

CLOSE ENCOUNTERS AMONG THE PRIMES¹

J. B. Friedlander* and H. Iwaniec**

**Department of Mathematics, University of Toronto, M5S 2E4 Canada*

***Department of Mathematics, Rutgers University, New Brunswick NJ 08903, USA*

e-mails: frdlndr@math.toronto.edu; iwaniec@comcast.net

(Received 26 December 2013; after final revision 27 April 2014;

accepted 4 May 2014)

ABSTRACT. This paper was written, apart from one technical correction, in July and August of 2013. The, then very recent, breakthrough of Y. Zhang [18] had revived in us an intention to produce a second edition of our book “Opera de Cribro”, one which would include an account of Zhang’s result, stressing the sieve aspects of the method. A complete and connected version of the proof, in our style but not intended for journal publication, seemed a natural first step in this project.

Following the further spectacular advance given by J. Maynard (arXiv:1311.4600, Nov 20, 2013), we have had to re-think our position. Maynard’s method, at least in its current form, proceeds from GPY in quite a different direction than does Zhang’s, and achieves numerically superior results. Consequently, although Zhang’s contribution to the distribution of primes in arithmetic progressions certainly retains its importance, the fact remains that much of the material in this paper would no longer appear in a new edition of our book. Because this paper contains some innovations that we do not wish to become lost, we have decided to make the work publicly available in its original format. We are extremely pleased to be able to include it in the current volume, commemorating the one hundred and twenty-fifth anniversary of the birth of Srinivasa Ramanujan.

Key words : Primes; arithmetic progressions; sieves

¹Research of JF supported in part by NSERC Grant A5123 and that of HI supported in part by NSF Grant DMS-1101574. We thank Leo Goldmakher and Pedro Pontes for their help with the physical appearance of the paper.

CONTENTS

1. Introduction	635
2. The GPY Restricted Sieve	638
2.1. Introduction	638
2.2. Shuffling the Sieve Terms	640
2.3. Choosing the Sieve Weights	642
2.4. Estimation of $G'_s(D)$	646
3. Proof of the Main Theorem	651
3.1. Sums over primes	651
3.2. Combinatorial arrangements	653
3.3. Primes in residue classes to divisor-rich moduli	655
3.4. Completing the proofs of the theorems	658
4. Bilinear Forms in Residue Classes	659
4.1. Classical results	659
4.2. Bilinear forms over product moduli	664
5. The Dispersion Method	668
5.1. Statement of results	668
5.2. Evaluation of \mathcal{D}_1	669
5.3. First estimation of \mathcal{D}_{11}	673
5.4. Second estimation of \mathcal{D}_{11}	677
5.5. Evaluation of \mathcal{D}_2	678
5.6. Estimation of \mathcal{D}_{21}	679
5.7. Evaluation of \mathcal{D}_3	679
6. Cropped Divisors in Residue Classes	680
6.1. Statement of results	680
6.2. Extracting the main term	682
6.3. Weyl's shift	682
6.4. Compacting variables	684
6.5. Correlation of Kloosterman sums	685
6.6. Completing the proof of Proposition 6.1	686
References	688

1. Introduction

The twin prime conjecture, dating from antiquity, predicts that there are infinitely many pairs of consecutive odd integers, both of which are prime. Letting p_n denote the n -th prime number, we can write this as

$$(1.1) \quad p_{n+1} - p_n = 2$$

infinitely often. One knows from the Prime Number Theorem the much weaker statement that

$$(1.2) \quad \liminf_{n \rightarrow \infty} \frac{p_{n+1} - p_n}{\log p_n} \leq 1,$$

because $\log p_n$ is the average gap. Even replacing the right hand side of (1.2) by smaller constants took a painfully long time with many contributors, beginning with P. Erdős [5] and continuing with [1], [14], [16], ...

Two major breakthroughs have occurred during the past few years. In the first of these, D. Goldston, J. Pintz, and C. Y. Yıldırım [13], building on earlier work [12], proved that indeed

$$(1.3) \quad \liminf_{n \rightarrow \infty} \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

Quite recently, Y. Zhang [18] has proven that the gaps between consecutive primes are absolutely bounded infinitely often, specifically that

$$(1.4) \quad p_{n+1} - p_n \leq 70\,000\,000$$

infinitely often. The proof draws on a wealth of ideas from previous sources while combining significant innovations of its own.

In the authors' book [11] we gave an account of the GPY results, following their general outline but introducing some innovations of our own, mainly for the purpose of clarifying the sieve ideas. It is our goal in the current writing to proceed in analogous fashion with the work [18] of Zhang.

There is a tremendous amount of interest in lowering the numerical bound in (1.4), particularly in the pending Polymath8 project ². Although we are aware of many possibilities, it is not our intention to pursue them here. Nevertheless, we should specify some bound; we prove the following

²now available at arXiv :1402.0811

Prime Gaps Theorem. *There are infinitely many primes p_n with*

$$(1.5) \quad p_{n+1} - p_n \leq 2\,448\,798.$$

In other words, we have:

Prime Cousins Theorem. *There exists a positive even number $c \leq 2\,448\,798$ such that p and $p + c$ are both primes infinitely often.*

Work on the small prime gaps problem has, ever since [1], depended crucially on the Bombieri-Vinogradov theorem for the distribution of prime numbers in residue classes to large moduli. This theorem (which has acted as a substitute for the Generalized Riemann Hypothesis in many applications) gives the bound

$$(1.6) \quad \sum_{q \leq Q} \max_{(a,q)=1} \left| \psi(x; q, a) - \frac{x}{\varphi(q)} \right| \ll x(\log x)^{-A}$$

valid for $x \geq 2$, $Q \leq x^{\frac{1}{2}}(\log x)^{-B}$ and any $A \geq 1$, where both B and the implied constant may depend on A .

In the fundamental paper [13] the authors found that the exponent $\frac{1}{2}$ was a key threshold. Had it been anything smaller than $\frac{1}{2}$, their arguments would have failed to obtain (1.3). On the other hand, could it be replaced by anything larger, say $\frac{1}{2} + \delta$, they could prove the much stronger bounded gaps result

$$(1.7) \quad p_{n+1} - p_n < c(\delta)$$

infinitely often with a positive constant $c(\delta)$ depending on δ .

The surprising, nevermind conditional, result (1.7) of Goldston-Pintz-Yıldırım rekindled enhanced interest in extending (1.6) type estimates with $Q = x^{\frac{1}{2}+\delta}$, going beyond the Riemann Hypothesis threshold. There had already been a number of earlier results of such quality, by E. Bombieri, E. Fouvry, J. B. Friedlander and H. Iwaniec in the series of papers [7], [6], [2], [3], [4]. These results did not hold uniformly over all reduced classes $a \pmod{q}$ but required one to fix an $a \neq 0$ and then sum over moduli $q \sim Q$, $(q, a) = 1$. As such, they could not be directly married to the GPY argument.

In the GPY scheme, one considers an admissible set (see the lines between (2.2) and (2.3))

$$\mathcal{H} = \{h_1, \dots, h_k\}$$

of positive integers and then applies a sieve to attempt to show that, for many n 's, the set $\{n - h_1, \dots, n - h_k\}$ contains more than one prime, hence at least two primes! In

applying a sieve to the polynomial $(X - h_1) \cdots (X - h_k)$, one has to consider, in the resulting congruence sums (see p. 125 of [11]) more than a single class, yet nowhere near as many as $\varphi(q)$ of them, and these relatively few classes are tied together in arithmetic fashion by the Chinese Remainder Theorem. Even so, the problem of proving a (1.6) type bound for the sum over all moduli $q \sim Q$ in a range $Q = x^{\frac{1}{2} + \delta}$ is still open. Zhang succeeds by considering a sum wherein the moduli run only over smooth numbers, that is, those with no large prime factors, and $a \pmod{q}$ is restricted to the roots of

$$\prod_{i \neq j} (X + h_i - h_j) \equiv 0 \pmod{q}.$$

The fact is that, in the GPY sieve, the restriction of the support of the sieve weights to smooth numbers does not greatly alter the strength of the sieve bounds. This latter phenomenon had been noted and utilized by Y. Motohashi and J. Pintz [17], and it was also known by others. We have articulated the relevant features in various demonstrations of sieve capability in a broader context. Zhang, however, was the one who made the whole thing work. The added flexibility in being able to factor the moduli in almost arbitrary proportion enabled Zhang to improve the bounds of the exponential sums which are required for this approach.

The paper is organized as follows. In Section 2 we study the restricted GPY sieve. We can borrow much of our arguments in this part from our book [11]. In Section 3 we state Proposition 3.4, which is the modified extension of the bound (1.6), and we begin its proof. We decompose sums over primes by means of combinatorial arrangements as did Zhang (which was in turn patterned after that in [2] where Heath-Brown's formula is chosen for the task, as it had been in [7]). This leads us to multiple Dirichlet convolution forms which are reduced to two basic types. The first of these, a bilinear form, is placed in a general context in Section 4 and then estimated in Section 5 using the dispersion method of Linnik à la [7] and [2], but with Zhang's modulus factorization playing a key role. Then, in Section 6 we consider smooth croppings, to various sizes, of the divisor function $\tau_3(n)$ in residue classes $a \pmod{q}$. The sums of these cropped divisors, together with the bilinear forms from Section 4, are building blocks for covering the sums over primes. We provide uniform estimates for single moduli q and every fixed reduced class $a \pmod{q}$ based on arguments originating in [10], but again with a key modification due to Zhang which takes advantage of q being a smooth modulus. Actually, one could exploit averaging over these special moduli q , and this would be

one of the sources for a lowering of the constant in (1.5), cf. E. Fouvry, E. Kowalski, and P. Michel [9].

2. The GPY Restricted Sieve

2.1. Introduction. The starting point in Zhang’s paper [18] is the lovely construction of the sifted sum by Goldston-Pintz-Yıldırım, which itself is based on Selberg’s sieve weights. Zhang applies to their construction the same weights λ_d , but with an extra restriction of the support to “smooth” numbers, that is; the numbers d which have no large prime divisors, say

$$(2.1) \quad d \mid P(z),$$

where $P(z)$ denotes the product of all primes $p < z$, and z is relatively small. If z is not extremely small, it turns out that such a restriction sacrifices very little in the GPY estimates. This phenomenon appears in any sieve having large dimension, and it is particularly pronounced in the Selberg Λ^2 -sieve, see Section 7.10 of [11]. On the other hand, the restriction (2.1) introduces a much-desired flexibility in the arrangement of bilinear forms out of primes and sieve weights. This extra feature permits stronger estimates for the exponential sums which control the remainder terms. The gain amply compensates for the loss caused by (2.1).

Much of the material required exists already in the book [11] in thinly disguised form; therefore, we shall make frequent reference to this material. We begin by recalling the GPY construction (in our format).

Let \mathcal{H} be a finite set of $k \geq 2$ positive integers,

$$(2.2) \quad Q(X) = \prod_{h \in \mathcal{H}} (X - h)$$

and $\omega(d)$ the number of roots of $Q(x) \equiv 0 \pmod{d}$. We say that the set \mathcal{H} is *admissible* if $\omega(p) < p$ for every prime p . In particular, it shows that \mathcal{H} does not contain two consecutive integers. Let Δ be a positive squarefree integer which is divisible by every prime factor of

$$(2.3) \quad \det \mathcal{H} = \prod_{h \neq h'} (h - h') = \prod_{h \in \mathcal{H}} |Q'(h)|.$$

Every prime $p \leq k$ divides Δ . If $p \nmid \Delta$, then $\omega(p) = k$. Associated with the set \mathcal{H} , we introduce two positive constants

$$(2.4) \quad \gamma(\mathcal{H}) = \prod_{p|\Delta} \left(1 - \frac{\omega(p)}{p}\right), \quad H(\mathcal{H}) = \prod_p \left(1 - \frac{\omega(p)}{p}\right) \left(1 - \frac{1}{p}\right)^{-k}.$$

We consider the sequence $\mathcal{A} = (a_n)$ with

$$(2.5) \quad a_n = \sum_{h \in \mathcal{H}} \Lambda(n - h) - \log n$$

for $n > \max \mathcal{H}$. Our goal is to prove that a_n is positive for infinitely many n . For every such n it follows that at least two of the shifted integers $n - h$ with $h \in \mathcal{H}$ are prime powers. Actually, the higher powers will make a negligible contribution, so we obtain gaps between primes bounded by the diameter of \mathcal{H} . We are not able to achieve this goal by summing the elements themselves, but can do so by weighting the sequence $\mathcal{A} = (a_n)$ in the style which is familiar from Selberg’s lower-bound sieve (see Section 7.6 of [11]). Specifically, we consider the following sum:

$$(2.6) \quad W(D, N) = \sum_{(Q(n), \Delta)=1} F(n/N) a_n \left(\sum_{d|Q(n)} \lambda_d \right),$$

where (λ_d) is an upper-bound sieve of level D . We shall specify the sieve later and it will be a Λ^2 -sieve. Here $F(t)$ is any fixed smooth function compactly supported in \mathbb{R}^+ , $F(t) \geq 0$. Therefore the summation over n in (2.6) is restricted to a segment $n \asymp N$. Of course, this smoothing device is introduced only for technical simplification; it has no bearing on the GPY construction (throughout, we shall ignore the dependence on F in any implied constants).

We shall prove that, for a judiciously chosen sieve (λ_d) ,

$$(2.7) \quad W(D, N) \asymp N(\log N)^{1-k}$$

provided $k = |\mathcal{H}|$ is sufficiently large; $k = 175561$ is fine. Hence, we shall derive (see Section 3.4) the following results

Main Theorem. *For all N sufficiently large in terms of the set \mathcal{H} we have*

$$(2.8) \quad \sum_{\substack{N < n < 2N \\ (Q(n), P(w))=1}} \sum_{h \neq h'} \Lambda(n - h) \Lambda(n - h') \asymp N(\log N)^{2-k},$$

where $w = N^{1/b}$ with b a sufficiently large constant which depends on k .

The extra restriction $(Q(n), P(w)) = 1$ in (2.8) means that every number $n - h$ with $h \in \mathcal{H}$ has no prime divisors smaller than $N^{1/b}$, hence it has at most b prime factors. Ignoring this restriction we deduce from (2.8) that there exists $c \in \mathcal{H} - \mathcal{H}$, $c \neq 0$, such that $\pi_c(N)$, the number of primes p in $N \leq p < 2N$ with $p + c$ being a prime, satisfies the lower bound

$$(2.9) \quad \pi_c(N) \gg N(\log N)^{-k}$$

for a positive proportion of N 's. A more precise expectation, the Hardy-Littlewood conjecture, predicts that for any even number $c \neq 0$ we have $\pi_c(N) \sim \mathfrak{S}N(\log N)^{-2}$ as $N \rightarrow \infty$, where \mathfrak{S} is a positive constant depending on c .

2.2. Shuffling the Sieve Terms. Given any sieve-weighted sum, its terms need to be rearranged so that the evaluation job boils down to that of the congruence sums for the sifting sequence. In this subsection we do so with (2.6) for the sequence (2.5). We also execute some easy parts of the relevant summations to clean and isolate those things which are more intricate. Throughout, we assume that λ_d are supported on squarefree numbers $d < D$, $(d, \Delta) = 1$ and

$$(2.10) \quad |\lambda_d| \leq \tau_3(d).$$

First, inserting (2.5) into (2.6) we get

$$(2.11) \quad W(D, N) = \sum_{h \in \mathcal{H}} V_h(D, N) - U(D, N) \log N$$

where we set

$$(2.12) \quad V_h(D, N) = \sum_{(Q(n), \Delta)=1} \Lambda(n - h) F\left(\frac{n}{N}\right) \left(\sum_{d|Q(n)} \lambda_d \right),$$

$$(2.13) \quad U(D, N) = \sum_{(Q(n), \Delta)=1} F\left(\frac{n}{N}\right) \frac{\log n}{\log N} \left(\sum_{d|Q(n)} \lambda_d \right).$$

The latter sum can be evaluated quickly by splitting into residue classes $\alpha \pmod{\Delta}$ and $\beta \pmod{d}$. We have

$$\begin{aligned} \sum_{\substack{n \equiv \alpha \pmod{\Delta} \\ n \equiv \beta \pmod{d}}} F\left(\frac{n}{N}\right) \frac{\log n}{\log N} &= \frac{1}{\Delta d} \int F\left(\frac{t}{N}\right) \frac{\log t}{\log N} dt + O(1) \\ &= \frac{N}{\Delta d} \int F(t) \left(1 + \frac{\log t}{\log N}\right) dt + O(1). \end{aligned}$$

Moreover, the number of classes $\alpha \pmod{\Delta}$ with $(Q(\alpha), \Delta) = 1$ is equal to $\gamma(\mathcal{H})\Delta$ and the number of classes $\beta \pmod{d}$ with $d \mid Q(\beta)$ is equal to $\tau_k(d)$. Hence

$$(2.14) \quad U(D, N) = \gamma(\mathcal{H})G(D)\{\hat{F}(0)N + \tilde{F}(0)N/\log N\} + O(D(\log D)^{3k})$$

where $\hat{F}(0), \tilde{F}(0)$ are the integrals of $F(t), F(t) \log t$, respectively, and

$$(2.15) \quad G(D) = \sum_{(d, \Delta)=1} \lambda_d \tau_k(d)/d.$$

Next, we are going to evaluate $V_h(D, N)$ for every h separately. Given $h \in \mathcal{H}$, let $Z_h(X)$ denote the polynomial

$$(2.16) \quad Z_h(X) = X^{-1}Q(X+h) = \prod_{h' \neq h} (X+h-h').$$

By shifting n to $n+h$ and using $F((n+h)/N) = F(n/N) + O(h/N)$, we get

$$V_h(D, N) = \sum_{(nZ_h(n), \Delta)=1} \Lambda(n)F(n/N) \sum_{d \mid Z_h(n)} \lambda_d + O(D(\log D)^{3k}).$$

Note that the contribution of n 's which are not coprime with d is negligible, because n is a prime power and $d < D$. Splitting the summation over n into residue classes $\alpha \pmod{\Delta}, \beta \pmod{d}$ such that

$$(2.17) \quad (\alpha Z(\alpha), \Delta) = 1, \quad (\beta, d) = 1, \quad d \mid Z_h(\beta),$$

we obtain

$$\sum_d \lambda_d \sum_{\alpha(\Delta)} \sum_{\beta(d)} \sum_{\substack{n \equiv \alpha(\Delta) \\ n \equiv \beta(d)}} \Lambda(n)F(n/N).$$

If the inner sum is replaced by its expected value

$$(2.18) \quad \frac{1}{\varphi(q)} \sum_{(n,q)=1} \Lambda(n)F(n/N), \quad q = \Delta d,$$

we get a free summation over all α, β satisfying (2.17). The number of α 's is the same as before, namely $\gamma(\mathcal{H})\Delta$, but the number of β 's drops to $\tau_{k-1}(d)$. Moreover, (2.18) is well-approximated by $\hat{F}(0)N/\varphi(q)$ up to a very small error term $(\log q)/\varphi(q)$. Hence we can write

$$(2.19) \quad V_h(D, N) = \gamma(\mathcal{H})G'(D)\hat{F}(0)N + R_h(D, N) + O(D(\log D)^{3k}),$$

where

$$(2.20) \quad G'(D) = \frac{\Delta}{\varphi(\Delta)} \sum_{(d,\Delta)=1} \lambda_d \tau_{k-1}(d) / \varphi(d)$$

and $R_h(D, N)$ is the remainder

$$(2.21) \quad \sum_{\substack{\alpha(\Delta) \\ (Z_h(\alpha), \Delta)=1}}^* \sum_{d < D} \lambda_d \sum_{\substack{\beta(d) \\ d|Z_h(\beta)}}^* \left(\sum_{\substack{n \equiv \alpha(\Delta) \\ n \equiv \beta(d)}} \Lambda(n) F\left(\frac{n}{N}\right) - \frac{1}{\varphi(\Delta d)} \sum_{(n, \Delta d)=1} \Lambda(n) F\left(\frac{n}{N}\right) \right).$$

Note that the main term in (2.19) is the same one for every $h \in \mathcal{H}$, and we have $k = |\mathcal{H}|$ of these terms. Adding $V_h(D, N)$ and subtracting $U(D, N)$ as in (2.11), we arrive at the following formula for the GPY sifted sum:

$$(2.22) \quad W(D, N) = \gamma(\mathcal{H}) N \hat{F}(0) \left\{ k G'(D) - G(D) (\log N + O(1)) \right\} + \sum_{h \in \mathcal{H}} R_h(D, N) + O(D(\log D)^{3k}).$$

The remainders $R_h(D, N)$, which are given by (2.21) must satisfy, for every $h \in \mathcal{H}$,

$$(2.23) \quad R_h(D, N) \ll N(\log N)^{-A}$$

with some $A > k - 1$ (for practical purposes, for every A), so they do not strike the main term. It is critical to have (2.23) for $D = N^\theta$ with $\theta > \frac{1}{2}$, and we shall achieve this level of moduli, but only for a special choice of the sieve weights λ_d .

At the same time, we also require the upper-bound sieve (λ_d) of level D to be such that

$$(2.24) \quad k G'(D) - G(D) \log N > \eta (\log N)^{1-k},$$

with a small positive constant η , so that the lower bound in (2.7) will hold (the corresponding upper bound follows easily by any reasonable sieve of dimension $k - 1$).

2.3. Choosing the Sieve Weights. So far our transformations are valid for any upper-bound sieve (λ_d) of level D , actually for any sequence (λ_d) of real numbers satisfying (2.10), but we are interested in special (λ_d) which yield (2.24). Here we are faced with sieves of dimension $k - 1$ in the case of $G'(D)$ and of dimension k in the case of $G(D)$. It turns out that k must be quite large to give the positive lower bound (2.24),

and the Λ^2 -sieve of Selberg works best in large dimensions. The Λ^2 -sieve assumes that λ_d are given in the form

$$(2.25) \quad \lambda_d = \sum_{[d_1, d_2]=d} \rho_{d_1} \rho_{d_2}$$

where (ρ_d) is any sequence of real numbers with

$$(2.26) \quad \rho_1 = 1, \quad \rho_d = 0 \quad \text{if } d \geq \sqrt{D}.$$

For a complete account of the Λ^2 -sieve, see Chapter 7 of [11]. Now, $kG'(D) - G(D) \log N$ becomes a quadratic form in the sieve constituents ρ_d , $1 \leq d < \sqrt{D}$. It would be a complicated process to determine the maximum of this quadratic form, however, it suffices to have results which are manageable in applications. To this end we follow the strategy of minimizing $G(D)$ alone, hoping that the resulting value of $G'(D)$ will be pretty good as well. Let us say a few words about how we proceeded in [11] with the arithmetical transformations of such forms. First, by the linear substitution (7.180), the quadratic form $G(D)$ is diagonalized to (7.181) while $G'(D)$ is expressed in the new variables in (7.186). At this point, we make the special choice (7.187) of the new variables y_d :

$$y_d = \frac{1}{Y(D)} \left(\log \frac{\sqrt{D}}{d} \right)^l, \quad 1 \leq d < \sqrt{D},$$

where l is any positive integer at our disposal and $Y(D)$ is the normalizing factor. This choice yields (7.188) and (7.189) with $Y(D)$ given by (7.190). We write

$$Y(D) = \sum_{m < \sqrt{D}} h(m) \left(\log \frac{\sqrt{D}}{m} \right)^l$$

where $h(d)$ is the multiplicative function supported on squarefree numbers d coprime with Δ , $h(p) = k/(p - k)$ if $p \nmid \Delta$. Hence, the original sieve constituents become

$$\rho_d = \mu(d) \frac{d}{\tau_k(d)} \sum_{\substack{b < \sqrt{D} \\ b \equiv 0 \pmod{d}}} h(b) y_b = \frac{\mu(d) d}{Y(D) \tau_k(d)} \sum_{\substack{d < \sqrt{D} \\ b \equiv 0 \pmod{d}}} h(b) \left(\log \frac{\sqrt{D}}{b} \right)^l,$$

see (7.182) of [11]. By the arguments of van Lint-Richert (see the bottom lines on p. 92 of [11]) we get

$$\begin{aligned}
 Y(D) &= \sum_{a|d} \sum_{\substack{m < \sqrt{D} \\ (m,d)=a}} h(m) \left(\log \frac{\sqrt{D}}{m} \right)^l = \sum_{a|d} h(a) \sum_{\substack{m < \sqrt{D}/a \\ (m,d)=1}} h(m) \left(\log \frac{\sqrt{D}}{am} \right)^l \\
 &\geq \left(\sum_{a|d} h(a) \right) \sum_{\substack{m < \sqrt{D}/d \\ (m,d)=1}} h(m) \left(\log \frac{\sqrt{D}}{dm} \right)^l = \mu(d)\rho_d Y(D).
 \end{aligned}$$

Hence $|\rho_d| \leq 1$, which, together with (2.25) implies (2.10).

Remarks. The above choice of ρ_d is not optimal for $G(D)$, however it is used in the original work of GPY because it produces a very good value of the whole quadratic form $kG'(D) - G(D) \log N$. After the GPY work, B. Conrey considered $\mu(d)\rho_d$ as a continuous function of d and applied variational calculus to find the best choice, which turned out to be a combination of two Bessel functions J_{k-2} (in unpublished notes of 2005)³.

We have evaluated the sums $G(D), G'(D)$ asymptotically in (7.193), (7.194) respectively, again from [11]. Here, we quote these results:

$$(2.27) \quad G(D) = \gamma(k, l) \frac{H(\mathcal{H})}{\gamma(\mathcal{H})} (\log \sqrt{D})^{-k} \left\{ 1 + O\left(\frac{1}{\log D}\right) \right\},$$

$$(2.28) \quad G'(D) = \frac{2l + 1}{(l + 1)(k + 2l + 1)} G(D) \{ \log D + O(1) \},$$

where $\gamma(k, l) = k! \binom{2l}{l} \binom{l+k}{l} / \binom{2l+k}{l}$. Hence, (2.22) becomes

$$(2.29) \quad W(D, N) = \gamma(k, l) H(\mathcal{H}) \hat{F}(0) \{ w(D, N) + O(1) \} N (\log \sqrt{D})^{-k} + \sum_{h \in \mathcal{H}} R_h(D, N)$$

where

$$(2.30) \quad w(D, N) = \mu(k, l) \log D - \log N$$

with $\mu(k, l) = k(2l + 1) / (l + 1)(k + 2l + 1)$, which we write as

$$(2.31) \quad \mu(k, l) = 2 \left/ \left(1 + \frac{1}{2l + 1} \right) \left(1 + \frac{2l + 1}{k} \right) \right.$$

³A more extensive analysis has now been given by Farkas, Pintz and Révész in the volume: Number Theory, Analysis and Combinatorics, de Gruyter (2014), pp 75–104.

It is clear that the factor $\mu(k, l)$ is always smaller than 2, so we need the sieve level $D = N^\theta$ with $\theta > \frac{1}{2}$ to be able to conclude that $w(D, N) \gg \log N$, and hence (2.7). The Bombieri-Vinogradov theorem can handle our remainders $R_h(D, N)$ only to level $D < N^{\frac{1}{2}}$, so it is not sufficient, yet it barely misses. Indeed, if l is large and k is considerably larger, then $\mu(k, l)$ is close to 2. A nice choice of l in terms of k is the integer with $\sqrt{k} \leq 2l + 1 < \sqrt{k} + 2$, giving

$$\mu(k, l) \geq 2 / \left(1 + \frac{2}{\sqrt{k}} + \frac{2}{k} \right).$$

As alluded in the introduction, we shall be able to achieve the sieve level $D = N^\theta$ with an absolute exponent $\theta > \frac{1}{2}$ for the sieve weights λ_d supported on numbers having no large prime divisors, say

$$(2.32) \quad d \mid P(z), \quad z = D^{1/2s}, \quad s \geq 1.$$

Let $G_s(D), G'_s(D)$ denote the corresponding sums (2.15), (2.20) with the extra restriction (2.32). There is no extra trouble to evaluate these sums. The extra restriction (2.32) can be automatically secured by assuming (temporarily) in the arithmetical transformations (7.180)-(7.194) of [11], that Δ is divisible by every prime in the segment $z \leq p < \sqrt{D}$. Of course, such a Δ would be huge (temporarily), but it does not matter as long as we are handling algebraic identities without estimating anything. Therefore (7.188) of [11] becomes

$$(2.33) \quad Y_s^2(D)G_s(D) = \sum_{\substack{d < \sqrt{D} \\ \Delta d \mid P(z)}} \frac{\tau_k(d)}{f(d)} \left(\log \frac{\sqrt{D}}{d} \right)^{2l}$$

and (7.189) of [11] becomes

$$(2.34) \quad Y_s^2(D)G'_s(D) = \frac{\Delta}{\varphi(\Delta)} \sum_{\substack{d < \sqrt{D} \\ \Delta d \mid P(z)}} \left(\frac{d}{\varphi(d)} \right)^2 \frac{\tau_{k-1}(d)}{f(d)} \left(\sum_{\substack{n < \sqrt{D}/d \\ \Delta dn \mid P(z)}} \frac{1}{\varphi(n)} \left(\log \frac{\sqrt{D}}{dn} \right)^l \right)^2$$

with $Y_s(D)$, the normalizing factor, given by

$$(2.35) \quad Y_s(D) = \sum_{\substack{d < \sqrt{D} \\ \Delta d \mid P(z)}} \frac{\tau_k(d)}{f(d)} \left(\log \frac{\sqrt{D}}{d} \right)^l.$$

Here $f(d)$ is the multiplicative function with $f(p) = p - k \geq 1$, and $l \geq 0$ is an integer at our disposal. For $z = \sqrt{D}$, that is, for $s = 1$, we have the formula (7.192) of [11];

$$(2.36) \quad Y_1(D) = Y(D) = \frac{l!}{(l+k)!} \frac{\gamma(\mathcal{H})}{H(\mathcal{H})} (\log \sqrt{D})^{k+l} \left\{ 1 + O\left(\frac{1}{\log D}\right) \right\}.$$

In the next subsection we shall estimate $G'_s(D)$ from below for all $s \geq 1$. Of course, we have the obvious upper bound $G'_s(D) \leq G'_1(D) = G'(D)$, but it turns out that $G'_s(D)$ loses very little as compared to $G'_1(D)$, if s is relatively small; precisely we have

$$(2.37) \quad G'_s(D) \geq \left(1 - 2s \left(1 - \frac{1}{s} \right)^k \right) G'(D), \quad \text{if } s \geq 2.$$

Here D is assumed to be large. Note that $(1 - 1/s)^k < e^{-k/s}$. For $G_s(D)$ and $Y_s(D)$ we can use the obvious upper bounds $G_s(D) \leq G_1(D) = G(D)$ and $Y_s(D) \leq Y_1(D) = Y(D)$. Hence, for the sum (2.6) with λ_d restricted by (2.32), we obtain (2.29) with

$$(2.38) \quad w(D, N) \geq C_k(\theta, s) \log N$$

provided

$$(2.39) \quad C_k(\theta, s) = 2\theta(1 - 2se^{-k/s}) \left(1 + \frac{2}{\sqrt{k}} + \frac{2}{k} \right)^{-1} - 1 > 0.$$

If $z = D^{1/2s}$ with s sufficiently large, that is, if the support of the sieve (λ_d) is smooth enough, we shall achieve a level $D = N^\theta$ with $\theta = \theta(s)$ exceeding $\frac{1}{2}$ by an absolute constant. Then, taking k large, it is evident without resorting to numerical computation, that $C_k(\theta, s)$ is positive. We shall be allowed to take (see (3.34)) $s = 4494$ and $\theta = 105/209$, numbers which are sufficient to show the positivity of $C_k(\theta, s)$ for

$$k = (419)^2 = 175561.$$

Note that for this choice the loss of $2s(1 - 1/s)^k$ in the lower bound (2.37), which is caused by the sieve-weight restriction (2.32), is numerically minuscule, $2se^{-k/s} < 10^{-13}$.

It remains to prove the lower bound (2.37), which we do in the next subsection.

2.4. Estimation of $G'_s(D)$. Recall that $G'_s(D)$ is given by (2.34) with $z = D^{1/2s}$ and that $G'(D) = G'_1(D)$. We begin the proof of (2.37) by generalizing Corollary A.6 of [11] to accommodate the restriction $m \mid P(z)$.

Let $\kappa \geq 1$ and $l \geq 0$ be integers and $g(m)$ be a multiplicative function supported on squarefree integers such that

$$(2.40) \quad 0 \leq g(p) = \frac{\kappa}{p} + O\left(\frac{1}{p^2}\right).$$

Denote, for $x > 1, z \geq 2$,

$$(2.41) \quad S_l(x, z) = \sum_{\substack{m \leq x \\ \Delta m | P(z)}} g(m) \left(\log \frac{x}{m}\right)^l.$$

In Corollary A.6 of [11] we proved that for $x = z \geq 2$ (in which case the summation is restricted only by $(m, \Delta) = 1$)

$$(2.42) \quad S_l(x, x) = C(\log x)^{\kappa+l} \{1 + O(1/\log x)\}$$

where C is a positive constant which depends on Δ, l and the function g . Hence, we derive:

Lemma 2.1. *For $x > 1$ and $z \geq 2$ we have*

$$(2.43) \quad S_l(x, z) = h(s)S_l(z, z)\{1 + O(1/\log z)\},$$

where $s = \log x / \log z$ and $h(s)$ is the continuous solution to the differential-difference equation

$$(2.44) \quad h(s) = s^{\kappa+l}, \quad \text{if } 0 < s \leq 1,$$

$$(2.45) \quad sh'(s) = (\kappa + l)h(s) - \kappa h(s - 1), \quad \text{if } s > 1.$$

Proof. For $l = 0$ we get (2.43) by Theorem A.8 of [11], which we write in the form

$$S_0(x, z) = h_0(s)S_0(z, z)\{1 + O(1/\log z)\},$$

where $h_0(s)$ is the solution of (2.44)-(2.45) with $l = 0$. We derive the general case $l \geq 1$ by partial summation as follows;

$$\begin{aligned} S_l(x, z) &= \int_1^x \left(\log \frac{x}{y}\right)^l dS_0(y, z) = - \int_1^x S_0(y, z) d\left(\log \frac{x}{y}\right)^l \\ &= - \int_1^x \left\{ h_0\left(\frac{\log y}{\log z}\right) + O\left(\frac{1}{\log z}\right) \right\} d\left(\log \frac{x}{y}\right)^l S_0(z, z). \end{aligned}$$

Changing the variable $y = z^t = x^{t/s}$ we get

$$(2.46) \quad S_l(x, z) = I(s)S_0(z, z)(\log z)^l \{1 + O(1/\log z)\},$$

where

$$(2.47) \quad I(s) = - \int_0^s h_0(t) d(s-t)^l.$$

For $0 < s \leq 1$ this gives

$$(2.48) \quad I(s) = l \int_0^s t^\kappa (s-t)^{l-1} dt = \frac{\kappa!l!}{(\kappa+l)!} s^{\kappa+l} = I(1)h(s).$$

For $s > 1$ we write the integral (2.47) in the form

$$sI(s) = l \int_0^s h_0(t)(s-t)^l dt + l \int_0^s h_0(t)t(s-t)^{l-1} dt.$$

Differentiating, we derive (assume $l > 1$)

$$(sI(s))' = lI(s) - l \int_0^s h_0(t)t d(s-t)^{l-1} = lI(s) + l \int_0^s (h_0(t)t)'(s-t)^{l-1} dt.$$

Hence,

$$\begin{aligned} (sI(s))' &= lI(s) + l \int_1^s ((\kappa+1)h_0(t) - \kappa h_0(t-1))(s-t)^{l-1} dt + l \int_0^1 (h_0(t)t)'(s-t)^{l-1} dt \\ &= (\kappa+l+1)I(s) - \kappa I(s-1). \end{aligned}$$

Thus, $I(s)$ satisfies the differential-difference equation (2.45). (For $l = 1$ the arguments are similar.) Then, (2.46) with $x = z \geq 2$ yields

$$S_l(z, z) = I(1)S_0(z, z)(\log z)^l \{1 + O(1/\log z)\}.$$

Combining this with (2.46) we obtain (2.43). □

Note that (2.43) can be written in the following form (use (2.42))

$$(2.49) \quad S_l(x, z) = H(s)S_l(x, x)\{1 + O(1/\log x)\}, \quad \text{if } x \geq 2, z \geq 2,$$

where $H(s) = s^{-\kappa-l}h(s)$, so $H(s)$ is the continuous solution of

$$(2.50) \quad H(s) = 1, \quad \text{if } 0 < s \leq 1,$$

$$(2.51) \quad s^{\kappa+l+1}H'(s) = -\kappa(s-1)^{\kappa+l}H(s-1), \quad \text{if } s > 1.$$

Obviously, $H(s)$ is positive and decreasing, so $H(s) \leq 1$.

Lemma 2.2. *For $s > 1$ we have*

$$(2.52) \quad H(s) > 1 - \frac{\kappa s}{\kappa+l+1} \left(1 - \frac{1}{s}\right)^{\kappa+l+1}.$$

Proof. By (2.51) we get

$$\begin{aligned}
 1 - H(s) &= \kappa \int_1^s \left(1 - \frac{1}{t}\right)^{\kappa+l} t^{-1} H(t-1) dt \leq \kappa \int_1^s \left(1 - \frac{1}{t}\right)^{\kappa+l} t^{-1} dt \\
 &\leq \kappa s \int_1^s \left(1 - \frac{1}{t}\right)^{\kappa+l} t^{-2} dt = \frac{\kappa s}{\kappa + l + 1} \left(1 - \frac{1}{s}\right)^{\kappa+l+1}. \quad \square
 \end{aligned}$$

Remarks. One can show that $H(s)$ satisfies the integral equation

$$(2.53) \quad s^{\kappa+l+1} q(s) H(s) = c + \kappa \int_{s-1}^s x^{\kappa+l} q(x+1) H(x) dx$$

for $s \geq 1$, where $c = e^{\kappa\gamma}(\kappa + l)!/l!$, and

$$(2.54) \quad q(s) = \frac{1}{l!} \int_0^\infty \exp\left(-sz + \kappa \int_0^z (1 - e^{-u}) u^{-1} du\right) z^l dz.$$

This is the solution of the equation

$$(2.55) \quad (s^{\kappa+l+1} q(s))' = \kappa s^{\kappa+l} q(s+1), \quad s > 0,$$

which is adjoint to the equation (2.51), see Appendix B of [11]. We have, by (B.13) and (B.12) of [11])

$$(2.56) \quad \begin{aligned} q(s) &\sim cs^{-\kappa-l-1}, & \text{as } s \rightarrow 0, \\ q(s) &\sim s^{-l-1}, & \text{as } s \rightarrow \infty. \end{aligned}$$

Hence (2.53) shows that

$$(2.57) \quad H(s) \sim cs^{-\kappa}, \quad \text{as } s \rightarrow \infty.$$

Our neat lower bound (2.52) can be improved significantly if $s > \kappa + l$ by using (2.53), but we do not need anything in this range.

Now we are ready to prove (2.37) for any $s \geq 2$. By (2.41), (2.42), and (2.49), the innermost sum in (2.34) is equal to

$$CH \left(\frac{\log \sqrt{D}/d}{\log z} \right) \left(\log \frac{\sqrt{D}}{d} \right)^l \left(\log \frac{\sqrt{D}}{d} + O(1) \right)$$

where $H(s)$ is the function defined in (2.50) and (2.51) for $\kappa = 1$ (in the range $\sqrt{D}/2 < d < \sqrt{D}$, the claim is trivial). Hence $Y_s^2(D)G'_s(D)$ in (2.34) is equal to

$$\frac{C^2 \Delta}{\varphi(\Delta)} \sum_{\substack{d < \sqrt{D} \\ \Delta d | P(z)}} \left(\frac{d}{\varphi(d)} \right)^2 \frac{\tau_{k-1}(d)}{f(d)} H^2 \left(\frac{\log \sqrt{D}/d}{\log z} \right) \left(\log \frac{\sqrt{D}}{d} \right)^{2l} \left(\log \frac{\sqrt{D}}{d} + O(1) \right)^2.$$

All we need to know about $H(s)$ is (2.50) and the lower bound

$$H^2(s) > 1 - \frac{2s}{l+2} \left(1 - \frac{1}{s}\right)^{l+2} \quad \text{if } s > 1,$$

which follows by (2.52). Hence, the above formula yields the following lower bound:

$$(2.58) \quad Y_s^2(D)G'_s(D) \geq \frac{C^2\Delta}{\varphi(\Delta)} \sum_{\substack{d < \sqrt{D} \\ \Delta d | P(z)}} \left(\frac{d}{\varphi(d)}\right)^2 \frac{\tau_{k-1}(d)}{f(d)} \{\dots\} + O((\log D)^{k+2l})$$

where the content of $\{\dots\}$ is equal to

$$\left(\log \frac{\sqrt{D}}{d}\right)^{2l+2} - \frac{2}{l+2} (\log z)^{-1} \left(\log \frac{\sqrt{D}}{d}\right)^{l+1} \left(\log \frac{\sqrt{D}}{zd}\right)^{l+2}$$

with the negative term being present only for $d < \sqrt{D}/z$. In this negative term we use $\log(\sqrt{D}/d) \leq (1 - 1/s)^{-1} \log \sqrt{D}$, obtaining

$$\{\dots\} \geq \left(\log \frac{\sqrt{D}}{d}\right)^{2l+2} - \frac{2s}{l+2} \left(1 - \frac{1}{s}\right)^{-l-1} (\log \sqrt{D})^l \left(\log \frac{\sqrt{D}}{zd}\right)^{l+2}$$

with the negative term being present only for $d < \sqrt{D}/z$. Inserting this into (2.58) we get two sums, over $d < \sqrt{D}$ in the positive terms and over $d < \sqrt{D}/z$ in the negative terms. The positive sum is of type (2.41) with $\kappa = k - 1$ and l replaced by $2l + 2$. The negative sum is also of type (2.41) with the same $\kappa = k - 1$, but with l replaced by $l + 2$. Let $K(s)$ and $L(s)$ denote the corresponding functions defined in (2.50) and (2.51). Applying (2.49) and comparing the resulting lower bound for $Y_s^2(D)G'_s(D)$ with the asymptotic value of $Y_1^2(D)G'_1(D)$ we conclude that

$$G'_s(D) > \chi(s)G'(D)\{1 + O(1/\log D)\},$$

where

$$\chi(s) = K(s) - \frac{2s}{l+2} \left(1 - \frac{1}{s}\right)^{-l-1} \left(1 - \frac{1}{s}\right)^{k+l+1} L(s).$$

Here, the factor $(1 - 1/s)^{k+l+1}$ in front of $L(s)$ comes by way of re-scaling:

$$Y_1^2(D/z^2)G'_1(D/z^2) = Y_1^2(D)G'_1(D) \left(\frac{\log \sqrt{D}/z}{\log \sqrt{D}}\right)^{k+l+1} \left\{1 + O\left(\frac{1}{\log D}\right)\right\}.$$

Finally, by the lower bound (2.52) for $K(s)$ and the upper bound $L(s) \leq 1$, we find that

$$\chi(s) > 1 - \frac{(k-1)s}{k+2l+2} \left(1 - \frac{1}{s}\right)^{k+2l+2} - \frac{2s}{l+2} \left(1 - \frac{1}{s}\right)^k > 1 - 2s \left(1 - \frac{1}{s}\right)^k.$$

This completes the proof of (2.37) (D is assumed to be large).

3. Proof of the Main Theorem

3.1. Sums over primes. The most common technique for handling sums over primes uses combinatorial formulas of sieve exclusion-inclusion type; see Chapters 17,18 of [11], where many variations on this theme are illustrated with applications. In the context of the dispersion method, the first use of such an identity is due to Linnik in his foundational work [15]. In this paper we develop a smooth version of Heath-Brown’s formula.

Let $\psi(u)$ be a smooth function supported on $|u| \leq 2$ with

$$(3.1) \quad \psi(u) = 1 \quad \text{if } |u| \leq 1.$$

Let $K \geq 1$ and $M \geq 1$. For any integer n with $1 \leq n \leq 2M^K$ we have

$$(3.2) \quad \Lambda(n) = - \sum_{1 \leq J \leq K} (-1)^J \binom{K}{J} \sum_{l_1 \cdots l_J m_1 \cdots m_J = n} \mu(m_1) \cdots \mu(m_J) \cdot \psi\left(\frac{m_1}{M}\right) \cdots \psi\left(\frac{m_J}{M}\right) \log l_1.$$

This can be seen from the following lines:

$$\begin{aligned} \frac{\zeta'}{\zeta}(s) \left(1 - \zeta(s) \sum_m \mu(m) \psi\left(\frac{m}{M}\right) m^{-s}\right)^K &= \sum_{n > 2M^K} a_n n^{-s} \\ &= \frac{\zeta'}{\zeta}(s) + \sum_{1 \leq J \leq K} (-1)^J \binom{K}{J} \zeta'(s) \zeta(s)^{J-1} \left(\sum_m \mu(m) \psi\left(\frac{m}{M}\right) m^{-s}\right)^J. \end{aligned}$$

We shall sum arithmetic functions over primes with smooth weights. For technical convenience the sum is often split into dyadic boxes. Here is how one can perform such a split in a general context without compromising the smoothness. Suppose $F = \phi * \cdots * \phi * \Phi$ is given as the multiplicative convolution:

$$(3.3) \quad F(x) = \int \cdots \int \phi(t_1) \cdots \phi(t_r) \Phi(t_1 \cdots t_r/x) d^*t,$$

where $d^*t = (t_1 \cdots t_r)^{-1} dt_1 \cdots dt_r$. Choose $\phi(y)$ and $\Phi(y)$, smooth and supported on $1 \leq y \leq 2$, with

$$(3.4) \quad \int \phi(y) y^{-1} dy = 1.$$

Let $N \geq 1$. Note that if $1 \leq n \leq 2^r N$, then

$$\int_{1/2}^{2^r N} \phi\left(\frac{n}{y}\right) \frac{dy}{y} = \int \phi(y)y^{-1} dy = 1.$$

Hence, for any numbers $n_1, \dots, n_s \geq 1, 1 \leq s \leq r$, we have

(3.5)

$$F\left(\frac{n_1 \cdots n_s}{N}\right) = \int \cdots \int \int_{1/2}^{2^r N} \cdots \int_{1/2}^{2^r N} \phi\left(t_1 \frac{n_1}{y_1}\right) \phi\left(\frac{n_1}{y_1}\right) \cdots \phi\left(t_s \frac{n_s}{y_s}\right) \phi\left(\frac{n_s}{y_s}\right) \cdot \phi(t_{s+1}) \cdots \phi(t_r) \Phi\left(\frac{t_1 \cdots t_r}{y_1 \cdots y_s} N\right) d^*t d^*y.$$

This integral representation provides simultaneously a separation of the variables n_1, \dots, n_r and their partition into smooth, dyadic-type boxes. Indeed, the integration variables satisfy

(3.6)
$$\frac{1}{2} < t_j < 2, \quad \text{for } 1 \leq j \leq r,$$

(3.7)
$$\frac{1}{2} < y_j < 2^r N, \quad \text{for } 1 \leq j \leq s,$$

(3.8)
$$2^{-r-1} N < y_1 \cdots y_s < 2^r N,$$

while the projected summation variables n_1, \dots, n_s are well located in the segments

(3.9)
$$y_j < n_j < 2y_j, \quad 1 \leq j \leq s.$$

Note that the measure of the set (3.6)-(3.8) is bounded by

(3.10)
$$(\log 4)^r (\log 2^{r+1} N)^{s-1} \log 2^{2r+1} \ll (3r/2)^r (\log 2N)^{s-1}.$$

Of course, one can take the convolution $F = \phi_1 * \cdots * \phi_r * \Phi$ with a different test function for every variable, however, our choice $\phi_1 = \cdots = \phi_r = \phi$ does the job fine.

We shall use the expression (3.5) with $r = 2K$ and $s = 2J$, so for reference, in this case we denote the domain of integration (3.6)-(3.8) by $\mathcal{M}_{JK}(N)$. The measure of $\mathcal{M}_{JK}(N)$ (with respect to the multiplicative group) satisfies

(3.11)
$$|\mathcal{M}_{JK}(N)| \leq 2(3K)^{2K} (\log 2N)^{2J-1}.$$

Applying (3.5) to (3.2) we reach the following “decomposition” formula for smoothly cropped prime powers.

Proposition 3.1. *Let $F = \phi * \dots * \phi * \Phi$ be the multiplicative convolution of $2K$ copies of a smooth function $\phi(y)$ supported in $1 \leq y \leq 2$ and normalized by (3.4), and $\Phi(y)$ supported in $1 \leq y \leq 2$. Let*

$$(3.12) \quad 1 \leq N \leq 2(M/4)^K.$$

Then, for any $n \geq 1$ we have

$$(3.13) \quad \Lambda^{(n)}F\left(\frac{n}{N}\right) = - \sum_{1 \leq J \leq K} (-1)^J \binom{K}{J} \int_{\mathcal{M}_{JK}(N)} \left(\prod_{2J < j \leq 2K} \phi(t_j) \right) \Phi\left(\frac{t_1 \dots t_{2K}}{y_1 \dots y_{2J}} N\right) A_n(t_1, \dots, t_{2J}; y_1, \dots, y_{2J}) \frac{dt_1 \dots dt_{2K}}{t_1 \dots t_{2K}} \frac{dy_1 \dots dy_{2J}}{y_1 \dots y_{2J}}$$

where $A_n(\cdot)$ is given by the Dirichlet convolution of $2J$ factors

$$(3.14) \quad A_n(\cdot) = \sum_{l_1 \dots l_J m_1 \dots m_J = n} \mu(m_1) \dots \mu(m_J) \psi\left(\frac{m_1}{M}\right) \dots \psi\left(\frac{m_J}{M}\right) (\log l_1) \prod_{1 \leq j \leq J} \phi\left(\frac{l_j}{y_j}\right) \phi\left(\frac{l_j t_j}{y_j}\right) \phi\left(\frac{m_j}{y_{J+j}}\right) \phi\left(\frac{m_j}{y_{J+j}} t_{J+j}\right).$$

Remarks. The function $F(x)$ is supported in the segment $\frac{1}{2} \leq x \leq 2^{2K}$, therefore $\Lambda^{(n)}F(n/N)$ captures the prime powers inside $\frac{N}{2} < n < 2^{2K}N$ with weights smoothly vanishing at the endpoints.

The Fourier transform $\hat{F}(s)$ at $s = 0$ will appear as a factor in the main terms. For $F = \phi * \dots * \phi * \Phi$ given by (3.3) we have

$$\hat{F}(0) = \left(\int \phi(y) dy \right)^r \int \Phi(y) y^{-2} dy.$$

Given ϕ and $r \geq 1$ we shall take Φ to be normalized so that

$$(3.15) \quad \hat{F}(0) = 1.$$

3.2. Combinatorial arrangements. For clarity of exposition, let us arrange the y -ordinates of the points in the domain of integration $\mathcal{M}_{JK}(N)$ (see (3.6)-(3.8)) into the decreasing sequence $\{y_1, y_2, \dots, y_{2J}\} = \{Y_1, Y_2, \dots, Y_{2J}\}$, say, with

$$(3.16) \quad Y_1 \geq Y_2 \geq \dots \geq Y_{2J}.$$

By (3.8) we get

$$(3.17) \quad Y_1 Y_2 \dots Y_{2J} \asymp N.$$

Lemma 3.2. *Let $0 < \eta < 1/40$. Then, at least one of the following three cases hold:*

(C1) *we have*

$$Y_1 \gg N^{\frac{5}{8}-\eta},$$

(C2) *there is a subproduct of (3.17), say X , with*

$$N^{\frac{3}{8}+\eta} \ll X \ll N^{\frac{5}{8}-\eta},$$

(C3) *we have $J > 1$,*

$$Y_1 \geq Y_2 \geq Y_3 \gg N^{\frac{1}{4}-\eta} \quad \text{and} \quad Y_1 Y_2 Y_3^{\frac{5}{4}} \gg N^{\frac{65}{64}-\frac{13}{8}\eta}.$$

The implied constants in (C1), (C2), (C3) depend only on the implied constant in (3.17).

Obviously, Lemma 3.2 follows from the next lemma, which asserts similar but neat inequalities.

Lemma 3.3. *Let $0 < \eta < 1/40$ and ν_1, \dots, ν_r be positive numbers with*

$$(3.16') \quad \nu_1 \geq \dots \geq \nu_r$$

$$(3.17') \quad \nu_1 + \dots + \nu_r = 1.$$

Then, at least one of the following three cases hold:

(C1') *we have*

$$\nu_1 > \frac{5}{8} - \eta,$$

(C2') *there is a subsum of (3.17'), say μ , with*

$$\frac{3}{8} + \eta < \mu < \frac{5}{8} - \eta,$$

(C3') *we have $r \geq 3$,*

$$\nu_1 \geq \nu_2 \geq \nu_3 \geq \frac{1}{4} - 2\eta \quad \text{and} \quad \nu_1 + \nu_2 + \frac{5}{4}\nu_3 \geq \frac{65}{64} - \frac{13}{8}\eta.$$

Proof. Negating (C1'), (C2'), we are going to derive (C3'). First we find that $\nu_1 \leq \frac{3}{8} + \eta$. Hence $r \geq 3$. Let $j \geq 2$ be the largest such that

$$\mu = \nu_2 + \dots + \nu_j < \frac{5}{8} - \eta.$$

Since $\nu_2 + \dots + \nu_r = 1 - \nu_1 \geq \frac{5}{8} - \eta$, it follows that $2 \leq j < r$. Then, by the negation of (C2'), we have $\mu \leq \frac{3}{8} + \eta$. Since $\mu + \nu_{j+1} \geq \frac{5}{8} - \eta$, we get

$$\nu_3 \geq \nu_{j+1} \geq \frac{1}{4} - 2\eta.$$

Hence we get $\nu_2 + \nu_3 \geq 2\nu_3 \geq \frac{1}{2} - 4\eta > \frac{3}{8} + \eta$, and by the negation of (C2') this lower bound improves to $\nu_2 + \nu_3 \geq \frac{5}{8} - \eta$. Hence,

$$\nu_1 + \nu_2 + \frac{5}{4}\nu_3 \geq 2\nu_2 + \frac{5}{4}\nu_3 \geq \frac{13}{8}(\nu_2 + \nu_3) \geq \frac{65}{64} - \frac{13}{8}\eta,$$

so the conditions (C3') hold. □

3.3. Primes in residue classes to divisor-rich moduli. Given $z \geq 2$ we say that a squarefree number q is z -divisor-rich, or simply *divisor-rich*, if it has divisors of any size up to a factor $\leq z$. For example, if $q \mid P(z)$ then q is z -divisor-rich.

Let $Z(X) \in \mathbb{Z}[X]$ be a polynomial which has no fixed prime divisors, that is, for every prime p the number of roots of $Z(x) \equiv 0 \pmod{p}$ is $\omega(p) < p$. The following result is a refined version of Theorem 2 of Zhang [18].

Proposition 3.4. *Let $\mathcal{Q} = \mathcal{Q}(z)$ be the set of z -divisor-rich numbers. We have*

$$(3.18) \quad \sum_{\substack{q \in \mathcal{Q} \\ q \leq Q}} \sum_{\substack{a \pmod{q} \\ q \mid Z(a)}}^* \left| \sum_{n \equiv a \pmod{q}} \Lambda(n) F\left(\frac{n}{N}\right) - \frac{N}{\varphi(q)} \right| \ll N(\log N)^{-A},$$

provided

$$(3.19) \quad z^{\frac{86}{207}} Q \leq N^{\frac{104}{207} - \varepsilon}.$$

Here $F(x)$ is a convolution function of type (3.3) normalized by $\hat{F}(0) = 1$, while ε, A are any positive numbers and the implied constant depends only on ε, A , the function $F(x)$ and the polynomial $Z(X)$.

Remarks. One can replace the smooth cropping $F(n/N)$ by the sharp cutting $n \leq N$, but this is not required in applications. However, the restriction to divisor-rich moduli is essential in Zhang's arguments.

The proof of Proposition 3.4 depends on several results on the distribution of general sequences over residue classes which are established in Sections 4 to 6. In this section we derive (3.18) by combining these results.

The sum over a in (3.18) makes the result look, at first glance, deceptively strong, but actually, there are not so many residue classes involved; $\omega(q) \ll \tau(q)^k$. Hence, it suffices to show that

$$(3.20) \quad \sum_{\substack{q \in \mathcal{Q} \\ q \leq Q}} \frac{1}{\omega(q)} \sum_{\substack{a \pmod{q} \\ q \mid Z(a)}}^* \left| \sum_{n \equiv a \pmod{q}} \Lambda(n) F\left(\frac{n}{N}\right) - \frac{N}{\varphi(q)} \right| \ll N(\log N)^{-A},$$

Clearly, (3.20) implies Proposition 3.4, by applying Cauchy’s inequality and the easy bound

$$\sum_{q \leq Q} \tau(q)^{2k} \left(\sum_{\substack{n \equiv a(q) \\ n \leq 2N}} \Lambda(n) + \frac{N}{\varphi(q)} \right) \ll N(\log N)^{4k}.$$

Moreover, the Bombieri-Vinogradov Theorem (see (1.6) or (9.81) of [11]) covers the range $q \leq N^{\frac{1}{2} - \frac{\varepsilon}{4}}$.

Now, for the proof of (3.20) we can assume that N is sufficiently large in terms of ε , and that q runs over $\mathcal{Q} = \mathcal{Q}(z)$ in the segment $Q < q \leq 2Q$ with

$$(3.21) \quad N^{\frac{1}{2} - \frac{\varepsilon}{3}} \leq Q \leq N^{\frac{104}{207} - \varepsilon} z^{-\frac{86}{207}}.$$

Therefore (3.19) implies

$$(3.22) \quad z \leq N^{1/172}.$$

We are going to apply the decomposition (3.13) with $K = 5$ and $M = 4N^{\frac{1}{5}}$. It suffices to show that

$$(3.23) \quad \sum_{\substack{q \sim Q \\ q \in \mathcal{Q}}} \frac{1}{\omega(q)} \sum_{\substack{a \pmod{q} \\ q|Z(a)}}^* \left| \sum_{n \equiv a(q)} A_n - \frac{1}{\varphi(q)} \sum_{(n,q)=1} A_n \right| \ll N(\log N)^{-A}$$

for every sequence $A_n = A_n(t_1, \dots, t_{2J}; y_1, \dots, y_{2J})$ of type (3.14). Recall that

$$(3.24) \quad \frac{1}{2} < y_j < 2^{10}N, \quad 1 \leq j \leq 2J,$$

$$(3.25) \quad 2^{-11}N < y_1 \cdots y_{2J} < 2^{10}N,$$

and $\frac{1}{2} < t_j < 2, 1 \leq j \leq 2J$, with $1 \leq J \leq 5$. Let η be given by

$$(3.26) \quad z^{\frac{3}{2}} Q^{\frac{11}{4}} = N^{\frac{11}{8} + \eta - \varepsilon}.$$

By (3.19) and (3.22) one can check that $\varepsilon/12 \leq \eta < 1/114$, a fortiori $0 < \eta < 1/40$. According to Lemma 3.2 we can arrange the set $\{y_1, y_2, \dots, y_{2J}\}$ into a decreasing sequence $Y_1 \geq Y_2 \geq \dots \geq Y_{2J} > \frac{1}{2}$ which satisfies at least one of the three conditions (C1), (C2), (C3).

If (C1) holds, then $Y_1 \gg N^{3/5}$, so the corresponding variable $l_j \asymp Y_1$ appears in $A_n, n = l_1 \dots l_J m_1 \dots m_J$, with a smooth crop function and the inner sum in (3.23) is bounded by $NY_1^{-1} \ll N^{2/5}$. Summing over q we see that in this case (3.23) is bounded by

$$N^{\frac{2}{5}} Q \ll N^{\frac{9}{10} + \frac{1}{414}} \ll N(\log N)^{-A}.$$

Suppose (C2) holds, so the sequence $Y_1 \geq Y_2 \geq \dots \geq Y_{2J}$ can be partitioned into two subsequences whose products, say N_1, N_2 , satisfy

$$N^{\frac{3}{8}+\eta} \ll N_1, N_2 \ll N^{\frac{5}{8}-\eta}, \quad N_1 N_2 \asymp N.$$

Obviously, each of the two subsequences has an element, Y' and Y'' respectively, with $Y', Y'' \gg N^{3/80}$ and the corresponding variables $m' \asymp Y', m'' \asymp Y''$ appear in A_n with a smooth crop function multiplied, or not, by the Möbius function. Therefore, the S-W condition holds for each of the two convolution sequences, see the examples (4.7) and (4.8). We can assume, by exchanging N_1 and N_2 if necessary, that

$$(3.27) \quad N^{\frac{3}{8}+\eta} \ll N_1 \ll N^{\frac{1}{2}}.$$

Note that $Q \gg N_1 N^{-\varepsilon/3}$. We can factor the divisor-rich moduli $q \sim Q$ into $q = rs$ with $r \asymp R, s \asymp S, RS = Q$, such that

$$(3.28) \quad z^{-1} N_1 \ll RN^\varepsilon \ll N_1.$$

Then (3.23) is bounded by

$$\sum_r \max_{(a,r)=1} \sum_s \max_{(b,s)=1} \left| \sum_{\substack{n \equiv a \pmod{r} \\ n \equiv b \pmod{s}}} A_n - \frac{1}{\varphi(rs)} \sum_{(n,rs)=1} A_n \right|.$$

For η given by (3.26), the lower bound of (3.27) becomes

$$(3.29) \quad z^{\frac{3}{2}} Q^{\frac{11}{4}} \ll N_1 N^{1-\varepsilon}.$$

The above two estimates verify the conditions of Corollary 4.6, with x, N replaced by N, N_1 , respectively. Therefore, (4.26) of Corollary 4.6 applies, showing that (3.23) is bounded by $N(\log N)^{-A}$, as requested.

Now we are left with the third condition (C3). In this case we have $Y_1 \geq Y_2 \geq Y_3 \geq 8N^{\frac{1}{5}} = 2M$, so the corresponding variables $n_1 \asymp Y_1, n_2 \asymp Y_2, n_3 \asymp Y_3$, appear in A_n with smooth crop functions. The product of the remaining variables is $d = n/n_1 n_2 n_3$, so $dY_1 Y_2 Y_3 \asymp N$. For η given by (3.26), the second estimate of (C3) becomes

$$(3.30) \quad Y_1 Y_2 Y_3^{\frac{5}{4}} \gg N^{\frac{65}{64}} (z^{\frac{3}{2}} Q^{\frac{11}{4}} N^{\varepsilon - \frac{11}{8}})^{-13/8}.$$

Hence, using (3.19) one can derive the following upper bound for Q

$$(3.31) \quad Q \ll \left(\frac{Y_3}{z}\right)^{\frac{1}{8}} \left(\frac{N}{d}\right)^{\frac{1}{2}} N^{-\frac{155}{64}\varepsilon}.$$

This verifies the condition (6.10) of Corollary 6.3, with L, M, N, x replaced by Y_1, Y_2, Y_3, N respectively, which proves (3.23) in this case.

Having covered all cases, we complete the proof of (3.20) and Proposition 3.4.

3.4. Completing the proofs of the theorems. Recall the formula (2.29) in which $w(D, N)$ satisfies the lower bound (2.38). Here $D = N^\theta$ is the level of distribution for primes in residue classes which must ensure that the remainder terms in (2.29) are negligible, meaning

$$(3.32) \quad R_h(D, N) \ll N(\log N)^{-A}.$$

Recall that $z = D^{1/2s}$ is the level of the sieve support restriction. Proposition 3.4 ensures (3.32), provided

$$(3.33) \quad z^{\frac{86}{207}} D \leq N^{\frac{104}{207} - \varepsilon}.$$

Putting $D = N^{\frac{1}{2} + \delta}$ we definitely get (3.33) if

$$(3.34) \quad \delta < \frac{1}{144} \left(1 - \frac{43}{s} \right).$$

On the other hand, the positivity condition (2.39) requires

$$(3.35) \quad \delta > \frac{1}{\sqrt{k}} + \frac{1}{k} + (1 + 2\delta)se^{-k/s}.$$

We choose $s = 4494$ and find that $\delta = 1/418$ satisfies (3.34). Then we find that $k = (419)^2 = 175561$ satisfies (3.35). This completes the proof of (2.7).

Applying the inequality

$$\sum_j c_j \leq 1 + \sum_{i < j} c_i c_j,$$

which holds for any c_j with $0 \leq c_j \leq 1$, we derive by (2.5), (2.6), (2.7) that

$$(3.36) \quad \sum_{(Q(n), \Delta)=1} F(n/N) \sum_{h \neq h'} \Lambda(n-h)\Lambda(n-h') \left(\sum_{\substack{d|Q(n) \\ d|P(z)}} \lambda_d \right) \gg N(\log N)^{2-k}.$$

Here the contribution of n 's with $Q(n)$ having a very small prime divisor is relatively insignificant, see Section 10.3 of [11]. Specifically, by (10.7) of [11] we get

$$\sum_{(Q(n), \Delta)=1} F(n/N) \left(\sum_{\substack{p|Q(n) \\ p < w}} 1 \right) \left(\sum_{\substack{d|Q(n) \\ d|P(z)}} \lambda_d \right) \ll \frac{\log w}{\log N} N(\log N)^{-k}$$

for every $w \geq 2$. Hence, if $w = N^{1/b}$ with b sufficiently large, this contribution can be ignored and (3.36) holds for the sum with the extra restriction $(Q(n), P(w)) = 1$. Having this extra restriction, we can remove the sieve weights by the following trivial estimation:

$$\sum_{\substack{d|Q(n) \\ d|P(z)}} \lambda_d \leq \sum_{d|Q(n)} \tau_3(d) \leq \tau_4(Q(n)) \leq 4^{bk}.$$

This completes the proof of the lower bound of (2.8). To get an upper bound we replace $\Lambda(n - h)\Lambda(n - h')$ by $(\log n)^2$ and apply any reasonable sieve which controls the condition $(Q(n), P(w)) = 1$. This completes the proof of the Main Theorem.

For the proof of the Prime Gaps Theorem we require a specific admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$. One can easily see that any set of k distinct primes larger than k is admissible. We choose the set of consecutive primes $k < h_1 < h_2 < \dots < h_k$, where h_1 is the first exceeding k and h_k is the largest of the bunch. This satisfies

$$(3.37) \quad \pi(h_k) = k + \pi(k),$$

so $h_k \sim k \log k$. For $k = 175561$ we find the following numbers:

$$\begin{aligned} h_1 &= 175573 = p_{15954}, \\ \pi(h_k) &= k + \pi(k) = 175561 + 15953 = 191\,514, \\ h_k &= p_{191514} = 2\,624\,371, \\ h_k - h_1 &= 2\,448\,798. \end{aligned}$$

This completes the proof of the Prime Gaps Theorem.

4. Bilinear Forms in Residue Classes

4.1. **Classical results.** Counting prime numbers in residue classes $a \pmod{q}$ with $(a, q) = 1$ has been transformed in Section 3 to estimates for bilinear forms of type

$$(4.1) \quad \sum_{mn \equiv a \pmod{q}} \alpha_m \beta_n$$

for special sequences $\mathcal{A} = (\alpha_m)$, $\mathcal{B} = (\beta_n)$ with

$$(4.2) \quad |\alpha_m| \leq \tau(m)^j,$$

$$(4.3) \quad |\beta_n| \leq \tau(n)^j.$$

Here and hereafter, j denotes a bounded number which can be different on each occurrence. The available technology is capable of dealing with such bilinear forms in great

generality if at least one of the sequences is known to be uniformly distributed over the reduced residue classes to moduli which are relatively small. Specifically, we shall assume and appeal several times to the following properties of (β_n) .

S-W Condition. Let $\mathcal{B} = (\beta_n)$ be a sequence of complex numbers with $|\beta_n| \leq \tau(n)^j$ for some $j \geq 1$. We say that \mathcal{B} satisfies the *Siegel-Walfisz condition* in the segment $1 \leq n \leq N$ if, for every $k \geq 1$ and $(l, k) = 1$, we have

$$(4.4) \quad \sum_{\substack{n \leq N \\ n \equiv l \pmod{k}}} \beta_n - \frac{1}{\varphi(k)} \sum_{\substack{n \leq N \\ (n,k)=1}} \beta_n \ll N(\log 2N)^{-A}$$

with any $A \geq 1$, the implied constant depending only on A and j .

Proposition 4.1. *If (β_n) satisfies the S-W condition in the segment $1 \leq n \leq N$, then*

$$(4.5) \quad \sum_{q \leq Q} \sum_{a(q)}^* \left| \sum_{\substack{n \leq N \\ n \equiv a \pmod{q}}} \beta_n - \frac{1}{\varphi(q)} \sum_{\substack{n \leq N \\ (n,q)=1}} \beta_n \right|^2 \ll (Q + N(\log 2N)^{-A}) N(\log 2N)^{2j}$$

for any $Q \geq 1$ and any $A \geq 1$ the implied constant depending only on A and j .

Proposition 4.1 follows from Theorem 9.14 of [11]; see the end of its proof on p. 168 to get the above statement without any restriction on Q . Of course, the best use of (4.5) is for $Q \ll N(\log 2N)^{-A}$.

In practice we often need (4.5) for the subsequence of (β_n) restricted by the coprimality condition $(n, r) = 1$. This can be deduced easily at the expense of losing a factor $\tau(r)^2$ in the bound (4.5). To this end we verify (4.4) for the subsequence by using the Möbius formula twice as follows:

$$\begin{aligned} \sum_{(n,r)=1} \xi_n &= \sum_{\substack{d|r \\ d \leq y}} \mu(d) \sum_{n \equiv 0 \pmod{d}} \xi_n + O\left(\sum_{\substack{d|r \\ d > y}} \left(\frac{N}{d} \sum_n |\xi_n|^2\right)^{\frac{1}{2}}\right) \\ &= \sum_{\substack{bc|r \\ bc \leq y}} \mu(b) \sum_{(n,b)=1} \xi_n + O\left(\tau(r) \left(\frac{N}{y} \sum_n |\xi_n|^2\right)^{\frac{1}{2}}\right). \end{aligned}$$

Here and below we assume that $n \leq N$ and $y \geq 1$ is at our disposal. Hence, for $(r, k) = 1$ we get

$$\Delta(r) = \sum_{\substack{(n,r)=1 \\ n \equiv l(k)}} \beta_n - \frac{1}{\varphi(k)} \sum_{(n,rk)=1} \beta_n = \sum_{\substack{bc|r \\ bc \leq y}} \mu(b) \left(\sum_{\substack{(n,b)=1 \\ n \equiv l(k)}} \beta_n - \frac{1}{\varphi(k)} \sum_{(n,bk)=1} \beta_n \right) + O\left(\tau(r)y^{-\frac{1}{2}}N(\log 2N)^{2j}\right).$$

Splitting into primitive classes $n \equiv \beta \pmod{b}$ and applying (4.4) with k replaced by bk we get the bound

$$\begin{aligned} \Delta(r) &\ll \sum_{\substack{bc|r \\ bc \leq y}} \varphi(b)N(\log 2N)^{-4A} + \tau(r)y^{-\frac{1}{2}}N(\log 2N)^{2j} \\ &\ll \tau(r)N(y(\log 2N)^{-4A} + y^{-\frac{1}{2}}(\log 2N)^{2j}). \end{aligned}$$

Of course, the assumption $(r, k) = 1$ can be ignored by applying the result for the maximal divisor of r which is co-prime with k . On choosing the best y , we are led to

Corollary 4.2. *Suppose $\mathcal{B} = (\beta_n)$, with $|\beta_n| \leq \tau(n)^j$, satisfies the S-W condition in the segment $1 \leq n \leq N$. Then for any $r \geq 1$, $k \geq 1$, and $(l, k) = 1$, we have*

$$(4.6) \quad \sum_{\substack{n \leq N \\ n \equiv l(k) \\ (n,r)=1}} \beta_n - \frac{1}{\varphi(k)} \sum_{\substack{n \leq N \\ (n,rk)=1}} \beta_n \ll \tau(r)N(\log 2N)^{-A}$$

with any $A \geq 1$, the implied constant depending only on A and j .

Hence, indeed, (4.5) holds for the sum restricted by $(n, r) = 1$ with an extra factor $\tau(r)^2$ on the right side.

To ease applications, we also introduce into (4.5) the factor $\tau(q)^j$ at a minor expense. Specifically, applying Cauchy’s inequality and the direct bound

$$\sum_{q \leq Q} \tau(q)^{2j} \sum_{a(q)}^* \left(\sum_{\substack{n \leq N \\ n \equiv a(q)}} \tau(n)^j + \frac{1}{\varphi(q)} \sum_{\substack{n \leq N \\ (n,q)=1}} \tau(n)^j \right)^2 \ll (Q + N)(\log 2QN)^{4j+1}$$

together with (4.5), we derive the following more practical result:

Corollary 4.3. *If (β_n) satisfies the S-W condition in the segment $1 \leq n \leq N$, then*

$$\sum_{q \leq Q} \tau(q)^j \sum_{a(q)}^* \left| \sum_{\substack{n \leq N \\ n \equiv a(q) \\ (n,r)=1}} \beta_n - \frac{1}{\varphi(q)} \sum_{\substack{n \leq N \\ (n,qr)=1}} \beta_n \right|^2 \\ \ll \tau(r)^2 (Q + \sqrt{QN} + N(\log 2N)^{-A}) N(\log 2QN)^{4j+1}$$

for any positive integer r , any $Q \geq 1$, and any $A \geq 1$, the implied constant depending only on A and j .

Examples. Obviously, the S-W condition holds for $\beta_n = f(n/N)$ where $f(x)$ is a continuous function compactly supported on \mathbb{R}^+ . By the Prime Number Theorem the S-W condition holds for

$$(4.7) \quad \beta_n = \mu(n) f(n/N).$$

The S-W condition is pretty flexible. Suppose that $\mathcal{B} = (\beta_n)$ is constructed as the convolution of two cropped sequences (η_r) with $|\eta_r| \leq \tau(r)^j$, and (γ_m) with $|\gamma_m| \leq \tau(m)^j$. Specifically, with $R \geq 1$, $M \geq 1$, take

$$(4.8) \quad \beta_n = \sum_{\substack{rm=n \\ r \leq R, m \leq M}} \eta_r \gamma_m .$$

By Corollary 4.2 it follows that, if (γ_m) satisfies the S-W condition in the segment $1 \leq m \leq M$, then (β_n) satisfies the S-W condition in the segment $1 \leq n \leq N = RM$, provided $\log M \gg \log 2N$. We have already used these observations along the lines of Case (C2) in the proof of Proposition 3.4.

If (β_n) satisfies the S-W condition in the segment $1 \leq n \leq N$, then it is natural to expect that the sum (4.1) is well-approximated by

$$\frac{1}{\varphi(q)} \sum_{\substack{m \leq M, n \leq N \\ (mn,q)=1}} \alpha_m \beta_n,$$

so our goal is to estimate the difference. The challenge is to get meaningful results which are valid for moduli q as large as possible. Detecting the congruence $mn \equiv a \pmod{q}$ by Dirichlet characters, and then applying the large sieve inequality produces a remarkable result (see Theorem 9.16 of [11]):

Proposition 4.4. *Let $\mathcal{A} = (\alpha_m)$, $\mathcal{B} = (\beta_n)$ be sequences of complex numbers with $|\alpha_m| \leq \tau(m)^j$ and $|\beta_n| \leq \tau(n)^j$. Suppose \mathcal{B} satisfies the S-W condition in the segment $1 \leq n \leq N$. Then we have*

$$(4.9) \quad \sum_{q \leq Q} \max_{(a,q)=1} \left| \sum_{\substack{m \leq M, n \leq N \\ mn \equiv a(q)}} \alpha_m \beta_n - \frac{1}{\varphi(q)} \sum_{\substack{m \leq M, n \leq N \\ (mn,q)=1}} \alpha_m \beta_n \right| \ll x(\log x)^{-A}$$

where $Q = x^{\frac{1}{2}}(\log x)^{-B}$, $x = MN \geq 2$, provided $M, N \geq x^\epsilon$ with $\epsilon > 0$. Here ϵ, A, j are any positive numbers, $B = B(A)$ depends on A , and the implied constant depends only on ϵ, A , and j .

Remarks. Using the same tools, one can establish the more robust estimate

$$(4.10) \quad \sum_{q \leq Q} \max_{(a,q)=1} \sum_{\substack{m \leq M \\ (m,q)=1}} \left| \sum_{\substack{n \leq N \\ mn \equiv a(q)}} \beta_n - \frac{1}{\varphi(q)} \sum_{\substack{n \leq N \\ (n,q)=1}} \beta_n \right| \ll x(\log x)^{-A},$$

subject to the same conditions as for (4.9) with the extra condition $N \geq M$. *Hint:* First define $a = a(q)$ for which the maximum is attained. Then, relax the absolute value symbol $|*|$ by introducing factors $c_m(q) = \pm 1$ and follow the arguments in the proof of Theorem 9.16 of [11]. Apply the mean value theorem for the character sums $\sum \chi(m)c_m(q)$ and the large sieve inequality for the character sums $\sum \chi(n)\beta_n$.

It is the use of Dirichlet characters in conjunction with the large sieve inequality that sets the limit $q \leq Q = x^{\frac{1}{2}}(\log x)^{-B}$ in (4.9). In the works [7], [2], different methods were developed. These are based on the ideas of the dispersion method of Ju. V. Linnik [15] and exponential sum bounds derived from the Riemann hypothesis for varieties and from the spectral theory of automorphic forms. This produced results for moduli q significantly exceeding the barrier $x^{\frac{1}{2}}$, provided the residue class is fixed, that is, as q varies a does not move.

After the work of Goldston-Pintz-Yıldırım[13], it became desirable to have similar results with $a \pmod q$ varying in a specific fashion. Precisely, $a \pmod q$ is required to run over the roots of the polynomial

$$(4.11) \quad Z(X) = \prod_{\substack{h' \in \mathcal{H} \\ h' \neq h}} (X + h' - h),$$

where \mathcal{H} is a fixed set of integers. Because $Z(X)$ is the product of linear polynomials, there would be no issue if the modulus $q = p$ was prime; simply a runs over the finite

set of numbers $h - h'$ which do not change with p . The problem becomes intricate for composite moduli q since the residue class $a \pmod{q}$ can be very large. Yitang Zhang [18] has shown how to deal with it by a clever application of the Chinese Remainder Theorem before opening the dispersion. His estimates hold only for special moduli (free of large prime divisors), but it is all that he needs to make a happy ending to this fascinating saga.

In the following sections we reproduce Zhang's estimates along his lines (which are reminiscent of the lines in our earlier works), but we proceed differently in details to enhance the presentation. Like Y. Zhang, we do not strive for the strongest quantitative results. In fact, we shall sacrifice strength to gain even a bit of clarity.

4.2. Bilinear forms over product moduli. We are interested in the bilinear form (4.1) to modulus q which is squarefree and admits a well-controlled factorization. We assume that

$$(4.12) \quad q = rs \text{ squarefree,}$$

where r, s run over the dyadic segments $R \leq r < 2R$, $S \leq s < 2S$ with $R \geq 1$, $S \geq 1$ to be restricted later. Therefore, r, s are squarefree, co-prime and $q \asymp RS = Q$. By the Chinese Remainder Theorem one can cover uniquely all the reduced classes modulo q by the pairs of reduced classes $a \pmod{r}$ and $b \pmod{s}$.

To every s as above, we associate an arithmetic function $\gamma_s(b)$ such that

$$(4.13) \quad \gamma_s(b) \text{ is periodic in } b \text{ of period } s,$$

$$(4.14) \quad |\gamma_s(b)| \leq 1, \quad \gamma_s(b) = 0 \text{ if } (b, s) \neq 1,$$

$$(4.15) \quad \sum_{b \pmod{s}} |\gamma_s(b)| \leq \rho(s),$$

where $\rho(s)$ is a multiplicative function bounded by $\tau(s)^j$ for some $j \geq 0$. Actually, we allow $\gamma_s(b)$ to depend on r, a . However, one should remember that $\gamma_s(b)$ are defined only if rs is squarefree, a condition which will be often tacitly assumed in our exposition, but recalled only occasionally.

Given the squarefree number r and $(a, r) = 1$, our goal is to evaluate the bilinear forms

$$(4.16) \quad \sum_{mn \equiv a \pmod{r}} \alpha_m \beta_n z(mn)$$

for quite general coefficients α_m, β_n satisfying (4.2), (4.3). Here the extra factor

$$(4.17) \quad z(mn) = \sum_s \gamma_s(mn)$$

is not a new device; it appears also between the lines in [18], but not in such an abstract shape. Naturally one expects that (4.16) is well-approximated by

$$(4.18) \quad \frac{1}{\varphi(r)} \sum_{(mn,r)=1} \alpha_m \beta_n z(mn),$$

hence our goal is to estimate the difference. In fact, the coefficients α_m play only a structural role, but nothing particular. Therefore, we shall succeed by using only the upper bound (4.2). We are going to treat

$$(4.19) \quad E(r, a) = \sum_{\substack{m \sim M \\ (m,r)=1}} \tau(m)^j \left| \sum_{\substack{n \\ mn \equiv a(r)}} \beta_n z(mn) - \frac{1}{\varphi(r)} \sum_{\substack{n \\ (n,r)=1}} \beta_n z(mn) \right|.$$

Recall that the restriction $m \sim M$ means $M \leq m < 2M$. From now on we also assume that $\mathcal{B} = (\beta_n)$ is supported on $N \leq n < 2N$.

We shall be obliged to perform an averaging of $E(r, a)$ over r only in a few spots and to exploit the S-W condition for \mathcal{B} essentially to match the relevant (predicted) main terms. Here are the basic results:

Proposition 4.5. *Let $R \geq 1, S \geq 1, M \geq 1, N \geq 2$. Suppose $|\beta_n| \leq \tau(n)^j$ and that $\mathcal{B} = (\beta_n)$ satisfies the S-W condition in the interval $N \leq n < 2N$. Then we have*

$$(4.20) \quad \sum_{\substack{r \sim R \\ r \text{ squarefree}}} \tau(r)^j \max_{(a,r)=1} E(r, a) \ll x(\log x)^{-A}$$

with $x = MN$, for any $A \geq 1$, subject to the following restrictions:

$$(4.21) \quad (R + S)x^\varepsilon < N < (RS)^{\frac{3}{2}},$$

$$(4.22) \quad RN < x^{1-\varepsilon} \quad \text{and} \quad R^{\frac{5}{4}} S^{\frac{11}{4}} N^{\frac{1}{2}} < x^{1-\varepsilon},$$

$$(4.23) \quad RSN < x^{1-\varepsilon} \quad \text{or} \quad R^{\frac{5}{4}} S^3 N^{\frac{1}{2}} < x^{1-\varepsilon}.$$

Here ε is any positive number and the implied constant in (4.20) depends only on ε, A, j .

Corollary 4.6. *Let $\mathcal{A} = (\alpha_m), \mathcal{B} = (\beta_n)$ be sequences of real numbers with*

$$\begin{aligned} |\alpha_m| &\leq \tau(m)^j, & M \leq m < 2M, & & M \geq 1, \\ |\beta_n| &\leq \tau(n)^j, & N \leq n < 2N, & & N \geq 2, \end{aligned}$$

and $\alpha_m = 0, \beta_n = 0$ elsewhere. Let $x = MN, Q = RS$ with $R \geq 1, S \geq 1$, and $z \geq 2$. Suppose the following restrictions hold:

$$(4.24) \quad N \ll x^{\frac{1}{2}}, \quad z^{-1}N \ll x^\varepsilon R \ll N$$

$$(4.25) \quad zQ \ll x^{\frac{9}{17}-\varepsilon}, \quad z^{\frac{3}{2}}Q^{\frac{11}{4}} \ll Nx^{1-3\varepsilon}$$

with some $\varepsilon > 0$ and some implied constants which may depend on ε . Suppose $\mathcal{B} = (\beta_n)$ satisfies the *S-W condition* in $N \leq n < 2N$. Then we have

$$(4.26) \quad \sum_{r \sim R} \max_{(a,r)=1} \sum_{\substack{s \sim S \\ rs \text{ squarefree}}} \max_{(b,s)=1} \left| \sum_{\substack{mn \equiv a(r) \\ mn \equiv b(s)}} \alpha_m \beta_n - \frac{1}{\varphi(rs)} \sum_{(mn,rs)=1} \alpha_m \beta_n \right| \ll x(\log x)^{-A}$$

for any $A \geq 1$, the implied constant depending only on ε, A, j .

Proof of Corollary 4.6. First observe that we can assume

$$(4.27) \quad Q = RS > x^{\frac{1}{2}}(\log x)^{-B}$$

for some $B = B(A) > 0$ by applying Proposition 4.4 in the opposite case (this requires $M, N \geq x^\varepsilon$, which hold due to (4.24)).

The point b at which the maximum is attained depends on r, a , and s . Given r squarefree, $r \sim R, (a, r) = 1$ and s squarefree, $s \sim S, (r, s) = 1$, we define the function $\gamma(b)$ in $b \pmod s$ by $\gamma(b) = \pm 1$ if b maximizes the absolute value among $(b, s) = 1$ and we put $\gamma(b) = 0$ otherwise. Here ± 1 is the sign of the innermost difference. Obviously $\gamma(b) = \gamma_s(b) = \gamma_s(b; r, a)$ satisfies the conditions (4.13)-(4.15) with $\rho(s) = 1$. Now we can write the above sum in the following form:

$$\sum_{r \sim R} \max_{(a,r)=1} \sum_{s \sim S} \left(\sum_{mn \equiv a(r)} \alpha_m \beta_n \gamma_s(mn) - \frac{1}{\varphi(rs)} \sum_{(mn,rs)=1} \alpha_m \beta_n \sum_{b(s)}^* \gamma_s(b) \right).$$

Changing the order of summation we estimate the sum (4.26) by

$$(4.28) \quad \sum_{r \sim R} \max_{(a,r)=1} \sum_{\substack{m \sim M \\ (m,r)=1}} \tau(m)^j \left| \sum_{mn \equiv a(r)} \beta_n z(mn) - \frac{1}{\varphi(r)} \sum_{(n,r)=1} \beta_n \sum_{(s,mn)=1} \frac{1}{\varphi(s)} \sum_{b(s)}^* \gamma_s(b) \right|.$$

Recall our convention that $\gamma_s(b)$ is defined only if rs is squarefree, so (4.28) has the hidden condition $(r, s) = 1$.

We are approaching the sum (4.19). The main term in (4.28) is not exactly the one we have assumed in (4.19), but it is very close. Indeed, if we write

$$\sum_{b(s)}^* \gamma_s(b) = \sum_{b(s)}^* \gamma_s(mb)$$

in (4.28) and go back to (4.17) for $z(mn)$ in the main term of (4.19), then it is clear that the difference between the main terms is

$$\frac{1}{\varphi(r)} \sum_s \sum_{b(s)}^* \gamma_s(mb) \left(\sum_{\substack{(n,r)=1 \\ n \equiv b(s)}} \beta_n - \frac{1}{\varphi(s)} \sum_{(n,rs)=1} \beta_n \right).$$

This difference contributes to (4.28) at most

$$\begin{aligned} \Delta(R, S, M, N) = & \sum_r \frac{1}{\varphi(r)} \max_{(a,r)=1} \sum_s \sum_{b(s)}^* \left(\sum_m \tau(m)^j |\gamma_s(mb)| \right) \left| \sum_{\substack{(n,r)=1 \\ n \equiv b(s)}} \beta_n - \frac{1}{\varphi(s)} \sum_{(n,rs)=1} \beta_n \right|. \end{aligned}$$

Note that

$$\begin{aligned} \sum_s \sum_{b(s)}^* \left(\sum_m \tau(m)^j |\gamma_s(mb)| \right)^2 & \leq \sum_s \sum_{m' \equiv m(s)} (\tau(m') \tau(m))^j \\ & \ll M^2 (\log 2M)^{4j}, \end{aligned}$$

because $s \sim S$ and $S \ll M^{\frac{1}{2}}$ by the conditions (4.24), (4.25). Hence, by Cauchy's inequality

$$\Delta(R, S, M, N)^2 \ll M^2 (\log 2M)^{4j} \sum_r \frac{1}{\varphi(r)} \Delta_r(S, N),$$

where

$$\Delta_r(S, N) = \sum_{s \sim S} \sum_{b(s)}^* \left| \sum_{\substack{(n,r)=1 \\ n \equiv b(s)}} \beta_n - \frac{1}{\varphi(s)} \sum_{(n,rs)=1} \beta_n \right|^2 \ll \tau(r)^2 N^2 (\log N)^{-A}$$

by Corollary 4.3 (note that $S \ll N^{\frac{1}{2}}$ by the conditions (4.24), (4.25)). This shows that

$$(4.29) \quad \Delta(R, S, M, N) \ll MN (\log N)^{-A},$$

which is admissible for (4.26).

To complete the proof of Corollary 4.6 it remains to check that the conditions (4.24), (4.25) imply the conditions (4.21)-(4.23) so that (4.20) can be applied. The task is not hard. First, (4.25) implies two corresponding estimates:

$$SN \ll x^{\frac{9}{17}}, \quad R^{\frac{5}{4}} S^{\frac{11}{4}} N^{\frac{1}{2}} \ll x^{1-\frac{3\epsilon}{2}},$$

because $z \gg NR^{-1}x^{-\epsilon}$ by (4.24). The second estimate, together with (4.27), yields $S \ll R$, so that (4.21) holds by (4.24) and (4.27). Next, we have $RN \ll N^2x^{-\epsilon} \ll x^{1-\epsilon}$ by (4.24), so that (4.22) holds. Finally, suppose that (4.23) is false. Then,

$$(SN)^{17} = (SN^2)^5 (S^3 N^{\frac{7}{4}})^4 \gg (x^\epsilon RSN)^5 (x^{\frac{5\epsilon}{4}} R^{\frac{5}{4}} S^3 N^{\frac{1}{2}})^4 \geq x^{9+\epsilon},$$

which gives the desired contradiction. □

5. The Dispersion Method

5.1. Statement of results. We shall prove Proposition 4.5 by the dispersion method. First, applying Cauchy’s inequality to (4.19) we get

$$(5.1) \quad E^2(r, a) \ll \mathcal{D}(r, a) M(\log M)^{4j},$$

where

$$(5.2) \quad \mathcal{D}(r, a) = \sum_{(m,r)=1} f\left(\frac{m}{M}\right) \left| \sum_{mn \equiv a(r)} \beta_n z(mn) - \frac{1}{\varphi(r)} \sum_{(n,r)=1} \beta_n z(mn) \right|^2$$

and f is a smooth function compactly supported on \mathbb{R}^+ .

We shall be able to handle the dispersion $\mathcal{D}(r, a)$ for any single class $a \pmod r$, except for matching the main terms. Therefore, we state the results in two stages.

Proposition 5.1. *Let r be squarefree, $(a, r) = 1$. Suppose $S \geq 1$, $M \geq 1$, $N \geq 2$, and*

$$(5.3) \quad (r + S)x^\epsilon \ll N \ll (rS)^{\frac{3}{2}}, \quad \text{with } x = MN.$$

Then, we have

$$(5.4) \quad \mathcal{D}(r, a) \ll \frac{M}{r} \sum_{k < 2S} \frac{\tau(k)^j}{k} \left(\mathcal{E}_k(r) + \frac{1}{\varphi(r)} \mathcal{E}_{kr}(1) \right) (\log N)^{2j} \\ + \{r^{-\frac{3}{2}} M^{\frac{1}{2}} N^2 + r^{-\frac{3}{4}} S^{\frac{11}{4}} N^{\frac{3}{2}}\} x^\epsilon + \min\{r^{-\frac{3}{2}} S^{\frac{1}{2}} M^{\frac{1}{2}} N^2, r^{-\frac{3}{4}} S^3 N^{\frac{3}{2}}\} x^\epsilon,$$

where

$$(5.5) \quad \mathcal{E}_k(r) = \sum_{c < 2S} \frac{\tau(c)^j}{c} \sum_{\nu(cr)}^* \left(\sum_{\substack{(n,k)=1 \\ n \equiv \nu(cr)}} \beta_n - \frac{1}{\varphi(cr)} \sum_{(n,ckr)=1} \beta_n \right)^2$$

with implied constants depending on ε and j . Moreover, if the sequence (β_n) satisfies the *S-W* condition in the segment $N \leq n < 2N$, then

$$(5.6) \quad \sum_{r < R} \mathcal{E}_k(r) \ll \tau(k) N^2 (\log x)^{-A}$$

for $R \ll Nx^{-\varepsilon}$, with any $A \geq 1$, the implied constant depending on A , ε , and j .

Now we can derive Proposition 4.5 from Proposition 5.1. By Cauchy’s inequality, the square of the sum (4.20) is bounded by

$$RM(\log R)^{4j} (\log M)^{4j} \sum_{r \sim R} \max_{(a,r)=1} \mathcal{D}(r, a).$$

Applying (5.4) and (5.6) to the above, we find

$$\begin{aligned} & RMR^{-1}Nx(\log x)^{-A} + RM\{R^{\frac{1}{2}}M^{\frac{1}{2}}N^2 + R^{\frac{1}{4}}S^{\frac{11}{4}}N^{\frac{3}{2}}\}x^\varepsilon \\ & + RM \min\{R^{-\frac{1}{2}}S^{\frac{1}{2}}M^{\frac{1}{2}}N^2, R^{\frac{1}{4}}S^3N^{\frac{3}{2}}\}x^\varepsilon \ll x^2(\log x)^{-2A} \end{aligned}$$

subject to the conditions (4.21)-(4.23). This proves (4.20).

Now we proceed to the proof of Proposition 5.1, which is the gist of this project. We open the square in (5.2) and arrange the dispersion into three parts:

$$(5.7) \quad \mathcal{D}(r, a) = \mathcal{D}_1 - 2\mathcal{D}_2 + \mathcal{D}_3$$

with the purpose of evaluating each part asymptotically so that the resulting main terms match and cancel out.

5.2. Evaluation of \mathcal{D}_1 . We start from \mathcal{D}_1 , because it is the hardest part and because the other two parts $\mathcal{D}_2, \mathcal{D}_3$ will make use of the arguments which appear in our treatment of \mathcal{D}_1 . We have

$$(5.8) \quad \begin{aligned} \mathcal{D}_1 &= \sum_{(m,r)=1} f\left(\frac{m}{M}\right) \left(\sum_{mn \equiv a(r)} \beta_n \left(\sum_s \gamma_s(mn) \right) \right)^2 \\ &= \sum_{n \equiv n'(r)} \sum \beta_n \beta_{n'} \sum_s \sum_{s'} \sum_{mn \equiv a(r)} f\left(\frac{m}{M}\right) \gamma_s(mn) \gamma_{s'}(mn'). \end{aligned}$$

Note that the coprimality conditions $(m, rss') = (n, rs) = (n', rs') = 1$ and $(ss', r) = 1$ are redundant with our conditions imposed on $\gamma_s(n)$. We split the summation over m into residue classes

$$\mu \equiv \overline{an[s, s']}[s, s'] + \omega\bar{r}r \pmod{r[s, s']},$$

where ω runs over the reduced classes $(\text{mod } [s, s'])$, and for each class we apply the Poisson formula

$$\sum_{m \equiv \mu \pmod{r[s, s']}} f(m/M) = \frac{M}{r[s, s']} \sum_h \hat{f}\left(\frac{hM}{r[s, s']}\right) e\left(\frac{-\mu h}{r[s, s']}\right)$$

giving

$$(5.9) \quad \mathcal{D}_1 = \frac{M}{r} \sum_{n \equiv n' \pmod{r}} \beta_n \beta_{n'} \sum_s \sum_{s'} \sum_{\omega \pmod{[s, s']}} \gamma_s(\omega n) \gamma_{s'}(\omega n') / [s, s'] \cdot \sum_h \hat{f}\left(\frac{hM}{r[s, s']}\right) e\left(-ah \frac{\overline{n[s, s']}}{r} - \frac{\omega h \bar{r}}{[s, s']}\right).$$

The main term will emerge from the zero frequency, and we denote its contribution by

$$(5.10) \quad \mathcal{D}_{10} = \hat{f}(0) \frac{M}{r} \sum_{n \equiv n' \pmod{r}} \beta_n \beta_{n'} \sum_s \sum_{s'} \frac{1}{[s, s']} \sum_{\omega \pmod{[s, s']}} \gamma_s(\omega n) \gamma_{s'}(\omega n').$$

We denote by \mathcal{D}_{11} the contribution of the terms with $h \neq 0$ to (5.9), so we have

$$(5.11) \quad \mathcal{D}_1 = \mathcal{D}_{10} + \mathcal{D}_{11}.$$

The complete sum over $\omega \pmod{[s, s']}$ in \mathcal{D}_{10} depends only on the residue class of $n'\bar{n}$ modulo

$$(5.12) \quad c = (s, s').$$

Precisely, if $(n, s) = (n', s') = 1$ and $n' \equiv en \pmod{c}$, then

$$(5.13) \quad \sum_{\omega \pmod{[s, s']}} \gamma_s(\omega n) \gamma_{s'}(\omega n') = \sum_{\omega \pmod{[s, s']}} \gamma_s(\omega) \gamma_{s'}(\omega e) = [s, s'] \sigma_{s, s'}(e),$$

the last equation being the definition of $\sigma_{s, s'}(e)$.

Proof of (5.13). This is an easy exercise in using the periodicity of $\gamma_s(b)$. Write $\omega = \alpha s' c^{-1} + \beta s$ with $\alpha \pmod{s}$ and $\beta \pmod{s' c^{-1}}$ being primitive classes. Then, we have

$$\begin{aligned} \gamma_s(\omega n) \gamma_{s'}(\omega n') &= \gamma_s(\alpha s' e^{-1} n) \gamma_{s'}((\alpha s' c^{-1} + \beta s) n') \\ &= \gamma_s(\alpha s' c^{-1} n) \gamma_{s'}(\alpha s' c^{-1} e n + \beta s n'). \end{aligned}$$

Here, summing over β we can replace n' by en . This proves (5.13). \square

Note that

$$(5.14) \quad [s, s'] \sum_{e(c)} \sigma_{s,s'}(e) = \gamma(s)\gamma(s'),$$

where c and $\sigma_{s,s'}(e)$ are defined by (5.12), (5.13) respectively, and

$$(5.15) \quad \gamma(s) = \sum_{\omega(s)} \gamma_s(\omega).$$

By (5.13) we get

$$\mathcal{D}_{10} = \hat{f}(0) \frac{M}{r} \sum_s \sum_{s'} \sum_{e(c)} \sigma_{s,s'}(e) \sum_{\substack{n' \equiv n(r) \\ n' \equiv en(c)}} \beta_n \beta_{n'}.$$

Do not forget the hidden coprimality conditions $(n, rs) = (n', rs') = 1$. The sum of $\beta_n \beta_{n'}$ can be arranged as follows:

$$\begin{aligned} \sum_{\substack{n' \equiv n(r) \\ n' \equiv en(c)}} \beta_n \beta_{n'} &= \frac{1}{\varphi(rc)} \left(\sum_{(n,rs)=1} \beta_n \right) \left(\sum_{(n',rs')=1} \beta_{n'} \right) \\ &+ \sum_{u(r)}^* \sum_{v(c)}^* \left(\sum_{\substack{n \equiv u(r) \\ n \equiv v(c)}} \beta_n - \frac{1}{\varphi(rc)} \sum_{(n,rs)=1} \beta_n \right) \left(\sum_{\substack{n' \equiv u(r) \\ n' \equiv ev(c)}} \beta_{n'} - \frac{1}{\varphi(rc)} \sum_{(n',rs')=1} \beta_{n'} \right). \end{aligned}$$

Hence, we obtain

$$(5.16) \quad \mathcal{D}_{10} = \mathcal{D}_{10}^* + \mathcal{D}_{10}^+,$$

where \mathcal{D}_{10}^* is the main term

$$(5.17) \quad \mathcal{D}_{10}^* = \hat{f}(0) \frac{M}{r\varphi(r)} \sum_s \sum_{s'} \frac{\gamma(s)\gamma(s')}{\varphi(c)[s, s']} \left(\sum_{(n,rs)=1} \beta_n \right) \left(\sum_{(n',rs')=1} \beta_{n'} \right),$$

and the remaining part is given by

$$(5.18) \quad \begin{aligned} \mathcal{D}_{10}^+ &= \hat{f}(0) \frac{M}{r} \sum_{u(r)}^* \sum_s \sum_{s'} \sum_{v(c)}^* \sum_{v'(c)}^* \sigma_{s,s'}(v/v') \\ &\cdot \left(\sum_{\substack{n \equiv u(r) \\ n \equiv v(c) \\ (n,rs)=1}} \beta_n - \frac{1}{\varphi(rc)} \sum_{(n,rs)=1} \beta_n \right) \left(\sum_{\substack{n \equiv u(r) \\ n \equiv v'(c) \\ (n,rs')=1}} \beta_n - \frac{1}{\varphi(rc)} \sum_{(n,rs')=1} \beta_n \right). \end{aligned}$$

The main term \mathcal{D}_{10}^* will be matched with the main terms of $\mathcal{D}_2, \mathcal{D}_3$ but the remaining part \mathcal{D}_{10}^+ can only be estimated completely after summation over $r \sim R$, because we have to apply the property (4.4) for moduli cr .

By the inequality $2|AB| \leq A^2 + B^2$ we find that the quadruple sum over $s, s', v(c), v'(c)$ in (5.18) is bounded by

$$\sum_s \sum_{c|s} \sum_{v(c)}^* \sum_{(s',s)=c} \sum_{v'(c)} |\sigma_{s,s'}(v'/v)| \left(\sum_{\substack{n \equiv u(r) \\ n \equiv v(c) \\ (n,rs)=1}} \beta_n - \frac{1}{\varphi(rc)} \sum_{(n,rs)=1} \beta_n \right)^2.$$

Here the sum over $v' \pmod{c}$ is bounded by (see (5.14), (4.15))

$$\begin{aligned} \sum_{e(c)}^* |\sigma_{s,s'}(e)| &\leq \frac{1}{[s, s']} \sum_{\omega([s, s'])} \sum_{e(c)}^* |\gamma_s(\omega)\gamma_{s'}(\omega e)| \\ &= \frac{1}{[s, s']} \left(\sum_{\omega(s)} |\gamma_s(\omega)| \right) \left(\sum_{\omega(s')} |\gamma_{s'}(\omega)| \right) \leq \frac{\rho(s)\rho(s')}{[s, s']}. \end{aligned}$$

Now put $s = ck, s' = cl$. Summing over $k, l < 2S$ we conclude the following clear estimate for \mathcal{D}_{10}^+ :

$$|\mathcal{D}_{10}^+| \leq |\hat{f}(0)|M \sum_{c,k,l < 2S} \sum_{\nu(cr)}^* \frac{\rho^2(c)\rho(kl)}{ckl} \left(\sum_{\substack{(n,crk)=1 \\ n \equiv \nu(cr)}} \beta_n - \frac{1}{\varphi(cr)} \sum_{(n,crk)=1} \beta_n \right)^2.$$

Hence,

$$(5.19) \quad \mathcal{D}_{10}^+ \ll \frac{M}{r} \left(\sum_{l < 2S} \frac{\tau(l)^j}{l} \right) \sum_{k < 2S} \frac{\tau(k)^j}{k} \mathcal{E}_k(r),$$

which yields the first part on the right side of (5.4).

Using Corollary 4.3, we can estimate $\mathcal{E}(r)$ on average over $r < R$ as follows:

$$\begin{aligned} \sum_{r < R} \mathcal{E}_k(r) &\leq \frac{2 \log 2S}{C} \sum_{q \leq CR} \tau(q)^{j+1} \sum_{\nu(q)}^* \left(\sum_{\substack{n \leq N \\ n \equiv \nu(q) \\ (n,k)=1}} \beta_n - \frac{1}{\varphi(q)} \sum_{\substack{n \leq N \\ (n,qk)=1}} \beta_n \right)^2 \\ &\ll \frac{\tau(k)}{C} (CR + \sqrt{CRN} + N(\log N)^{-A}) N(\log 2QN)^{4j+2}, \end{aligned}$$

where $1 \leq C \leq 2S$. Since j is fixed and A is arbitrary, we obtain (5.6).

5.3. **First estimation of \mathcal{D}_{11} .** Recall that \mathcal{D}_{11} is given by (5.9) with $h \neq 0$. We shall apply two methods for the estimation of \mathcal{D}_{11} . We begin with the one which is more complex. The second method will be easier because it makes use of the basic transformations in the first one.

We split the summation over n', n according to the difference, say

$$(5.20) \quad n' = n + u.$$

Therefore, for every given u , the variable n' will be uniquely determined by n . Let \mathcal{U} be the set of shifts u , so $\mathcal{U} \subseteq (-N, N)$, and in case of \mathcal{D}_{11} we have $u \equiv 0 \pmod{r}$. However, this property of u does not play any role, except that it ensures $|\mathcal{U}| \leq 2Nr^{-1} + 1 \leq 3Nr^{-1}$. Because the cardinality of \mathcal{U} is the only factor in the result, we shall be able to treat similarly the corresponding sum \mathcal{D}_{21} in subsection 5.6. Now, by (5.9) we obtain

$$(5.21) \quad \begin{aligned} |\mathcal{D}_{11}| &\leq \frac{2M}{r} \sum_{u \in \mathcal{U}} \sum_{(n,r)=1} \sum_{(n',r)=1} |\beta_n \beta_{n'}| \left| \sum_s \sum_{s'} \sum_{\omega} \sum_{h \neq 0} \right| \\ &\ll r^{-2} MN^{\frac{3}{2}} (\log N)^j T(u)^{\frac{1}{2}} \end{aligned}$$

for some $u \in \mathcal{U}$ with

$$T(u) = \sum_{(n,r)=1} f\left(\frac{n}{N}\right) \left| \sum_s \sum_{s'} \sum_{\omega} \sum_{h \neq 0} \right|^2.$$

Here, the coefficients $\beta_n, \beta_{n'}$ are gone; they are replaced with a nice smooth function f supported in $[\frac{1}{2}, 3]$. Squaring out, we get

$$(5.22) \quad \begin{aligned} T(u) &= \sum_{(n,r)=1} f\left(\frac{n}{N}\right) \sum_s \sum_{s'} \sum_{s_1} \sum_{s'_1} \frac{1}{[s, s'] [s_1, s'_1]} \\ &\quad \sum_{\omega \in ([s, s'])} \gamma_s(\omega n) \gamma_{s'}(\omega n') \sum_{\omega_1 \in ([s_1, s'_1])} \gamma_{s_1}(\omega_1 n) \gamma_{s'_1}(\omega_1 n') \\ &\quad \sum_{hh_1 \neq 0} \hat{f}\left(\frac{hM}{r[s, s']}\right) \bar{\hat{f}}\left(\frac{h_1M}{r[s_1, s'_1]}\right) e\left(\frac{-a\bar{n}}{r} (h\overline{[s, s']} - h_1\overline{[s_1, s'_1]}) - \frac{\omega h \bar{r}}{[s, s']} + \frac{\omega_1 h_1 \bar{r}}{[s_1, s'_1]}\right). \end{aligned}$$

Put

$$(5.23) \quad d = h[s_1, s'_1] - h_1[s, s'].$$

Let $T^=(u)$ be the partial sum of $T(u)$ with $d = 0$, and $T^\neq(u)$ the remaining sum.

We estimate $T^=(u)$ by summing all terms with absolute value. Given n with $(n, ss_1) = 1$ and $(n', s's'_1) = 1$ (recall (5.20)) we find that the sum over $\omega \pmod{[s, s']}$ is bounded by $\rho(s)\rho(s')$ and the sum over $\omega_1 \pmod{[s_1, s'_1]}$ is bounded by $\rho(s_1)\rho(s'_1)$. Hence

$$T^=(u) \ll N \sum_s \sum_{s'} \sum_{s_1} \sum_{s'_1} \frac{\rho(s)\rho(s')\rho(s_1)\rho(s'_1)}{[s, s'][s_1, s'_1]} \sum_{h \neq 0}^* \left| \hat{f} \left(\frac{hM}{r[s, s']} \right) \right|,$$

where \sum^* is restricted by the requirement that $h_1 = h[s_1, s'_1]/[s, s']$ is an integer. Hence, $[s_1, s'_1] = k[s, s']/(h, [s, s'])$ where k is an integer. We get $1 \leq k \leq 4S^2$ and

$$T^=(u) \ll S^\varepsilon N \sum_s \sum_{s'} \sum_{h \neq 0} \sum_{1 \leq k \leq 4S^2} \frac{(h, [s, s'])}{k[s, s']^2} \left| \hat{f} \left(\frac{hM}{r[s, s']} \right) \right| \tau_3(k)$$

because, for given s, s', h, k , the number of s_1, s'_1 is $\leq \tau_3(kss') \ll S^\varepsilon$. Hence

$$T^=(u) \ll S^\varepsilon N \sum_{l < 4S^2} \sum_{h \neq 0} \frac{(h, l)}{l^2} \left| \hat{f} \left(\frac{hM}{rl} \right) \right| (\log 2|h|)^4.$$

This easily yields

$$(5.24) \quad T^=(u) \ll rNM^{-1}(rS)^\varepsilon.$$

Now we proceed to $T^\neq(u)$ which is given by (5.22) with the summation condition $d = h[s_1, s'_1] - h_1[s, s'] \neq 0$. We put

$$(5.25) \quad \omega n \equiv \alpha(s), \quad \omega n' \equiv \alpha'(s'), \quad \omega_1 n \equiv \alpha_1(s_1), \quad \omega_1 n' \equiv \alpha'_1(s'_1),$$

and we extract from (5.22) the following formula for $T^\neq(u)$:

$$(5.26) \quad \sum_{(ss's_1s'_1, r)=1} \sum_{\substack{\alpha(s) \\ \alpha_1(s_1)}} \sum_{\alpha'(s')} \sum_{\alpha'_1(s'_1)} \left(\sum_{\substack{\alpha_1(s_1) \\ \alpha'_1(s'_1)}} \gamma_{s_1}(\alpha_1)\gamma_{s'_1}(\alpha'_1)/[s_1, s'_1] \right) \left(\sum_{\substack{hh_1 \neq 0 \\ d \neq 0}} \hat{f} \left(\frac{hM}{r[s, s']} \right) \hat{f} \left(\frac{h_1M}{r[s_1, s'_1]} \right) \right) \\ \sum_{\substack{(n, rss_1)=1 \\ (n', rs's'_1)=1}} f \left(\frac{n}{N} \right) e \left(\frac{-ad}{r} \overline{n[s, s'][s_1, s'_1]} - \frac{\omega h \bar{r}}{[s, s']} + \frac{\omega_1 h_1 \bar{r}}{[s_1, s'_1]} \right),$$

where $\omega \pmod{[s, s']}$ and $\omega_1 \pmod{[s_1, s'_1]}$ are determined in terms of n and $n' = n + u$ by the congruences (5.25).

The innermost sum over n in (5.26) is an incomplete Kloosterman sum of modulus rw with $w = [s, s', s_1, s'_1]$. We shall see this clearly after completing the sum and factoring

into prime moduli. To complete the sum, we split n into residue classes modulo rw and apply Poisson's formula. We find

$$(5.27) \quad \sum_n f\left(\frac{n}{N}\right) e\left(\frac{n}{N}\right) = \frac{N}{rw} \sum_l \hat{f}\left(\frac{lN}{rw}\right) V_l(rw)$$

where $V_l(rw)$ is the complete sum

$$V_l(rw) = \sum_{\nu \pmod{rw}} e\left(\frac{-\nu l}{rw} - \frac{ad}{r} \overline{\nu[s, s'] [s_1, s'_1]} - \frac{\omega h \bar{r}}{[s, s']} - \frac{\omega_1 h_1 \bar{r}}{[s_1, s'_1]}\right).$$

Here $\omega \pmod{[s, s']}$ and $\omega_1 \pmod{[s_1, s'_1]}$ are determined in terms of ν and $\nu' = \nu + u$ by the congruences (5.25) with n, n' substituted by ν, ν' , respectively. The complete sum factors into

$$V_l(rw) = \mathcal{J}_l(w) K(l\bar{w}, ad\overline{[s, s'] [s_1, s'_1]}; r)$$

where $K(*, *, r)$ is the Kloosterman sum to modulus r , and

$$\mathcal{J}_l(w) = \sum_{\substack{\nu \pmod{w} \\ (\nu, ss_1)=1 \\ (\nu', s'_1)=1}} e\left(\frac{-\nu l \bar{r}}{[s, s', s_1, s'_1]} - \frac{\omega h \bar{r}}{[s, s']} + \frac{\omega_1 h_1 \bar{r}}{[s_1, s'_1]}\right).$$

By Weil's estimate we get

$$|K(l\bar{w}, *, r)| \leq (l, r)^{\frac{1}{2}} r^{\frac{1}{2}} \tau(r).$$

For $l = 0$ we can do better by using estimates for Ramanujan's sums

$$|K(0, d; r)| \leq (d, r).$$

Next $\mathcal{J}_l(w)$ factors into sums to distinct prime moduli $p \mid w$. If $p \parallel ss's_1s'_1$, then the local sums appear in the form

$$\sum_{\nu \pmod{p}} e_p\left(-\nu l \overline{rw/p} + A\bar{\nu} + A'\bar{\nu}'\right)$$

with $AA' = 0$ and $\nu' = \nu + u$. These local sums are Kloosterman sums (if $A = 0$, then shift ν to $\nu - u$) which are bounded by $2(l, p)^{\frac{1}{2}} p^{\frac{1}{2}}$. If $p^2 \mid ss's_1s'_1$, then we use the trivial bound p . Hence, we get

$$|\mathcal{J}_l(w)| \leq \tau(w) \prod_{p \parallel ss's_1s'_1} (l, p)^{\frac{1}{2}} p^{\frac{1}{2}} \prod_{p^2 \mid ss's_1s'_1} p \leq \tau(w) (l, w)^{\frac{1}{2}} (ss's_1s'_1)^{\frac{1}{2}}.$$

For $l = 0$ we do better; $|\mathcal{J}_0(w)| \leq \tau(w)w$. Multiplying the above local estimates we obtain

$$(5.28) \quad |V_l(rw)| \leq \tau(rw)(l, rw)^{\frac{1}{2}} r^{\frac{1}{2}} (2S)^2$$

for every l , and for $l = 0$ we have

$$(5.29) \quad |V_0(rw)| \leq \tau(w)w(d, r).$$

Introducing (5.28) and (5.29) into (5.27) we get

$$(5.30) \quad \sum_n f\left(\frac{n}{N}\right) e(\) \ll N \frac{(d, r)}{r} \tau(w) + S^2 r^{\frac{1}{2}} \tau_3(rw)$$

where d is defined in (5.23). Introducing (5.30) into (5.26) we get

$$(5.31) \quad T^\neq(u) \ll \sum_s \sum_{s'} \sum_{s_1} \sum_{s'_1} \frac{\rho(s)\rho(s')}{[s, s']} \frac{\rho(s_1)\rho(s'_1)}{[s_1, s'_1]} \\ \sum_{\substack{hh_1 \neq 0 \\ d \neq 0}} \left| \hat{f}\left(\frac{hM}{r[s, s']}\right) \hat{f}\left(\frac{h_1M}{r[s_1, s'_1]}\right) \right| \left(N \frac{(d, r)}{r} + S^2 r^{\frac{1}{2}} \right) \tau_3(rw).$$

It is easy to see that the second term $S^2 r^{\frac{1}{2}}$ contributes at most (ignoring $d \neq 0$)

$$(5.32) \quad S^2 r^{\frac{1}{2}} \left(\frac{r}{M}\right)^2 \sum_s \sum_{s'} \sum_{s_1} \sum_{s'_1} \rho(s)\rho(s')\rho(s_1)\rho(s'_1)\tau_3(rw) \ll r^{\frac{5}{2}} S^6 M^{-2} (rS)^\epsilon.$$

The first term $N(d, r)/r$ contributes at most (keeping $d \neq 0$)

$$(rS)^\epsilon \frac{N}{r} \sum_{S \leq a, a_1 \leq 4S^2} (aa_1)^{-1} \sum_{\substack{hh_1 \neq 0 \\ ha_1 \neq h_1a}} \left(1 + \frac{|h|M}{ra}\right)^{-4} \left(1 + \frac{|h_1|M}{ra_1}\right)^{-4} (ha_1 - h_1a, r).$$

One can show by crude arguments that this is $\ll (rS)^\epsilon rS^4 NM^{-2}$, so it is smaller than (5.32) because of (5.3). Adding the above bound (5.32) for $T^\neq(u)$ to the bound (5.24) for $T^=(u)$ we find that

$$(5.33) \quad T(u) \ll (rNM^{-1} + r^{\frac{5}{2}} S^6 M^{-2})(rS)^\epsilon.$$

Hence, by (5.21) we obtain

$$(5.34) \quad \mathcal{D}_{11} \ll \left(r^{-\frac{3}{2}} N^2 M^{\frac{1}{2}} + r^{-\frac{3}{4}} S^3 N^{\frac{3}{2}}\right) (rS)^\epsilon.$$

5.4. **Second estimation of \mathcal{D}_{11} .** Recall again that \mathcal{D}_{11} is given by (5.9) with $h \neq 0$. Now we put the variable s' outside and apply Cauchy's inequality getting (in place of (5.21))

$$(5.35) \quad \mathcal{D}_{11} \ll r^{-2} S M N^{\frac{3}{2}} (\log N)^j T(u; s')^{\frac{1}{2}}$$

where $T(u; s')$ is given by the same Fourier series as $T(u)$ in (5.22), but with $s' = s'_1$ being fixed, so no longer summation variables. As before, we split

$$(5.36) \quad T(u; s') = T^=(u; s') + T^\neq(u; s')$$

according to $d = h[s_1, s'_1] - h_1[s, s'] = h[s_1, s'] - h_1[s, s']$ vanishing or not vanishing.

By the same arguments applied to $T^=(u)$ we derive

$$(5.37) \quad T^=(u; s') \ll r S^{-1} N M^{-1} (rS)^\epsilon.$$

Next, we proceed with $T^\neq(u; s')$ along the same lines as we did with $T^\neq(u)$, keeping in mind that $s' = s'_1$ are fixed. We arrive again at the product of Kloosterman sums, say $V_l(rw)$ with $w = [s, s', s_1, s'_1] = [s, s', s_1]$. Note that $w \ll S^3$, while we had before $w \ll S^4$. This deficit of s'_1 in w will translate into an extra saving of a factor $S^{1/2}$ in estimates for the relevant Kloosterman sums. Specifically, the complete sum $V_l(rw)$ factors into the sum of modulus r and the sum $\mathcal{J}_l(w; s')$ of modulus $w = [s, s', s_1]$. The one of modulus r gets the same bounds, but $\mathcal{J}_l(w; s')$ gets better bounds;

$$|\mathcal{J}_l(w; s')| \leq \tau(w) \prod_{p|ss's_1} (l, p)^{\frac{1}{2}} p^{\frac{1}{2}} \prod_{p^2|ss's_1} p \leq \tau(w) (l, w)^{\frac{1}{2}} (ss's_1)^{\frac{1}{2}}$$

and $|\mathcal{J}_0(w)| \leq \tau(w)w$. Hence we derive

$$|V_l(rw)| \leq \tau(rw) (l, rw)^{\frac{1}{2}} r^{\frac{1}{2}} (2S)^{\frac{3}{2}},$$

which is better by a factor $(2S)^{\frac{1}{2}}$ than (5.28), and

$$|V_0(rw)| \leq \tau(w)w(d, r),$$

which is the same as (5.29). Next, following along the same lines, we get

$$(5.38) \quad \begin{aligned} T^\neq(u) &\ll \sum_s \sum_{s_1} \frac{\rho(s)\rho(s')}{[s, s']} \frac{\rho(s_1)\rho(s'_1)}{[s_1, s'_1]} \\ &\sum_{\substack{hh_1 \neq 0 \\ d \neq 0}} \sum \left| \hat{f}\left(\frac{hM}{r[s, s']}\right) \hat{f}\left(\frac{h_1M}{r[s_1, s'_1]}\right) \right| \left(N \frac{(d, r)}{r} + S^{\frac{3}{2}} r^{\frac{1}{2}} \right) \tau_3(rw) \\ &\ll r^2 S^2 M^{-2} (r^{-1}N + r^{\frac{1}{2}} S^{\frac{3}{2}}) (rS)^\epsilon \ll r^{\frac{5}{2}} S^{\frac{7}{2}} M^{-2} (rS)^\epsilon \end{aligned}$$

because of (5.3). Adding this bound to (5.37) we obtain

$$(5.39) \quad T(u; s') \ll (rS^{-1}NM^{-1} + r^{\frac{5}{2}}S^{\frac{7}{2}}M^{-2})(rS)^\varepsilon.$$

Hence, by (5.35) we obtain

$$(5.40) \quad \mathcal{D}_{11} \ll (r^{-\frac{3}{2}}S^{\frac{1}{2}}N^2M^{\frac{1}{2}} + r^{-\frac{3}{4}}S^{\frac{11}{4}}N^{\frac{3}{2}})(rS)^\varepsilon.$$

Note that, by comparison, the first term in our first bound (5.34) is better by a factor $S^{\frac{1}{2}}$; however, the second term in our second bound (5.40) is better by a factor $S^{\frac{1}{4}}$.

5.5. Evaluation of \mathcal{D}_2 . Recall that \mathcal{D}_2 stands for the cross terms in the dispersion (5.7),

$$(5.41) \quad \mathcal{D}_2 = \frac{1}{\varphi(r)} \sum_{(n',r)=1} \beta_{n'} \sum_s \sum_{s'} \sum_{(m,r)=1} f\left(\frac{m}{M}\right) \sum_{mn \equiv a(r)} \beta_n \gamma_s(mn) \gamma_{s'}(mn').$$

We execute the summation over m by splitting into residue classes modulo $r[s, s']$ as in \mathcal{D}_1 . Apply Poisson's formula for every class, and pull out the contribution of the frequency $h = 0$, getting

$$(5.42) \quad \mathcal{D}_2 = \mathcal{D}_{20} + \mathcal{D}_{21}$$

with

$$\mathcal{D}_{20} = \hat{f}(0) \frac{M}{r\varphi(r)} \sum_n \sum_{n'} \beta_n \beta_{n'} \sum_s \sum_{s'} \frac{1}{[s, s']} \sum_{\omega([s, s'])} \gamma_s(\omega n) \gamma_{s'}(\omega n').$$

Do not forget the hidden co-primality conditions $(n, rs) = (n', rs') = 1$. The inner sum depends only on n'/n modulo $c = (s, s')$, see (5.13). Hence

$$\mathcal{D}_{20} = \hat{f}(0) \frac{M}{r\varphi(r)} \sum_s \sum_{s'} \sum_{e(c)} \sigma_{s, s'}(e) \sum_{n' \equiv en(c)} \beta_n \beta_{n'}.$$

Following the arguments applied to \mathcal{D}_{10} we arrive at

$$(5.43) \quad \mathcal{D}_{20} = \mathcal{D}_{20}^* + \mathcal{D}_{20}^+,$$

where \mathcal{D}_{20}^* is the main term which agrees with \mathcal{D}_{10}^* given by (5.17) and

$$(5.44) \quad \mathcal{D}_{20}^+ = \hat{f}(0) \frac{M}{r\varphi(r)} \sum_s \sum_{s'} \sum_{v(c)}^* \sum_{v'(c)}^* \sigma_{s, s'}(v'/v) \left(\sum_{\substack{(n,rs)=1 \\ n \equiv v(c)}} \beta_n - \frac{1}{\varphi(c)} \sum_{(n,rs)=1} \beta_n \right) \left(\sum_{\substack{(n',rs')=1 \\ n' \equiv v'(c)}} \beta_{n'} - \frac{1}{\varphi(c)} \sum_{(n',rs')=1} \beta_{n'} \right).$$

Following the arguments applied to \mathcal{D}_{10}^+ we conclude the clear estimate

$$|\mathcal{D}_{20}^+| \leq |\hat{f}(0)| \frac{M}{r\varphi(r)} \sum_{c,k,l < 2S} \frac{\rho^2(c)\rho(kl)}{ckl} \sum_{\substack{\nu(c) \\ n \equiv \nu(c)}}^* \left(\sum_{\substack{(n,crk)=1 \\ n \equiv \nu(c)}} \beta_n - \frac{1}{\varphi(c)} \sum_{(n,crk)=1} \beta_n \right)^2.$$

Hence,

$$(5.45) \quad \mathcal{D}_{20}^+ \ll \frac{M}{r\varphi(r)} \left(\sum_{l < 2S} \frac{\tau(l)^j}{l} \right) \sum_{k < 2S} \frac{\tau(k)^j}{k} \mathcal{E}_{kr}(1),$$

which yields the second part on the right side of (5.4).

5.6. Estimation of \mathcal{D}_{21} . Recall that \mathcal{D}_{21} is derived from (5.41) after applying Poisson's formula and dropping the contribution of the frequency $h = 0$. Therefore

$$(5.46) \quad \mathcal{D}_{21} = \frac{M}{r\varphi(r)} \sum_{(nn',r)=1} \beta_n \beta_{n'} \sum_s \sum_{s'} \sum_{\omega([s,s'])} \gamma_s(\omega n) \gamma_{s'}(\omega n') / [s, s'] \\ \sum_{h \neq 0} \hat{f} \left(\frac{hM}{r[s, s']} \right) e \left(-ah \frac{\overline{n[s, s']}}{r} - \frac{\omega h \bar{r}}{[s, s']} \right).$$

Note that this is the same formula as that for \mathcal{D}_{11} , except that here the condition $n' \equiv n \pmod{r}$ is gone, but its absence is paid back by the factor $1/\varphi(r)$. Now, the corresponding set of shifts \mathcal{U} is the full segment $|u| < 2N$ without the condition $u \equiv 0 \pmod{r}$, which does not matter for the arguments. Having said that, we get the same estimates for \mathcal{D}_{21} as those for \mathcal{D}_{11} . That is, (5.34) and (5.40) are satisfied by \mathcal{D}_{21} .

5.7. Evaluation of \mathcal{D}_3 . Recall that \mathcal{D}_3 is derived by squaring the main term in the dispersion (5.7):

$$(5.47) \quad \mathcal{D}_3 = \frac{1}{\varphi(r)^2} \sum_s \sum_{s'} \sum_{\substack{(n,rs)=1 \\ (n',rs')=1}} \beta_n \beta_{n'} \sum_{(m,r)=1} f \left(\frac{m}{M} \right) \gamma_s(mn) \gamma_{s'}(mn').$$

Clearly \mathcal{D}_3 is just the average of $\mathcal{D}_2 = \mathcal{D}_2(a)$ given by (5.41) over the classes $a \pmod{r}$, $(a, r) = 1$. Since the results for $\mathcal{D}_2 = \mathcal{D}_2(a)$ do not depend on a we get the same main term

$$(5.48) \quad \mathcal{D}_{30}^* = \mathcal{D}_{20}^* = \mathcal{D}_{10}^*$$

given by (5.17) and the same existing bounds for the rest of \mathcal{D}_3 . Actually, one can get much stronger estimates, but we do not need them.

Adding and subtracting all the parts of the dispersion, we complete the proof of Proposition 5.1 and its consequences Proposition 4.5 and Corollary 4.6.

6. Cropped Divisors in Residue Classes

6.1. Statement of results. It has been established in [10] that the divisor function $\tau_3(k)$ is distributed uniformly over the residue classes $a \pmod{q}$, $(a, q) = 1$ to modulus q which is fixed, yet relatively large. One of the results of [10] yields

$$\sum_{\substack{k \leq x \\ k \equiv a \pmod{q}}} \tau_3(k) - \frac{1}{\varphi(q)} \sum_{\substack{k \leq x \\ (k, q) = 1}} \tau_3(k) \ll q^{-1} x^{1-\vartheta}$$

with $\vartheta > 0$ a small constant, provided $q \leq x^{\frac{1}{2} + \frac{1}{231}}$. In practice one needs similar results for cropped divisors

$$\tau_F(k) = \sum_{lmn=k} F\left(\frac{l}{L}, \frac{m}{M}, \frac{n}{N}\right)$$

where $F(x, y, z)$ is a nice continuous function, compactly supported in $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$ and $L, M, N \geq 1$ are at our disposal. Here the three divisors run over segments $l \asymp L$, $m \asymp M$, $n \asymp N$. This extra localization is welcome for applications. Actually, we also treated in [10] cropped divisor functions of various shapes in a subsidiary capacity.

In this section we follow closely the arguments of [10] and enhance them by a device used by Y. Zhang [18] in order to cover a bit larger, yet vital range of L, M, N . For sheer curiosity, and perhaps for clarity, we let our arguments handle technical barriers somewhat differently than in [10] and [18]. First, we fix a cropping function f which is of \mathcal{C}^4 class and compactly supported in \mathbb{R}^+ . Let $q > 1$ be squarefree and $(a, q) = 1$. We shall evaluate the sum

$$(6.1) \quad S(L, M, N) = \sum_{lmn \equiv a \pmod{q}} f\left(\frac{l}{L}\right) f\left(\frac{m}{M}\right) f\left(\frac{n}{N}\right)$$

which is expected to be well-approximated by

$$(6.2) \quad S_0(L, M, N) = \frac{1}{\varphi(q)} \sum_{(lmn, q) = 1} f\left(\frac{l}{L}\right) f\left(\frac{m}{M}\right) f\left(\frac{n}{N}\right).$$

Our goal is to estimate the error term

$$(6.3) \quad E(L, M, N) = S(L, M, N) - S_0(L, M, N).$$

As in [18], we assume that the modulus q has a relatively small factor which will help us cover slightly wider ranges of L, M, N .

Proposition 6.1. *Let $L \geq M \geq N \geq 1$ with $LMN = x$ and $q > 1$ be squarefree, $(a, q) = 1$. Suppose $s \mid q$ and*

$$(6.4) \quad sN < \min(q, x/q).$$

Then

$$(6.5) \quad E(L, M, N) \ll x^\varepsilon \left(\frac{x}{N}\right)^{\frac{1}{2}} \left(\frac{q}{s}\right)^{\frac{1}{4}} + x^\varepsilon (xsN)^{\frac{1}{3}}$$

for any $\varepsilon > 0$, the implied constant depending only on ε .

If $s = 1$, then (6.5) is covered by Theorem 4 of [10]. We wish the bound (6.5) to be meaningful (non-trivial); specifically, we want

$$(6.6) \quad E(L, M, N) \ll q^{-1}x^{1-\varepsilon}$$

for some $\varepsilon > 0$, regardless of how small it is. By (6.5) we get (6.6) if

$$(6.7) \quad s^{\frac{1}{2}}q^{\frac{5}{2}}x^{\varepsilon-1} < sN < x^{-\varepsilon} \min(q, x^2q^{-3}).$$

Corollary 6.2. *Let q be squarefree and $(a, q) = 1$. Suppose q has divisors of any size up to a factor $z \geq 2$. Then (6.6) holds for $L \geq M \geq N \geq 1$ with $LMN = x$ if*

$$(6.8) \quad q \leq \left(\frac{N}{z}\right)^{\frac{1}{8}} x^{\frac{1}{2}-\varepsilon}.$$

Proof. We can assume that $z \leq N \leq qx^{-\varepsilon}$, otherwise the result is classical (easy to prove). Since $N \leq x^{\frac{1}{3}}$ and $q \leq x^{\frac{5}{8}}$, we have $N < x^{-\varepsilon} \min(q, x^2q^{-3})$. Therefore, there exists $s \mid q$ with

$$\frac{\min}{x^\varepsilon zN} \leq s < \frac{\min}{x^\varepsilon N}.$$

For this choice of s the conditions (6.7) are satisfied if

$$q^{\frac{5}{2}}x^{\varepsilon-1} < N \left(\frac{\min}{x^\varepsilon zN}\right)^{\frac{1}{2}},$$

that is, if $x^{3\varepsilon}zq^4 < Nx^2$ and $x^\varepsilon zq^8 < Nx^4$. However, the first condition follows from the second one. This completes the proof. □

Since Corollary 6.2 holds for every class $a \pmod{q}$ with $(a, q) = 1$, the result can be automatically generalized as follows:

Corollary 6.3. *Let q be squarefree and $(ad, q) = 1$. Suppose q has divisors of any size up to a factor $z \geq 2$. Then we have*

$$(6.9) \quad \sum_{d|mn \equiv a(q)} \sum_{l|L} \sum_{m|M} \sum_{n|N} f\left(\frac{l}{L}\right) f\left(\frac{m}{M}\right) f\left(\frac{n}{N}\right) = S_0(L, M, N) + O(q^{-1}d^{-1}x^{1-\varepsilon})$$

for any $L \geq M \geq N \geq 1$ with $dLMN = x$, if

$$(6.10) \quad q \leq \left(\frac{N}{z}\right)^{\frac{1}{8}} \left(\frac{x}{d}\right)^{\frac{1}{2}} x^{-\varepsilon}.$$

6.2. Extracting the main term. We begin the proof of Proposition 6.1 by applying Poisson’s formula in the largest variable $l \equiv a\overline{mn} \pmod{q}$, getting

$$(6.11) \quad S(L, M, N) = \frac{L}{q} \sum_h \hat{f}\left(\frac{hL}{q}\right) \sum_{mn} f\left(\frac{m}{M}\right) f\left(\frac{n}{N}\right) e\left(ah\frac{\overline{mn}}{q}\right).$$

Here and hereafter, we use the convention that if the multiplicative inverse of d modulo q appears, then it is assumed automatically, without writing, that d is restricted to $(d, q) = 1$. Of course, one must recall this restriction when things go out of sight.

The main term emerges from the zero frequency, which yields

$$(6.12) \quad \hat{f}(0) \frac{L}{q} \sum_{(mn, q)=1} f\left(\frac{m}{M}\right) f\left(\frac{n}{N}\right).$$

This is not exactly $S_0(L, M, N)$, but close. Indeed, we have

$$\begin{aligned} \sum_{(l, q)=1} f\left(\frac{l}{L}\right) &= \sum_{\delta|q} \mu(\delta) \sum_l f\left(\frac{\delta l}{L}\right) = \sum_{\delta|q} \mu(\delta) \left(\hat{f}(0) \frac{L}{\delta} + O(1)\right) \\ &= \hat{f}(0) L \frac{\varphi(q)}{q} + O(\tau(q)). \end{aligned}$$

Hence (6.12) becomes $S_0(L, M, N) + O(MN\tau(q)/\varphi(q))$, and we are left with

$$(6.13) \quad S(L, M, N) = S_0(L, M, N) + S^*(L, M, N) + O(MN\tau(q)/\varphi(q)),$$

where $S^*(L, M, N)$ denotes the part of (6.11) with $h \neq 0$. This will be estimated in several steps.

6.3. Weyl’s shift. First we are going to replace $f(m/M)$ by $f((m + bh)/M)$, where b is a positive integer at our disposal; it will be chosen later from a finite set of integers of suitable type. We write

$$(6.14) \quad S^*(L, M, N) = S_b(L, M, N) - \partial S_b(L, M, N),$$

where

$$(6.15) \quad S_b(L, M, N) = \frac{L}{q} \sum_n f\left(\frac{n}{N}\right) \sum_m \sum_{h \neq 0} \hat{f}\left(\frac{hL}{q}\right) f\left(\frac{m+bh}{M}\right) e\left(ah \frac{\overline{mn}}{q}\right)$$

and $\partial S_b(L, M, N)$ is similar, but with the inner sum over $h \neq 0$ replaced by

$$(6.16) \quad \sum_h \hat{f}\left(\frac{hL}{q}\right) \left(f\left(\frac{m+bh}{M}\right) - f\left(\frac{m}{M}\right)\right) e\left(ah \frac{\overline{mn}}{q}\right).$$

Applying Poisson’s formula to (6.16) (a reversed application) we find that the sum (6.16) is equal to

$$(6.17) \quad \frac{q}{L} \sum_{l \equiv a\overline{mn} \pmod{q}} g\left(\frac{l}{L}\right),$$

where $g(u)$ is such that

$$\hat{g}\left(\frac{hL}{q}\right) = \hat{f}\left(\frac{hL}{q}\right) \left(f\left(\frac{m+bh}{M}\right) - f\left(\frac{m}{M}\right)\right)$$

for every real h , that is;

$$(6.18) \quad \hat{g}(v) = \hat{f}(v) \left(f\left(\frac{m}{M} + v \frac{bq}{LM}\right) - f\left(\frac{m}{M}\right)\right).$$

By partial integration and the mean-value theorem we derive the following estimates:

$$|\hat{g}(v)| + |\hat{g}(v)''| \ll \frac{bq}{LM} (1+v^2)^{-2}, \quad g(u) \ll \frac{bq}{LM} (1+u^2)^{-1}.$$

Hence, (6.17) yields

$$\partial S_b(L, M, N) \ll \frac{bq}{LM} \sum_{\substack{lmn \equiv a \pmod{q} \\ m < M, n < N}} (1+l/L)^{-2} \ll x^\varepsilon bN,$$

by using $\tau_3(k) \ll k^\varepsilon$. Inserting this estimate into (6.14) we get

$$(6.19) \quad S^*(L, M, N) = S_b(L, M, N) + O(x^\varepsilon bN)$$

for any $\varepsilon > 0$, the implied constant depending on ε . Note that $S^*(L, M, N)$ does not depend on b , whereas the right side of (6.19) has received the extra parameter b for free.

In (6.15) we translate m to $m - bh$ and we write h/q in the lowest terms, getting

$$S_b(L, M, N) = \frac{L}{q} \sum_{q_0 q_1 = q} \sum_{\substack{h \neq 0 \\ (h, q_1) = 1}} \hat{f}\left(\frac{hL}{q_1}\right) \sum_{(mn, q_0) = 1} f\left(\frac{m}{M}\right) f\left(\frac{n}{N}\right) e\left(\frac{ah}{q_1} \overline{(m+bhq_0)n}\right).$$

Here the co-primality condition $(m, q_0) = 1$ is harmless, but $(n, q_0) = 1$ should be relaxed before applying Cauchy’s inequality. To this end we utilize the Möbius formula, giving

$$S_b(L, M, N) = \frac{L}{q} \sum_{q_0 q_1 = q} \sum_{\eta | q_0} \mu(\eta) \sum_{\substack{h \neq 0 \\ (h, q_1) = 1}} \hat{f}\left(\frac{hL}{q_1}\right) \sum_{(m, q_0) = 1} f\left(\frac{m}{M}\right) \sum_n f\left(\frac{n\eta}{N}\right) e\left(\frac{a_0 \overline{q_0 \eta}}{q_1} \overline{(m h q_0 + b)n}\right).$$

6.4. Compacting variables. Our next step (inspired by Burgess’ idea) is to build one larger variable out of m and h . Putting $a_0 \equiv a_0 \overline{q_0 \eta} \pmod{q_1}$ and for any $x \pmod{q_1}$

$$\nu(x) = \sum_{\substack{h \neq 0 \\ (h, q_1) = 1}} \hat{f}\left(\frac{hL}{q_1}\right) \sum_{\substack{(m, q_0) = 1 \\ m \equiv x h q_0 \pmod{q_1}}} f\left(\frac{m}{M}\right),$$

so we can write

$$S_b(L, M, N) = \frac{L}{q} \sum_{q_0 q_1 = q} \sum_{\eta | q_0} \mu(\eta) \sum_{x \pmod{q_1}} \nu(x) \sum_n f\left(\frac{n\eta}{N}\right) e\left(\frac{a_0 \overline{q_0 \eta}}{q_1} \overline{(x + b)n}\right).$$

Note that

$$(6.20) \quad \sum_{x \pmod{q_1}} \nu(x)^2 \ll \frac{M}{L} q_1^{1+\varepsilon}$$

because the number of solutions of $h'm \equiv hm' \pmod{q_1}$ in $1 \leq h, h' \leq H$ and $1 \leq m, m' \leq M$, for every $H, M \geq 1$ is bounded by $HM(1 + HM/q_1)q_1^\varepsilon$. Now we choose a finite set \mathcal{B} of positive integers and sum $S_b(L, M, N)$ over $b \in \mathcal{B}$. Then we apply Cauchy’s inequality and (6.20) getting

$$(6.21) \quad \sum_{b \in \mathcal{B}} S_b(L, M, N) \ll q^\varepsilon \sum_{q_0 q_1 = q} \left(\frac{LM}{q_0 q}\right)^{\frac{1}{2}} \sum_{\eta | q_0} T(\mathcal{B}, N/\eta)^{\frac{1}{2}},$$

where

$$T(\mathcal{B}, N) = \sum_{x \pmod{q_1}} \left| \sum_n f\left(\frac{n}{N}\right) \sum_{b \in \mathcal{B}} e\left(\frac{a_0 \overline{q_0 \eta}}{q_1} \overline{(x + b)n}\right) \right|^2.$$

Note that we do not display (for brevity) all the floating parameters upon which our objects depend. For instance, $T(\mathcal{B}, N)$ depends on the modulus q_1 and also on q_0, η via $a_0 \equiv a_0 \overline{q_0 \eta} \pmod{q_1}$. Squaring out we find

$$(6.22) \quad T(\mathcal{B}, N) = T^=(\mathcal{B}, N) + T^\neq(\mathcal{B}, N),$$

where

$$(6.23) \quad T^{\neq}(\mathcal{B}, N) = \sum_{b \neq b'} \sum_n \sum_{n'} f\left(\frac{n}{N}\right) f\left(\frac{n'}{N}\right) \sum_{x(q_1)} e\left(\frac{a_0}{q_1} \left(\overline{(x+b-b')n} - \overline{xn'}\right)\right)$$

and $T^=(\mathcal{B}, N)$ is the corresponding sum with $b = b'$. In this case the inner sum reduces to the Ramanujan sum (remember $(nn', q_1) = 1$ by our convention)

$$\sum_{x(q_1)}^* e\left((n' - n)\frac{x}{q}\right) \ll (n' - n, q_1).$$

Summing over n, n' , we obtain

$$(6.24) \quad T^=(\mathcal{B}, N) \ll \tau(q)N(q_1 + N)|\mathcal{B}|.$$

The other sum $T^{\neq}(\mathcal{B}, N)$ is much more difficult to estimate; we apply the Riemann hypothesis for varieties in the concealed form of bounds for 3-dimensional exponential sums.

6.5. Correlation of Kloosterman sums. In (6.23) we execute the summation over n, n' by Poisson's formula, obtaining

$$(6.25) \quad T^{\neq}(\mathcal{B}, N) = \left(\frac{N}{q_1}\right)^2 \sum_{b \neq b'} \sum_v \sum_w \hat{f}\left(\frac{vN}{q_1}\right) \hat{f}\left(\frac{wN}{q_1}\right) K(b - b', a_0v, a_0w; q_1),$$

where $K(u, v, w; q_1)$ denotes the complete sum modulo q_1 in three variables (a kind of correlation of Kloosterman sums)

$$(6.26) \quad K(u, v, w; k) = \sum_{x, y, z \pmod{k}} e_k\left(\overline{(x+u)y} - \overline{xz} + vz + wy\right).$$

Proposition 6.4. *Let $k \geq 1$ be squarefree and u, v, w be integers. We have*

$$(6.27) \quad K(u, v, w; k) \ll (u, k)^{\frac{1}{2}}(vw, k)^{-\frac{1}{2}}k^{\frac{3}{2}+\varepsilon},$$

where ε is any positive number, and the implied constant depends only on ε . Moreover, for $u \equiv 0 \pmod{k}$ we have

$$(6.28) \quad |K(0, v, w; k)| \leq k(v + w, k)/(vw, k).$$

Proof. By multiplicativity we can assume that $k = p$ is prime. We shall compute $K(u, v, w; p)$ in all cases, except for $p \nmid uvw$. If $p \mid u$ we get

$$\begin{aligned}
 K(0, v, w; p) &= \sum_x \sum_y \sum_z e_p(\overline{xy} - \overline{xz} + vz + wy) \\
 &= R(v + w, p)p - R(v, p)R(w, p) = \begin{cases} R(v + w, p), & \text{if } p \mid vw, \\ R(v + w, p)p - 1, & \text{if } p \nmid vw, \end{cases}
 \end{aligned}$$

where $R(a, p)$ is the Ramanujan sum; $|R(a, p)| < (a, p)$. Hence

$$|K(0, v, w; p)| < p(v + w, p)/(vw, p)$$

which proves (6.28). Now let $p \nmid u, p \mid v$. We get

$$\begin{aligned}
 K(u, 0, w; p) &= \sum_x \sum_y \sum_z e_p(\overline{(x + u)y} - \overline{xz} + wy) \\
 &= - \sum_y \sum_{x \neq 0} e_p(\overline{(x + u)y} + wy) = K(\overline{u}, w; p) + R(w, p),
 \end{aligned}$$

where $K(\overline{u}, w; p)$ denotes the classical Kloosterman sum. Similarly we evaluate $K(u, v, 0; p) = K(\overline{u}, v; p) + R(v, p)$. In both cases, using Weil's bound we obtain

$$|K(u, v, w; p)| \leq 3(v + w, p)^{\frac{1}{2}} p^{\frac{1}{2}}, \quad \text{if } p \nmid u, p \mid vw.$$

This bound easily satisfies (6.27). Finally, if $p \nmid uvw$ we cannot compute $K(u, v, w; p)$, but we have the bound (see the Appendix in [10] by Birch and Bombieri)

$$K(u, v, w; p) \ll p^{\frac{3}{2}}.$$

Clearly, this bound satisfies (6.27). Combining these results one completes the proof of Proposition 6.4. □

Remark. Sums of type (6.26) have been recently considered among much more general sums in a conceptual framework by [8].

6.6. Completing the proof of Proposition 6.1. Now, we proceed to the estimation of $T^\neq(\mathcal{B}, N)$, starting from the formula (6.25), by applying Proposition 6.4. This constitutes the core of the proof of Proposition 6.1. At this point, we are going to exploit the divisor s of q and, as Zhang [18], we take the set \mathcal{B} of positive integers which are divisible by s ;

$$\mathcal{B} = \{b = cs; 1 \leq c \leq C\},$$

where $C = |\mathcal{B}|$ will be chosen later to optimize the results. The case $s = 1$ (our original setup in [10]) does lead to pretty strong results. However, by allowing s to be a nontrivial divisor of q , Zhang succeeded to produce improvements, crucial for this application, which are due to the estimates for the Ramanujan sums being better than those for the correlation Kloosterman sums (compare (6.28) and (6.27)).

Put $s_1 = s/(s, q_0)$ and $r = q/[s, q_0] = q(s, q_0)/sq_0$. We have

$$q_1 = rs_1, \quad s_1 \mid s, \quad (r, s) = 1.$$

The complete sum of modulus q_1 in (6.25) factors into the complete sum of modulus r and the complete sum of modulus s_1 , but with the exponentials (additive characters) twisted by the factors $\overline{s_1} \pmod{r}$ and $\overline{r_1} \pmod{s_1}$, respectively. By (6.27) we get

$$K(*, *, *, r) \ll (c - c', r)^{\frac{1}{2}}(vw, r)^{-\frac{1}{2}}r^{\frac{3}{2}+\varepsilon},$$

and by (6.28) we get

$$K(*, *, *, s_1) \ll s_1(v + w, s_1)/(vw, s_1).$$

Inserting these estimates into (6.25) and summing over $1 \leq c \neq c' \leq C$ we obtain

$$(6.29) \quad T^{\neq}(\mathcal{B}, N) \ll q^\varepsilon \frac{C^2 N^2}{q_1} \sum_v \sum_w \left| \hat{f}\left(\frac{vN}{q_1}\right) \hat{f}\left(\frac{wN}{q_1}\right) \right| \frac{(v + w, s_1)}{(vw, s_1)} \left(\frac{r}{(vw, r)}\right)^{\frac{1}{2}}.$$

The terms $v = w = 0$ contribute $|\hat{f}(0)|^2$. The terms $v = 0, w \neq 0$, and the terms $v \neq 0, w = 0$ contribute

$$2|\hat{f}(0)| \sum_{w \neq 0} \left| \hat{f}\left(\frac{wN}{q_1}\right) \right| \frac{(w, s_1)}{s_1} \ll \frac{q_1 \tau(s_1)}{Ns_1}.$$

The terms $v = -w \neq 0$ contribute at most

$$2r^{\frac{1}{2}}s_1 \sum_{w \neq 0} \left| f\left(\frac{wN}{q_1}\right) \right|^2 \ll r^{\frac{1}{2}}s_1q_1/N.$$

The terms $vw(v + w) \neq 0$ contribute at most

$$r^{\frac{1}{2}} \sum_{vw(v+w) \neq 0} \sum \left| \hat{f}\left(\frac{vN}{q_1}\right) \hat{f}\left(\frac{wN}{q_1}\right) \right| (v + w, s_1).$$

Here, we apply the following estimates for the Fourier transforms

$$|\hat{f}(v)\hat{f}(w)| \ll (1 + |v|)^{-2}(1 + |w|)^{-4} \leq (1 + |v + w|)^{-2}(1 + |w|)^{-2}.$$

Hence, the contribution of terms $vw(v+w) \neq 0$ is bounded by

$$r^{\frac{1}{2}} \sum_{w>0} \left(1 + \frac{wN}{q_1}\right)^{-2} \sum_{t>0} \left(1 + \frac{tN}{q_1}\right)^{-2} (t, s_1) \ll \tau(s_1) r^{\frac{1}{2}} q_1^2 N^{-2}.$$

Inserting the above estimates into (6.29) we get

$$T^{\neq}(\mathcal{B}, N) \ll q^\varepsilon \frac{C^2 N^2}{q_1} \left\{ 1 + \frac{q_1}{Ns_1} + \frac{r^{\frac{1}{2}} s_1 q_1}{N} + \frac{r^{\frac{1}{2}} q_1^2}{N^2} \right\} \ll C^2 s^{-\frac{1}{2}} q^{\frac{3}{2} + \varepsilon},$$

because $sN < q$ by (6.4). Adding (6.24), we obtain

$$T(\mathcal{B}, N) \ll q^\varepsilon (CNq + C^2 s^{-\frac{1}{2}} q^{\frac{3}{2}}).$$

Next, inserting this into (6.21) we have

$$\frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} S_b(L, M, N) \ll q^\varepsilon \left(\frac{LMN}{C} \right)^{\frac{1}{2}} + q^\varepsilon \left(\frac{q}{s} \right)^{\frac{1}{4}} (LM)^{\frac{1}{2}}.$$

Adding this to (6.19) we find

$$S^*(L, M, N) \ll x^\varepsilon \left(\left(\frac{x}{C} \right)^{\frac{1}{2}} + \left(\frac{q}{s} \right)^{\frac{1}{4}} \left(\frac{x}{N} \right)^{\frac{1}{2}} + CsN \right).$$

Here C is any positive integer, but, of course, the result holds for any real number $C \geq 1$. We take $C = (x/S^2 N^2)^{1/3} \geq 1$ by (6.4), getting

$$S^*(L, M, N) \ll x^\varepsilon \left(\frac{q}{s} \right)^{\frac{1}{4}} \left(\frac{x}{N} \right)^{\frac{1}{2}} + x^\varepsilon (xsN)^{\frac{1}{3}}.$$

Hence, by (6.13) we arrive at (6.5), which completes the proof of Proposition 6.1.

REFERENCES

- [1] E. Bombieri and H. Davenport, *Small differences between prime numbers*, Proc. Roy. Soc. Ser. A **293** (1966), 1–18.
- [2] E. Bombieri, J. B. Friedlander, and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. **156** (1986), 203–251.
- [3] ———, *Primes in arithmetic progressions to large moduli. II*, Math. Ann. **277** (1987), 361–393.
- [4] ———, *Primes in arithmetic progressions to large moduli. III*, J. Amer. Math. Soc. **2** (1989), 215–224.
- [5] P. Erdős, *The difference of consecutive primes*, Duke Math. J. **6** (1940), 438–441.
- [6] E. Fouvry, *Autour du théorème de Bombieri-Vinogradov*, Acta Math. **152** (1984), 219–244.
- [7] E. Fouvry and H. Iwaniec, *Primes in arithmetic progressions*, Acta Arith. **42** (1983), 197–218.
- [8] E. Fouvry, E. Kowalski, and P. Michel, *Algebraic twists of modular forms and Hecke orbits* (2012), arXiv:1207.0617.
- [9] ———, *On the exponent of distribution of the ternary divisor function* (2013), arXiv:1304.3199.

- [10] J. B. Friedlander and H. Iwaniec, *Incomplete Kloosterman sums and a divisor problem*, Ann. of Math. (2) **121** (1985), 319–350.
- [11] ———, *Opera de cribro*, American Mathematical Society Colloquium Publications, vol. 57, American Mathematical Society, Providence, RI, 2010.
- [12] D. A. Goldston and C. Y. Yıldırım, *Higher correlations of divisor sums related to primes. III. Small gaps between primes*, Proc. Lond. Math. Soc. (3) **95** (2007), 653–686.
- [13] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, *Primes in tuples. I*, Ann. of Math. (2) **170** (2009), 819–862.
- [14] M. N. Huxley, *On the differences of primes in arithmetical progressions*, Acta Arith. **15** (1968/1969), 367–392.
- [15] Ju. V. Linnik, *The dispersion method in binary additive problems*, American Mathematical Society, Providence, R.I., 1963.
- [16] H. Maier, *Small differences between prime numbers*, Michigan Math. J. **35** (1988), 323–344.
- [17] Y. Motohashi and J. Pintz, *A smoothed GPY sieve*, Bull. Lond. Math. Soc. **40** (2008), 298–310.
- [18] Y. Zhang, *Bounded gaps between primes*, Ann. Math **179** (2014), 1121–1174.