F_\ell(\tfrac{1}{2}+it,\chi)|^2dt$$

$$\ll \int_{\mathbb{R}}\int_{\mathbb{R}}\sum_{Q_1<q_1\le 2Q_1}\sideset{}{^*}\sum_{\chi(q_1)}\int_{-2T}^{2T}\left|\sum_{n\le M}\frac{\chi(n)}{n^{1/2+i(t+u_1)}}\right|^2$$

$$\times\prod_{\substack{1\le \ell\le J_j\\ \ell\ne j}}|F_\ell(\tfrac{1}{2}+it,\chi)|^2\frac{dtdu_1du_2}{(1+u_1^2)(1+u_2^2)}$$

$$\ll \int_{-\infty}^{\infty}\sum_{Q_1<q_1\le 2Q_1}\sideset{}{^*}\sum_{\chi(q_1)}\int_{-2T}^{2T}\left|\sum_{n\le M}\frac{\chi(n)}{n^{1/2+i(t+u)}}\right|^2\prod_{\substack{1\le \ell\le J_j\\ \ell\ne j}}|F_\ell(\tfrac{1}{2}+it,\chi)|^2\frac{dtdu}{1+u^2}.$$

Applying the mean value theorem (Lemma 2.3.27) gives that the above is bounded by

$$\ll\left(\frac{XM}{N_i}+Q_1^2T\right)(\log X)^{4k_0^2+2}$$
$$\ll\left(\frac{X}{N_i}(1+(Q_1T/N_i)^{1+\delta_1})+Q_1^2T\right)(\log X)^{4k_0^2+2}. \tag{A.2.9}$$

By the definition of δ_1 and our assumption $N_i>(Q_1T)^{1/2+\delta}\log^B X$, we have that

$$\frac{(Q_1T)^{1+\delta_1}}{N_i^{2+\delta_1}}<\frac{(Q_1T)^{1+\delta_1-(2+\delta_1)(1/2+\delta)}}{(\log X)^{B(2+\delta_1)}}<\frac{1}{(\log X)^{B(2+\delta_1)}}.$$

Applying (A.2.7), we have that (A.2.9) is bounded by

$$\ll \left(\frac{X}{N_i} + Q_1^2 T\right) (\log X)^{4k_0^2+2}$$
$$\ll \left(\frac{X}{\log^B X} + X^{1-\beta}(\log X)^{-A/2+8k_0^2+3D/2+2}\right) (\log X)^{4k_0^2+2},$$

which is acceptable as $B > 0$ is taken sufficiently large in terms of A, k_0 and D, and thus we may assume (A.2.8) from now on.

Therefore combining Assumptions 2 (A.2.6) and 3 (A.2.8), if F_i has coefficient 1 or $\log n$ we have that

$$N_i \leq (Q_1 T)^{1/2+\delta} \log^B X \leq X^{(1/2+\delta)(1-\theta)+o(1)} \leq X^{1/2-\delta}$$

as long as $\delta > 0$ is sufficiently small. Our application of the Heath-Brown identity ensures that $N_i \ll X^{1/3}$ for the F_i with coefficient $\mu(n)$, so that $K \geq 3$.

We would like to apply results on large values of Dirichlet polynomials (Lemma A.1.1), however the set $\mathcal{P}(\chi, T, \mathbf{U})$ is not necessarily well-spaced. We first construct

$$\mathcal{Z}(\chi, T, \mathbf{U}) := \{n \in \mathbb{Z} : [n, n+1] \cap \mathcal{P}(\chi, T, \mathbf{U}) \neq \emptyset\} =: \{n_1, \ldots, n_r\},$$

say, such that $n_1 < \cdots < n_r$. For each $1 \leq i \leq r$ choose one point $t_i \in [n_i, n_i + 1] \cap \mathcal{P}(\chi, T, \mathbf{U})$. This set of points $\{t_1, \ldots, t_r\}$ is not yet necessarily well-spaced; if n_i and n_{i+1} are consecutive integers it may be the case that

$|t_i - t_{i+1}| < 1$. We now split this set of points $\{t_1, \ldots, t_r\}$ into two well-spaced sets by separately considering the odd and even indexed points. In particular we define

$$\mathcal{R}_m(\chi, T, \mathbf{U}) := \{t_i : 1 \le i \le r, i \equiv m \mod 2\}$$

for $m = 0, 1$ and

$$\mathcal{R}_m(T, \mathbf{U}) := \bigcup_{Q_1 < q_1 \le 2Q_1} \bigcup_{\substack{\chi(q_1) \\ \chi \text{ primitive}}} \{(\chi, t) : t \in \mathcal{R}_m(\chi, T, \mathbf{U})\} \qquad \text{(A.2.10)}$$

again for $m = 0, 1$. We are now considering the polynomial $G_j(s, \chi)$ over well-spaced sets of points, so we can now apply results on large values of Dirichlet polynomials to prove a bound for the size of $|\mathcal{R}_m(T, \mathbf{U})|$ for $m = 0, 1$, which will enable us to complete the proof.

Lemma A.2.4. *Let $A > 0$ and $k_0 \ge 3$ be an integer. Let $U := U_1 \cdots U_{2k}$ satisfy Assumption 1. Then*

$$|\mathcal{R}_m(T, \mathbf{U})| \ll \frac{X}{U^2 (\log X)^{A + 4k_0 + 3D + 4}},$$

for $m = 0, 1$.

Proof. Let $m = 0$ or 1. We separate into three cases according to the sizes of U_i with $i \in \mathcal{I}$.

Case 1. We assume there is $i \in \mathcal{I}$ such that $U_i > \sqrt{N_i}/(\log^B X)$.

In this case, we let r be a positive integer such that $U_i^{2r} \ge Q_1^2 T$. We now apply Lemma A.1.1 with $A(s,\chi) = F_i(s,\chi)^r$ which has length N_i^r, and we take $V = U_i^r$. The coefficients of $F_i(s,\chi)^r$ are bounded by $\ll (\log n)^r d_r(n)$, so we have that

$$|\mathcal{R}_m(T,\mathbf{U})| \ll \left(\frac{N_i^r}{U_i^{2r}} + \frac{Q_1^2 T}{U_i^{2r}}\right)(\log X)^{3r^2+24}$$
$$\ll \left((\log X)^{2rB} + 1\right)(\log X)^{3r^2+6r+18}.$$

Now, recalling that $U \leq \sqrt{X}/(\log^C X)$, we have

$$|\mathcal{R}_m(T,\mathbf{U})| \ll \frac{X}{U^2(\log X)^{2C-2rB-3r^2-6r-18}}.$$

Taking $C > 0$ sufficiently large in terms of B and r gives us the required bound.

Case 2. We have that $U_i \leq \sqrt{N_i}/(\log^B X)$ for all $i \in \mathcal{I}$ and there is some $j \in \mathcal{I}$ such that $U_j \leq X^{\beta/2}/(\log^B X)$.

We again apply Lemma A.1.1, but this time with $A(s,\chi) = \prod_{\ell \neq j} F_\ell(s,\chi)$ which has length X/N_j, and take $V = U/U_j$. The coefficients of $\prod_{\ell \neq j} F_\ell(s,\chi)$ are bounded by $\ll (\log n) d_{2k_0-1}(n)$ so that

$$|\mathcal{R}_m(T,\mathbf{U})| \ll \left(\frac{XU_j^2}{U^2 N_j} + \frac{Q_1^2 T U_j^2}{U^2}\right)(\log X)^{12k_0^2+24}$$
$$\ll \frac{X}{U^2}\left(\frac{1}{\log^{2B} X} + \frac{(\log X)^{-A/2+8k_0^2+3D/2+2}}{\log^{2B} X}\right)(\log X)^{12k_0^2+24},$$

where we have used (A.2.7). This is acceptable provided B is sufficiently large in terms of A, k_0 and D.

Case 3. We have $U_i \in [X^{\beta/2}/(\log^B X), \sqrt{N_i}/(\log^B X)]$ for all $i \in \mathcal{I}$.

For now, we assume instead that

$$|\mathcal{R}_m(T,\mathbf{U})| \geq \frac{X}{U^2(\log X)^{A+4k_0+3D+4}}.$$

Once again we apply Lemma A.1.1 with $A(s,\chi) = \prod_{\ell \neq i} F_\ell(s,\chi)$ for each $i \in \mathcal{I}$, so that

$$\begin{aligned} |\mathcal{R}_m(T,\mathbf{U})| &\ll \left(\frac{XU_i^2}{N_iU^2} + \frac{Q_1^2TXU_i^6}{N_iU^6}\right)(\log X)^{12k_0^2+24} \\ &\ll \frac{X}{U^2(\log X)^{2B-12k_0^2-24}} + \frac{X^{2-\beta}U_i^6(\log X)^{-A/2+20k_0^2+3D/2+26}}{N_iU^6}. \end{aligned}$$

As B is sufficiently large in terms of A, k_0 and D the first term does not provide a contradiction. However, the second term gives that

$$\frac{U^4}{U_i^6} \ll \frac{X^{1-\beta}(\log X)^{A/2+28k_0^2+3D/2+30}}{N_i}.$$

Taking the product over all $i \in \mathcal{I}$, we have that

$$\prod_{i\in\mathcal{I}} \frac{1}{N_i} \ll \frac{1}{N_1 \cdots N_{J_j}} \prod_{i \notin \mathcal{I}} N_i \ll X^{k\delta^2-1}.$$

Therefore, with $I := |\mathcal{I}|$, we have that

$$U^{4I-6} \ll X^{I(1-\beta)+k\delta^2-1}(\log X)^{I(A/2+28k_0^2+3D/2+30)}.$$

We also have that $U \geq \prod_{i \in \mathcal{I}} U_i \geq X^{I\beta/2}(\log X)^{2BI}$, so that

$$X^{I\beta(2I-3)}(\log X)^{4BI(2I-3)} \ll X^{I(1-\beta)+J_j\delta^2-1}(\log X)^{I(A/2+28k_0^2+3D/2+30)}$$
$$\ll X^{I(1-\beta)+J_j\delta^2-1+\delta^2}.$$

In particular, comparing the powers of X, we must have that $2\beta I(I-1) \leq I - 1 + (J_j+1)\delta^2$ and therefore

$$\beta \leq \frac{1}{2I} + \frac{\delta^2(J_j+1)}{2I(I-1)}.$$

Recalling that $I \geq 3$ and taking δ sufficiently small in terms of ε and J_j, this contradicts that $\beta \geq 1/6 + \varepsilon/2$. Therefore, we must have in this case that

$$|\mathcal{R}_m(T, \mathbf{U})| \leq \frac{X}{U^2(\log X)^{A+4k_0+3D+4}},$$

as required. □

We are now able to complete the proof of Lemma A.2.3 and therefore the proof of Lemma 2.2.6.

Proof of Lemma A.2.3. Fix an integer $1 \leq j \leq L$. We first split the domain of integration according to the size of the factors $F_i(s,\chi)$ of $G_j(s,\chi)$. As in (A.2.5), we fix some integers $U_1, \ldots, U_{J_j}$ such that $1 \leq U_1 \ll \sqrt{N_1}\log N_1$ and $1 \leq U_j \ll \sqrt{N_i}$ for $j = 2, \ldots, J_j$, and set $U := U_1 \cdots U_{J_j}$. Define $\mathcal{P}(\chi, T, \mathbf{U})$

to be

$$\left\{t \in \mathbb{R} : T \leq |t| + 1 \leq 2T, U_j \leq |F_i(\tfrac{1}{2} + it, \chi)| \leq 2U_i, 1 \leq i \leq J_j\right\}.$$

Then we have that

$$\begin{aligned}&\sum_{Q_1 < q_1 \leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{T \leq |t|+1 \leq 2T} |G_j(\tfrac{1}{2} + it, \chi)|^2 dt \\ &\ll (\log X)^{4k_0} \sum_{Q_1 < q_1 \leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{\mathcal{P}(\chi, T, \mathbf{U})} |G_j(\tfrac{1}{2} + it, \chi)|^2 dt,\end{aligned}$$

since there are up to $O(\log^{2k_0} X)$ sets $\mathcal{P}(\chi, T, \mathbf{U})$ to consider. We further split $\mathcal{P}(\chi, T, \mathbf{U})$ into the well-spaced sets $\mathcal{R}_m(T, \mathbf{U})$ for $m = 0, 1$ as in (A.2.10). We have that

$$\begin{aligned}&\sum_{Q_1 < q_1 \leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \int_{\mathcal{P}(\chi, T, \mathbf{U})} |G_j(\tfrac{1}{2} + it, \chi)|^2 dt \\ &\ll \sum_{Q_1 < q_1 \leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \sum_{i=1}^{r} |G_j(\tfrac{1}{2} + it_i, \chi)|^2 \\ &= \sum_{Q_1 < q_1 \leq 2Q_1} \sideset{}{^*}\sum_{\chi(q_1)} \left(\sum_{i \leq r/2} |G_j(\tfrac{1}{2} + it_{2i}, \chi)|^2 + \sum_{i \leq r/2} |G_j(\tfrac{1}{2} + it_{2i-1}, \chi)|^2 \right) \\ &\ll U^2 \left(|\mathcal{R}_0(T, \mathbf{U})| + |\mathcal{R}_1(T, \mathbf{U})| \right).\end{aligned}$$

By Lemma A.2.4, this is

$$\ll \frac{X}{(\log X)^{A + 4k_0 + 3D + 4}},$$

which immediately gives the required bound. □

Bibliography

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

[2] E. Bombieri. On twin almost primes. *Acta Arith.*, 28(2):177–193, 1975/76.

[3] S. Chowla. The Riemann hypothesis and Hilbert's tenth problem. *Norske Vid. Selsk. Forh. (Trondheim)*, 38:62–64, 1965.

[4] H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000. Revised and with a preface by Hugh L. Montgomery.

[5] N. Debouzy. Twins almost prime under a Elliott-Halberstam's conjecture. *arXiv e-prints*, page arXiv:1907.06393, July 2019.

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

[7] N. Evans. Correlations of almost primes. *Mathematical Proceedings of the Cambridge Philosophical Society*, page 1–44, 2022.

[8] J. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010.

[9] A. Fujii, P. X. Gallagher, and H. L. Montgomery. Some hybrid bounds for character sums and Dirichlet L-series. In *Topics in number theory (Proc. Colloq., Debrecen, 1974)*, pages 41–57. Colloq. Math. Soc. János Bolyai, Vol. 13, 1976.

[10] P. X. Gallagher. A large sieve density estimate near $\sigma = 1$. *Invent. Math.*, 11:329–339, 1970.

[11] D. A. Goldston, S. W. Graham, J. Pintz, and C. Y. Yıldırım. Small gaps between products of two primes. *Proc. Lond. Math. Soc. (3)*, 98(3):741–774, 2009.

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

[13] J. Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Societé mathematique de France*, 24:199–220, 1896.

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

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

[16] G. Harman, N. Watt, and K. Wong. A new mean-value result for Dirichlet L-functions and polynomials. *Q. J. Math.*, 55(3):307–324, 2004.

[17] D. R. Heath-Brown. Prime numbers in short intervals and a generalized Vaughan identity. *Canadian J. Math.*, 34(6):1365–1377, 1982.

[18] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[19] M. Jutila. Zero-density estimates for L-functions. *Acta Arith.*, 32(1):55–62, 1977.

[20] O. Klurman, A. P. Mangerel, and J. Teräväinen. Multiplicative functions in short arithmetic progressions. *arXiv e-prints*, page arXiv:1909.12280, Sept. 2019.

[21] D. Koukoulopoulos. Primes in short arithmetic progressions. *Int. J. Number Theory*, 11(5):1499–1521, 2015.

[22] A. F. Lavrik. On the twin prime hypothesis of the theory of primes by the method of I. M. Vinogradov. *Soviet Math. Dokl.*, 1:700–702, 1960.

[23] J. D. Lichtman. Averages of the Möbius Function on Shifted Primes. *Q. J. Math.*, 73(2):729–757, 11 2021.

[24] J. D. Lichtman and J. Teräväinen. On the Hardy–Littlewood–Chowla conjecture on average. *Forum Math. Sigma*, 10:e57, 2022.

[25] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math. (2)*, 183(3):1015–1056, 2016.

[26] K. Matomäki, M. Radziwiłł, and T. Tao. An averaged form of Chowla's conjecture. *Algebra Number Theory*, 9(9):2167–2196, 2015.

[27] K. Matomäki, M. Radziwiłł, and T. Tao. Correlations of the von Mangoldt and higher divisor functions I. Long shift ranges. *Proc. Lond. Math. Soc. (3)*, 118(2):284–350, 2019.

[28] K. Matomäki, M. Radziwiłł, and T. Tao. Correlations of the von Mangoldt and higher divisor functions II: divisor correlations in short ranges. *Math. Ann.*, 374(1-2):793–840, 2019.

[29] K. Matomäki and J. Teräväinen. Almost primes in almost all short intervals II. *arXiv e-prints*, page arXiv:2207.05038, July 2022.

[30] J. Maynard. Small gaps between primes. *Ann. of Math. (2)*, 181(1):383–413, 2015.

[31] F. Mertens. Ueber einige asymptotische Gesetze der Zahlentheorie. *J. Reine Angew. Math.*, 77:289–338, 1874.

[32] H. Mikawa. On prime twins. *Tsukuba J. Math.*, 15(1):19–29, 1991.

[33] H. L. Montgomery. *Topics in multiplicative number theory*. Lecture Notes in Mathematics, Vol. 227. Springer-Verlag, Berlin-New York, 1971.

[34] H. L. Montgomery. *Ten lectures on the interface between analytic number theory and harmonic analysis*. Number 84. American Mathematical Soc., 1994.

[35] H. L. Montgomery and R. C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2007.

[36] D. H. J. Polymath. Variants of the Selberg sieve, and bounded intervals containing many primes. *Res. Math. Sci.*, 1:Art. 12, 83, 2014.

[37] T. Tao. *An epsilon of room, I: real analysis*, volume 1. American Mathematical Soc., 2010.

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

[39] E. C. Titchmarsh. A divisor problem. *Rend. Circ. Math. Palermo 54 (1930)*, 54(1):414–429, 1930.

[40] C.-J. Vallee-Poussin. de la (1896), recherches analytiques sur la théorie des nombres premiers, i–iii. *Ann. Soc. Sci. Bruxelles*, 20:183–256.

[41] J. G. Van der Corput. Sur la démonstration de l'hypothèse de Goldbach pour les nombres impairs, donnée par Vinogradow. *Bull. Soc. Math. France*, 67:87–99, 1939.

[42] R. C. Vaughan. *The Hardy-Littlewood method*, volume 80 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge-New York, 1981.

[43] D. Wolke. Über das Primzahl-Zwillingsproblem. *Math. Ann.*, 283(4):529–537, 1989.

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