

NEW LARGE VALUE ESTIMATES FOR DIRICHLET POLYNOMIALS

LARRY GUTH AND JAMES MAYNARD

ABSTRACT. We prove new bounds for how often Dirichlet polynomials can take large values. This gives improved estimates for a Dirichlet polynomial of length N taking values of size close to $N^{3/4}$, which is the critical situation for several estimates in analytic number theory connected to prime numbers and the Riemann zeta function. As a consequence, we deduce a zero density estimate $N(\sigma, T) \leq T^{30(1-\sigma)/13+o(1)}$ and asymptotics for primes in short intervals of length $x^{17/30+o(1)}$.

1. INTRODUCTION

In this paper we prove new bounds for the frequency of large values of Dirichlet polynomials. This gives improved estimates for a Dirichlet polynomial of length N taking values of size close to $N^{3/4}$. Our main result is the following.

Theorem 1.1 (Large values estimate). *Suppose (b_n) is a sequence of complex numbers with $|b_n| \leq 1$, and $(t_r)_{r \leq R}$ is a sequence of 1-separated points in $[0, T]$ such that*

$$\left| \sum_{n=N}^{2N} b_n n^{it_r} \right| \geq V$$

for all $r \leq R$. Then we have

$$R \leq T^{o(1)} \left(N^2 V^{-2} + N^{18/5} V^{-4} + T N^{12/5} V^{-4} \right).$$

The bound of Theorem 1.1 can be compared with the bound

$$(1.1) \quad R \leq T^{o(1)} \left(N^2 V^{-2} + T \min(NV^{-2}, N^4 V^{-6}) \right)$$

coming from combining the classical Mean Value Theorem for Dirichlet polynomials and the Montgomery-Halász-Huxley large values estimate. Together these previous results give bounds at least as strong as Theorem 1.1 when V is smaller than $N^{7/10}$ or bigger than $N^{8/10}$, but Theorem 1.1 gives a stronger bound when $N^{7/10+\epsilon} < V < N^{8/10-\epsilon}$ and $N \leq T^{5/6-\epsilon}$. There are various improvements of the large values estimate which will supersede ours when V is a bit larger than $N^{3/4}$ (see [Iv, Chapter 11], for example), but Theorem 1.1 represents the first substantive improvement on the bound (1.1) when $N^{7/10+\epsilon} \leq V \leq N^{3/4}$ (when the Mean Value Theorem gave

the previous best bounds) or on bounds when V is close to $N^{3/4}$ (where the previous best bounds gave only very small improvements on (1.1)).

The key interest in this result is that for many applications in analytic number theory, it is the case when $V \approx N^{3/4}$ which is the critical limiting scenario. In this situation the Mean Value Theorem and Large Values Estimate both give a bound of roughly $N^{1/2} + TN^{-1/2}$ whereas Theorem 1.1 gives roughly $N^{3/5} + TN^{-3/5}$. Therefore we obtain an improvement for N smaller than $T^{10/11}$, and correspondingly we expect Theorem 1.1 to lead to a quantitative improvement to any result where this covers the limiting situation.

One well-studied situation where the limiting case is improved by Theorem 1.1 is zero-density estimates for the Riemann Zeta function $\zeta(s)$. Let $N(\sigma, T)$ be the number of zeroes of $\zeta(s)$ in the rectangle $\Re(s) \geq \sigma$ and $|\Im(s)| \leq T$. After early work by Carlson [Ca], Ingham [In] proved the bound

$$(1.2) \quad N(\sigma, T) \leq T^{\frac{3(1-\sigma)}{2-\sigma} + o(1)},$$

ultimately relying on the Mean Value Theorem for Dirichlet polynomials. Huxley [Hu], building on work of Montgomery [M3] and Halász [Ha] and ultimately relying on the Montgomery-Halász large values estimate, proved the bound

$$(1.3) \quad N(\sigma, T) \leq T^{\frac{3(1-\sigma)}{3\sigma-1} + o(1)}.$$

This improves on (1.2) for $\sigma > 3/4$ (corresponding to when the term N^4V^{-6} is smaller than NV^{-2} for $V = N^\sigma$ in $\min(NV^{-2}, N^4V^{-6})$ in (1.1)).

When $\sigma \approx 3/4$ the bounds (1.2) and (1.3) coincide, and the critical situation turns out to be related to values of size $V = N^{3/4}$ of Dirichlet polynomials of length $N = T^{4/5}$, where both estimates give $R \leq T^{3/5+o(1)}$. In this situation Theorem 1.1 gives an improved estimate of $R \leq T^{13/25+o(1)}$. Incorporating Theorem 1.1 into the zero density machinery gives the following result.

Theorem 1.2 (Zero density estimate). *Let $N(\sigma, T)$ denote the number of zeros ρ of $\zeta(s)$ with $\Re(\rho) \geq \sigma$ and $|\Im(\rho)| \leq T$. Then we have*

$$N(\sigma, T) \leq T^{15(1-\sigma)/(3+5\sigma)+o(1)}.$$

Combining this with Ingham's estimate when $\sigma \leq 7/10$, we obtain

$$(1.4) \quad N(\sigma, T) \leq T^{30(1-\sigma)/13+o(1)}.$$

The exponent $30/13$ improves on the previous exponent of $12/5$ due to Huxley [Hu]. This has the following corollaries for the distribution of primes in short intervals.

Corollary 1.3 (Count of primes in short intervals). *Let $y \in [x^{17/30+\epsilon}, x^{0.99}]$. Then we have*

$$\pi(x+y) - \pi(x) = \frac{y}{\log x} + O_\epsilon\left(y \exp(-\sqrt[4]{\log x})\right).$$

The exponent $\frac{17}{30}$ improves on the previous exponent $\frac{7}{12}$ due to Huxley [Hu].

Corollary 1.4 (Count of primes in ‘almost-all’ short intervals). *Let $y \in [X^{2/15+\epsilon}, X^{0.99}]$. Then for all but $O(X \exp(-\sqrt[4]{\log x}))$ choices of $x \in [X, 2X] \cap \mathbb{N}$ we have*

$$\pi(x+y) - \pi(x) = \frac{y}{\log x} + O_\epsilon\left(y \exp(-\sqrt[4]{\log x})\right).$$

The exponent $\frac{2}{15}$ improves on the previous exponent $\frac{1}{6}$ due to Huxley [Hu].

We expect there to be various further applications of Theorem 1.1 (and the underlying ideas) to improving quantitative estimates related to the primes and similar objects.

1.1. Background. Given an integer $N \in \mathbb{N}$ and a sequence $(b_n)_{N < n \leq 2N}$, a Dirichlet polynomial is a trigonometric sum of the form

$$(1.5) \quad D(t) = \sum_{N < n \leq 2N} b_n e^{it \log n}.$$

We say such a polynomial has length N , and we study the question how often a Dirichlet polynomial can be large in the interval $[0, T]$. The precise question can be formulated as follows.

Main Question. *Suppose that $D(t)$ is a Dirichlet polynomial of the form (1.5) with $|b_n| \leq 1$ for all n . Suppose that $W \subset [0, T]$ is a 1-separated set of points t where $|D(t)| > N^\sigma$. What is the largest possible cardinality of W , in terms of N , T , and σ ?*

There are many examples of Dirichlet polynomials where $|W| \gg N^{2-2\sigma}$ when $\sigma \in (1/2, 1)$ and $N \leq T$.¹

Montgomery’s large value conjecture [M2, page 142, Conjecture 2] predicts that under some natural conditions this lower bound is essentially tight. Bourgain [B5]

¹For example, we can form such a Dirichlet polynomial as follows. If w is a smooth 1-bounded function supported on $[1, 2]$, $\epsilon > 0$ is a sufficiently small fixed constant and $M^\epsilon < H < M/T^\epsilon$, then it follows from Poisson summation and repeated integration by parts that

$$\left| \sum_h w\left(\frac{h}{H}\right) (M+h)^{-it} \right| = \begin{cases} \gg H, & \text{if } |t| \leq M/(10H), \\ \ll \epsilon^2 M^2 / (t^2 H), & \text{if } M/(\epsilon H) \ll |t| \leq M. \end{cases}$$

Therefore (taking $H = N^\sigma$ and H -separated choices of $M \in [N, 2N]$) if $t_1, \dots, t_J \in [T, T+N]$ are $J = \epsilon N^{1-\sigma} + O(1)$ points which are $\epsilon^{-1} N^{1-\sigma}$ -separated, the Dirichlet polynomial

$$D(t) := \sum_{j=1}^J \sum_h w\left(\frac{h}{N^\sigma}\right) \left(N + (j-1) \left\lfloor \frac{N}{J} \right\rfloor + h\right)^{i(t_j - t)}$$

will take a value of size $\gg N^\sigma$ on an interval of length $\gg N^{1-\sigma}$ around each of the $\gg \epsilon N^{1-\sigma}$ points t_j . Thus we can find $\gg_\epsilon N^{2-2\sigma}$ 1-separated points where it takes a value of size $\gg N^\sigma$.

gave a counterexample to an earlier (stronger) formulation given in terms of the ℓ^2 norm of the coefficients b_n .

Conjecture 1.5 (Montgomery's large value conjecture). *Let $\sigma > 1/2$ and $D(t) = \sum_{N < n \leq 2N} b_n e^{it \log n}$ with $|b_n| \leq 1$. Suppose $W \subset [0, T]$ is a 1-separated set such that $|D(t)| > N^\sigma$ for $t \in W$. Then there is a constant $C(\sigma)$ such that*

$$|W| \leq C(\sigma) T^{o(1)} N^{2-2\sigma}.$$

Using a simple orthogonality argument, it has long been known that for any $T \geq N$,

$$(1.6) \quad |W| = O(TN^{1-2\sigma}),$$

which corresponds to the term NV^{-2} in $\min(NV^{-2}, N^4V^{-6})$ in (1.1).

This basic estimate has been improved in several regimes. An integer power of a Dirichlet polynomial is a Dirichlet polynomial, and applying (1.6) to powers of D gives improved bounds when N is fairly small compared with T . For $\sigma \leq 3/4$, the best previous bounds on our main question came from this approach. In particular, if $\sigma \leq 3/4$ and if $N \in [T^{2/3}, T]$, then (1.6) represented the best known bound.

In the late 60s, Montgomery [M], building on ideas of Halász [Ha] and Turán [HT], showed that Conjecture 1.5 is true if σ is sufficiently large. Indeed, (1.1) gives Conjecture 1.5 when $N^\sigma > N^{1/2}T^{1/4}$. The large value method gives very strong information for large σ , but it gives no information if $\sigma \leq 3/4$ as Montgomery explains in [M2, page 141]. For σ closer to 1, there have been several refinements of the underlying ideas (see, for example, [Bo, HB3, Ju]).

If one knows some more structure about the set of large values of a Dirichlet polynomial, then one can hope to have improved bounds. For example, another important result is Heath-Brown's work about the behavior of Dirichlet polynomials on difference sets:

Theorem 1.6 (Heath-Brown, [HB]). *Suppose that \mathcal{T} is a 1-separated set of points in an interval of length T . Let $|a_n| \leq T^{o(1)}$ be a complex sequence. Then*

$$\sum_{t_1, t_2 \in \mathcal{T}} \left| \sum_{n=N}^{2N} a_n n^{i(t_1 - t_2)} \right|^2 \leq T^{o(1)} \left(|\mathcal{T}|^2 N + |\mathcal{T}| N^2 + |\mathcal{T}|^{5/4} T^{1/2} N \right).$$

In the range $N \in [T^{2/3}, T]$, the last term can be ignored and the right-hand side is essentially sharp². Theorem 1.6 gives strong information about our main question if the set W has a lot of additive structure, such as arithmetic progressions.

We will make this precise using the idea of additive energy, which we define as follows. For a finite set W , let

²if $N \geq T^{2/3}$ then $|\mathcal{T}|^{5/4} T^{1/2} N \leq |\mathcal{T}|^{5/4} N^{7/4} = (|\mathcal{T}|^2 N)^{1/4} (|\mathcal{T}| N^2)^{3/4} \leq |\mathcal{T}|^2 N + |\mathcal{T}| N^2$.

$$(1.7) \quad E(W) := \#\{w_1, w_2, w_3, w_4 \in W : |w_1 + w_2 - w_3 - w_4| \leq 1\}.$$

A simple consequence of Theorem 1.6 is the following (we give a proof, as well as refined bounds, in Section 11).

Lemma 1.7. *Let $N \in [T^{2/3}, T]$, $\sigma > 1/2$ and $D(t) = \sum_{N < n \leq 2N} b_n n^{it}$ with $|b_n| \leq 1$. Suppose $W \subset [0, T]$ is a 1-separated set such that $|D(t)| > N^\sigma$ for $t \in W$. Then*

$$E(W) \leq |W|^3 N^{1-2\sigma+o(1)} + |W|^2 N^{2-2\sigma+o(1)}.$$

If $E(W)$ is very large, say $E(W) > |W|^3 T^{-o(1)}$, then $E(W)$ is larger than the first term on the right hand side, so it must be bounded by the second term. This implies Conjecture 1.5 for Dirichlet polynomials D with $E(W) \geq |W|^3 T^{-o(1)}$. Thus we can obtain improved bounds for Dirichlet polynomials whose large value set has a lot of additive structure.

We introduce a new method that gives good bounds for Dirichlet polynomials on sets of small energy. We outline our method in the next section.

Remark. *Many earlier results on large values of Dirichlet polynomials (such as those stemming from the Mean Value Theorem or the Montgomery-Huxley-Halász large values estimate) allowed one to relax the ℓ^∞ constraint $|b_n| \leq 1$ on the coefficients to a weaker ℓ^2 constraint $(\sum_{n=N}^{2N} |b_n|^2)^{1/2} = O(N^{1/2})$. Our argument crucially relies on the stronger ℓ^∞ assumption to bound the additive energy $E(W)$ efficiently, and we do not know how to obtain Theorem 1.1 if we only have ℓ^2 bounds on the coefficients. The remark after Lemma 11.4 highlights where we use this assumption.*

1.2. Notations and conventions. We write $A \ll B$ to mean that $|A| \leq CB$ for an absolute constant C , and $A \ll_z B$ to denote that the constant C may depend on the parameter z . We write $A \asymp B$ to mean $A \ll B$ and $B \ll A$ both hold, and $A \sim B$ to mean $B < A \leq 2B$. Similarly, we write $A \lesssim B$ to mean that for any $\epsilon > 0$, there is a constant $C(\epsilon) > 0$ depending only on ϵ such that $|A| \leq C(\epsilon)T^\epsilon B$ for all large T , and $A \lesssim_z B$ to denote a dependency on a parameter z . Asymptotic quantities such as $o(1)$ are interpreted as $T \rightarrow \infty$.

We use $e(x) := e^{2\pi i x}$ to denote the complex exponential. Summations will run over integers unless specified otherwise.

1.3. Acknowledgements. LG would like to thank Yuqiu Fu for many interesting discussions about this problem. LG is supported by a Simons Investigator Award. JM is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 851318). We would like to thank the anonymous referees for a number of helpful comments and corrections. For the purposes of open access, the authors have applied a CC-BY license to any author accepted manuscript arising from this submission.

2. SKETCH OUTLINE

For a heuristic description of our argument, let us consider a critical case of previous zero density estimates, where $\sigma = 3/4$ and we wish to improve upon the bound $|W| \lesssim TN^{-1/2}$ for any set W of separated points in an interval of length T such that $|D(t)| > N^{3/4}$ when $t \in W$. We focus on the situation when $T = N^{1+\delta}$ for some small constant $\delta > 0$ (a corresponding improvement for larger values of T then follows by subdivision), and suppress many technical details (such as the presence of smoothing in most summations) for exposition.

Let M be the $|W| \times N$ matrix with entries

$$M_{t,n} = n^{it}$$

for $t \in W$ and $n \sim N$. We see that if $\mathbf{b} = (b_n)_{n \sim N}$ then

$$D(t) = \sum_n b_n n^{it} = (M\mathbf{b})_t,$$

and so we have (recalling $|b_n| \leq 1$)

$$N^{3/2}|W| = N^{2\sigma}|W| < \sum_{t \in W} |D(t)|^2 = \|M\mathbf{b}\|_2^2 \leq \|M\|^2 \|\mathbf{b}\|_2^2 \leq N\|M\|^2,$$

where $\|M\|$ is the matrix (operator) norm of M . Thus we wish to improve upon the bound $\|M\| \lesssim T^{1/2}$ which follows from the Mean Value Theorem (or the Large Values Estimate), and we might hope that $\|M\| \approx |N|^{1/2}$. We note that

$$\|M\| = s_1(M)$$

where $s_1(M)$ is the largest singular value of M , which is also the square-root of the largest eigenvalue of M^*M . Substituting this in above, we find

$$|W| \leq N^{-1/2} s_1(M)^2.$$

A simple bound for $s_1(M)$ is then given by the trace of powers of M^*M ; for any $r \in \mathbb{N}$

$$s_1(M) \leq \text{tr}((M^*M)^r)^{1/2r}.$$

We might hope $\text{tr}((M^*M)^r) \approx |W|N^r$, which would imply $s_1(M) \leq N^{1/2}|W|^{1/2r}$. Unfortunately it appears to be very difficult to estimate these traces accurately for large r , and we would only improve upon the bound $|W| \lesssim TN^{-1/2}$ if $(TN^{-1/2})^{1/r} \leq TN^{-1}$, which requires r to be large. Nevertheless, a variant of this bound for $r = 3$ (which reduces the contribution from some of the trivial terms) allows one to show a bound similar to

$$s_1(M)^6 \lesssim \left| \sum_{\substack{t_1, t_2, t_3 \in W \\ |t_j - t_k| > T^\sigma \forall j \neq k}} \sum_{n_1, n_2, n_3 \sim N} n_1^{i(t_1 - t_2)} n_2^{i(t_2 - t_3)} n_3^{i(t_3 - t_1)} \right|.$$

The right hand side would be $\text{tr}((M^*M)^3)$ if we didn't have the lower bounds on $|t_i - t_j|$ (and so we avoid the $t_1 = t_2 = t_3$ terms). To improve upon the bound $|W| \lesssim TN^{-1/2}$ we wish to show the right hand side is a bit smaller than T^3 .

The natural approach to estimating the right hand side would be to apply the approximate functional equation (or Poisson summation and stationary phase) to each of the inner sums, which would yield a bound of roughly

$$\frac{N^3}{T^{3/2}} \left| \sum_{t_1, t_2, t_3 \in W} \sum_{m_1, m_2, m_3 \sim T/N} m_1^{i(t_1 - t_2)} m_2^{i(t_2 - t_3)} m_3^{i(t_3 - t_1)} c_{t_1 - t_2} c_{t_2 - t_3} c_{t_3 - t_1} \right|$$

for some coefficients c_t of size 1. Unfortunately the $c_{t_i - t_j}$ coefficients link the variables t_1, t_2 and t_3 , and this makes it difficult to show any cancellation over these summations. Without any cancellation in the t_1, t_2, t_3 sums we would be limited to a bound of $|W|^3 N^{3/2} \approx T^3$ even if we obtained square-root cancellation in the sums over m_1, m_2, m_3 , and so this would fail to give the desired improvement on T^3 (but could potentially give results if $\sigma > 3/4$).

A key observation is that it can be beneficial to avoid the use of stationary phase. Applying Poisson summation without simplifying the Fourier integrals yields a bound of roughly

$$\begin{aligned} & N^3 \sum_{t_1, t_2, t_3 \in W} \sum_{|m_1|, |m_2|, |m_3| \sim T/N} \int_{[1, 2]^3} e^{-2\pi i N \mathbf{m} \cdot \mathbf{u}} \left(\frac{u_1}{u_3}\right)^{it_1} \left(\frac{u_2}{u_1}\right)^{it_2} \left(\frac{u_3}{u_1}\right)^{it_3} d\mathbf{u} \\ & \leq N^3 \sum_{|\mathbf{m}| \sim T/N} \left| \int_{[1, 2]^3} e^{-2\pi i N \mathbf{m} \cdot \mathbf{u}} R\left(\frac{u_1}{u_3}\right) R\left(\frac{u_2}{u_1}\right) R\left(\frac{u_3}{u_2}\right) d\mathbf{u} \right|, \end{aligned}$$

where $R(x) := \sum_{t \in W} |x|^{it}$. By not simplifying the integrals over \mathbf{u} directly we have given up a factor of $T^{3/2}$, but now the variables t_1, t_2, t_3 are nicely separated from one another, and we can hope to show cancellation in these sums by showing the R function exhibits cancellation. However, even square-root cancellation in $R(x)$ would only win back a factor of $|W|^{3/2}$ which is less than the $T^{3/2}$ factor we gave up by not applying stationary phase. Therefore we need to exploit simultaneous cancellation in the integral over \mathbf{u} and the R functions.

The next important observation is that the arguments of the R functions only lie in a two dimensional subvariety $z_1 z_2 z_3 = 1$. If we change variables to $v_1 = u_1/u_3$ and $v_2 = u_2/u_3$ then the integral above is roughly

$$\int_{[1/2, 2]^2} \left(\int_{[1, 2]} e^{-2\pi i N(m_1 v_1 + m_2 v_2 + m_3) u_3} du_3 \right) R(v_1) R\left(\frac{v_2}{v_1}\right) R\left(\frac{1}{v_2}\right) dv_1 dv_2.$$

There is a large amount of cancellation in the inner integral unless $m_1v_1 + m_2v_2 + m_3 \approx 0$, and this allows us to win back a factor of T . We are then reduced to bounding expressions of the form

$$(2.1) \quad \frac{N^3}{T} \sum_{|\mathbf{m}| \sim T/N} \left| \int_{\substack{(v_1, v_2) \in [1/2, 2]^2 \\ m_1v_1 + m_2v_2 + m_3 = 0}} R(v_1) R\left(\frac{v_2}{v_1}\right) R\left(\frac{1}{v_2}\right) dv_1 \right|.$$

If we had uniform square-root cancellation in the R functions, we would now get a bound of $T^2|W|^{3/2}$, which comfortably beats the desired bound of T^3 when $|W| \approx TN^{-1/2}$, and so we would get a corresponding improvement to the large value estimates and zero density results.

Of course, we cannot expect to prove uniform square-root cancellation in the R function for arbitrary sets W . Nevertheless, one can show that R does exhibit cancellation on average; $\|R(v)\|_{L^2([1/2, 2])}^2 \lesssim |W|$ and $\|R(v)\|_{L^4([1/2, 2])}^4 \lesssim E(W)$, where $E(W)$ from (1.7) counts approximate additive quadruples in W . Treating the \mathbf{m} summation trivially, this would give a bound

$$(2.2) \quad T^2|W|^{1/2}E(W)^{1/2},$$

which would beat the target of T^3 if $E(W)$ is a bit smaller than $T^2/|W| \approx N^2|W|^3/T^2$. Thus we obtain good bounds whenever $E(W)$ is small. In particular, when $T \approx N^{1+\delta}$ with δ small, this means that we can handle any set W except those for which the additive energy is close to the maximal possible value of $|W|^3$.

We would like to complement this with an argument for when $E(W)$ is large. As mentioned in the introduction, Heath-Brown's result becomes useful in this situation. In this case, Lemma 1.7 implies that $E(W) \ll |W|^2N^{1/2} \approx |W|^3N/T$, which isn't quite strong enough to cover all ranges. A refinement of this lemma can improve this to $E(W) \ll |W|^3N^2/T^2$ (at least when δ is small and $|W| \approx TN^{-1/2}$), which then allows us to get a small improvement in all cases except when $E(W) \approx |W|^3N^2/T^2$.

To overcome this final obstacle when $E(W) \approx |W|^3N^2/T^2$, we go back to (2.1), and exploit the averaging over \mathbf{m} . After a change of variables we need to consider

$$\frac{N^3}{T} \sum_{|m_1|, |m_2|, |m_3| \sim T/N} \int_{v_1 \in [1/2, 2]} \left| R(v_1) R\left(\frac{m_1v_1 + m_3}{m_2v_1}\right) R\left(\frac{m_1v_1 + m_3}{m_2}\right) \right| dv_1.$$

The only way the bound above could be tight is if there is a sparse set U (of measure roughly $TN^{-3/2}$) on which $R(u)$ is large (taking values of size roughly $N^{1/2}$), and such that for many $u \in U$ and many $|m_1|, |m_2|, |m_3| \sim T/N$ we have $(m_1u + m_3)/m_2 \in U$ and $(m_1u + m_3)/(m_2u) \in U$. We show that there cannot be a small set U which has this property of being approximately closed under many affine

transformations $u \mapsto (m_1 u + m_2)/m_3$, and this ultimately leads to an improvement of (2.2) to roughly

$$TN|W|^{1/2}E(W)^{1/2}.$$

This now gives an improvement on the desired bound of T^3 whenever $E(W)$ is smaller than $|W|^3$, and so the bounds obtained on $E(W)$ stemming from Heath-Brown's result above are now sufficient to give an improvement for any size of $E(W)$. Therefore we obtain an improvement on the bound $|W| \lesssim TN^{-1/2}$ for arbitrary W .

3. REDUCTION TO MAIN PROPOSITION

In this section we reduce Theorem 1.1 to a similar but technically more convenient proposition where we have inserted a smoothing in the n variable (which aids the later Fourier analysis) and used subdivision to specialize to the case $N = T^{4/5}$ (which is when the final two terms in Theorem 1.1 coincide). Throughout the rest of the paper we fix a smooth function $w : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ supported on $[1, 2]$ with $\|w^{(j)}\|_{\infty} = O_j(1)$ for all $j \in \mathbb{Z}_{\geq 0}$ and with $w(t) = 1$ for $t \in [6/5, 9/5]$.

Proposition 3.1. *Let $\sigma \in [7/10, 8/10]$ and $\epsilon > 0$. Let b_n be a sequence of complex numbers with $|b_n| \leq 1$ and W be a set of T^ϵ -separated points in an interval of length $T = N^{6/5}$ such that*

$$\left| \sum_n w\left(\frac{n}{N}\right) b_n n^{it} \right| \geq N^\sigma$$

for all $t \in W$. Then we have

$$|W| \leq TN^{(12-20\sigma)/5+o_\epsilon(1)}.$$

Proof of Theorem 1.1 assuming Proposition 3.1. As mentioned in the introduction, the result follows from (1.1) if $V \leq N^{7/10+o(1)}$ or if $V \geq N^{8/10-o(1)}$, so we may assume $V \in [4N^{7/10}, N^{8/10}]$, in which case $N^2V^{-2} \leq N^{18/5}V^{-4}$. Similarly, Theorem 1.1 follows from (1.1) if $N \geq T$ since then $R \leq T^{o(1)}N^2V^{-2}$, so we may assume $N < T$. By splitting $D(t)$ into 3 separate pieces (and using the triangle bound), we see that it suffices to show the result when $b_n = 0$ unless $n \in [6N/5, 9N/5]$. Since the function w is 1 on $[6/5, 9/5]$, we then have that $b_n = b_n w(n/N)$, so we may insert the weight $w(n/N)$. Finally, having inserted the smooth weights we now relax the vanishing condition on the b_n . Thus, letting $V = N^\sigma$, it suffices to show that whenever $\sigma \in [7/10, 8/10]$, $N \ll T$, (a_n) is a 1-bounded complex sequence and \mathcal{W} is a set of 1-separated points such that

$$\left| \sum_n w\left(\frac{n}{N}\right) a_n n^{it} \right| \geq N^\sigma$$

for each $t \in \mathcal{W}$, we have

$$|\mathcal{W}| \leq T^{o(1)}(N^{18/5-4\sigma} + TN^{12/5-4\sigma}).$$

We now fix $\eta > 0$ and choose $\mathcal{W}' \subset \mathcal{W}$ so that \mathcal{W}' is T^η -separated and $|\mathcal{W}'| \geq |\mathcal{W}|/T^\eta$ (this can be achieved by picking the smallest element of \mathcal{W} and then repeatedly choosing the next smallest element which is at least T^η away from all picked elements). If $T \leq N^{6/5}$, then we apply Proposition 3.1 to bound \mathcal{W}' directly (taking $\epsilon = \eta/2$ so that \mathcal{W}' is $N^{6\epsilon/5}$ -separated since $M \ll T$), which implies that

$$|\mathcal{W}| \leq T^\eta |\mathcal{W}'| \leq T^\eta N^{(18-20\sigma)/5+o_\eta(1)}.$$

Letting $\eta \rightarrow 0$ sufficiently slowly then gives the result in this case. If instead $T > N^{6/5}$, then we divide \mathcal{W}' into $\lceil T/N^{6/5} \rceil$ subsets \mathcal{W}'_j each supported on an interval of length $N^{6/5}$, and we apply Proposition 3.1 to bound each \mathcal{W}'_j separately. This gives

$$|\mathcal{W}| \leq T^\eta \sum_{j \leq \lceil T/N^{6/5} \rceil} |\mathcal{W}'_j| \leq T^{1+\eta} N^{(12-20\sigma)/5+o_\eta(1)}.$$

Letting $\eta \rightarrow 0$ sufficiently slowly then gives the result in this case too. \square

4. THE MATRIX M_W AND ITS SINGULAR VALUES

Now we begin to work on the proof of Proposition 3.1. We will work with the smoothed version $D_N(t)$ of $D(t)$ from Proposition 3.1

$$(4.1) \quad D_N(t) := \sum_n w\left(\frac{n}{N}\right) b_n n^{it},$$

where w is the smooth bump supported on $[1, 2]$ defined in Section 3. Similarly, given a set $W \subseteq \mathbb{R}$, let M_W be the $|W| \times N$ matrix with smoothed entries

$$(4.2) \quad (M_W)_{t,n} = w(n/N) n^{it},$$

where $t \in W$ and $n \sim N$.

Lemma 4.1 (Large values of Dirichlet polynomials controlled by singular values). *Let M_W be the matrix defined in (4.2), and $s_1(M_W)$ its largest singular value. If $|D_N(t)| \geq N^\sigma$ on W and if $|b_n| \leq 1$, then we have*

$$|W| \ll N^{1-2\sigma} s_1(M_W)^2.$$

Proof. Let \mathbf{b} be the vector with components b_n . Then note that for each t in W ,

$$D_N(t) = \sum_n w(n/N) b_n n^{it} = (M_W \mathbf{b})_t.$$

Therefore we can relate the behavior of D_N on W (for arbitrary \mathbf{b}) to properties of the matrix M_W , in particular its singular values. We write $s_j(M_W)$ for the j^{th} singular value of M_W , with the convention that $s_1(M_W) \geq s_2(M_W) \geq \dots \geq s_k(M_W)$ and $k = \min(|W|, N)$ is the number of singular values. Let M_W have singular value decomposition $M_W = U\Sigma V$, so that Σ is a rectangular matrix with $\Sigma_{ii} = s_i(M_W)$ and $\Sigma_{ij} = 0$ if $i \neq j$, and U, V are unitary matrices.

If $|D_N(t)| \geq N^\sigma$ on W , then we see

$$\begin{aligned} |W|N^{2\sigma} &\leq \sum_{t \in W} |D_N(t)|^2 = (M_W \mathbf{b})^* M_W \mathbf{b} = (V \mathbf{b})^* \Sigma^2 V \mathbf{b} \\ &\leq s_1(M_W)^2 \|V \mathbf{b}\|_{\ell^2}^2 \\ &= s_1(M_W)^2 \|\mathbf{b}\|_{\ell^2}^2. \end{aligned}$$

Finally, if $|b_n| \leq 1$ then $\|\mathbf{b}\|_{\ell^2}^2 \ll N$. Substituting this into the expression above and rearranging now gives the result. \square

Now $s_1(M_W)$ is equal to the square root of the largest eigenvalue value of the $|W| \times |W|$ matrix $M_W M_W^*$, with entries

$$(M_W M_W^*)_{t_1, t_2} = \sum_n w \left(\frac{n}{N} \right)^2 n^{i(t_1 - t_2)}.$$

A simple bound for $s_1(M_W)$ is therefore to use the trace: for any integer $r \geq 1$ we have

$$s_1(M_W)^2 = s_1(M_W M_W^*) \leq \left(\sum_{j=1}^k s_j(M_W M_W^*)^r \right)^{1/r} = \text{tr}((M_W M_W^*)^r)^{1/r}.$$

One might guess that $s_j(M_W) \lesssim N^{1/2}$ for all j , in which case we would have

$$\text{tr}((M_W M_W^*)^r)^{1/r} \lesssim |W|^{1/r} N.$$

If one could establish such a sharp bound on $\text{tr}((M_W^* M_W)^r)$ for large r , this would give Conjecture 1.5. Unfortunately we do not know how to obtain good bounds when $r \geq 4$, so we work with $r = 3$. In this case, even a sharp bound

$$\text{tr}((M_W M_W^*)^3)^{1/3} \lesssim |W|^{1/3} N$$

would only yield $|W| \lesssim N^{3-3\sigma}$, which is worse than the bounds established by previous works. To get around this issue, we note that if we are in the extreme scenario when

$$\frac{1}{k} \sum_{i=1}^k s_i(M_W)^6 = \left(\frac{1}{k} \sum_{i=1}^k s_i(M_W)^2 \right)^3$$

then in fact we must have that all the singular values are the same, and so

$$s_1(M_W) = \text{tr}((M_W M_W^*)^3)^{1/6} k^{-1/6},$$

a significant improvement on the bound $\text{tr}((M_W M_W^*)^3)^{1/6}$ we had before. Similarly, we would expect that if $\text{tr}((M_W M_W^*)^3)$ is close to $\text{tr}(M_W M_W^*)^3/k^2$, then we would also get an improved bound on $s_1(M_W)$ since most of the contribution to $\text{tr}((M_W M_W^*)^3)$ would be coming from the many other singular values. The following lemma makes this precise, stating that we can essentially replace $\text{tr}((M_W M_W^*)^3)$ with the difference

$$\text{tr}((M_W M_W^*)^3) - \frac{\text{tr}(M_W M_W^*)^3}{k^2}$$

for the purposes of bounding $s_1(M_W)$.

Lemma 4.2 (Bound for singular values in terms of traces). *Let A be an $m \times n$ complex matrix. Then we have*

$$s_1(A) \leq 2 \left(\text{tr}((AA^*)^3) - \frac{\text{tr}(AA^*)^3}{m^2} \right)^{1/6} + 2 \left(\frac{\text{tr}(AA^*)}{m} \right)^{1/2}.$$

Proof. Recall that $\text{tr}((AA^*)^j) = \sum_{i=1}^m \lambda_i^j$ where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of the $m \times m$ matrix AA^* , which are real and non-negative, and that $s_1(A) = \max_i \lambda_i^{1/2}$. We see that it is sufficient to show for any non-negative reals x_1, \dots, x_k

$$(4.3) \quad x_1 \leq 2 \left(\sum_{i=1}^k x_i^6 - \frac{(\sum_{i=1}^k x_i^2)^3}{k^2} \right)^{1/6} + 2 \left(\frac{\sum_{i=1}^k x_i^2}{k} \right)^{1/2}.$$

By Hölder's inequality, we have that $\sum_{i=2}^k x_i^6 \geq (\sum_{i=2}^k x_i^2)^3 / (k-1)^2 \geq (\sum_{i=2}^k x_i^2)^3 / k^2$. Thus

$$\begin{aligned} x_1^6 &= \sum_{i=1}^k x_i^6 - \sum_{i=2}^k x_i^6 \leq \sum_{i=1}^k x_i^6 - \frac{(\sum_{i=2}^k x_i^2)^3}{k^2} \\ &\leq \left(\sum_{i=1}^k x_i^6 - \frac{(\sum_{i=1}^k x_i^2)^3}{k^2} \right) + 3x_1^2 \frac{(\sum_{i=1}^k x_i^2)^2}{k^2} \\ &\leq \max \left(4 \left(\sum_{i=1}^k x_i^6 - \frac{(\sum_{i=1}^k x_i^2)^3}{k^2} \right), 4x_1^2 \frac{(\sum_{i=1}^k x_i^2)^2}{k^2} \right). \end{aligned}$$

This in turn implies

$$x_1^6 \leq \max \left(4 \left(\sum_{i=1}^k x_i^6 - \frac{(\sum_{i=1}^k x_i^2)^3}{k^2} \right), 8 \frac{(\sum_{i=1}^k x_i^2)^3}{k^3} \right),$$

which gives (4.3). \square

Thus we wish to estimate $\text{tr}(M_W M_W^*)$ and $\text{tr}((M_W M_W^*)^3)$. In both cases we expand the trace and use Poisson summation as a first step. In anticipation of this, we introduce the function

$$(4.4) \quad h_t(u) := w(u)^2 u^{it},$$

which appears in the sums defining the coefficients of $M_W M_W^*$. We first record a basic tail estimate for the Fourier transform \widehat{h}_t .

Lemma 4.3 (Non-stationary phase). *Let $h_t(u) = w(u)^2 u^{it}$. Then we have*

(1) *For any integer $j \geq 0$ we have*

$$\widehat{h}_t(\xi) \ll_j (1 + |t|)^j / |\xi|^j.$$

(2) *For any integer $j \geq 0$ we have*

$$\widehat{h}_t(\xi) \ll_j (1 + |\xi|)^j / |t|^j.$$

Proof. Since $\|w^{(j)}\|_\infty \ll_j 1$ for all $j \geq 0$, we have that $\|h_t^{(j)}\|_\infty \ll_j 1 + |t|^j$ for all $j \geq 0$. Thus, by integration by parts (and using that w is compactly supported), we have that

$$\widehat{h}_t(\xi) = \int e(-\xi u) h_t(u) du = \frac{1}{(2\pi i \xi)^j} \int e(-\xi u) h_t^{(j)}(u) du \ll_j \frac{1 + |t|^j}{|\xi|^j}.$$

Similarly, if $g_\xi(u) = e(-\xi u) w(u)^2$ then $\|g_\xi^{(j)}\|_\infty \ll_j 1 + |\xi|^j$, so integration by parts gives

$$\widehat{h}_t(\xi) = \int g_\xi(u) u^{it} du = \frac{(-1)^j}{(it + 1) \cdots (it + j)} \int g_\xi^{(j)}(u) u^{it+j} du \ll_j \frac{1 + |\xi|^j}{|t|^j}. \quad \square$$

Lemma 4.4 (Hilbert-Schmidt Norm estimate). *If $W \subset \mathbb{R}$ is a finite set with $|W| \leq N^{O(1)}$, then*

$$\text{tr}(M_W M_W^*) = N|W| \|w\|_{L^2}^2 + O(N^{-100}).$$

Proof. Expanding the trace, we see that

$$\text{tr}(M_W M_W^*) = \sum_{t \in W} \sum_n w(n/N)^2 = |W| \sum_{n \in \mathbb{Z}} h_0(n/N).$$

Because h_0 is a smooth compactly supported function, the sum $\sum_n h_0(n/N)$ is very close to the integral $N \int_{\mathbb{R}} h_0(u) d\xi = N \|w\|_{L^2}^2$. We can get a precise estimate using Poisson summation, which gives (separating the term $m = 0$)

$$\sum_n h_0(n/N) = N \sum_m \widehat{h}_0(Nm) = N\widehat{h}_0(0) + O\left(N \sum_{m \neq 0} |\widehat{h}_0(Nm)|\right).$$

The first term on the right hand side is $N\|w\|_{L^2}^2$ and the second term is $O(N^{-100})$ by Lemma 4.3. \square

Lemma 4.5 (Expansion of the cubic trace). *Let W be T^ϵ -separated. Then we have*

$$\mathrm{tr}((M_W M_W^*)^3) = N^3 |W| \|w\|_{L^2}^6 + \sum_{m \in \mathbb{Z}^3 \setminus \{0\}} I_m + O_\epsilon(T^{-100}),$$

where

$$I_m := N^3 \sum_{t_1, t_2, t_3 \in W} \widehat{h}_{t_1-t_2}(m_1 N) \widehat{h}_{t_2-t_3}(m_2 N) \widehat{h}_{t_3-t_1}(m_3 N).$$

Proof. First we expand $\mathrm{tr}((M_W M_W^*)^3)$ as the sum S , given by

$$\begin{aligned} S &= \sum_{n_1, n_2, n_3 \in \mathbb{Z}} \sum_{t_1, t_2, t_3 \in W} w\left(\frac{n_1}{N}\right)^2 w\left(\frac{n_2}{N}\right)^2 w\left(\frac{n_3}{N}\right)^2 n_1^{i(t_1-t_2)} n_2^{i(t_2-t_3)} n_3^{i(t_3-t_1)} \\ &= \sum_{t_1, t_2, t_3 \in W} \sum_{n_1, n_2, n_3 \in \mathbb{Z}} h_{t_1-t_2}\left(\frac{n_1}{N}\right) h_{t_2-t_3}\left(\frac{n_2}{N}\right) h_{t_3-t_1}\left(\frac{n_3}{N}\right), \end{aligned}$$

where, as in (4.4), we have $h_t(u) = w(u)^2 u^{it}$. We now perform Poisson summation in n_1, n_2, n_3 , which gives

$$S = N^3 \sum_{m_1, m_2, m_3 \in \mathbb{Z}} \sum_{t_1, t_2, t_3 \in W} \widehat{h}_{t_1-t_2}(m_1 N) \widehat{h}_{t_2-t_3}(m_2 N) \widehat{h}_{t_3-t_1}(m_3 N) = \sum_{m \in \mathbb{Z}^3} I_m.$$

Finally, we separate the term $m_1 = m_2 = m_3 = 0$, which contributes

$$I_0 = N^3 \sum_{t_1, t_2, t_3 \in W} \widehat{h}_{t_1-t_2}(0) \widehat{h}_{t_2-t_3}(0) \widehat{h}_{t_3-t_1}(0).$$

Since W is T^ϵ -separated, $\widehat{h}_{t_1-t_2}(0) \ll_\epsilon T^{-200}$ if $t_1 \neq t_2$ by Lemma 4.3. Thus the terms in the I_0 above are negligible unless $t_1 = t_2 = t_3$, and so

$$I_0 = N^3 \sum_{t \in W} \widehat{h}_0(0)^3 + O_\epsilon(T^{-100}) = N^3 |W| \|w\|_{L^2}^6 + O_\epsilon(T^{-100}).$$

Putting this together gives the result. \square

Putting together Lemmas 4.1-Lemma 4.5, and noting that the $N^3 |W| \|w\|_{L^2}^3$ term cancels with $\mathrm{tr}(M_W M_W^*)^3 / |W|^2$, gives the following.

Proposition 4.6. *Let W be T^ϵ -separated, and let $|b_n| \leq 1$ be such that $|D_N(t)| > N^\sigma$ for all $t \in W$. Then we have*

$$|W| \ll_\epsilon N^{2-2\sigma} + N^{1-2\sigma} \left(\sum_{m \in \mathbb{Z}^3 \setminus \{0\}} I_m \right)^{1/3},$$

where I_m is the quantity defined in Lemma 4.5.

The first term above corresponds to the best possible estimate is $|W| \ll N^{2-2\sigma}$ of Conjecture 1.5, and so our task is reduced to getting a good bound for the sum of I_m .

5. THE PIECES OF THE SUM S

Recall from Proposition 4.6, we have

$$|W|^3 \ll N^{6-6\sigma} + N^{3-6\sigma} \sum_{m \in \mathbb{Z}^3 \setminus \{0\}} I_m,$$

where

$$I_m = N^3 \sum_{t_1, t_2, t_3 \in W} \widehat{h}_{t_1-t_2}(m_1 N) \widehat{h}_{t_2-t_3}(m_2 N) \widehat{h}_{t_3-t_1}(m_3 N).$$

To get started, we note a few cases when $|\widehat{h}_{t_1-t_2}(mN)|$ is easy to understand via Lemma 4.3. Since W is T^ϵ -separated, we see that if $t_1 \neq t_2$, by Lemma 4.3 we have

$$(5.1) \quad |\widehat{h}_{t_1-t_2}(0)| \ll_\epsilon T^{-100}.$$

On the other hand, if $t_1 = t_2$, then we have

$$(5.2) \quad \widehat{h}_{t_1-t_2}(0) = \widehat{h}_0(0) = \int w^2(u) du \asymp 1.$$

If $t_1 = t_2$ but $m \neq 0$, then by Lemma 4.3 we have

$$(5.3) \quad |\widehat{h}_{t_1-t_2}(mN)| \ll m^{-100} N^{-100}.$$

Finally, if $m > T^{1+\epsilon}/N$, then since W is contained in an interval of length T , we have by Lemma 4.3 and taking $j = \lceil 200/\epsilon \rceil + 100$

$$(5.4) \quad |\widehat{h}_{t_1-t_2}(mN)| \ll_\epsilon \frac{T^{100}}{(mN)^{100}} \left(\frac{T}{T^{1+\epsilon}} \right)^{200/\epsilon} \ll T^{-100} m^{-100}.$$

With this in mind, we divide the sum into pieces

$$(5.5) \quad \sum_{m \in \mathbb{Z}^3 \setminus \{0\}} I_m = S_1 + S_2 + S_3,$$

where S_1 contains the terms where exactly one m_i is non-zero, S_2 contains the terms where exactly two m_i are non-zero, and S_3 contains the terms where all three m_i are non-zero.

We will see that S_1 is negligible. In the next section, we will bound S_2 using Heath-Brown's theorem, Theorem 1.6. The main part of the paper is concerned with studying S_3 , which contains most of the terms and is most difficult.

Proposition 5.1 (S_1 bound). *We have*

$$S_1 = O_\epsilon(T^{-10}).$$

Proof. By symmetry, we see that

$$S_1 \leq 3N^3 \sum_{t_1, t_2, t_3 \in W} \sum_{m_3 \neq 0} |\widehat{h}_{t_1-t_2}(0) \widehat{h}_{t_2-t_3}(0) \widehat{h}_{t_3-t_1}(m_3N)|.$$

By (5.4) (using the trivial bound $|\widehat{h}_t(\xi)| \ll 1$ for the other factors), terms with $|m_3| > T^{1+\epsilon}/N$ contribute

$$\ll_\epsilon N^3 |W|^3 \sum_{m > T^{1+\epsilon}/N} T^{-100} m^{-100} \ll T^{-10}.$$

Thus we may restrict attention to terms with $|m_3| < T^{1+\epsilon}/N$. Next we consider terms with $t_1 \neq t_2$. Using (5.1) to bound $|\widehat{h}_{t_1-t_2}(0)|$ (and the trivial bound $\widehat{h}_t \ll 1$ for the remaining factors), we see the terms with $t_1 \neq t_2$ and $|m_3| < T^{1+\epsilon}/N$ contribute

$$\ll_\epsilon N^3 |W|^3 \frac{T^{1+\epsilon}}{N} T^{-100} \ll T^{-10}.$$

Similarly, the terms with $t_2 \neq t_3$ contribute $O_\epsilon(T^{-10})$. The remaining terms have $t_1 = t_2 = t_3$. For these terms we apply (5.3) to bound $|\widehat{h}_{t_3-t_1}(m_3N)|$, which shows that the terms with $t_1 = t_2 = t_3$ also contribute $O_\epsilon(T^{-10})$. This gives the result. \square

6. THE CONTRIBUTION OF S_2

The aim of this section is to establish the following bound for the sum S_2 , which ultimately relies on Heath-Brown's estimate, Theorem 1.6.

Proposition 6.1 (S_2 bound). *For any choice of $k \in \mathbb{N}$*

$$S_2 \lesssim_{\epsilon, k} N^2 |W|^2 + TN |W|^{2-1/k} + N^2 |W|^2 \left(\frac{T^{1/2}}{|W|^{3/4}} \right)^{1/k}.$$

The proof of this proposition relies on the following consequence of stationary phase, which is part of the well-known ‘reflection principle’ for Dirichlet polynomials, or the approximate functional equation (values of a Dirichlet polynomial of length N at $t \in [T, 2T]$ are determined by values of a Dirichlet polynomial of length T/N).

Lemma 6.2 (Approximate functional equation). *For every t with $|t| \sim T_0 \geq T^\epsilon$, we have*

$$\left| \sum_{m \neq 0} \widehat{h}_t(mN) \right| \ll \frac{1}{T_0^{1/2}} \int_{u \lesssim 1} \left| \sum_{m \lesssim T_0/N} m^{-i(t+u)} \right| du + O_\epsilon(T^{-100}).$$

Although somewhat standard, we will give a detailed proof of Lemma 6.2 below. Let us first use it to bound S_2 .

Proof of Proposition 6.1 assuming Lemma 6.2. Recall that S_2 is the sum of those I_m where exactly two m_i are non-zero. By symmetry, we have

$$S_2 = 3N^3 \sum_{m_1, m_2 \neq 0} \sum_{t_1, t_2, t_3 \in W} \widehat{h}_{t_1-t_2}(m_1N) \widehat{h}_{t_2-t_3}(m_2N) \widehat{h}_{t_3-t_1}(0).$$

If $t_1 \neq t_3$, then (5.1) shows that the last factor $\widehat{h}_{t_3-t_1}(0)$ is $O_\epsilon(T^{-100})$, and so using the bound $\widehat{h}_t(u) \ll (1 + |t|^2)/|u|^2$ from Lemma 4.3 for the remaining factors, these terms contribute $O_\epsilon(T^{-10})$ in total. Therefore we have

$$S_2 = 3N^3 \widehat{h}_0(0) \sum_{m_1, m_2 \neq 0} \sum_{t_1, t_2 \in W} \widehat{h}_{t_1-t_2}(m_1N) \widehat{h}_{t_2-t_1}(m_2N) + O_\epsilon(T^{-10}).$$

Since $h_t(u) = w(u)^2 u^{it}$, we have $h_{-t}(u) = \overline{h_t(u)}$, and so $\widehat{h}_{-t}(\xi) = \overline{\widehat{h}_t(-\xi)}$. In particular,

$$\widehat{h}_{t_2-t_1}(m_2N) = \overline{\widehat{h}_{t_1-t_2}(-m_2N)}.$$

Therefore, we can simplify the last equation to get

$$S_2 = 3N^3 \widehat{h}_0(0) \sum_{t_1, t_2 \in W} \left| \sum_{m \neq 0} \widehat{h}_{t_1-t_2}(mN) \right|^2 + O_\epsilon(T^{-10}).$$

Remark. *Heath-Brown’s theorem, Theorem 1.6, gives a good estimate for the sum $\sum_{t_1, t_2 \in W} \left| \sum_{m \in \mathbb{Z}} \widehat{h}_{t_1-t_2}(mN) \right|^2$. However we cannot apply it immediately because we need to be careful to leave out the term with $m = 0$. This term is related to I_0 which we handled carefully in the previous section.*

If $t_1 = t_2$, then $\sum_{m \neq 0} \widehat{h}_{t_1-t_2}(mN)$ is negligible by (5.3) and (5.4). So, splitting the sum dyadically according to the size of $t_1 - t_2$, we find

$$S_2 \lesssim N^3 \sup_{\substack{M=2^j \\ T^\epsilon/N < M < 2T/N}} \sum_{\substack{t_1 \neq t_2 \in W \\ |t_1 - t_2| \sim MN}} \left| \sum_{m \neq 0} \widehat{h}_{t_1 - t_2}(mN) \right|^2 + O_\epsilon(T^{-10}).$$

If $|t_1 - t_2| \sim MN$, then $\sum_{m \neq 0} \widehat{h}_{t_1 - t_2}(mN)$ can be approximated by a Dirichlet polynomial of length M . Indeed, by Lemma 6.2, for such t_1, t_2 we have

$$\left| \sum_{m \neq 0} \widehat{h}_{t_1 - t_2}(mN) \right| \lesssim \frac{1}{M^{1/2}N^{1/2}} \int_{|u| \lesssim 1} \left| \sum_{m \lesssim M} m^{-i(t_1 - t_2 - u)} \right| du + O_\epsilon(T^{-100}).$$

Squaring and summing over $t_1, t_2 \in W$ with $|t_1 - t_2| \sim NM$ gives

$$S_2 \lesssim \sup_{\substack{M \leq 2T/N \\ |u| \lesssim 1}} \frac{N^2}{M} \sum_{\substack{t_1 \neq t_2 \in W \\ |t_1 - t_2| \sim MN}} \left| \sum_{1 \leq m \lesssim M} m^{i(t_1 - t_2 - u)} \right|^2 + O_\epsilon(T^{-10}).$$

We can now drop the condition $|t_1 - t_2| \sim MN$ for an upper bound, and split the summation range $m \lesssim M$ into dyadic intervals. We note that the dyadic range when $M = 1$ gives a contribution which dominates the error term, so for notational convenience we can absorb the error term into the main sum. Thus we find

$$(6.1) \quad S_2 \lesssim_\epsilon \frac{N^2}{M} \sum_{t_1, t_2 \in W} \left| \sum_{m \sim M} a_m m^{i(t_1 - t_2)} \right|^2$$

for some choice of $M \lesssim T/N$ and some coefficients $|a_m| \leq 1$.

We apply Hölder's inequality to this sum, and rewrite the $2k^{\text{th}}$ power of the Dirichlet polynomial as the 2^{nd} power of a longer Dirichlet polynomial. For any choice of positive integer k , we find that

$$(6.2) \quad \sum_{t_1, t_2 \in W} \left| \sum_{m \sim M} a_m m^{i(t_1 - t_2)} \right|^2 \leq |W|^{2-2/k} \left(\sum_{t_1, t_2 \in W} \left| \sum_{m \lesssim M^k} b_m m^{i(t_1 - t_2)} \right|^2 \right)^{1/k}$$

for some coefficients $b_m \leq M^{o_k(1)}$ (by the divisor bound). Theorem 1.6 bounds sums of this type. We recall the statement.

Theorem (Heath-Brown). *Let \mathcal{T} be a 1-separated set of reals, contained in an interval of length T . Let $|a_n| \lesssim 1$ be a complex sequence. Then*

$$\sum_{t_1, t_2 \in \mathcal{T}} \left| \sum_{n \sim N} a_n n^{i(t_1 - t_2)} \right|^2 \lesssim |\mathcal{T}|^2 N + |\mathcal{T}| N^2 + |\mathcal{T}|^{5/4} T^{1/2} N.$$

This result implies that

$$(6.3) \quad \sum_{t_1, t_2 \in W} \left| \sum_{m \asymp M^k} b_m m^{i(t_1 - t_2)} \right|^2 \lesssim_k |W|^2 M^k + |W| M^{2k} + |W|^{5/4} T^{1/2} M^k.$$

Substituting (6.2) and (6.3) back into (6.1), we see that

$$(6.4) \quad \begin{aligned} S_2 &\lesssim_{\epsilon, k} \frac{N^2}{M} (|W|^2 M + M^2 |W|^{2-1/k} + |W|^2 M T^{1/2k} |W|^{-3/4k}) \\ &\lesssim_{\epsilon, k} N^2 |W|^2 + TN |W|^{2-1/k} + N^2 |W|^2 \left(\frac{T^{1/2}}{|W|^{3/4}} \right)^{1/k}. \end{aligned}$$

This gives the result. \square

Now we return to the proof of Lemma 6.2, which roughly says $\widehat{h}_t(mN)$ can be thought of as a smoothed version of $t^{-1/2} m^{it}$ supported on $m \asymp t/N$.

Proof of Lemma 6.2. Since $t \sim T_0 \geq T^\epsilon$, by Lemma 4.3 we have that $\widehat{h}_t(mN) \ll_\epsilon T^{-100} m^{-2}$ unless $|m| \lesssim T_0/N$. Thus it suffices to just consider terms with $|m| \leq M$ for some suitable $M = T_0^{1+o(1)}/N$ at the cost of an $O_\epsilon(T^{-100})$ error term. We focus on the terms with positive m ; the terms with negative m can be bounded analogously.

We expand the definition of \widehat{h}_t and truncate the integral using the support of w

$$\widehat{h}_t(mN) = \int_{-\infty}^{\infty} w(u)^2 u^{it} e(-mNu) du = \int_{1/m}^{2M/m} w(u)^2 u^{it} e(-mNu) du.$$

Let $H(s) := \int_0^\infty w(u)^2 u^{s-1} du$ be the Mellin transform of h_0 , which is entire and satisfies $H(s) \ll_j |s|^{-j}$ for any $j \in \mathbb{Z}_{>0}$ when $|\Re(s)| \leq 10$ (by repeated integration by parts). Applying Mellin inversion $(w(u)^2)^{-1} = (2\pi i)^{-1} \int_{1-i\infty}^{1+i\infty} H(s) u^{-s} ds$, we have that

$$\widehat{h}_t(mN) = \frac{1}{2\pi i} \int_{1/m}^{2M/m} \int_{1-i\infty}^{1+i\infty} H(s) u^{it-s} e(-mNu) ds du.$$

By the rapid decay of H we may truncate the s integral to $|s| \lesssim 1$ at the cost of a $O(T^{-100})$ error term. We then make a change of variables $s = 1 + ir$ and $v = Nmu$

$$\widehat{h}_t(mN) = \frac{1}{2\pi} \int_{|r| \lesssim 1} H(1 + ir) (mN)^{i(t-r)} \left(\int_N^{2NM} v^{-1+i(t-r)} e(-v) dv \right) dr.$$

Summing over $1 \leq m \leq M$ and applying the triangle bound gives

$$\begin{aligned} \sum_{m \leq M} \widehat{h}_t(mN) &= \frac{1}{2\pi} \int_{|r| \lesssim 1} H(1+ir) N^{i(t-r)} \sum_{1 \leq m \leq M} m^{i(t-r)} \int_N^{2NM} \frac{v^{i(t-r)} e(-v)}{v} dv dr \\ &\ll \int_{|r| \lesssim 1} \left| \sum_{1 \leq m \leq M} m^{i(t-r)} \right| \left| \int_N^{2NM} v^{-1+i(t-r)} e(-v) dv \right| dr. \end{aligned}$$

Integration by parts and the Van-der-Corput first and second derivative bounds (see [M2, Chapter 3, Lemmas 1 and 2]) then show that (for $|r| \geq 2$)

$$\begin{aligned} \int_V^{2V} v^{-1+ir} e(-v) dv &\ll \frac{1}{|r|}, & \text{if } V \leq |r|/20 \text{ or } V \geq 20|r|, \\ \int_V^{2V} v^{-1+ir} e(-v) dv &\ll \frac{1}{|r|^{1/2}}, & \text{if } |r|/20 \leq V \leq 20|r|. \end{aligned}$$

Since $|t-r| \asymp T_0$ when $r \lesssim 1$ (and recalling that $T^\epsilon \leq T_0$, $NM \lesssim T$), together these give

$$\int_N^{2NM} v^{-1+i(t-v)} e(-v) dv \ll T_0^{-1/2}.$$

This gives the result. \square

7. THE CONTRIBUTION OF S_3 : A KEY CANCELLATION

Now we begin to study S_3 , which is the most difficult term. Recall that

$$S_3 = \sum_{m_1, m_2, m_3 \neq 0} I_m,$$

where

$$I_m = N^3 \sum_{t_1, t_2, t_3 \in W} \widehat{h}_{t_1-t_2}(m_1N) \widehat{h}_{t_2-t_3}(m_2N) \widehat{h}_{t_3-t_1}(m_3N).$$

By Lemma 4.3, $|\widehat{h}_t(\xi)| \ll_j (1+|t|)^j/|\xi|^j$ for any $j \in \mathbb{Z}_{\geq 0}$, and so \widehat{h}_t is rapidly decaying when $|\xi|$ is much bigger than $|t|$, and hence I_m is negligible unless $|m| \lesssim T/N$. Thus

$$(7.1) \quad S_3 = \sum_{0 < |m_1|, |m_2|, |m_3| \lesssim T/N} I_m + O(T^{-100}).$$

The first step in our argument is an estimate for $|I_m|$. We introduce the function $R(v)$ which will play an important role in our analysis of S_3 :

$$(7.2) \quad R(v) := \sum_{t \in W} |v|^{it} = \widehat{W} \left(\frac{\log |v|}{-2\pi} \right),$$

where $\widehat{W}(\xi) := \sum_{t \in W} e^{-2\pi i t \xi}$ is the Fourier transform of the distribution with a delta function at each point of W . In this paper we will occasionally find it convenient to work with \widehat{W} , but will not work with the distribution directly.

Proposition 7.1 (Cancellation within the I_m integrals). *We have*

$$|I_m| \ll N^3 \int_{\substack{|m_1 v_1 + m_2 v_2 + m_3| \lesssim \frac{1}{N} \\ v_1 \asymp v_2 \asymp 1}} \left| R(v_1) R\left(\frac{v_2}{v_1}\right) R(v_2) \right| dv_1 dv_2 + O(T^{-200}).$$

Moreover, if $|m_1| \leq |m_2| \leq |m_3|$, then $|I_m| = O(T^{-200})$ unless $|m_2| \asymp |m_3|$.

Proof. To simplify notation, let $w_1(\mathbf{u}) := w(u_1)^2 w(u_2)^2 w(u_3)^2$. Expanding the definition of \widehat{h}_t as an integral and swapping the order of summation and integration, we have

$$(7.3) \quad \begin{aligned} I_m &= N^3 \sum_{t_1, t_2, t_3 \in W} \int_{\mathbb{R}^3} e(-N\mathbf{m} \cdot \mathbf{u}) w_1(\mathbf{u}) u_1^{i(t_1 - t_2)} u_2^{i(t_2 - t_3)} u_3^{i(t_3 - t_1)} d\mathbf{u} \\ &= N^3 \int_{\mathbb{R}^3} e(-N\mathbf{m} \cdot \mathbf{u}) w_1(\mathbf{u}) R\left(\frac{u_1}{u_3}\right) R\left(\frac{u_2}{u_1}\right) R\left(\frac{u_3}{u_2}\right) d\mathbf{u}. \end{aligned}$$

In (7.3), the R functions depend on u_1/u_3 , u_2/u_1 and u_3/u_2 . We therefore rewrite the integral using these variables. We define v_1 and v_2 by

$$v_1 := \frac{u_1}{u_3}, \quad v_2 := \frac{u_2}{u_3}.$$

We rewrite the integral I_m in terms of the variables v_1, v_2, u_3 . When we change variables, the R factors depend on v_1, v_2 but not on u_3 .

$$R\left(\frac{u_1}{u_3}\right) R\left(\frac{u_2}{u_1}\right) R\left(\frac{u_3}{u_2}\right) = R(v_1) R\left(\frac{v_2}{v_1}\right) R\left(\frac{1}{v_2}\right)$$

The exponential factor also works out in a nice way in the new variables:

$$e^{-N\mathbf{m} \cdot \mathbf{u}} = e^{-N(m_1 u_1 + m_2 u_2 + m_3 u_3)} = e^{-N(m_1 v_1 + m_2 v_2 + m_3) u_3}.$$

A Jacobian computation shows that

$$du_1 du_2 du_3 = u_3^2 dv_1 dv_2 du_3.$$

So in the new variables, our integral I_m becomes

$$N^3 \int_{\mathbb{R}^3} e(-N(m_1 v_1 + m_2 v_2 + m_3) u_3) w_2(u_3, v_1, v_2) R(v_1) R\left(\frac{v_2}{v_1}\right) R\left(\frac{1}{v_2}\right) dv_1 dv_2 du_3,$$

where

$$(7.4) \quad w_2(u_3, v_1, v_2) := u_3^2 w(u_3)^2 w(v_1 u_3)^2 w(v_2 u_3)^2.$$

Since the R factors do not involve u_3 , we rewrite our formula to do the u_3 integral first:

$$N^3 \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}} e(-N(m_1 v_1 + m_2 v_2 + m_3) u_3) w_2(u_3, v_1, v_2) du_3 \right) R(v_1) R\left(\frac{v_2}{v_1}\right) R\left(\frac{1}{v_2}\right) dv_1 dv_2.$$

A key observation in our proof is that we can analyze the norm of this inner integral very accurately using non-stationary phase. Recalling the definition (7.4) of w_2 , we see that for any $j \in \mathbb{Z}_{\geq 0}$, $w_2(u_3, v_1, v_2)$ has j^{th} derivative with respect to u_3 bounded by $O_j(1)$ (since w is supported on $[1, 2]$ with $\|w^{(\ell)}\|_{\infty} \ll_{\ell} 1$ for all $\ell \in \mathbb{Z}_{\geq 0}$). Thus for any $\eta > 0$, the inner integral is $O_{\eta}(T^{-300})$ unless $|m_1 v_1 + m_2 v_2 + m_3| \leq T^{\eta}/N$ by repeated integration by parts. In general, the inner integral has size $\ll 1$. In addition, $w_2(u_3, v_1, v_2)$ vanishes unless $v_1, v_2 \in [1/2, 2]$, because $w(u)$ is supported on $u \in [1, 2]$. Therefore, the inner integral vanishes unless $v_1 \in [1/2, 2]$ and $v_2 \in [1/2, 2]$. Using these bounds for the inner integral and then using the triangle inequality, we see that since $\eta > 0$ was arbitrary

$$|I_m| \ll N^3 \int_{\substack{|m_1 v_1 + m_2 v_2 + m_3| \lesssim \frac{1}{N} \\ v_1, v_2 \in [1/2, 2]}} \left| R(v_1) R\left(\frac{v_2}{v_1}\right) R\left(\frac{1}{v_2}\right) \right| dv_1 dv_2 + O(T^{-200}).$$

Since $|R(v)| = |R(1/v)|$, we can replace $R(1/v_2)$ by $R(v_2)$. (This is not really important, but it makes later computations cleaner.)

Finally, this integral vanishes unless we can find $v_1 \asymp 1$ and $v_2 \asymp 1$ so that $m_1 v_1 + m_2 v_2 + m_3$ is almost zero. If $|m_1| \leq |m_2| \leq |m_3|$, this can only happen if $|m_2| \asymp |m_3|$. This gives the last claim in the proposition. \square

Remark. *The cancellation in the inner integral when $|m_1 v_1 + m_2 v_2 + m_3|$ is not $\lesssim 1/N$ is one of the key observations in our proof. This cancellation is specific to Dirichlet polynomials as opposed to more general trigonometric polynomials. For instance, one may consider a ‘generalized’ Dirichlet polynomial of the form $\tilde{D}(t) = \sum_{n \sim N} b_n e^{it\phi(n)}$, where the function $\phi(n)$ has smoothness and convexity properties similar to those of $\log n$. One can follow the argument above, but the inner integral will have the form $\int e(Ng_{\mathbf{m}, v_1, v_2}(u_3)) du_3$ for some function $g_{\mathbf{m}, v_1, v_2}(u_3)$. In general, the function $g_{\mathbf{m}, v_1, v_2}(u_3)$ will not be linear (or monomial) in u_3 and so we would not be in the special situation where the whole integral is either oscillating or stationary. Thus one would expect to have to use stationary*

phase around stationary points in the integral, and the bounds would not work out as they do here.

Because of the last claim in Proposition 7.1, we can restrict attention to m with $0 < |m_1| \leq |m_2| \asymp |m_3|$. The domain of integration can be rewritten in the form

$$\left| v_2 - \frac{m_1 v_1 + m_3}{-m_2} \right| \lesssim \frac{1}{|m_2|N} \asymp \frac{1}{|m_3|N}.$$

So the domain of integration is essentially the $\frac{1}{N|m_3|}$ -neighborhood of the curve $v_2 = \frac{m_1 v_1 + m_3}{-m_2}$. Therefore, $|I_m|$ is morally bounded by

$$\frac{N^3}{N|m_3|} \int_{v_1 \asymp 1} \left| R(v_1) R\left(\frac{m_1 v_1 + m_3}{-m_2 v_1}\right) R\left(\frac{m_1 v_1 + m_3}{-m_2}\right) \right| dv_1.$$

We can make this rigorous by using a smoothed version of R . Define a smoothed version of $|R(u)|$ in terms of compactly supported bump functions $\tilde{\psi}_1, \tilde{\psi}_2$ and a parameter $M \geq 1$ by

$$(7.5) \quad \tilde{R} = \tilde{R}_{\tilde{\psi}_1, \tilde{\psi}_2, M}(u) := \left(\int NM \tilde{\psi}_1(NM(u-u')) \tilde{\psi}_2(u') |R(u')|^2 du' \right)^{1/2}.$$

The following proposition gives an expansion of S_3 in terms of such integrals.

Proposition 7.2 (Expansion of S_3). *There is a choice of $1 \leq M_1 \leq M \lesssim T/N$ and a choice of non-negative bump functions $\tilde{\psi}_1, \tilde{\psi}_2$ with $\tilde{\psi}_1(x)$ supported on $|x| \lesssim 1$ and satisfying $\|\tilde{\psi}_1^{(j)}\|_\infty \lesssim_j 1$ for all $j \in \mathbb{Z}_{\geq 0}$ and $\tilde{\psi}_2(x)$ supported on $x \asymp 1$ and satisfying $\tilde{\psi}_2^{(j)} \ll_j 1$ for all $j \in \mathbb{Z}_{\geq 0}$ and with $\tilde{\psi}_1(0) = \tilde{\psi}_2(1) = 1$, such that*

$$S_3 \lesssim \frac{N^2}{M} \sum_{\substack{|m_1| \sim M_1 \\ |m_2|, |m_3| \asymp M}} \tilde{I}_m + O(T^{-100}),$$

where

$$\tilde{I}_m := \int_{v_1 \asymp 1} \left| R(v_1) \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2 v_1}\right) \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2}\right) \right| dv_1.$$

Proof. Recall from (7.1) that S_3 is bounded by

$$S_3 \leq \sum_{0 < |m_1|, |m_2|, |m_3| \lesssim T/N} |I_m| + O(T^{-100}).$$

From (7.3), we have

$$I_m = N^3 \int_{\mathbb{R}^3} e(-N\mathbf{m} \cdot \mathbf{u}) w_1(\mathbf{u}) R\left(\frac{u_1}{u_3}\right) R\left(\frac{u_2}{u_1}\right) R\left(\frac{u_3}{u_2}\right) d\mathbf{u}.$$

From the definition (7.2) for $R(v)$, we see that $R(1/v) = \bar{R}(v)$. Therefore we see that $I_{(m_1, m_2, m_3)} = \bar{I}_{(m_2, m_1, m_3)}$, and similarly for any other transposition of (m_1, m_2, m_3) . Thus $|I_m|$ is invariant under any permutation of (m_1, m_2, m_3) , and so we can reduce to the case $|m_1| \leq |m_2| \leq |m_3|$ at the cost of a factor of 6. By Proposition 7.1, such terms are negligible unless $|m_2| \asymp |m_3|$. Thus, by choosing dyadic scales to maximize the right hand side, we find that there is an $M_1 \leq M \lesssim T/N$ such that

$$(7.6) \quad S_3 \lesssim \sum_{\substack{|m_1| \sim M_1 \\ |m_2| \asymp M \\ |m_3| \asymp M}} |I_m| + O(T^{-100}).$$

By Proposition 7.1, we have for $m_2 \asymp M$

$$|I_m| \ll N^3 \int_{v_1 \asymp 1} |R(v_1)| \left(\int_{\substack{v_2 \asymp 1 \\ |v_2 - \frac{m_1 v_1 + m_3}{-m_2}| \lesssim \frac{1}{MN}}} \left| R\left(\frac{v_2}{v_1}\right) R(v_2) \right| dv_2 \right) dv_1 + O(T^{-200}).$$

Using Cauchy-Schwarz (and a change of variables $v_2 \mapsto v_2 v_1$ for the first factor), we bound the inner integral by

$$\left(\int_{\substack{v_2 \asymp 1 \\ MN|v_2 - \frac{m_1 v_1 + m_3}{-m_2}| \lesssim 1}} |R(v_2)|^2 dv_2 \right)^{1/2} \left(\int_{\substack{v_2 \asymp 1 \\ MN|v_2 - \frac{m_1 v_1 + m_3}{-m_2}| \lesssim 1}} |R(v_2)|^2 dv_2 \right)^{1/2}.$$

We can now choose a smooth bump function $\tilde{\psi}_2(v_2)$ which majorizes both the integration constraints $v_2 \asymp 1$ and satisfies the support and derivative conditions of the proposition. Similarly, we can choose a bump function $\tilde{\psi}_1$ such that $\tilde{\psi}_1(MN(\frac{m_1 v_1 + m_3}{-m_2 v_1} - v_2))$ majorizes the integration constraint $MN|v_2 - \frac{m_1 v_1 + m_3}{-m_2 v_1}| \lesssim 1$ in the first factor above and $\tilde{\psi}_1(MN(\frac{m_1 v_1 + m_3}{-m_2} - v_2))$ majorizes the corresponding constraint in the second factor and $\tilde{\psi}_1$ satisfies the support and derivative constraints of the proposition. Recalling the definition (7.5) of \tilde{R} , we then see that for this choice of $\tilde{\psi}_1, \tilde{\psi}_2$ and M , the product of integrals above is

$$\ll \frac{1}{MN} \tilde{R}\left(\frac{m_1 v_1 + m_3}{-m_2 v_1}\right) \tilde{R}\left(\frac{m_1 v_1 + m_3}{-m_2}\right).$$

Thus we find that

$$|I_{m_1, m_2, m_3}| \ll \frac{N^2}{M} \tilde{I}_{m_1, -m_2, m_3} + O(T^{-100}).$$

Finally, since we are summing over m_2 with $|m_2| \asymp M$, we can replace $-m_2$ with m_2 without changing the overall sum. Substituting this into our expression (7.6) for S_3 above then gives the result. \square

8. BASIC ESTIMATE FOR THE LOW ENERGY CASE

In this section, we begin to estimate S_3 using Proposition 7.2. Recall that $R(v) = \sum_{t \in W} |v|^{it}$. The best bound for $|R(v)|$ we can hope for is square root cancellation: $|R(v)| \ll |W|^{1/2}$. If indeed $|R(v)| \approx |W|^{1/2}$ for all $v \asymp 1$, then we get $S_3 \lesssim N^2 M^2 |W|^{3/2} \lesssim T^2 |W|^{3/2}$. We will see more generally that this bound holds whenever the energy of W is very small.

Proposition 8.1 (S_3 controlled by energy). *If W is a T^ϵ -separated set contained in an interval of length T , then*

$$S_3 \lesssim_\epsilon T^2 |W|^{1/2} E(W)^{1/2}.$$

Remark. *If we look at the critical case when $T = N^{5/4}$ and $\sigma = 3/4$, then this estimate (together with our bounds for S_2) gives an improvement to the basic orthogonality estimate (1.6) when $E(W)$ is significantly below $|W|^{7/3}$. In the special case when $E(W) \approx |W|^2$, this Proposition is enough to prove our main theorem, Theorem 1.1. On the other hand, when $E(W)$ is very large in terms of $|W|$, then we will get good estimates using Lemma 1.7. However there is an intermediate range of energy that is not yet covered. In the next two sections, we will develop a strengthening of this proposition that gives new estimates for a wider range of energies.*

We begin with some basic lemmas about the moments of R .

Lemma 8.2 (L^2 bound). *Let W be a T^ϵ -separated set contained in an interval of length T . Then*

$$\int_{v \asymp 1} |R(v)|^2 dv \ll_\epsilon |W|.$$

Proof. Let $\psi_1(v)$ be a smooth bump function which majorizes the range of integration of the integral in the lemma and is supported on $v \asymp 1$ (so satisfies $\|\psi_1^{(j)}\|_\infty \ll_j 1$ for all $j \in \mathbb{Z}_{\geq 0}$). Then we have

$$\int_{v \asymp 1} |R(v)|^2 dv \leq \int \psi_1(v) |R(v)|^2 dv.$$

We substitute the definition of $R(v)$ from (7.2), and let $\psi_2(\tau) := 2\pi e^{-2\pi\tau} \psi_1(e^{-2\pi\tau})$. Then, making a change of variables $v = e^{-2\pi\tau}$ gives

$$\int \psi_1(v)|R(v)|^2 dv = \int \psi_2(\tau)|\widehat{W}(\tau)|^2 d\tau = \sum_{t_1, t_2 \in W} \widehat{\psi}_2(t_1 - t_2).$$

Note that ψ_2 is a smooth bump around the origin with $\|\psi_2^{(j)}\|_\infty \ll_j 1$ for all $j \in \mathbb{Z}_{\geq 0}$, so $|\widehat{\psi}_2(\xi)| \ll_j |\xi|^{-j}$ for any $j \in \mathbb{Z}_{\geq 0}$. Since W is T^ϵ -separated, if $t_1 \neq t_2$ we have that $\widehat{\psi}_2(t_1 - t_2) \ll_\epsilon T^{-100}$, so the terms with $t_1 \neq t_2$ are negligible. The terms with $t_1 = t_2$ contribute $\ll |W|$ to the sum above. Thus the total sum is $O_\epsilon(|W|)$, as required. \square

Lemma 8.3 (L^4 bound). *Let $M, \widetilde{\psi}_1, \widetilde{\psi}_2$ be as in Proposition 7.2. Then we have*

$$\int_{v \asymp 1} |\widetilde{R}(v)|^4 dv \lesssim E(W) \quad \text{and} \quad \int_{v \asymp 1} |R(v)|^4 dv \lesssim E(W).$$

Proof. From the definition (7.5) of \widetilde{R} , recalling that $\widetilde{\psi}_1, \widetilde{\psi}_2 \ll 1$ are supported on $|x| \lesssim 1$ and $|x| \asymp 1$ respectively, and Cauchy-Schwarz, we have

$$\int_{v \asymp 1} |\widetilde{R}(v)|^4 dv \lesssim N^2 M^2 \int_{\substack{v \asymp 1 \\ |u-v| \lesssim 1/NM \\ |u'-v| \lesssim 1/NM}} |R(u')|^4 du' dudv \lesssim \int_{u' \asymp 1} |R(u')|^4 du'.$$

Therefore it suffices to prove the result for R . We recall that $R(v) = \widehat{W}(\log |v|)$, so by a change of variables $v = e^\tau$ we see that it suffices to show

$$\int_{\tau \ll 1} |\widehat{W}(\tau)|^4 d\tau \lesssim E(W).$$

Let $\eta > 0$ and let ψ_1 be a smooth bump supported on $\tau \ll 1$ such that $\psi_1(\tau/T^\eta)$ majorizes the range of integration. Then we see that

$$\begin{aligned} \int_{\tau \asymp 1} |\widehat{W}(\tau)|^4 d\tau &\leq \int \psi_1\left(\frac{\tau}{T^\eta}\right) |\widehat{W}(\tau)|^4 d\tau \\ &= \sum_{t_1, t_2, t_3, t_4 \in W} \int \psi_1\left(\frac{\tau}{T^\eta}\right) e(\tau(t_1 + t_2 - t_3 - t_4)) d\tau \\ &= T^\eta \sum_{t_1, t_2, t_3, t_4 \in W} \widehat{\psi}_1\left(T^\eta(t_3 + t_4 - t_1 - t_2)\right). \end{aligned}$$

Since $\widehat{\psi}_1$ decays rapidly, we may restrict the summation to $|t_1 + t_2 - t_3 - t_4| \leq 1$ at the cost of an $O_\eta(T^{-100})$ error term. The remaining terms contribute $\ll T^\eta E(W)$. Thus, letting $\eta \rightarrow 0$ we obtain

$$\int_{\tau \asymp 1} |\widehat{W}(\tau)|^4 d\tau \lesssim E(W). \quad \square$$

Proposition 8.1 follows quickly from Lemma 8.2 and Lemma 8.3, but before we establish this we record a Fourier decay estimate for \widetilde{R} which will be needed in later sections. If f is a version of $|R|^2$ smoothed over intervals of length $1/B$, the Fourier transform of f will decay rapidly beyond B , and since R is essentially constant on intervals of length $1/T$, f cannot be too small on a $1/T$ -neighbourhood of 1.

Lemma 8.4 (Fourier decay of smoothenings of R). *Let ψ_1, ψ_2, ψ_3 be smooth non-negative bump functions satisfying:*

- (1) $\psi_1(t)$ and $\psi_2(t)$ are supported on $t \asymp 1$, and $\psi_3(t)$ is supported on $|t| \lesssim 1$.
- (2) $\psi_1(1) = \psi_2(1) = \psi_3(0) = 1$.
- (3) For any $j \in \mathbb{Z}_{\geq 0}$ we have $\psi_1^{(j)}, \psi_2^{(j)} \ll_j 1$, $\psi_3^{(j)} \lesssim_j 1$.

Let $T^\epsilon \leq B \lesssim T$ and

$$f(u) := \psi_1(u) \int B\psi_3(B(u-u'))\psi_2(u')|R(u')|^2 du'.$$

Then for all $j \in \mathbb{Z}_{\geq 0}$ we have

$$\widehat{f}(\xi) \lesssim_j \frac{T^j}{|\xi|^j} \sup_u f(u).$$

Proof. Let

$$f_1(u) := \int B\psi_3(B(u-u'))\psi_2(u')|R(u')|^2 du', \quad g(u) := \psi_2(u)|R(u)|^2.$$

Then $f_1(u)$ is a convolution of $B\psi_3(Bu)$ and $g(u)$, so $\widehat{f}_1(\xi)$ has Fourier transform $\widehat{\psi}_3(\xi/B)\widehat{g}(\xi)$. Since $|R(u)|^2 \leq |W|^2$ and ψ_2 is supported on $|x| \asymp 1$, we see that $\widehat{g}(\xi) \ll |W|^2$. The derivative bounds on ψ_3 imply that $\widehat{\psi}_3(\xi) \lesssim_j B^j/(1+|\xi|)^j$ for all $j \in \mathbb{Z}_{\geq 0}$. Thus $\widehat{f}_1(\xi) \lesssim_j |W|^2 B^j/(1+|\xi|)^j$ for $j \in \mathbb{Z}_{\geq 0}$. Since $f(u) = \psi_1(u)f_1(u)$ we have $\widehat{f} = \widehat{\psi}_1 * \widehat{f}_1$ and $\widehat{\psi}_1(\xi) \ll_j (1+|\xi|)^{-j}$ for $j \in \mathbb{Z}_{\geq 0}$ from the derivative bound on ψ_1 . This gives

$$\widehat{f}(\xi) \lesssim_j |W|^2 \left(\frac{B}{|\xi|}\right)^j \quad \text{for all } j \in \mathbb{Z}_{\geq 0}.$$

Since $W \subseteq [0, T]$, we have that $|R(u)|^2 \gg |W|^2$ if u is a sufficiently small multiple of $1/T$ from 1. Since $\psi_2(1) = \psi_3(0) = 1$ we also have that $\psi_2(u) \gg 1$ if u is sufficiently close to 1 and $\psi_3(B(1-u)) \gg 1$ if u is a sufficiently small multiple of $1/B$ from 1. Since $B \lesssim T$, we find from restricting u to a neighbourhood of 1 of width a small multiple of $\min(1/T, 1/B)$, that

$$f(1) = \int B\psi_3(B(1-u))\psi_2(u)|R(u)|^2 du \gtrsim \frac{B}{T}|W|^2.$$

Thus for any $j \in \mathbb{Z}_{\geq 0}$,

$$\widehat{f}(\xi) \lesssim_j |W|^2 \left(\frac{B}{|\xi|}\right)^j \lesssim \frac{TB^{j-1}}{|\xi|^j} f(1) \lesssim_j \frac{T^j}{|\xi|^j} \sup_u f(u). \quad \square$$

We now return to prove Proposition 8.1.

Proof of Proposition 8.1. Starting with Proposition 7.2, we have for some $M_1 \leq M \lesssim T/N$ and some suitable bumps $\tilde{\psi}_1, \tilde{\psi}_2$

$$S_3 \lesssim \sum_{\substack{|m_1| \sim M_1 \\ |m_2|, |m_3| \asymp M}} \frac{N^2}{M} \int_{v_1 \asymp 1} \left| R(v_1) \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2 v_1}\right) \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2}\right) \right| dv_1.$$

Using Hölder's inequality, we find that the integral over v_1 is bounded by

$$\left(\int_{v_1 \asymp 1} |R(v_1)|^2 dv_1 \right)^{1/2} \left(\int_{v_1 \asymp 1} \left| \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2 v_1}\right) \right|^4 dv_1 \right)^{1/4} \\ \times \left(\int_{v_1 \asymp 1} \left| \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2}\right) \right|^4 dv_1 \right)^{1/4}.$$

In the second integral we change of variables to $u = \frac{m_1 v_1 + m_3}{m_2 v_1}$ with Jacobian factor $\asymp 1$. In the third integral change of variables to $u = \frac{m_1 v_1 + m_3}{m_2}$ with a Jacobian factor of norm $\asymp M/M_1$. Since \tilde{R} is supported on $u \asymp 1$ (from the support of $\tilde{\psi}_2$), we obtain

$$S_3 \ll N^2 M^2 \left(\int_{v_1 \asymp 1} |R(v_1)|^2 dv_1 \right)^{1/2} \left(\int_{u \asymp 1} |\tilde{R}(u)|^4 du \right)^{1/2}$$

Using $M \lesssim T/N$ and Lemmas 8.2 and 8.3, we find

$$S_3 \lesssim T^2 |W|^{1/2} E(W)^{1/2}. \quad \square$$

When $E(W) \approx |W|^2$, the bound from Proposition 8.1 is the best bound for S_3 we know how to prove and corresponds to square-root cancellation in the R function. For larger $E(W)$, however, we can improve the bound for S_3 . Let us indicate the general direction here, and then we will develop the tool we need in the next section.

Ignoring some technical smoothing, we morally have

$$S_3 \lesssim \sum_{\substack{|m_1| \sim M_1 \\ |m_2|, |m_3| \asymp M}} \frac{N^2}{M} \int_{v_1 \asymp 1} \left| R(v_1) R\left(\frac{m_1 v_1 + m_3}{m_2 v_1}\right) R\left(\frac{m_1 v_1 + m_3}{m_2}\right) \right| dv_1.$$

We can split up the sum over \mathbf{m} and integral over v_1 into pieces where the first R factor has size $\sim A_1$, the second R factor has size $\sim A_2$, and the third R factor has size $\sim A_3$. For simplicity, suppose that $A_1 = A_2 = A_3 = A$, which we expect to be the critical case, and focus on the value of A that dominates the integral. If $A \approx |W|^{1/2}$, then we get the bound corresponding to minimal energy. If A is

larger, then $|R(v)| \sim A$ for only a small subset $U_A \subset \{v \asymp 1\}$ by Lemma 8.2. Our splitting of the summation and integration would mean we have the conditions $v_1 \in U_A$, $(m_1 v_1 + m_3)/(m_2 v_1) \in U_A$ and $(m_1 v_1 + m_3)/m_2 \in U_A$ and typically we would think it to be rare for these three sparse conditions to simultaneously hold. If the simple analysis in Proposition 8.1 was sharp, it would mean that for many $v \in U_A$ and many m_1, m_2, m_3 , we have $\frac{m_1 v + m_3}{m_2} \in U_A$. We will see that a small set U cannot be approximately invariant under this large set of affine transformations. In the next section, we will prove a precise estimate in this spirit, and then we will use it to give stronger bounds for S_3 when the energy is greater than $|W|^2$.

9. SUMMING OVER AFFINE TRANSFORMATIONS

Given $M > 0$ and a compactly supported smooth function f , we define

$$J(f) := \sup_{0 < M_1, M_2, M_3 < M} \int \left(\sum_{|m_1| \sim M_1, m_2 \sim M_2, |m_3| \ll M_3} f\left(\frac{m_1 u + m_3}{m_2}\right) \right)^2 du,$$

which is an average of sums of affine transformations of f . The aim of this section is to establish the following general bound for $J(f)$.

Proposition 9.1 (Equidistribution over affine transformations). *Suppose that $f(u)$ is non-negative and supported on $u \asymp 1$ and that $|\widehat{f}(\xi)| \lesssim_j (T/|\xi|)^j \sup_u |f(u)|$ for all $j \in \mathbb{Z}_{\geq 0}$. Then*

$$J(f) \lesssim M^6 \left(\int f(u) du \right)^2 + M^4 \int f(u)^2 du.$$

To avoid any possible confusion over the dependencies of the implied constants in the statement above, we emphasize that this is saying that if $f : \mathbb{R} \rightarrow [0, \infty)$ and $T \in \mathbb{R}$ have the property that there is a constant $C > 0$ such that f is supported on $[1/C, C]$ and the property that for every $\epsilon > 0$ and $j \in \mathbb{Z}_{\geq 0}$ there is a constant $c(j, \epsilon)$ such that $|\widehat{f}(\xi)| \leq c(j, \epsilon) T^\epsilon (T/|\xi|)^j \|f\|_\infty$ (for all $\xi \in \mathbb{R}$), then we can conclude that for every $\delta > 0$ there is a constant $c'(\delta) > 0$, depending only on C and all the constants $c(j, \epsilon)$, such that

$$J(f) \leq c'(\delta) T^\delta (M^6 \|f\|_1^2 + M^4 \|f\|_2^2).$$

We emphasize that in this section all implied constants from our \ll and \lesssim notation may depend on the implied constants C and $c(j, \epsilon)$ from the assumptions on f .

Remark. *By a simple Cauchy-Schwarz argument, we can bound the left hand side by $M^6 \int |f(u)|^2 du$. This bound is tight if $f(u)$ is a smooth bump on $u \sim 1$. The Proposition improves on this bound when $f(u)$ is sparse. To get a sense of what it means, it is good to imagine the example that $f = \chi_U$ where U is a union of*

(smoothed) $1/T$ -intervals. The proposition says that if $|U|$ is much smaller than 1, then the sets $\{\frac{m_1 U + m_3}{m_2}\}_{m_1, m_2, m_3 \sim M}$ cannot overlap too much.

There are two key examples which correspond to the two terms on the right-hand side. If U is a random subset, then the left-hand side is comparable to $M^6 (\int f(u) du)^2$. If U is the $1/T$ -neighborhood of the set of rational numbers r/s with $r, s \sim B$ (for some $B < T^{1/2}$), then the left-hand side is comparable to $M^4 \int |f(u)|^2 du$. Indeed, if $B > M$ and $u = r/s + O(1/T)$ where $r, s \sim B$ and $s = de$ with $d \sim M$, then we see from restricting to the terms with $m_1 = d$ that

$$\sum_{m_1, m_2, m_3 \sim M} f\left(\frac{m_1 u + m_2}{m_3}\right) \geq \sum_{m_2, m_3 \sim M} f\left(\frac{r + em_2}{em_3} + O(|u - r/s|)\right) \gg M^2$$

whenever $|u - r/s| \leq c/T$ for a sufficiently small constant c , since $(r + em_2)/(em_3)$ is also a rational with numerator and denominator of size B . Thus the integrand of $J(f)$ is of size M^4 on a set of measure $\gtrsim B^2/T \asymp \|f\|_2^2$. Similarly, if $B \leq M$ and $u = r/s + O(1/T)$ we can restrict the summand to $s|m_3$ and $\gcd(m_1, rm_3/s + m_2) \sim M/B$ and achieve the same bound.

The following lemma is the main technical result used to prove Proposition 9.1, which is based on a fairly long Fourier analytic argument.

Lemma 9.2 (Iterative bound for $J(f)$). *Let f be as in Proposition 9.1. Then there is a bump function $\psi(x)$ supported on $|x| \lesssim 1$ such that*

$$J(f) \lesssim M^6 \left(\int f(u) du \right)^2 + \left(M^4 \int f(u)^2 du \right)^{1/2} J(\tilde{f})^{1/2},$$

where \tilde{f} is defined in terms of $\psi(x)$ by

$$\tilde{f}(u) := T \int \psi(T(u - u')) f(u') du'.$$

As above, the bump function $\psi(x)$ may depend on the implied constants in the assumptions on f . If f satisfies $\widehat{f}(\xi) \leq c(j, \epsilon) T^\epsilon (T/|\xi|)^j \sup_u f(u)$ and is supported on $[1/C, C]$, then Lemma 9.2 claims that there is a non-negative 1-bounded function $\psi(x)$ such that for all $\delta > 0$ and $j \in \mathbb{Z}_{\geq 0}$ there are constants $c'(\delta)$ and $c'(j, \delta)$ (depending only on C and the constants $c(j, \epsilon)$) such that $\psi(x)$ is supported on $|x| \leq c'(\delta) T^\delta$, $|\psi^{(j)}(\xi)| \leq c'(j, \delta) T^\delta$ and such that

$$J(f) \leq c'(\delta) T^\delta (M^6 \|f\|_1^2 + M^2 \|f\|_2 J(\tilde{f})^{1/2}).$$

Before we prove Lemma 9.2, we first show how to deduce Proposition 9.1 from it.

Proof of Proposition 9.1 assuming Lemma 9.2. We wish to show that for any $\epsilon > 0$ there is a $C(\epsilon) > 0$ such that

$$(9.1) \quad J(f) \leq C(\epsilon)T^\epsilon \left(M^6 \left(\int f(u)du \right)^2 + M^4 \int f(u)^2 du \right),$$

for any function f satisfying the assumptions of Proposition 9.1 (with the constant $C(\epsilon)$ depending only on ϵ and the implied constants of the assumptions on f).

We wish to prove this by downwards induction on ϵ . As a base case, the result clearly holds for $\epsilon = 100$. By induction, we can assume that (9.1) holds with $3\epsilon/2$ in place of ϵ and seek to establish (9.1). We first apply Lemma 9.2 to give

$$J(f) \lesssim M^6 \left(\int f(u)du \right)^2 + \left(M^4 \int f(u)^2 du \right)^{1/2} J(\tilde{f})^{1/2}.$$

Since $f(u)$ is supported on $u \asymp 1$ and $\psi(x)$ is supported on $x \lesssim 1$, we see that $\tilde{f}(u)$ is also supported on $u \asymp 1$. Moreover, \tilde{f} has Fourier transform $\widehat{\tilde{f}}(\xi) \widehat{\psi}(\xi/T)$, so also satisfies the rapid decay assumption of Proposition 9.1. Thus \tilde{f} satisfies the assumptions of Proposition 9.1 (with implied constants depending only on the implied constants from the assumptions on f) and so we may apply the induction hypothesis to $J(\tilde{f})$, which shows

$$J(\tilde{f}) \ll_\epsilon T^{3\epsilon/2} \left(M^6 \left(\int \tilde{f}(u)du \right)^2 + M^4 \int \tilde{f}(u)^2 du \right).$$

Since \tilde{f} is a smoothed version of f , we can bound $\int \tilde{f}(u)du \lesssim \int f(u)du$ and $\int \tilde{f}(u)^2 du \lesssim \int f(u)^2 du$. Substituting these bounds back into our bound for $J(f)$, we find

$$J(f) \lesssim_\epsilon T^{3\epsilon/4} \left(M^6 \left(\int f(u)du \right)^2 + M^4 \int f(u)^2 du \right).$$

Thus (9.1) holds for $C(\epsilon)$ sufficiently large, which completes the induction. \square

We now return to the proof of Lemma 9.2.

Proof of Lemma 9.2. The most interesting situation is when $M_1 = M_2 = M_3 = M$. We encourage the reader to keep this case in mind on first reading.

By rescaling we may assume that $\sup_u f(u) \asymp 1$. Crudely, the rapid decay of \widehat{f} implies that $f'(u) \lesssim T^2$, so $\int f(u)du \gtrsim 1/T^2$ and

$$\begin{aligned} \sum_{m_3 \in \mathbb{Z}} f\left(\frac{m_1 v + m_3}{m_2}\right) &= \sum_{\substack{m_3 \\ |m_1 v + m_3| < m_2}} \left(\int_{m_3}^{m_3+1} f\left(\frac{m_1 v + u}{m_2}\right) du + O\left(\frac{\|f'\|_\infty}{m_2}\right) \right) \\ &\lesssim M \int f(u)du + T^2, \end{aligned}$$

which implies the result if $M > T^4$. Thus we may assume that $M \leq T^4$. We may assume that all the implied constants in the assumptions on f are sufficiently small compared with T since otherwise the result is trivial. Moreover, since $\int f(u)du \gtrsim T^{-2}$ and $M \leq T^4$, any $O(T^{-100})$ additive factors are negligible.

We let $\psi_1(x)$ be a smooth bump supported on $|x| \ll 1$ so that $\psi_1(m_3/M_3)$ majorizes the summation condition $|m_3| \ll M_3$. Thus we can bound the inner sum in $J(f)$ by

$$g(u) := \sum_{|m_1| \sim M_1, m_2 \sim M_2} \sum_{m_3} \psi_1\left(\frac{m_3}{M_3}\right) f\left(\frac{m_1 u + m_3}{m_2}\right).$$

Squaring and integrating over u , and then applying Plancherel gives (for the choice of M_1, M_2, M_3 achieving the supremum)

$$(9.2) \quad J(f) \leq \int |g(u)|^2 du = \int |\widehat{g}(\xi)|^2 d\xi.$$

We wish to estimate $\widehat{g}(\xi)$. We have

$$\widehat{g}(\xi) = \sum_{|m_1| \sim M_1, m_2 \sim M_2} \int \sum_{m_3} \psi_1\left(\frac{m_3}{M_3}\right) f\left(\frac{m_1 u + m_3}{m_2}\right) e(-\xi u) du.$$

We do a change of variables: $\tilde{u} = u + \frac{m_3}{m_1}$, so that $f\left(\frac{m_1 u + m_3}{m_2}\right) = f\left(\frac{m_1 \tilde{u}}{m_2}\right)$. In the new variables, we get

$$\widehat{g}(\xi) = \sum_{|m_1| \sim M_1, m_2 \sim M_2} \left(\int f\left(\frac{m_1 \tilde{u}}{m_2}\right) e(-\xi \tilde{u}) d\tilde{u} \right) \left(\sum_{m_3 \in \mathbb{Z}} \psi_1\left(\frac{m_3}{M_3}\right) e\left(\frac{m_3}{m_1} \xi\right) \right).$$

The integral in parentheses is $\frac{m_2}{m_1} \widehat{f}\left(\frac{m_2}{m_1} \xi\right)$. The first key point in our analysis is that we can explicitly do the last sum by Poisson summation. It is equal to $M_3 \sum_{\ell} \widehat{\psi}_1\left(M_3\left(\ell - \frac{\xi}{m_1}\right)\right)$. So all together we have

$$\widehat{g}(\xi) = \sum_{|m_1| \sim M_1} M_3 \sum_{\ell} \widehat{\psi}_1\left(M_3\left(\ell - \frac{\xi}{m_1}\right)\right) \sum_{m_2 \sim M_2} \frac{m_2}{m_1} \widehat{f}\left(\frac{m_2}{m_1} \xi\right).$$

Since $\widehat{\psi}_1$ is rapidly decaying, $\widehat{\psi}_1\left(M_3\left(\ell - \frac{\xi}{m_1}\right)\right)$ is negligible unless $|\xi - \ell m_1| \lesssim \frac{M_1}{M_3}$. Therefore, we have

$$(9.3) \quad |\widehat{g}(\xi)| \leq \sum_{|m_1| \sim M_1} \sum_{\ell: |\xi - m_1 \ell| \lesssim \frac{M_1}{M_3}} M_3 \left| \sum_{m_2 \sim M_2} \frac{m_2}{m_1} \widehat{f}\left(\frac{m_2}{m_1} \xi\right) \right| + O(T^{-100}).$$

We have to estimate $\int_{\mathbb{R}} |\widehat{g}(\xi)|^2 d\xi$. For a small $\eta > 0$, we break up the domain of integration into the region $|\xi| \leq T^\eta M_1/M_3$, the region $T^\eta M_1/M_3 < |\xi| \leq T^6$ and the remainder:

$$(9.4) \quad \int_{\mathbb{R}} |\widehat{g}(\xi)|^2 d\xi = \underbrace{\int_{|\xi| \leq T^\eta \frac{M_1}{M_3}} |\widehat{g}(\xi)|^2 d\xi}_I + \underbrace{\int_{T^\eta \frac{M_1}{M_3} < |\xi| \leq T^6} |\widehat{g}(\xi)|^2 d\xi}_{II} + \underbrace{\int_{T^6 < |\xi|} |\widehat{g}(\xi)|^2 d\xi}_{III}.$$

If $|\xi| > T^6$, then since $\widehat{f}(\xi)$ is rapidly decaying for $|\xi| > T$ and $M \leq T^4$, we see that $\widehat{f}(m_2\xi/m_1)$ is negligible and $|\widehat{g}(\xi)| \ll T^{-100}|\xi|^{-2}$. Thus

$$(9.5) \quad III = O(T^{-100}).$$

If $|\xi| \leq T^\eta M_1/M_3$, then in (9.3), the only terms that contribute have $|\ell| \ll T^\eta/M_3 \ll T^\eta$. Thus, by Cauchy-Schwarz we obtain from (9.3)

$$\begin{aligned} |\widehat{g}(\xi)|^2 &\ll T^{2\eta} M_1 \sum_{|m_1| \sim M_1} M_3^2 \left| \sum_{m_2 \sim M_2} \frac{m_2}{m_1} \widehat{f}\left(\frac{m_2}{m_1}\xi\right) \right|^2 + O(T^{-100}) \\ &\ll T^{2\eta} M_2^4 M_3^2 \sup_{\xi} |\widehat{f}(\xi)|^2. \end{aligned}$$

We absorbed the $O(T^{-100})$ error term in our bound since $\widehat{f}(0) \gtrsim T^{-2}$. Therefore

$$(9.6) \quad \begin{aligned} I &= \int_{|\xi| \leq T^\eta M_1/M_3} |\widehat{g}(\xi)|^2 d\xi \ll T^{3\eta} M_1 M_2^4 M_3 \sup_{\xi} |\widehat{f}(\xi)|^2 \\ &\ll T^{3\eta} M^6 \left(\int f(u) du \right)^2. \end{aligned}$$

Now suppose that $T^\eta M_1/M_3 < |\xi| \leq T^6$. We return to (9.3) and consider the number of terms in the outer double sum. Let $s = m_1 \ell$. In this range, s must be a non-zero integer in the $\lesssim M_1/M_3$ neighborhood of ξ as soon as T is sufficiently large in terms of η . The number of such integers s is $\lesssim 1 + M_1/M_3$. Since $|\xi| \leq T^6$ and s is non-zero, each such integer s has $\lesssim 1$ factorizations as $s = m_1 \ell$. All together the number of terms in the outer double sum is $\lesssim 1 + M_1/M_3$. Therefore, we can use Cauchy-Schwarz in (9.3) to bound $|\widehat{g}(\xi)|^2$ by

$$\lesssim_{\eta} \left(1 + \frac{M_1}{M_3}\right) \sum_{|m_1| \sim M_1} \sum_{\substack{\ell \\ |\xi - \ell m_1| \lesssim M_1/M_3}} M_3^2 \left| \sum_{m_2 \sim M_2} \frac{m_2}{m_1} \widehat{f}\left(\frac{m_2}{m_1}\xi\right) \right|^2 + O(T^{-200}).$$

Therefore the term II is bounded by

$$\lesssim_{\eta} (M_1 M_3 + M_3^2) \sum_{|m_1| \sim M_1} \sum_{\ell} \int_{|\xi - \ell m_1| \lesssim M_1/M_3} \left| \sum_{m_2 \sim M_2} \frac{m_2}{m_1} \widehat{f}\left(\frac{m_2}{m_1} \xi\right) \right|^2 d\xi + O(T^{-100}).$$

Morally this integral does not depend on m_1 , and we can make this precise by changing variables. For each m_1, ℓ , we write $\xi = \ell m_1 + \frac{m_2}{M_3} \tau$ and do a change of variables to get (extending the range of integration slightly for an upper bound so we have a range independent of m_1)

$$II \lesssim_{\eta} (M_1 + M_3) \sum_{\ell} \int_{|\tau| \lesssim 1} \left| \sum_{m_2 \sim M_2} m_2 \widehat{f}\left(\ell m_2 + \frac{m_2}{M_3} \tau\right) \right|^2 d\tau + O(T^{-100}).$$

Since $\widehat{f}(\xi) = O(T^{-200})$ unless $|\xi| \lesssim T$, we can restrict the sum over ℓ to the range $|\ell| \lesssim T/M_2$ at the cost of a negligible error. We introduce a bump $\psi_2(x)$ supported on $|x| \lesssim 1$ so that $\psi_2(M_2 \ell/T)$ majorizes this summation condition, and bound the last expression by

$$\lesssim_{\eta} (M_1 + M_3) \Sigma_{II} + O(T^{-100}),$$

where

$$\Sigma_{II} := \sum_{\ell} \psi_2\left(\frac{M_2 \ell}{T}\right) \int_{|\tau| \lesssim 1} \left| \sum_{m_2 \sim M_2} m_2 \widehat{f}\left(\ell m_2 + \frac{m_2}{M_1} \tau\right) \right|^2 d\tau.$$

We write out \widehat{f} as an integral, expand out the square and bring the summation over ℓ and integration over τ on the inside to get

$$(9.7) \quad \Sigma_{II} = \int \int \sum_{m_2, m'_2 \sim M_2} m_2 m'_2 f(u) f(u') Z_1 Z_2 du' du,$$

where $Z_1 = Z_1(m_2, m'_2, u, u')$ and $Z_2 = Z_2(m_2, m'_2, u, u')$ are given by

$$\begin{aligned} Z_1 &:= \int_{|\tau| \lesssim 1} e\left(\tau \left(\frac{m'_2}{M_1} u' - \frac{m_2}{M_1} u\right)\right) d\tau, \\ Z_2 &:= \sum_{\ell} \psi_2\left(\frac{M_2 \ell}{T}\right) e\left(\ell \left(m'_2 u' - m_2 u\right)\right). \end{aligned}$$

Trivially we have $|Z_1| \lesssim 1$. By Poisson summation, and the rapid decay of $\widehat{\psi}_2$, we have

$$\begin{aligned}
Z_2 &= \frac{T}{M_2} \sum_j \widehat{\psi}_2\left(\frac{j - m'_2 u' + m_2 u}{M_2/T}\right) \\
&= \frac{T}{M_2} \sum_{|j - m'_2 u' + m_2 u| \lesssim M_2/T} \widehat{\psi}_2\left(\frac{j - m'_2 u' + m_2 u}{M_2/T}\right) + O(T^{-100}).
\end{aligned}$$

Substituting these back into our expression (9.7) for Σ_{II} and recalling $M \leq T^4$, we find

$$\Sigma_{II} \lesssim M_2 \int f(u) \sum_{m_2, m'_2 \sim M_2} \sum_j \int_{|u' - \frac{m_2 u + j}{m'_2}| \lesssim \frac{1}{T}} T f(u') du' du + O(T^{-80}).$$

This gives

$$II \lesssim_{\eta} M_2 (M_1 + M_3) \int f(u) \sum_{m_2, m'_2 \sim M_2} \sum_j \int_{|u' - \frac{m_2 u + j}{m'_2}| \lesssim \frac{1}{T}} T f(u') du' du + O(T^{-70}).$$

Since $\widehat{f}(\xi)$ rapidly decays for $|\xi| > T$, f is morally almost constant on intervals of length $1/T$. We let $\psi(x)$ be a smooth bump function supported on $x \lesssim 1$, and then define \widetilde{f} to be a slightly smoothed version of f given by

$$\widetilde{f}(u) := \int T \psi(T(u - u')) f(u') du',$$

as in the statement of Lemma 9.2

We can choose ψ such that the integral over u' in the bound for II above is bounded by $\widetilde{f}(\frac{m_2 u + j}{m'_2})$ and $\widetilde{f}(u)$ is supported on $u \asymp 1$ (recall that the range of support of f is assumed to be sufficiently small in terms of T). Thus we find (recalling that f is supported on $u \asymp 1$ and $M_1, M_2, M_3 \leq M$)

$$II \lesssim_{\eta} M^2 \int_{u \asymp 1} f(u) \sum_{m_2, m'_2 \sim M_2} \sum_j \widetilde{f}\left(\frac{m_2 u + j}{m'_2}\right) du + O(T^{-70}).$$

Now we apply Cauchy-Schwarz to get

$$(9.8) \quad II \lesssim_{\eta} \left[M^4 \int f(u)^2 du \right]^{1/2} \left[\int_{u \asymp 1} \left(\sum_{m_2, m'_2 \sim M_2, j \in \mathbb{Z}} \widetilde{f}\left(\frac{m_2 u + j}{m'_2}\right) \right)^2 du \right]^{1/2}.$$

(The $O(T^{-70})$ term it is clearly majorized by the above expression and so may be absorbed into the implied constant.) Since $u \asymp 1$ and $f(x)$ is supported on $x \asymp 1$, we may restrict the summation over j to $j \ll M_2$. Thus we see that the second term in square brackets is bounded by $J(\widetilde{f})$. Putting together (9.2), (9.4), (9.5),

(9.6) and (9.8) we find that for any $\eta > 0$, provided T is sufficiently large in terms of η , we have

$$J(f) \lesssim_{\eta} T^{3\eta} M^6 \left(\int f(u) du \right)^2 + \left(M^4 \int f(u)^2 du \right)^{1/2} J(\tilde{f})^{1/2}.$$

This now gives the result on letting $\eta \rightarrow 0$ sufficiently slowly with T . \square

10. FURTHER BOUNDS FOR S_3

In this section, we use our bounds for sums over affine transformations to improve our bound for S_3 . We will get the following estimate.

Proposition 10.1 (Refined S_3 bound). *If W is a T^ϵ -separated set contained in an interval of length T , then*

$$(10.1) \quad S_3 \lesssim_{\epsilon} T^2 |W|^{3/2} + TN |W|^{1/2} E(W)^{1/2}.$$

Remark. Comparing with Proposition 8.1, when $E(W) \approx |W|^2$ both propositions give $S_3 \lesssim T^2 |W|^{3/2}$. When $E(W)$ is large, Proposition 10.1 is better by a factor of N/T . If we look at the critical case when $T = N^{5/4}$ and $\sigma = 3/4$, then this estimate (together with our bounds for S_2) gives an improvement to the basic orthogonality estimate (1.6) when $E(W)$ is significantly below $|W|^3$. In other words, this Proposition gives an improvement unless the energy is essentially maximal. When $E(W)$ is very large, however, we will be able to get good estimates in the next section using Theorem 1.6.

Proof. Recall from Proposition 7.2 that $S_3 \lesssim S_{3,0} + O(T^{-100})$, where

$$S_{3,0} := \frac{N^2}{M} \int_{v_1 \asymp 1} |R(v_1)| \sum_{\substack{|m_1| \sim M_1 \\ |m_2|, |m_3| \asymp M}} \left| \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2 v_1}\right) \tilde{R}\left(\frac{m_1 v_1 + m_3}{m_2}\right) \right| dv_1$$

for some $1 \leq M_1 \leq M \lesssim T/N$ and some suitable smooth bumps $\tilde{\psi}_1, \tilde{\psi}_2$ which appear in the definition (7.5) of $\tilde{R} = \tilde{R}_{\tilde{\psi}_1, \tilde{\psi}_2, M}$. By Cauchy-Schwarz we have

$$S_{3,0} \leq \frac{N^2}{M} S_{3,1}^{1/2} S_{3,2}^{1/2},$$

where (using Lemma 8.2 to bound $S_{3,1}$)

$$S_{3,1} := \int_{v \asymp 1} |R(v)|^2 dv \ll_{\epsilon} |W|,$$

$$S_{3,2} := \int_{v \asymp 1} \left(\sum_{\substack{|m_1| \sim M_1, \\ |m_2|, |m_3| \asymp M}} \left| \tilde{R}\left(\frac{m_3 + m_1 v}{m_2 v}\right) \tilde{R}\left(\frac{m_3 + m_1 v}{m_2}\right) \right| \right)^2 dv.$$

By Cauchy-Schwarz again, we have that

$$S_{3,2} \ll S_{3,3}^{1/2} S_{3,4}^{1/2},$$

where

$$S_{3,3} := \int_{v \asymp 1} \left(\sum_{|m_1| \sim M_1, |m_2|, |m_3| \asymp M} \left| \tilde{R} \left(\frac{m_3 + m_1 v}{m_2 v} \right) \right|^2 \right)^2 dv,$$

$$S_{3,4} := \int_{v \asymp 1} \left(\sum_{|m_1| \sim M_1, |m_2|, |m_3| \asymp M} \left| \tilde{R} \left(\frac{m_3 + m_1 v}{m_2} \right) \right|^2 \right)^2 dv.$$

We bound $S_{3,3}$ and $S_{3,4}$ using Proposition 9.1. To bound $S_{3,4}$, we use $f(v) = \psi_1(v) |\tilde{R}(v)|^2$, where $\psi_1(v)$ is a smooth bump supported on $v \asymp 1$ taking a maximal value of 1 at $v = 1$ which majorizes the range of integration. To control $S_{3,3}$, we make a change of variables $u = 1/v$ and rewrite $S_{3,3}$ as

$$S_{3,3} \ll \int_{u \asymp 1} \left(\sum_{|m_1| \sim M_1, |m_2|, |m_3| \asymp M} \left| \tilde{R} \left(\frac{m_3 u + m_1}{m_2} \right) \right|^2 \right)^2 du.$$

Then we use $f(u) = \psi_1(u) |\tilde{R}(u)|^2$ again. Lemma 8.4 shows that $\hat{f}(\xi) \lesssim_j \|f\|_\infty T^j / |\xi|^j$ (taking ‘ ψ_2 ’ to be $\tilde{\psi}_2$, ‘ ψ_3 ’ to be $\tilde{\psi}_1$ and ‘ B ’ to be MN), and so f satisfies the Fourier decay conditions of Proposition 9.1. Moreover, f is clearly has support on $u \asymp 1$ from the support of ψ_1 . Thus Proposition 9.1 gives the bounds

$$S_{3,3}, S_{3,4} \lesssim M^6 \left(\int_{v \asymp 1} |\tilde{R}(v)|^2 dv \right)^2 + M^4 \int_{v \asymp 1} |\tilde{R}(v)|^4 dv.$$

Applying Lemmas 8.2 and 8.3, we get

$$S_{3,3}, S_{3,4} \lesssim_\epsilon M^6 |W|^2 + M^4 E(W).$$

The same bound holds for $S_{3,2}$ since $S_{3,2} \leq S_{3,3}^{1/2} S_{3,4}^{1/2}$. Then we get

$$\begin{aligned} S_{3,0} &\lesssim \frac{N^2}{M} S_{3,1}^{1/2} S_{3,2}^{1/2} \lesssim_\epsilon \frac{N^2}{M} |W|^{1/2} (M^6 |W|^2 + M^4 E(W))^{1/2} \\ &= N^2 M^2 |W|^{3/2} + N^2 M |W|^{1/2} E(W)^{1/2}. \end{aligned}$$

Since $M \lesssim T/N$ and $S_3 \lesssim S_{3,0} + O(T^{-100})$, we get

$$S_3 \lesssim_\epsilon T^2 |W|^{3/2} + TN |W|^{1/2} E(W)^{1/2}. \quad \square$$

11. ENERGY BOUND

In this section, we prove bounds related to the energy of W , which show that a Dirichlet polynomial cannot be too large on a set of large energy. These bounds ultimately rely on Heath-Brown's bound Theorem 1.6, and are closely related to (and refine) arguments in [HB2]. Recall that in equation (1.7), we defined the energy of a finite set $W \subset \mathbb{R}$ by

$$E(W) := \#\{t_1, t_2, t_3, t_4 \in W : |t_1 + t_2 - t_3 - t_4| \leq 1\}.$$

We will prove two bounds about the behavior of Dirichlet polynomials on sets of high energy. The first bound is Lemma 1.7. We recall the statement here.

Lemma. *Let $N \in [T^{2/3}, T]$, $\sigma > 1/2$ and $D(t) = \sum_{n \sim N} b_n n^{it}$ with $|b_n| \leq 1$. Suppose $W \subset [0, T]$ is a 1-separated set such that $|D(t)| > N^\sigma$ for $t \in W$. Then*

$$E(W) \leq |W|^3 N^{1-2\sigma+o(1)} + |W|^2 N^{2-2\sigma+o(1)}.$$

Combining Lemma 1.7 with our earlier results is enough to ultimately get an improvement on (1.1) in the key scenario $N = T^{4/5}$, $|W| = T^{3/5}$. The second bound is a little more complicated, but it leads to stronger estimates in our applications.

Proposition 11.1 (Bound for energy). *Suppose that $D(t) = \sum_{n \sim N} b_n n^{it}$ with $|b_n| \leq 1$. Suppose that W is a 1-separated set contained in an interval of length T , and that $|D(t)| \geq N^\sigma$ for $t \in W$. If $T^{3/4} \leq N \leq T$, then*

$$(11.1) \quad E(W) \lesssim |W| N^{4-4\sigma} + |W|^{21/8} T^{1/4} N^{1-2\sigma} + |W|^3 N^{1-2\sigma}.$$

Since the algebra is a little messy, we take a moment to process the bounds. For one thing, if $E(W)$ is very large, we get very strong bounds on σ . For instance, if $E(W) \approx |W|^3$, and if $N^{2/3} \leq T \leq N$, then Lemma 1.7 gives a sharp estimate: either $N^\sigma \lesssim N^{1/2}$ or $|W| N^{2\sigma} \lesssim N^2$. More important for our application is when we get an improvement on the basic orthogonality bound $|W| N^{2\sigma} \ll TN$. The full equations are a little messy, but if we plug in the key scenario $\sigma = 3/4$ and $N = T^{4/5}$, then Lemma 1.7 gives an improvement on the bound $|W| \lesssim T^{3/5}$ when $E(W) \geq |W|^{\frac{8}{3}+\epsilon}$ and Proposition 11.1 gives an improvement when $E(W) \geq |W|^{\frac{19}{8}+\epsilon}$ (for some $\epsilon > 0$).

If $|W| \approx TN^{1-2\sigma}$, then the bound in Proposition 11.1 would be $(N/T)^2 |W|^3 + (|W|^{5/8} T^{-6/8}) |W|^3 + |W|^4/T$. The first term will be the most important for us, and generally sets the limitations on our bounds. The second term will be negligible in practice (and could be improved with a bit more effort). The final term corresponds to the additive energy of a random set in an interval of length T .

The condition $N \geq T^{3/4}$ is used to simplify intermediate terms occurring in the proof of Proposition 11.1 and this range could be improved with some extra effort. For the purposes of Theorem 1.1 the key situation is when $N = T^{5/6}$ as in Proposition 3.1.

An immediate consequence of Proposition 11.1 is a good bound for the key term S_3 by substituting the bound of Proposition 11.1 (applied to $D_N(t)$) into Proposition 10.1.

Proposition 11.2 (S_3 Bound). *Let $N \geq T^{3/4}$. Then we have*

$$S_3 \lesssim T^2 |W|^{3/2} + T |W| N^{3-2\sigma} + T |W|^2 N^{3/2-\sigma} + T^{9/8} |W|^{29/16} N^{3/2-\sigma}.$$

Now we turn to the proof of Proposition 11.1.

Suppose that $D(t) = \sum_{n \sim N} b_n n^{it}$ and $|D(t)| \geq N^\sigma$ on W . Morally, we also have $|D(t)| \gtrsim N^\sigma$ if the distance from t to W is $\ll 1$. In particular, if $|t_1 + t_2 - t_3 - t_4| \leq 1$, then morally $|D(t_1 + t_2 - t_3)| \gtrsim N^\sigma$. Therefore we should expect that

$$E(W) \lesssim \sum_{t_1, t_2, t_3 \in W} \frac{|D(t_1 + t_2 - t_3)|^2}{N^{2\sigma}} = N^{-2\sigma} \sum_{n_1, n_2 \sim N} b_{n_1} \overline{b_{n_2}} R\left(\frac{n_1}{n_2}\right)^2 R\left(\frac{n_2}{n_1}\right).$$

We can make this heuristic argument literally true by working with a slightly smoothed version of D .

Lemma 11.3 (Dirichlet polynomials do not vary too fast). *Let $D(t)$ be as in Proposition 11.1. Then we have*

$$|D(t)| \ll \int_{|u-t| \lesssim 1} |D(u)| du + O(T^{-100}).$$

Proof. Let $\psi(x)$ be a smooth bump which is supported on $|2\pi x - \log N| \ll 1$ and is equal to 1 on $[(2\pi)^{-1} \log N, (2\pi)^{-1} \log 2N]$, and extend b_n to all $n \in \mathbb{Z}$ by setting $b_n = 0$ if it is not the case that $n \sim N$. Then we have

$$D(t) = \sum_{n \sim N} b_n n^{it} = \sum_n b_n n^{it} \psi\left(\frac{\log n}{2\pi}\right) = \int \widehat{\psi}(\xi) D(t - \xi) d\xi.$$

By the rapid decay of $\widehat{\psi}$ we may restrict to $|\xi| \lesssim 1$ at the cost of an error $O(T^{-100})$. \square

Lemma 11.4 (Energy controlled by discrete 3rd moment). *Let $D(t)$ and W be as in Proposition 11.1. Then we have*

$$E(W) \lesssim N^{-2\sigma} \sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^3.$$

Proof. Since $|D(t)| > N^\sigma$ for $t \in W$, we have

$$E(W) = \sum_{\substack{t_1, t_2, t_3, t_4 \in W \\ |t_1 + t_2 - t_3 - t_4| \leq 1}} 1 \leq N^{-2\sigma} \sum_{\substack{t_1, t_2, t_3, t_4 \in W \\ |t_1 + t_2 - t_3 - t_4| \leq 1}} |D(t_4)|^2.$$

By Lemma 11.3 and Cauchy-Schwarz, we have for $|t_1 + t_2 - t_3 - t_4| \leq 1$

$$|D(t_4)|^2 \ll \int_{|u-t_4| \lesssim 1} |D(u)|^2 du \ll \int_{|u-t_1+t_2-t_3| \lesssim 1} |D(u)|^2 du.$$

Since W is 1-separated, given t_1, t_2, t_3 there are $\ll 1$ choices of $t_4 \in W$ such that $|t_1 + t_2 - t_3 - t_4| \leq 1$. Thus we see that

$$\begin{aligned} E(W) &\ll N^{-2\sigma} \sum_{t_1, t_2, t_3 \in W} \int_{s \lesssim 1} |D(t_1 + t_2 - t_3 - s)|^2 ds \\ &= N^{-2\sigma} \sum_{n_1, n_2 \sim N} b_{n_1} \overline{b_{n_2}} \int_{s \lesssim 1} \left(\frac{n_2}{n_1}\right)^{is} R\left(\frac{n_1}{n_2}\right)^2 R\left(\frac{n_2}{n_1}\right) ds \\ &\lesssim N^{-2\sigma} \sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^3. \quad \square \end{aligned}$$

Remark. In the final line of the proof of Lemma 11.4 we made important use of the assumption $|b_n| \leq 1$. This is the only place in the paper where we make use of this ℓ^∞ bound rather than an ℓ^2 bound. If we only had a weaker bound on the coefficients available, we would ultimately require a version of Heath-Brown's Theorem with weaker assumptions on the coefficients b_n .

Next we note that Heath-Brown's theorem (Theorem 1.6) bounds $\sum_{n_1, n_2 \sim N} |R(\frac{n_1}{n_2})|^2$ fairly directly and, with a bit more work, can be used to bound $\sum_{n_1, n_2 \sim N} |R(\frac{n_1}{n_2})|^4$ too.

Lemma 11.5 (Discrete second moment). *For any $M \geq 1$,*

$$\sum_{n_1, n_2 \sim M} \left| R\left(\frac{n_1}{n_2}\right) \right|^2 \lesssim |W|^2 M + |W| M^2 + |W|^{5/4} T^{1/2} M.$$

Proof. We have that

$$\sum_{n_1, n_2 \sim M} \left| R\left(\frac{n_1}{n_2}\right) \right|^2 = \sum_{t_1, t_2 \in W} \left| \sum_{n \sim M} n^{i(t_1 - t_2)} \right|^2,$$

so by Theorem 1.6 this is

$$\lesssim |W|^2 M + |W| M^2 + |W|^{5/4} T^{1/2} M. \quad \square$$

Lemma 1.7 is now a quick consequence of our arguments so far.

Proof of Lemma 1.7. By Lemma 11.4 and the trivial bound $|R(x)| \leq |W|$ we have

$$E(W) \lesssim N^{-2\sigma} \sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \leq |W| N^{-2\sigma} \sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^2.$$

Lemma 11.5 (which is just Theorem 1.6) now shows for $N > T^{2/3}$ we have

$$E(W) \lesssim |W|^3 N^{1-2\sigma} + |W|^2 N^{2-2\sigma}. \quad \square$$

To do better we look at higher moments to avoid the potentially wasteful use of the trivial bound $|R(x)| \leq |W|$.

Lemma 11.6 (Discrete fourth moment). *For any $M \geq 1$,*

$$\sum_{n_1, n_2 \sim M} \left| R\left(\frac{n_1}{n_2}\right) \right|^4 \lesssim |W|^4 M + M^2 E(W) + E(W)^{3/4} |W| T^{1/2} M.$$

Proof. We split the sum in the R function according to the number of representations of u as approximately $t_1 - t_2$. Let $\lfloor x \rfloor$ denote the largest integer $\leq x$, and define

$$U_B := \left\{ u \in \mathbb{Z} : \#\{(t_1, t_2) \in W^2 : \lfloor t_1 - t_2 \rfloor = u\} \sim B \right\}.$$

Clearly U_B is empty if $B < 1/2$ or if $B > |W|$. Thus, using Cauchy-Schwarz

$$\begin{aligned} |R(x)|^4 &= \left| \sum_{t_1, t_2 \in W} x^{i(t_1 - t_2)} \right|^2 = \left| \sum_{B=2^j} \sum_{u \in U_B} \sum_{\substack{t_1, t_2 \in W \\ \lfloor t_1 - t_2 \rfloor = u}} x^{i(t_1 - t_2)} \right|^2 \\ &\lesssim \sum_{B=2^j \leq |W|} \left| \sum_{u \in U_B} \sum_{\substack{t_1, t_2 \in W \\ \lfloor t_1 - t_2 \rfloor = u}} x^{i(t_1 - t_2)} \right|^2. \end{aligned}$$

Taking $x = n_1/n_2$ and summing over $n_1, n_2 \sim M$ then gives

$$\begin{aligned} \sum_{n_1, n_2 \sim M} \left| R\left(\frac{n_1}{n_2}\right) \right|^4 &\lesssim \sup_{B \leq |W|} \sum_{n_1, n_2 \sim M} \left| \sum_{u \in U_B} \sum_{\substack{t_1, t_2 \in W \\ \lfloor t_1 - t_2 \rfloor = u}} \left(\frac{n_1}{n_2}\right)^{i(t_1 - t_2)} \right|^2 \\ &\leq \sup_{B \leq |W|} \sum_{u_1, u_2 \in U_B} \left(\sum_{\substack{t_1, t_3 \in W \\ \lfloor t_1 - t_3 \rfloor = u_1}} 1 \right) \left(\sum_{\substack{t_2, t_4 \in W \\ \lfloor t_2 - t_4 \rfloor = u_2}} 1 \right) \sup_{|s| \ll 1} \left| \sum_{n \sim M} n^{i(u_1 - u_2 + s)} \right|^2 \\ &\lesssim \sup_{B \leq |W|} B^2 \sum_{u_1, u_2 \in U_B} \sup_{|s| \ll 1} \left| \sum_{n \sim M} n^{i(u_1 - u_2 + s)} \right|^2. \end{aligned}$$

By using Lemma 11.3 to replace the supremum with an integral, and then applying Theorem 1.6, we find

$$\begin{aligned} \sum_{n_1, n_2 \sim M} \left| R\left(\frac{n_1}{n_2}\right) \right|^4 &\ll \sup_{B \leq |W|} B^2 \int_{t \lesssim 1} \sum_{u_1, u_2 \in U_B} \left| \sum_{n \sim M} n^{i(u_1 - u_2 + t)} \right|^2 dt \\ &\ll \sup_{B \leq |W|} B^2 \left(|U_B|^2 M + |U_B| M^2 + T^{1/2} |U_B|^{5/4} M \right). \end{aligned}$$

We have that $B|U_B| \leq |W|^2$ and $B^2|U_B| \leq E(W)$, so this gives

$$\sum_{n_1, n_2 \sim M} \left| R\left(\frac{n_1}{n_2}\right) \right|^4 \ll |W|^4 M + M^2 E(W) + E(W)^{3/4} |W| T^{1/2} M. \quad \square$$

To bound $\sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^3$, we could use Hölder:

$$\sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \leq \left(\sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^2 \right)^{1/2} \left(\sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^4 \right)^{1/2}$$

and then bound the two factors using Lemmas 11.5 and 11.6. However, this Hölder step is somewhat lossy. If n'_1/n'_2 is a rational number of small height, then the sum $\sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^p$ counts $|R(n'_1/n'_2)|^p$ many times – because there are many $n_1, n_2 \sim N$ with $n_1/n_2 = n'_1/n'_2$. The 4th moment tends to be dominated by n_1, n_2 with large $\gcd(n_1, n_2)$, but the 2nd moment tends to be dominated by n_1, n_2 with small $\gcd(n_1, n_2)$. Therefore, instead of doing Hölder immediately, we now split our argument according to the size of $\gcd(n_1, n_2)$.

Let $d = \gcd(n_1, n_2)$ and $n_1 = n'_1 d$, $n_2 = n'_2 d$ for some $n'_1, n'_2 \sim N/d$ with $\gcd(n'_1, n'_2) = 1$. Thus we have for any choice of parameter D (dropping the coprimality constraint when d is large)

$$(11.2) \quad E(W) \leq N^{-2\sigma} \sum_{d \leq D} \sum_{\substack{n'_1, n'_2 \sim N/d \\ \gcd(n'_1, n'_2) = 1}} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^3 + N^{-2\sigma} \sum_{d \geq D} \sum_{n'_1, n'_2 \sim N/d} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^3.$$

First we consider small d . When d is small enough, the distinct fractions n'_1/n'_2 are very well distributed and so it makes sense to compare our sum with $\int_{v \asymp 1} |R(v)|^3 dv$.

We recall that W is contained in an interval of length T . Morally, $|\widehat{W}(\tau)|$ is locally constant on intervals of length $1/T$. Since $R(v) = \widehat{W}(\log v / (-2\pi))$, we see that for $v \asymp 1$, $|R(v)|$ is morally locally constant at scale $1/T$. We make this precise in the following lemma:

Lemma 11.7. *For $v \asymp 1$,*

$$|R(v)| \ll T \int_{|v' - v| \lesssim 1/T} |R(v')| dv' + O(T^{-100}).$$

Proof. Since $v \asymp 1$, we can do a change of variables, $\tau = (-2\pi)^{-1} \log v$, and it suffices to prove that

$$|\widehat{W}(\tau)| \ll T \int_{|\tau' - \tau| \lesssim 1/T} |\widehat{W}(\tau')| d\tau' + O(T^{-100}).$$

We know that W is contained in an interval of length T ; call this $[T_0, T_0 + T]$. Let ψ be a smooth bump which is 1 on $[0, 1]$. Then we have

$$\widehat{W}(\tau) = \sum_{t \in W} e(-t\tau) = \sum_{t \in W} e(-t\tau) \psi\left(\frac{t - T_0}{T}\right) = \int \widehat{\psi}(\xi) \widehat{W}\left(\tau - \frac{\xi}{T}\right) e\left(\frac{-T_0\xi}{T}\right) d\xi.$$

By the rapid decay of $\widehat{\psi}$, we may restrict the integral to $\xi \lesssim 1$ at the cost of an $O(T^{-100})$ error term. Since $\widehat{\psi} \ll 1$ this then gives the result. \square

Lemma 11.8 (Small GCD terms). *We have*

$$\sum_{\substack{n_1, n_2 \sim N \\ \gcd(n_1, n_2) \leq D}} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \lesssim (DT + N^2) |W|^{1/2} E(W)^{1/2}.$$

Proof. Let $d = \gcd(n_1, n_2)$ and $n_1 = n'_1 d$, $n_2 = n'_2 d$ for some $n'_1, n'_2 \sim N/d$ with $\gcd(n'_1, n'_2) = 1$. By Lemma 11.7, we have

$$\sum_{\substack{n'_1, n'_2 \sim N/d \\ \gcd(n'_1, n'_2) = 1}} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^3 \ll T \int_{v \asymp 1} |R(v)|^3 \left(\sum_{\substack{n_1, n'_2 \sim N/d \\ \gcd(n'_1, n'_2) = 1 \\ |v - n'_1/n'_2| \lesssim 1/T}} 1 \right) dv.$$

Since the fractions n'_1/n'_2 are d^2/N^2 -separated, we have that the inner sum over n'_1, n'_2 on the right hand side is $\lesssim 1 + N^2/(d^2T)$. Thus we find

$$\begin{aligned} \sum_{d \leq D} \sum_{\substack{n'_1, n'_2 \sim N/d \\ \gcd(n'_1, n'_2) = 1}} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^3 &\lesssim \sum_{d \leq D} \left(T + \frac{N^2}{d^2}\right) \int_{v \asymp 1} |R(v)|^3 dv \\ &\lesssim \sum_{d \leq D} \left(T + \frac{N^2}{d^2}\right) \left(\int_{v \asymp 1} |R(v)|^2 dv\right)^{1/2} \left(\int_{v \asymp 1} |R(v)|^4 dv\right)^{1/2} \\ &\lesssim (DT + N^2) |W|^{1/2} E(W)^{1/2}. \end{aligned} \quad \square$$

We choose $D := N^2/T$, so this gives

$$(11.3) \quad \sum_{\substack{n_1, n_2 \sim N \\ \gcd(n_1, n_2) \leq D}} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \lesssim N^2 |W|^{1/2} E(W)^{1/2}.$$

Lemma 11.9 (Large GCD terms). *Let $D = N^2/T$ and $N \geq T^{3/4}$. Then we have*

$$\sum_{\substack{n_1, n_2 \sim N \\ \gcd(n_1, n_2) \geq D}} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \lesssim N|W|^3 + NT^{1/4}|W|^{21/8} + E(W)^{1/2}|W|^{1/2}N^2.$$

Proof. As in the previous lemma, we let $d = \gcd(n_1, n_2)$ and $n_1 = n'_1 d$, $n_2 = n'_2 d$. When d is large, we keep the discrete summation over n'_1, n'_2 and apply Cauchy-Schwarz directly, giving

$$\sum_{n'_1, n'_2 \sim N/d} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^3 \ll \left(\sum_{n'_1, n'_2 \sim N/d} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^2 \right)^{1/2} \left(\sum_{n'_1, n'_2 \sim N/d} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^4 \right)^{1/2}.$$

Now we can bound the factors on the right-hand side by Lemmas 11.5 and 11.6, with $M = N/d$. This gives

$$\begin{aligned} \sum_{n'_1, n'_2 \sim N/d} \left| R\left(\frac{n'_1}{n'_2}\right) \right|^3 &\lesssim \left(\frac{|W|N^2}{d^2} + \frac{|W|^2 N}{d} + \frac{|W|^{5/4} T^{1/2} N}{d} \right)^{1/2} \\ &\quad \times \left(\frac{N|W|^4}{d} + \frac{N^2 E(W)}{d^2} + \frac{E(W)^{3/4} |W| T^{1/2} N}{d} \right)^{1/2}. \end{aligned}$$

Summing over $d \geq D$, using Cauchy-Schwarz, and recalling that $D = N^2/T$ then gives

$$\begin{aligned} \sum_{\substack{n_1, n_2 \sim N \\ \gcd(n_1, n_2) \geq D}} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 &\lesssim \left(\frac{|W|N^2}{D} + |W|^2 N + |W|^{5/4} T^{1/2} N \right)^{1/2} \\ &\quad \times \left(N|W|^4 + \frac{N^2 E(W)}{D} + E(W)^{3/4} |W| T^{1/2} N \right)^{1/2}. \end{aligned}$$

Next we work on simplifying and organizing the algebra. Recall that we have $N \geq T^{3/4}$ and $D = N^2/T$. Therefore, we have $|W|^{5/4} T^{1/2} N \geq |W|T = |W|N^2/D$, and we can ignore the first term in the first factor. Thus the above expression is bounded by

$$(11.4) \quad \lesssim \left(|W|^2 N + |W|^{5/4} T^{1/2} N \right)^{1/2} \left(N|W|^4 + E(W)T + E(W)^{3/4} |W| T^{1/2} N \right)^{1/2}.$$

There are two main cases, depending on whether $|W| > T^{2/3}$ or not. If $|W| > T^{2/3}$ then $|W|^2 N > |W|^{5/4} T^{1/2} N$, and so the first factor is dominated by $|W|^2 N$. We turn to the second factor. If $|W| > T^{2/3}$, then $N|W|^4 > N|W|^{13/4} T^{1/2} \geq E(W)^{3/4} |W| T^{1/2} N$. Also, since $N \geq T^{3/4} > T^{1/3}$ and $E(W) \leq |W|^3$, $N|W|^4 >$

$|W|^3 T \geq E(W)T$. So the second factor is dominated by $N|W|^4$. Therefore, if $|W| > T^{2/3}$, (11.4) simplifies to

$$(11.5) \quad \sum_{\substack{n_1, n_2 \sim N \\ \gcd(n_1, n_2) \geq D}} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \lesssim N|W|^3.$$

Now suppose $|W| \leq T^{2/3}$. We see that $|W|^2 N \leq |W|^{5/4} T^{1/2} N$, so the first factor is dominated by $|W|^{5/4} T^{1/2} N$. Turning to the second factor, and recalling that $N \geq T^{3/4} > T^{1/2}$ and $E(W) \leq |W|^3$, we see that $E(W)^{3/4} |W| T^{1/2} N \geq E(W)T$. Thus, if $|W| \leq T^{2/3}$ we see that (11.4) simplifies to

$$(11.6) \quad \sum_{\substack{n_1, n_2 \sim N \\ \gcd(n_1, n_2) \geq D}} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \lesssim \left(|W|^{5/4} T^{1/2} N \right)^{1/2} \left(N|W|^4 + E(W)^{3/4} |W| T^{1/2} N \right)^{1/2} \\ \lesssim N T^{1/4} |W|^{21/8} + E(W)^{1/2} |W|^{1/2} N^2 \left(\frac{T^{1/2} |W|^{5/8}}{E(W)^{1/8} N} \right).$$

Since $E(W) \geq |W|^2$, $|W| \leq T^{2/3}$ and $N \geq T^{3/4}$, we see that $T^{1/2} |W|^{5/8} \leq E(W)^{1/8} N$, so the final term in (11.6) is $O(|W|^{1/2} E(W)^{1/2} N^2)$. Thus, combining (11.5) and (11.6), we find that provided $N \geq T^{3/4}$, regardless of the size of W , we have

$$\sum_{\substack{n_1, n_2 \sim N \\ \gcd(n_1, n_2) \geq D}} \left| R\left(\frac{n_1}{n_2}\right) \right|^3 \lesssim N|W|^3 + N T^{1/4} |W|^{21/8} + E(W)^{1/2} |W|^{1/2} N^2. \quad \square$$

Proof of Proposition 11.1. First we use Lemma 11.4 to give

$$E(W) \lesssim N^{-2\sigma} \sum_{n_1, n_2 \sim N} \left| R\left(\frac{n_1}{n_2}\right) \right|^3.$$

Splitting according to whether $\gcd(n_1, n_2) \leq D = N^2/T$ or not, we find by Lemma 11.8 and Lemma 11.9 that

$$E(W) \lesssim N^{-2\sigma} \left(N|W|^3 + N T^{1/4} |W|^{21/8} + E(W)^{1/2} |W|^{1/2} N^2 \right).$$

This rearranges to give

$$E(W) \lesssim |W| N^{4-4\sigma} + |W|^{21/8} T^{1/4} N^{1-2\sigma} + |W|^3 N^{1-2\sigma}. \quad \square$$

12. PROOF OF RESULTS ON LARGE VALUES OF DIRICHLET POLYNOMIALS

In this section we prove our main results on the large values of Dirichlet polynomials by assembling the tools in the previous sections.

Proof of Proposition 3.1. Suppose that $|D_N(t)| \geq N^\sigma$ on the set W contained in an interval of length $T = N^{6/5}$. By Proposition 4.6 and (5.5), we have

$$|W| \ll_\epsilon N^{2-2\sigma} + N^{1-2\sigma} \left(\sum_{m \in \mathbb{Z}^3 \setminus \{0\}} I_m \right)^{1/3} = N^{2-2\sigma} + N^{1-2\sigma} (S_1 + S_2 + S_3)^{1/3}.$$

By Proposition 5.1, S_1 is negligible. We bound S_2 by Proposition 6.1, and S_3 by Proposition 11.2. Therefore we get for any choice of $k \in \mathbb{N}$

$$\begin{aligned} |W|^3 N^{6\sigma-3} &\lesssim_\epsilon N^3 + S_2 + S_3 \\ &\lesssim_{\epsilon,k} N^3 + |W|^2 N^2 + TN|W|^{2-1/k} + N^2|W|^{2-3/4k} T^{1/2k} + T^2|W|^{3/2} \\ &\quad + T|W|N^{3-2\sigma} + T|W|^2 N^{3/2-\sigma} + T^{9/8}|W|^{29/16} N^{3/2-\sigma}. \end{aligned}$$

In this formula k comes from the bound for S_2 . It is a positive integer that we can choose. The last inequality rearranges to give

$$\begin{aligned} |W| &\lesssim_{\epsilon,k} N^{2-2\sigma} + N^{5-6\sigma} + T^{\frac{k}{k+1}} N^{(4-6\sigma)\frac{k}{k+1}} + N^{(5-6\sigma)\frac{4k}{4k+3}} T^{\frac{2}{4k+3}} + T^{4/3} N^{2-4\sigma} \\ (12.1) \quad &+ T^{1/2} N^{3-4\sigma} + TN^{9/2-7\sigma} + T^{18/19} N^{72/19-112\sigma/19}. \end{aligned}$$

We choose $k = 4$. We also simplify the formulas using $T = N^{6/5}$.

$$\begin{aligned} |W| &\lesssim_\epsilon T \left(N^{(4-10\sigma)/5} + N^{(19-30\sigma)/5} + N^{(74-120\sigma)/25} + N^{(298-480\sigma)/95} \right. \\ &\quad \left. + N^{(12-20\sigma)/5} + N^{(9-14\sigma)/2} + N^{(354-560\sigma)/95} \right) \\ &\lesssim_\epsilon TN^{(4-10\sigma)/5} + TN^{(12-20\sigma)/5} + TN^{(9-14\sigma)/2}. \end{aligned}$$

If $\sigma \in [7/10, 8/10]$ the first and third terms can be dropped and we get

$$|W| \lesssim_\epsilon TN^{(12-20\sigma)/5}. \quad \square$$

Remark. When N is not too close to T , the most important terms in our bound for $|W|$ in (12.1) are the terms $T^{4/3}N^{2-4\sigma}$ and $T^{1/2}N^{3-4\sigma}$, and our choice of $T = N^{6/5}$ comes from balancing these terms. In this range the terms coming from S_2 do not make a significant contribution and there is some flexibility with the choice of k ; the $T^{\frac{k}{k+1}}N^{(4-6\sigma)\frac{k}{k+1}}$ term is typically increasing in k and the $N^{(5-6\sigma)\frac{4k}{4k+3}}T^{\frac{2}{4k+3}}$ term is typically decreasing in k , so the optimal choice of k balances these terms. The

choice $k = 4$ is made as a simple choice that is sufficient for the S_2 terms to be dominated by the S_3 terms in the range under consideration.

When $N > T^{5/6}$ the process of going from Proposition 3.1 to Theorem 1.1 is somewhat wasteful since it actually bounds the number of large values in $[0, N^{6/5}]$. By using a variation of the above argument, the bound in Theorem 1.1 could be improved. We record one such improvement here.

Proposition 12.1 (Large values estimate for $N \geq T^{5/6}$). *Suppose $(b_n)_{n \sim N}$, $(t_r)_{r \leq R}$ are as in Theorem 1.1, and that $T^{5/6} \leq N \leq T$ and $V = N^\sigma$ with $\sigma \geq 7/10$. Then we have*

$$R \lesssim N^{2-2\sigma} + T^{1/2}N^{3-4\sigma} + \inf_{k \in \mathbb{N}} \left(T^{\frac{k}{k+1}} N^{(4-6\sigma)\frac{k}{k+1}} + N^{(5-6\sigma)\frac{4k}{4k+3}} T^{\frac{2}{4k+3}} \right).$$

In particular, we have

$$R \lesssim N^{2-2\sigma} + T^{1/2}N^{3-4\sigma} + T^{(30\sigma-21)/5} N^{(46-60\sigma)/5}.$$

Proposition 12.1 implies that the $N^{18/5}V^{-4}$ term in Theorem 1.1 can be replaced by $T^{1/2}N^{3-4\sigma} + T^{(30\sigma-21)/5}N^{(46-60\sigma)/5}$, which is smaller. When $\sigma = 3/4$, Proposition 12.1 improves on (1.1) by a factor of $(T/N)^{1/2}$. Since we anticipate the main uses of Theorem 1.1 to be when $N < T^{5/6}$ we content ourselves with the simpler formulation of Theorem 1.1.

Proof. Jutila's large values estimate [Ju, Theorem (1.4)] with $k = 3$ gives

$$R \lesssim N^{2-2\sigma} + TN^{(10-16\sigma)/3} + TN^{18-24\sigma},$$

which implies our bound for $\sigma \geq 39/50$ since the second and third terms above are then both smaller than $T^{1/2}N^{3-4\sigma}$. Thus we only need to consider $\sigma \in [7/10, 39/50]$. The bound now follows from (12.1), (1.1) and a little algebra. For $\sigma \in [7/10, 39/50]$ and $N \in [T^{5/6}, T]$ we have that

$$T^{1/2}N^{3-4\sigma} \gg T^{4/3}N^{2-4\sigma} + TN^{9/2-7\sigma} + T^{18/19}N^{72/19-112\sigma/19}.$$

Therefore the $T^{1/2}N^{3-4\sigma}$ term in (12.1) dominates the 5th, 7th and 8th terms in (12.1). We also have

$$T^{1/2}N^{3-4\sigma} \gg \min(TN^{1-2\sigma}, N^{5-6\sigma}).$$

Therefore, by combining (12.1) and (1.1) we find that

$$R \lesssim N^{2-2\sigma} + T^{1/2}N^{3-4\sigma} + \inf_{k \in \mathbb{N}} \left(T^{\frac{k}{k+1}} N^{(4-6\sigma)\frac{k}{k+1}} + N^{(5-6\sigma)\frac{4k}{4k+3}} T^{\frac{2}{4k+3}} \right).$$

This gives the first bound. For the final bound, we note that if $N^{6\sigma-4} \geq T$ then the result follows from (1.1), so we may assume that $N^{6\sigma-4} < T$ and so the first term in the infimum above is increasing with k . Similarly, if $N^{4-4\sigma} > T$ then $TN^{1-2\sigma} < T^{1/2}N^{3-4\sigma}$ so the result follows from (1.1). Therefore we may also assume that $N^{4-4\sigma} \leq T$. In this case $N^{15-18\sigma} \leq T^2$ and so in the second term in the infimum above is decreasing in k . Thus, for the expression to be less than $T^{(30\sigma-21)/5}N^{(46-60\sigma)/5}$ we require

$$\frac{73 - 138n - 90\sigma + 180n\sigma}{12(1-n)(10\sigma-7)} \leq k \leq \frac{-21 + 46n + 30\sigma - 60n\sigma}{2(1-n)(13-15\sigma)}$$

where $n := \log N / \log T$. The upper bound of this interval is always at least 1 for $n \in [5/6, 1)$, $\sigma \in [7/10, 39/50]$ (it is increasing in σ in this range so is smallest when $\sigma = 7/10$) and the length of this interval is

$$1 + \frac{5(6n-5)(-41 + 123\sigma - 90\sigma^2)}{12(1-n)(10\sigma-7)(13-15\sigma)}.$$

Thus the interval has length at least 1 whenever $\sigma \in [7/10, 39/50]$, and so there is a choice of $k \in \mathbb{N}$ giving the desired bound. \square

13. APPLICATIONS TO RIEMANN ZETA FUNCTION AND PRIME NUMBERS

13.1. Proof of Theorem 1.2. As noted in the introduction, Theorem 1.2 follows from Ingham's result (1.2) if $\sigma \leq 7/10$ and Huxley's result (1.3) if $\sigma \geq 8/10$, so we may assume that $\sigma \in [7/10, 8/10]$. Clearly it suffices to show the bound of Theorem 1.2 for zeros with imaginary part in $[T, 2T]$, since the result for $[0, T]$ then follows by considering $T/2, T/4, \dots$ in place of T .

We now briefly recall the classical zero-detecting methodology, referring the reader to [M3, Chapter 12] or [MP, Appendix C] for more complete details. Given a parameter $N = 2^j$, we let

$$b_n := \left(\sum_{\substack{d|n \\ d \leq 2T^{1/100}}} \mu(d) \right) \exp\left(-\frac{n}{T^{1/2}}\right), \quad D(s) := \sum_{n \sim N} b_n n^{-s}.$$

(The reader should note that this is not quite consistent with the notation $D(t)$ used earlier in the paper.)

Then, a non-trivial zero $\rho = \beta + i\gamma$ of $\zeta(s)$ with $\beta \geq \sigma$ and $\gamma \in [T, 2T]$ is called a 'Type I zero' if there is a choice of $N = 2^j \in [T^{1/100}, T^{1/2}(\log T)^2]$ such that $|D(\rho)| \geq 1/(3 \log T)$. If it is not a Type I zero then it is a 'Type II zero', and the number of Type II zeros is $\leq T^{2-2\sigma}(\log T)^{O(1)}$ by [MP, Lemma 24]. Thus it suffices to bound the number of Type I zeros. There are $O(\log T)$ choices of N so we focus on the value of N which gives the largest number of Type I zeros.

We now make a slight modification to D to remove the dependencies on the real parts. Let $\rho = \beta + i\gamma$ be a Type I zero with $\beta \geq \sigma$, and let $\psi(u)$ be a smooth

function equal to $e^{u(\sigma-\beta)}$ on $[\log N, \log 2N]$ and supported on $[(\log N)/2, 2 \log N]$ with $\|\psi^{(j)}(t)\|_\infty \ll_j t^{-j}$ for all $j \in \mathbb{Z}_{\geq 0}$. We then note that by Fourier expansion

$$D(\rho) = \sum_{n \sim N} b_n n^{-\sigma - i\gamma} \psi(\log n) = \int_{\mathbb{R}} \widehat{\psi}(\xi) \left(D(\sigma + i(\gamma - 2\pi\xi)) \right) d\xi.$$

Since $\widehat{\psi}$ is rapidly decreasing, we may truncate the integral to $\xi \lesssim 1$ at the cost of an $O(T^{-100})$ error term. Therefore we see that if ρ is a Type I zero, we have $|D(\sigma + i\gamma + i\xi)| \gtrsim 1$ for some $\xi \lesssim 1$. There are $O(\log T)$ non-trivial zeros $\rho = \beta + i\gamma$ with $\gamma \in [t, t+1]$ for any $t \in [T, 2T]$. Therefore we can find a 1-separated set of points $(s_r)_{r \leq R}$ in $[T, 2T]$ with $|D(\sigma + is_r)| \gtrsim 1$ and the number R of points satisfies $R \gtrsim N(\sigma, 2T) - N(\sigma, T)$. Let

$$\tilde{b}_n := \left(\frac{N}{n}\right)^\sigma b_n, \quad \tilde{D}(t) := \sum_{n \sim N} \tilde{b}_n n^{it} = N^\sigma D(\sigma + it).$$

Thus it suffices to show that if $N < T^{1/2+o(1)}$ and W is a 1-separated set in $[T, 2T]$ such that $|\tilde{D}(t)| \gtrsim N^\sigma$, we have $|W| \lesssim T^{15(1-\sigma)/(3+5\sigma)+o(1)}$.

First we note that if $N < T^{1/100}$ then $b_n = 0$ identically, so the result is trivial in this range. Thus we may assume $N > T^{1/100}$. We can then choose a value of $k \ll 1$ such that

$$(13.1) \quad T^{10/(6+10\sigma)} \leq N^k \leq T^{15/(6+10\sigma)}.$$

This is clearly possible if $N \leq T^{5/(6+10\sigma)}$, whereas if $N > T^{5/(6+10\sigma)}$ we can take $k = 2$ since $N < T^{1/2+o(1)}$ and $15/(6+10\sigma) > 1$ when $\sigma \leq 8/10$. Set

$$\alpha := \frac{15(1-\sigma)}{(3+5\sigma)(18/5-4\sigma)}.$$

If N, k are such that $N^k \leq T^\alpha$, then we apply Theorem 1.1 to the Dirichlet polynomial \tilde{D}^k . This gives

$$|W| \lesssim N^{2k(1-\sigma)} + N^{(18/5-4\sigma)k} + TN^{(12/5-4\sigma)k} \lesssim T^{15(1-\sigma)/(3+5\sigma)}.$$

In the final inequality above we applied the upper bound from (13.1) to the first term, the upper bound $N^k \leq T^\alpha$ to the second term and the lower bound from (13.1) to the final term.

If instead we have $N^k > T^\alpha$, then we apply the usual Mean Value Theorem to \tilde{D}^k . This gives

$$(13.2) \quad |W| \lesssim N^{2k-2k\sigma} + TN^{k-2k\sigma} \lesssim T^{15(1-\sigma)/(3+5\sigma)} + T^{1+(1-2\sigma)\alpha}.$$

Here we applied the upper bound from (13.1) to the first term and the lower bound from $N^k \geq T^\alpha$ to the second term. We then note that for $\sigma < 9/10$ we have

$$1 + (1 - 2\sigma)\alpha = \frac{129 - 195\sigma + 50\sigma^2}{2(3 + 5\sigma)(9 - 10\sigma)} = \frac{15(1 - \sigma)}{3 + 5\sigma} - \frac{250(\sigma - 3/4)^2 + 3/8}{2(3 + 5\sigma)(9 - 10\sigma)},$$

and so (13.2) simplifies to $|W| \lesssim T^{15(1-\sigma)/(3+5\sigma)}$, as required.

Remark. For the purposes of proving a zero density estimate of the form $N(\sigma, T) \lesssim T^{A(1-\sigma)}$ with A a fixed constant as small as possible, the critical case in our work is when $\sigma = 7/10$, $N = T^{5/13}$ (so we apply Theorem 1.1 to a Dirichlet polynomial of length $N^2 = T^{10/13}$), we subdivide $[0, T]$ into intervals of length $T_1 = T^{12/13}$ and where the set W of large values on each subinterval has $|W| \approx T_1^{2/3}$ and $E(W) \approx |W|^{5/2} \approx |W|^4/T_1$. In this critical situation our bounds for S_1 and S_2 are both best possible, and so any further improvement would have to come from the S_3 term. Our bound for $E(W)$ is also likely to be difficult to improve since a random set W would have $E(W) \approx |W|^4/T_1$. The argument of Section 9 is also essentially tight if the R function is taking $T_1^{2/3}$ values of size $|W|/M \approx T_1^{1/2}$ and the set of these values is highly concentrated on rationals with numerator and denominator of size $T_1^{1/3}$ (recall the remark after Proposition 9.1).

13.2. Proof of Corollary 1.3 and Corollary 1.4. These are well-known to follow quickly from (1.4), but for completeness we give a proof. By partial summation, it suffices to prove corresponding results for the Von Mangoldt function in place of the prime indicator function. By the explicit formula (see, for example [Da, Chapter 17]) we have for any choice of $2 \leq T \leq x$

$$\sum_{n \in [x, x+y]} \Lambda(n) = y - \sum_{|\rho| \leq T} \left(\frac{(x+y)^\rho - x^\rho}{\rho} \right) + O\left(\frac{x(\log x)^3}{T} \right).$$

We choose $T = xy^{-1} \exp(2\sqrt[4]{\log x})$ so the error term is $O(y \exp(-\sqrt[4]{\log x}))$, and note that the term in parentheses is $\int_x^{x+y} t^{\rho-1} dt \ll yx^{\Re(\rho)-1}$. Therefore, by considering $1/\log x$ -separated values of σ , we find that

$$\sum_{n \in [x, x+y]} \Lambda(n) = y + O\left(y(\log x) \sup_{\sigma} x^{\sigma-1} N(\sigma, T) \right) + O(y \exp(-\sqrt[4]{\log x})).$$

By combining (1.4) with a slightly stronger result (such as [Ju] or [M3, Theorem 12.1]) that loses at most logarithmic factors for σ closer to 1, we have the bound

$$(13.3) \quad N(\sigma, T) \ll T^{(30/13+o(1))(1-\sigma)} (\log T)^{O(1)}.$$

Using this and the Vinogradov-Korobov zero-free bound $N(\sigma, T) = 0$ for $\sigma \geq 1 - c(\log T)^{-2/3} (\log \log T)^{-1/3}$ for a suitable constant $c > 0$ (see [M3, Corollary 11.4]), we find that

$$\begin{aligned} \sup_{\sigma} x^{\sigma-1} N(\sigma, T) &\ll (\log T)^{O(1)} \sup_{\sigma \leq 1-c(\log T)^{-5/7}} \left(\frac{T^{30/13+o(1)}}{x} \right)^{1-\sigma} \\ &\ll_{\epsilon} \exp(-\sqrt[4]{\log x}) \end{aligned}$$

provided $T < x^{13/30-\epsilon/2}$. Recalling that $T = xy^{-1} \exp(2\sqrt[4]{\log x})$ and $y \geq x^{17/30+\epsilon}$, this gives Corollary 1.3.

For Corollary 1.4, we first let $\delta = X^{-13/15+\epsilon/2}$. By splitting $[x, x+y]$ into intervals of length δx and applying Cauchy-Schwarz, we see that

$$\int_X^{2X} \left(\sum_{n \in [x, x+y]} \Lambda(n) - y \right)^2 dx \ll \frac{y^2}{\delta^2 X^2} \int_X^{3X} \left(\sum_{n \in [x, x+\delta x]} \Lambda(n) - \delta x \right)^2 dx + O(\delta^2 X^3).$$

If the corollary was false, the left hand side would be significantly larger than $y^2 X \exp(-3\sqrt[4]{\log X})$, so it suffices to show that the integral on the right hand side is $\ll \delta^2 X^3 \exp(-3\sqrt[4]{\log X})$. Applying the explicit formula as above with $T = \delta^{-1} \exp(4\sqrt[4]{\log X})$, we see that it suffices to show that

$$(13.4) \quad \int_X^{3X} \left| \sum_{|\rho| < T} x^{\rho} \left(\frac{(1+\delta)^{\rho} - 1}{\rho} \right) \right|^2 dx \ll \delta^2 X^3 \exp(-3\sqrt[4]{\log X}).$$

Expanding the sum, and performing the integral over x , we obtain

$$\sum_{|\rho_1|, |\rho_2| \leq T} \left(\frac{(1+\delta)^{\rho_1} - 1}{\rho_1} \right) \overline{\left(\frac{(1+\delta)^{\rho_2} - 1}{\rho_2} \right)} \int_X^{3X} x^{\rho_1 + \overline{\rho_2}} dx \ll \delta^2 \sum_{|\rho_1|, |\rho_2| < T} \frac{X^{\Re(\rho_1) + \Re(\rho_2) + 1}}{|\rho_1 + \overline{\rho_2} + 1|}.$$

Since $X^{\Re(\rho_1) + \Re(\rho_2) + 1} \leq X^{2\Re(\rho_1) + 1} + X^{2\Re(\rho_2) + 1}$, and noting that (since there are $O(\log T)$ zeros in a horizontal strip of height 1) we have

$$\sum_{|\rho_2| < T} |1 + z + \overline{\rho_2}|^{-1} \ll (\log T)^2$$

for any $|z| < T$ with $\Re(z) \geq 0$. Thus we find that

$$\int_X^{3X} \left| \sum_{|\rho| < T} x^{\rho} \left(\frac{(1+\delta)^{\rho} - 1}{\rho} \right) \right|^2 dx \ll (\log X)^3 \delta^2 \sup_{\sigma} X^{2\sigma+1} N(\sigma, T).$$

As above, applying (1.4) and the zero-free region, we have that

$$\sup_{\sigma} X^{2\sigma+1} N(\sigma, T) \ll X^3 \sup_{\sigma \leq 1-c(\log T)^{-5/7}} \left(\frac{T^{30/13+o(1)}}{X^2} \right)^{1-\sigma} \ll_{\epsilon} X^3 \exp(-10\sqrt[4]{\log X}),$$

on recalling that $T = \delta^{-1} \exp(4\sqrt[4]{\log X}) \lesssim X^{13/15-\epsilon/3}$. Putting this together then gives (13.4), as required.

Remark. *By using a prime decomposition (such as the Heath-Brown Identity) and Mellin inversion, it is possible to relate the count of primes in short intervals directly to Dirichlet polynomials. The critical situation for both Corollary 1.3 and Corollary 1.4 is handling a product of six Dirichlet polynomials each of size roughly $x^{1/6}$ (this was the limiting case in the earlier work of Huxley [Hu] too). As in [HB4], by bounding this contribution corresponding to six almost equal sized primes using a sieve method, one could obtain an asymptotic estimate in the slightly larger range $y \in [x^{17/30-\epsilon}, x^{0.99}]$ and $y \in [X^{2/15-\epsilon}, X^{0.99}]$ at the cost of a worse error term of size roughly $O(\epsilon^4 y / \log x)$.*

REFERENCES

- [B5] J. Bourgain, Remarks on Montgomery’s conjectures on Dirichlet sums. Lecture Notes in Math., 1469 Springer-Verlag, Berlin, 1991, 153-165.
- [Bo] J. Bourgain, On large values estimates for Dirichlet polynomials and the density hypothesis for the Riemann zeta function. IMRN, (2000), no. 2, 133-146.
- [Ca] F. Carlson. Über die Nullstellen der Dirichletschen Reihen und der Riemannsches ζ -Funktion. Ark. Mat. Astron. Fys., 15(20):28 pp., 1921
- [Da] H. Davenport, *Multiplicative number theory*, third ed.. Springer-Verlag, New York, 2000.
- [Ha] G. Halász, Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen. Acta Math. Acad. Sci. Hungaricae 19 (1968), 365-403.
- [HT] G. Halász, and P. Turán, On the distribution of roots of Riemann zeta and allied functions, I. Journal of Number Theory 1.1 (1969), 121-137.
- [HB] D. R. Heath-Brown, A large values estimate for Dirichlet polynomials. Journal of the London Mathematical Society 2.1 (1979), 8-18.
- [HB2] D. R. Heath-Brown, The differences between consecutive primes, II. Journal of the London Mathematical Society, 2.19 (1979), 207-220.
- [HB3] D. R. Heath-Brown, Zero density estimates for the Riemann zeta-function and Dirichlet L-functions. Journal of the London Mathematical Society, 2.19 (1979), 221-232.
- [HB4] D. R. Heath-Brown, The number of primes in a short interval. Journal für die reine und angewandte Mathematik 389 (1988), 22-63.
- [Hu] M. N. Huxley. On the difference between consecutive primes. Invent. Math., 15:164-170, 1972.
- [In] A. E. Ingham. On the estimation of $N(\sigma, T)$. Quart. J. Math. Oxford Ser., 11:291-292, 1940.
- [Iv] A. Ivić. *The Riemann Zeta-Function*. Wiley, 1985.
- [Ju] M. Jutila, Zero-density estimates for L-functions. Acta Arith., 32 (1977), 55-62.
- [MP] J. Maynard and K. Pratt, Half-isolated zeros and zero-density estimates. IMRN, (2024), no 19, 12978-13014,
- [M] H. Montgomery, Mean and large values of Dirichlet polynomials. Invent. Math. 8.4 (1969), 334-345.
- [M2] H. Montgomery, *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*. No. 84. American Mathematical Soc., 1994.
- [M3] H. Montgomery. *Topics in Multiplicative Number Theory*. Lecture Notes in Mathematics, Vol. 227. Springer-Verlag, Berlin-New York, 1971.