

# HELSON'S CONJECTURE FOR SMOOTH NUMBERS

SETH HARDY AND MAX WENQIANG XU

ABSTRACT. Let  $\Psi(x, y)$  denote the count of  $y$ -smooth numbers below  $x$  and  $P(n)$  denote the largest prime factor of  $n$ . We prove that for  $f$  a Steinhaus random multiplicative function, the partial sums over  $y$ -smooth numbers enjoy better than squareroot cancellation, in the sense that

$$\mathbb{E} \left| \sum_{\substack{1 \leq n \leq x \\ P(n) \leq y}} f(n) \right| = o \left( \Psi(x, y)^{1/2} \right),$$

uniformly for  $(\log x)^{30} \leq y \leq x$ . Our bounds are quantitative and give a large saving when  $y$  isn't too close to  $x$ .

## 1. INTRODUCTION

**1.1. Motivation and statement of results.** Multiplicative functions are a fundamental object of study in number theory, and, in recent years, research on *random* multiplicative functions has been pursued as an avenue to better understand the behaviour of certain families of multiplicative functions. The study of random multiplicative functions was initiated by Wintner [27] in 1944, where he introduced the *Rademacher* random multiplicative function as a model for studying the partial sums of the Möbius function. In this paper, we concern ourselves only with the *Steinhaus* random multiplicative function,  $f: \mathbb{N} \rightarrow \mathbb{C}$ , which is defined by letting  $(f(p))_{p \text{ prime}}$  be independent and identically distributed random variables uniformly distributed on the complex unit circle  $\{z \in \mathbb{C} : |z| = 1\}$ , and extending to composite numbers  $n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$  by setting

$$f(n) = \prod_{i=1}^r f(p_i)^{\alpha_i}.$$

The Steinhaus random multiplicative function serves as a model for families of Dirichlet characters,  $n \rightarrow \chi(n)$  for  $\chi \bmod r$  chosen uniformly at random, in addition to continuous characters,  $n \rightarrow n^{it}$ , for  $t \in [T, 2T]$  chosen uniformly at random, where  $r$  and  $T$  are to be thought of as parameters tending to infinity. For  $f$  a Steinhaus random multiplicative function, we have “perfect orthogonality”, in the sense that

$$\mathbb{E} \left[ f(n) \overline{f(m)} \right] = \mathbf{1}_{n=m}.$$

This relation is mirrored by corresponding orthogonality relations for characters. It is of great interest to analytic number theorists to understand the statistical behaviour

of character sums,  $\sum_{n \leq x} \chi(n)$ , and zeta sums,  $\sum_{n \leq x} n^{it}$ . In light of the orthogonality relations, the second moments of these quantities are fairly straightforward to evaluate, most of all in the Steinhaus case where we immediately see that  $\mathbb{E}|\sum_{n \leq x} f(n)|^2 = [x]$ . A natural problem that follows is to understand the first moment,  $\mathbb{E}|\sum_{n \leq x} f(n)|$ , in addition to the character and zeta sum analogues, though understanding these quantities is a much harder problem. A naive first guess is that, ignoring the multiplicative dependence structure, the sum  $\sum_{n \leq x} f(n)$  may behave like the first absolute moment of a sum of mean-zero independent random variables with unit variance, and, if this were true, the Cauchy–Schwarz upper bound  $\mathbb{E}|\sum_{n \leq x} f(n)| \leq \sqrt{x}$  would be sharp up to a constant. This upper bound is often referred to as squareroot cancellation, and more generally gives  $\mathbb{E}|\sum_{n \in \mathcal{A}} f(n)| \leq \sqrt{|\mathcal{A}|}$  for an arbitrary set  $\mathcal{A}$ . Helson [16] conjectured that, for the full sum up to  $x$ , we should actually have more than squareroot cancellation, in the sense that  $\mathbb{E}|\sum_{n \leq x} f(n)| = o(\sqrt{x})$ . The answer to this conjecture remained unclear until the problem was fully resolved by Harper [12], who showed that Helson’s conjecture was true, proving the precise estimate

$$\mathbb{E}\left|\sum_{1 \leq n \leq x} f(n)\right| \asymp \frac{\sqrt{x}}{(\log \log x)^{1/4}}.$$

The analogous character and zeta upper bounds are also proved by Harper in [14] using a remarkable *derandomisation method*. More precisely, Harper explicitly relates these moments of character sums and zeta sums to (a certain stage in the proof of) the random multiplicative case. As is to be expected, the length of the sum in that case depends on the parameters  $r$  and  $T$  that determine the size of the family of characters. For example, it is shown that for any large prime  $r$  and for all  $x \leq r$  with  $x, \frac{r}{x} \rightarrow +\infty$ , one has

$$\frac{1}{r-1} \sum_{\chi \bmod r} \left| \sum_{1 \leq n \leq x} \chi(n) \right| = o(\sqrt{x}).$$

In that work, Harper also conjectured (see [14, Equation (1.2)]) that this phenomenon can be pushed further, in that the “conductor restriction”  $x \leq r$  can be weakened if one has a Möbius twist, an application of which would be to break the classical “squareroot barrier” in Möbius cancellation (see [26] for further discussion). We shall discuss later in the introduction (after Theorem 1.1) about how an analogous conjecture in our setting may have consequences for counting smooth numbers in small intervals.

The key idea behind Harper’s theorem is to notice that the partial sums are connected to a phenomenon known as *critical Gaussian multiplicative chaos* (GMC). To briefly elaborate, we define  $F_x(s) := \prod_{p \leq x} (1 - \frac{f(p)}{p^s})^{-1}$  to be the truncated Euler product. By a non-trivial conditioning argument and an application of Plancherel’s identity, the proof works by first showing that

$$\mathbb{E}\left|\sum_{n \leq x} f(n)\right| \approx \sqrt{\frac{x}{\log x}} \mathbb{E} \left( \int_{-1/2}^{1/2} |F_x(1/2 + it)|^2 dt \right)^{1/2}, \quad (1.1)$$

and then by showing that this expectation of the Euler product integral is smaller than naively anticipated. The driving force behind this is the fact that  $(\log |F(1/2 + it)|)_{t \in [-1/2, 1/2]}$  is well approximated by a Gaussian field with logarithmic correlations, and the exponent 2 is exactly the critical value in this case, all together meaning that these integrals are precisely the setting where one observes critical Gaussian multiplicative chaos. The reader is encouraged to consult [11, 15] for more detailed discussions. We also remark that there have been recent progress in understanding the exact limiting distribution of the partial sums in [6, 8].

It is natural to ask whether one still obtains “better than squareroot cancellation” when the full sum is replaced by a sum over a set  $\mathcal{A}$  that has interesting arithmetic structure. In previous work of the second author [28], the case where  $\mathcal{A} = \mathcal{A}(x, R)$  is a set of  $R$ -rough numbers up to  $x$ , i.e., all elements in the set have prime factors with size at least  $R$ , was studied. In that case, there is an interesting transition range, since we observe “better than squareroot” cancellation,

$$\mathbb{E} \left| \sum_{n \in \mathcal{A}(x, R)} f(n) \right| = o \left( \sqrt{|\mathcal{A}(x, R)|} \right),$$

whenever  $\log \log R$  is smaller than roughly  $\sqrt{\log \log x}$ , but the statement fails to hold when  $R$  is any larger. The case where  $\mathcal{A}$  is a short interval was studied in [1], and, together with the related work [2], these three papers observe some universality in the transition threshold which is ultimately related to the ballot problem.

In this paper, our attention lies on what happens when  $\mathcal{A}$  is the set of  $y$ -smooth numbers, i.e., all elements in  $\mathcal{A}$  only have prime factors  $p \leq y$ . Here and throughout the paper, let  $P(n)$  denote the largest prime factor of  $n$ , so that  $\Psi(x, y) = \#\{n \leq x : P(n) \leq y\}$  denotes the count of  $y$ -smooth numbers up to  $x$ . Note that the case where  $y = x$  corresponds to the full sum, for which we know from Harper[12] that we have better than squareroot cancellation. Analogously to the previous cases, one might wonder if there is a certain transition range where the “better than squareroot” cancellation disappears. Our main theorem suggests that this may not be the case.

**Theorem 1.1.** *Suppose that  $(\log x)^{30} \leq y \leq x$ . Then we have*

$$\mathbb{E} \left| \sum_{\substack{1 \leq n \leq x \\ P(n) \leq y}} f(n) \right| = o \left( \Psi(x, y)^{1/2} \right),$$

*uniformly for  $y$  in this range. Quantitative bounds can be found in Theorems 1.3 and 1.4.*

One motivation and a potential application for our result is to make further progress on the classical problem of counting  $y$ -smooth numbers in short intervals  $[x, x + h]$ . See recent developments and discussions in [23, 30]. The breakthrough work of Matomäki and Radziwiłł [21] shows the existence of  $x^\varepsilon$ -smooth numbers unconditionally in intervals of length  $h \gg_\varepsilon \sqrt{x}$ , coming close to solving an old problem that, for any fixed  $\varepsilon > 0$ , there exists an  $x^\varepsilon$ -smooth number in  $[x, x + \sqrt{x}]$  when  $x$  is sufficiently large. To see how our result may be useful, similarly to [14, Equation (1.2)] (though removing the

“deterministic” contribution from  $t \approx 0$ ), we might conjecture that for any fixed  $A > 0$  and small  $\varepsilon > 0$ , we have, very roughly speaking,

$$\frac{1}{T} \int_{\substack{|t| \leq T \\ t \neq 0}} \left| \sum_{\substack{n \leq x \\ P(n) \leq x^\varepsilon}} n^{-it} \right| dt \ll \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq x^\varepsilon}} f(n) \right|, \quad \forall x \leq T^A. \quad (1.2)$$

Importantly, we know from Theorem 1.1 that the right-hand side is  $o(\sqrt{\Psi(x, x^\varepsilon)})$ . Similar to as described in [14], though isolating the main term in this case, Perron’s formula suggests that

$$\frac{\Psi(x+h, x^\varepsilon) - \Psi(x, x^\varepsilon)}{h} \approx \frac{1}{2\pi x} \int_{-x/h}^{x/h} \sum_{\substack{n \leq x+h \\ P(n) \leq x^\varepsilon}} n^{-it} \left( \frac{x}{h} \right)^{it} dt \approx \frac{\Psi(x, x^\varepsilon)}{x} + E(x, x^\varepsilon, h),$$

where

$$E(x, x^\varepsilon, h) \ll \frac{1}{x} \int_{\substack{|t| \leq T \\ t \neq 0}} \left| \sum_{\substack{n \leq x+h \\ P(n) \leq x^\varepsilon}} n^{-it} \right| dt.$$

In particular, since  $\Psi(x, x^\varepsilon) \asymp_\varepsilon x$ , combining Theorem 1.1 and (1.2) (with  $A = 2$ ) gives the bound  $E(x, x^\varepsilon, \sqrt{x}) = o(\sqrt{\Psi(x, x^\varepsilon)/x}) = o(1)$ , which (very roughly) gives an asymptotic for the count in intervals of length  $h = \sqrt{x}$ . In fact, our quantitative result for  $y$  in this range, Theorem 1.4, would suggest further that this strategy could show the existence of  $x^\varepsilon$ -smooth numbers in intervals of length  $h \gg \sqrt{x}/(\log \log x)^{1/4}$ , similarly to how the analogous conjecture in [14] gives Möbius cancellation in sums of length  $h \gg \sqrt{x}/(\log \log x)^{1/4}$ .

Based on Theorem 1.1, it seems reasonable to conjecture that “better than squareroot cancellation” should hold uniformly for all  $y \leq x$ .

**Conjecture 1.2.** *Theorem 1.1 holds uniformly for any  $2 \leq y \leq x$ .*

It would be nice if one could solve Conjecture 1.2 completely, though our main theorem covers what we believe to be the most interesting range. Analogously to (1.1), we always work by connecting the partial sums to the Euler product. To illustrate how the analysis changes for smaller  $y$ , in the most extreme case, when  $y = 2$ , we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq 2}} f(n) \right| = \mathbb{E} \left| \sum_{n \leq \lfloor \frac{\log x}{\log 2} \rfloor} f(2)^n \right| = \int_0^1 \left| \sum_{n \leq \lfloor \frac{\log x}{\log 2} \rfloor} e(n\theta) \right| d\theta \ll \log \log x,$$

which is very small relative to  $\Psi(x, 2)^{1/2} \asymp (\log x)^{1/2}$ . This provides evidence towards our conjecture, and also points out that for small  $y$ , the problem is related to finding cancellation in exponential sums over lattices. The exponential sum estimate becomes challenging once  $y \gg (\log x)^{1+o(1)}$ , and new ideas are needed to cover this range.

In this paper, we observe two distinct sources for the “better than squareroot cancellation” phenomena in Theorem 1.1; the size of  $y$  being key to this analysis. First of all, for  $y$  being close to  $x$ , one should expect that the critical GMC phenomenon uncovered

by Harper [12] should give better than squareroot cancellation. We find this to be the case, and moreover, we are able to use critical GMC phenomenon to prove better than squareroot cancellation even when  $y$  is as small as  $e^{(\log \log x)^2}$ , say. We outline how this is done in Section 1.4, and a precise statement is given in Theorem 1.4. The critical GMC saving in our case specifically comes from exploiting the randomness of  $f(p)$  on the small primes (to be precise, primes smaller than roughly  $e^{\frac{1}{1-\alpha}}$  for  $\alpha = \alpha(x, y)$  the saddle point, see Section 1.2 for details), and, as one may expect, the saving is always quite small (like a power of  $\log \log x$ , or smaller).

Given the previous works [12, 28, 1], one may not expect to observe cancellation that is larger than some power of  $\log \log x$  for any reasonably large  $y$ . *However, we find that when  $y$  is smaller than roughly  $x^{\frac{1}{\log \log x}}$ , we can obtain an unexpectedly large saving (actually coming from the primes closer to  $y$ ).* In short, this saving comes from the fact that the dominant contribution to certain Euler product expectations comes from highly unlikely events (we explain this in detail in Section 1.3). The flavour of this might remind some readers of the *supercritical* GMC case, although this is not exactly what we observe, since our Euler products are shifted off the half-line and are no longer approximated by the exponential of a log-correlated Gaussian field. This novel phenomenon allows us to prove the following:

**Theorem 1.3.** *Let  $u = \frac{\log x}{\log y}$  and fix  $\varepsilon > 0$ . For  $e^{(\log \log x)^{1+\varepsilon}} \leq y \leq x^{\frac{1}{(\log \log x)^{1+\varepsilon}}}$ , we have*

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \sqrt{\Psi(x, y) \exp(-u(\log 2 + o(1)))}, \quad (1.3)$$

where the  $o(1)$  term goes to zero uniformly for  $y$  in this range. This result is derived from the following bound: In the larger range  $(\log x)^{30} \leq y \leq x^{\frac{1}{7 \log \log x}}$ , we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \Psi(x^2, y)^{1/4} \exp\left(\frac{10u \log u}{\log y} + O(u^{8/11})\right) (\log x)^{5/8} (\log y)^{13/8},$$

where all implied constants are uniform for  $y$  in this range. This saves a factor of at least  $e^{cu}$  (for some fixed  $c > 0$ ) over the trivial bound  $\sqrt{\Psi(x, y)}$  uniformly on the entire range  $(\log x)^{30} \leq y \leq x^{\frac{1}{7 \log \log x}}$ .

This theorem is proved in Section 5, and we stress that the proof *does not* make use of critical GMC phenomena, where the saving would typically be much smaller. We have not made a significant attempt to optimise the constants appearing in the range where the theorem holds.

**1.2. Smooth numbers and the transition range.** Before outlining our proofs, for the uninitiated reader, we give a short review of smooth numbers, restricting ourselves to the case where  $(\log x)^2 \leq y \leq x$ , say. We refer readers to [7] for more background.

We can quite accurately upper bound the number of  $y$ -smooth numbers below  $x$  simply by a simple application of Rankin's trick. Define the truncated Euler product  $\zeta(s, y) := \prod_{p \leq y} (1 - \frac{1}{p^s})^{-1}$  for any  $\operatorname{Re}(s) > 0$ . For any  $\sigma > 0$ , we have

$$\Psi(x, y) = \sum_{\substack{n \leq x \\ P(n) \leq y}} 1 \leq \sum_{P(n) \leq y} \left(\frac{x}{n}\right)^\sigma = x^\sigma \prod_{p \leq y} \left(1 - \frac{1}{p^\sigma}\right)^{-1} = x^\sigma \zeta(\sigma, y). \quad (1.4)$$

We now choose  $\sigma > 0$  so that it minimises the right-hand side. It turns out that this minimiser is unique, and is called the saddle point,  $\alpha = \alpha(x, y)$ . Since  $\sigma = \alpha(x, y)$  satisfies  $\frac{d}{d\sigma} \log(x^\sigma \zeta(\sigma, y))|_{\sigma=\alpha(x, y)} = 0$ , it follows that

$$\sum_{p \leq y} \frac{\log p}{p^{\alpha(x, y)} - 1} = \log x.$$

and it can be shown [20, Lemmas 1 and 2] that

$$\alpha(x, y) = 1 - \frac{\log(u \log(u+1))}{\log y} + O\left(\frac{1}{\log y}\right), \quad (1.5)$$

where  $u = \frac{\log x}{\log y}$ . The error term can be improved to  $o(\frac{1}{\log y})$  as soon as  $(\log x)^2 \leq y \leq x^{o(1)}$ , and we state a precisely result in Lemma 2.1. The bound obtained by Rankin's trick,  $\Psi(x, y) \leq x^\alpha \zeta(\alpha, y)$ , is close to optimal, seeing as

$$x^\alpha \zeta(\alpha, y) \ll \Psi(x, y) \sqrt{\log x \log y}, \quad (1.6)$$

and a precise comparison can be drawn from Lemma 2.2. In our analysis, in contrast to (1.1), we are interested in integrals of the quantity  $|F_y(\alpha/2 + it)|$ . In this case, the variance of  $\log |F_y(\alpha/2 + it)|$  is roughly  $\frac{1}{2} \sum_{p \leq y} \frac{1}{p^\alpha}$ , and letting  $z = e^{\frac{1}{1-\alpha}}$ , this can be written as

$$\begin{aligned} \frac{1}{2} \sum_{p \leq y} \frac{1}{p^\alpha} &= \frac{1}{2} \left( \sum_{p \leq z} \frac{1}{p^{1-\frac{1}{\log z}}} + \sum_{z < p \leq y} \frac{1}{p^\alpha} \right) \\ &\sim \frac{1}{2} \left( \log\left(\frac{1}{1-\alpha}\right) + \frac{y^{1-\alpha}}{(1-\alpha) \log y} \right), \end{aligned} \quad (1.7)$$

whenever  $1/2 \leq \alpha \leq 1 - \frac{1}{\log y}$ , say. Calculations of this type can be found in [22, Lemma 7.4]. Using the saddle point approximation (1.5), the first term in the parentheses is roughly  $\log \log y$ , and the second term is of size  $u$ . Therefore, the dominant contribution to this variance comes from the primes  $p \leq z$  whenever  $u$  is smaller than roughly  $\log \log x$ , whereas the dominant contribution comes from  $p > z$  whenever  $u$  is larger. This is why we see a phase transition at roughly  $y \approx x^{\frac{1}{\log \log x}}$ .

**1.3. Outline of the main result, Theorem 1.3.** In this section, we will explain the behaviour for moderately sized  $y$ , specifically  $(\log x)^4 \leq y \leq x^{\frac{1}{\log \log x}}$ , say. The typical way that one approaches these expectation problems is to use conditioning to relate the first absolute moment of our partial sums to the  $1/2$ 'th moment of the mean square of

a random Euler product, as in (1.1). For example, in our case, it would be possible to show (up to errors that we will ignore) that for any  $\beta > 1 - \frac{\log \log x}{\log y}$ , we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll x^{\beta/2} \mathbb{E} \left( \int_{-1/2}^{1/2} |F_y(\beta/2 + it)|^2 dt \right)^{1/2}, \quad (1.8)$$

where we recall that the random Euler product is defined as  $F_y(s) := \prod_{p \leq y} (1 - \frac{f(p)}{p^s})^{-1}$ . This is analogous to the Rankin's trick upper bound for  $\Psi(x, y)$  in equation (1.4), and we may suspect that the correct choice of  $\beta$  that minimises the right-hand side is  $\beta(x, y) = \alpha(x, y)$ , where  $\alpha(x, y)$  is the saddle point. Surprisingly, taking  $\beta = \alpha(x, y)$  will turn out to be sub-optimal, but we content ourselves with it for the time being, for both simplicity of exposition and since the optimal  $\beta$  will be taken to be quite close to the saddle point. Indeed, applying Hölder's inequality to the expectation on the right-hand side of (1.8), we find that

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll x^{\beta/2} \left( \int_{-1/2}^{1/2} \mathbb{E} |F_y(\beta/2 + it)|^2 dt \right)^{1/2} \ll x^{\beta/2} \zeta(\beta, y)^{1/2},$$

which is optimised by taking  $\beta = \alpha(x, y)$ , identically to Rankin's trick, and gives (roughly) squareroot cancellation. As in the work of Harper [12], the key idea is to notice that this use of Hölder's inequality is inefficient, and in fact, this application turns out to be far more wasteful than in the delicate work there.

To obtain a better bound, we return our attention to (1.8) and fix  $\beta = \alpha(x, y)$ . For compactness, we write  $\alpha$  in place of  $\alpha(x, y)$ . Writing  $|F_y(\alpha/2 + it)| = \exp(\operatorname{Re} \log F_y(\alpha/2 + it))$ , we obtain

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll x^{\alpha/2} \mathbb{E} \left( \int_{-1/2}^{1/2} \exp(2 \log |F_y(\alpha/2 + it)|) dt \right)^{1/2}. \quad (1.9)$$

To reiterate, the motivation for writing our Euler product in this way is that, for each fixed  $t \in [-1/2, 1/2]$ , the quantity  $\log |F_y(\alpha/2 + it)|$  is approximately Gaussian with zero mean and variance  $\approx \frac{1}{2} \sum_{p \leq y} \frac{1}{p^\alpha}$ , and these Gaussians should begin to decorrelate when  $t$  varies by  $\approx \frac{1}{\log y}$ . Letting  $G_j$  be independent random variables with distribution  $\mathcal{N}(0, \frac{1}{2} \sum_{p \leq y} \frac{1}{p^\alpha})$ , one might suspect that

$$\max_{t \in [-1/2, 1/2]} \log |F_y(\alpha/2 + it)| \approx \max_{|j| \leq \log y/2} G_j \approx \left( \log \log y \sum_{p \leq y} \frac{1}{p^\alpha} \right)^{1/2}, \quad (1.10)$$

with high probability. We contrast this with a moment calculation. Given that  $\log |F_y(\alpha/2 + it)|$  is approximately Gaussian with zero mean and variance  $\approx \frac{1}{2} \sum_{p \leq y} \frac{1}{p^\alpha}$ , the moment generating function is

$$\mathbb{E} |F(\alpha/2 + it)|^\gamma = \mathbb{E} \exp(\gamma \log |F_y(\alpha/2 + it)|) \asymp \exp\left(\frac{\gamma^2}{4} \sum_{p \leq y} \frac{1}{p^\alpha}\right) \asymp \zeta(\alpha, y)^{\frac{\gamma^2}{4}}. \quad (1.11)$$



Similar calculations for  $\alpha \approx 1$  can be found in [11, Euler Product Result 1]. In particular, one can find that the dominant contribution to  $\mathbb{E}|F(\alpha/2 + it)|^\gamma$  comes from increasingly large values of  $\log |F_y(\alpha/2 + it)|$ , for example when  $\gamma = 2$ , the dominant contribution comes from events where  $\log |F_y(\alpha/2 + it)| = \sum_{p \leq y} \frac{1}{p^\alpha} + O((\sum_{p \leq y} \frac{1}{p^\alpha})^{1/2})$ . Similar ideas are present in the analysis of moments of the Riemann zeta function, for example, in the works of Soundararajan [24] and Harper [9].

Therefore, when  $\alpha = 1$  (or is sufficiently close to 1), the dominant contribution to the second moment comes from events where  $\log |F_y(\alpha/2 + it)|$  is close to maximal (1.10). However, if  $\alpha$  is sufficiently smaller than 1, then the dominant contribution to the second moment comes from events where  $\log |F_y(\alpha/2 + it)|$  is significantly larger than the probable size of the maximum, that is, the *second moment is controlled by very unlikely events*. The calculation in (1.7) highlights that this transition happens roughly when  $y \approx x^{\frac{1}{\log \log x}}$ . Now, since we expect that the integrand is roughly fixed over intervals of length  $\approx \frac{1}{\log y}$ , the right-hand side of (1.9) is approximately

$$x^{\alpha/2} \mathbb{E} \left( \frac{1}{\log y} \sum_{|j| \leq \frac{\log y}{2}} \left| F_y \left( \alpha/2 + i \frac{j}{\log y} \right) \right|^2 dt \right)^{1/2}.$$

If the expectation is dominated by events where  $\max_{|j| \leq \frac{\log y}{2}} \log |F_y(\alpha/2 + ij/\log y)|$  is unusually large, then when such events occur, we would only expect to find  $\ll 1$  values of  $j$ , and all other values should be relatively small in comparison. Therefore, applying sub-additivity of  $x \rightarrow x^{1/2}$ , we have

$$\mathbb{E} \left( \frac{1}{\log y} \sum_{|j| \leq \frac{\log y}{2}} \left| F_y \left( \alpha/2 + i \frac{j}{\log y} \right) \right|^2 dt \right)^{1/2} \leq \frac{1}{(\log y)^{1/2}} \sum_{|j| \leq \frac{\log y}{2}} \mathbb{E} \left| F_y \left( \alpha/2 + i \frac{j}{\log y} \right) \right|,$$

which should be relatively efficient under such assumptions. Evaluating the right-hand side using (1.11), we deduce the bound

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll x^{\alpha/2} \zeta(\alpha, y)^{1/4} (\log y)^{1/2}.$$

By equation (1.6), this gives

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \sqrt{\frac{\Psi(x, y)}{\zeta(\alpha, y)^{\frac{1}{2}(1+o(1))}}},$$

up to some logarithmic factors, which is far better than squareroot cancellation. In fact, this strategy could instead prove a more general bound of the shape

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll x^{\beta/2} \zeta(\beta, y)^{1/4} (\log y)^{1/2},$$

for any  $\beta > 1 - \frac{\log \log x}{\log y}$ . Seeing as the above is  $((x^2)^\beta \zeta(\beta, y))^{1/4} (\log y)^{1/2}$ , we find that  $\beta = \alpha(x^2, y)$  is the optimal choice. Up to some errors, this gives the result seen in our



theorem, and more generally suggests that for all  $0 < q < 1$ , we should in fact have roughly

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right|^{2q} \ll_q \Psi(x^{1/q}, y)^{q^2} (\log y)^{1-q},$$

though we do not pursue such a result here.

We end this section by remarking that, compared to the previous works [12, 1, 28], there are new technical ideas needed to obtain our result in the range  $y \leq e^{(\log x)^{1/2+o(1)}}$ . The reason for the additional difficulty arises from the fact that, in this range,  $y$  is smaller than the saving that we hope to obtain (which is of size  $\approx e^{cu}$  for some constant  $c$ ). We explain more specifically why this is an issue at the end of Section 5.1. Let  $\mathcal{K} = \lceil \frac{\log y}{10 \log 5} \rceil$ . Roughly speaking, the remedy is to consider writing  $y$ -smooth  $n$  as

$$n = m_1 \dots m_{\mathcal{K}} s, \quad m_i := \prod_{\substack{p \in (y/5^i, y/5^{i-1}] \\ p^\alpha \parallel n}} p^\alpha \text{ for } 1 \leq i \leq \mathcal{K},$$

so that  $s$  is the part of  $n$  consisting only of prime factors below  $y/5^{\mathcal{K}} \approx y^{9/10}$ . For a typical  $y$ -smooth number, at least one of  $m_1, \dots, m_{\mathcal{K}}$  will be “large”, say  $\gg e^{cu}$  (the size of the saving). Therefore, we begin by splitting our original sum depending on the first of these parameters  $k$  such that  $m_k \geq Z$ . We then condition instead on  $f(p)$  for  $p \leq y/M^k$ , and, since  $m_k$  is large, we are able to approximate the outer sum over  $m_k$  by an integral with sufficient accuracy, and we can then follow the standard blueprint of applying Plancherel’s identity to obtain the Euler product. An accurate description of the splitting is given at the beginning of Section 5.3, since we find it useful to make further technical adjustments to the one given above.

**1.4. Completion of the proof of Theorem 1.1: Multiplicative Chaos.** To complete the introduction, we describe how one handles the somewhat delicate range when  $y$  is very close to  $x$ . Our analysis on this range actually allows us to “win” on a larger range of  $y$ , and we are able to prove the following:

**Theorem 1.4.** *For any  $e^{(\log \log x)^2} \leq y \leq x$ , say, we have*

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \frac{\Psi(x, y)^{1/2}}{\left( \log \min \left\{ \frac{1}{1-\alpha(x, y)}, \log x \right\} \right)^{1/4}},$$

where the implied constant is uniform in  $y$ .

*Remark.* In the range  $x^{\frac{1}{\log \log x}} \leq y \leq x$ , this result gives a saving of size  $(\log \log x)^{1/4}$ , similarly to [12].

To outline the main ideas behind this result, we begin by assuming that  $y$  is close to  $x$ , where we might expect to be close to the critical GMC regime studied in [12]. In that

work, having established an upper bound of the form (1.1), the key step involves showing that

$$\mathbb{E} \left( \int_{-1}^1 |F_x(1/2 + it)|^2 dt \right)^{1/2} \ll \left( \frac{\log x}{\sqrt{\log \log x}} \right)^{1/2}, \quad (1.12)$$

giving better than squareroot cancellation. Analogously to the previous section, we need to prove a statement along the lines of

$$\mathbb{E} \left( \int_{-1}^1 |F_y(\alpha/2 + it)|^2 dt \right)^{1/2} = o \left( \sqrt{\zeta(\alpha, y)} \right),$$

where  $\alpha = \alpha(x, y)$ . Of course, the main difference between our case and the work of Harper [12] is that our Euler products *are not* evaluated on the half-line. This means that the random process  $(\log |F_y(\alpha/2 + it)|)_{t \in [-1, 1]}$  does not necessarily behave like a log-correlated Gaussian field, and so it is not clear that we can connect our Euler products to critical Gaussian multiplicative chaos. The key observation is that the portion of the Euler product coming from  $p \leq e^{\frac{2}{1-\alpha}}$  *does* behave like it is evaluated on the half-line. To be precise, taking  $z = e^{\frac{2}{1-\alpha}}$ , we have

$$F_z(\alpha/2 + it) = F_z \left( 1/2 - \frac{1}{\log z} + it \right).$$

Note that the specific choice of 2 in the numerator of  $\frac{2}{1-\alpha}$  is unimportant<sup>1</sup> and is just chosen so that the second term is exactly  $\frac{1}{\log z}$ . The important point (which is shown rigorously in [12]) is that a shift of order  $\frac{1}{\log z}$  has no impact on the behaviour of these Euler products, so similarly to (1.12), we have

$$\mathbb{E} \left( \int_{-1}^1 \left| F_z \left( \frac{1}{2} + \frac{1}{\log z} + it \right) \right|^2 dt \right)^{1/2} \ll \left( \frac{\log z}{\sqrt{\log \log z}} \right)^{1/2} \asymp \left( \frac{1}{(1-\alpha) \sqrt{\log \frac{1}{1-\alpha}}} \right)^{1/2}.$$

Therefore, we can obtain better than squareroot cancellation on these partial Euler products over small primes. The key simple idea then is that we only win on products over small primes. By the triangle inequality, it suffices to handle sums roughly of the form

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ z < p \leq y}} f(n) \right| = \mathbb{E} \left| \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (z, y]}} f(m) \sum_{\substack{n \leq x/m \\ P(n) \leq z}} f(n) \right|,$$

We can apply Cauchy–Schwarz inequality to the conditional expectation at this stage in the proof, leaving only the randomness from these small primes  $p \leq z$ . In comparison to [12], the new input needed for this is a precise short intervals bound, which is provided by Lemma 2.5.

---

<sup>1</sup>In Section 4, we choose a slightly different constant to allow for direct use of a result from [12]. This is specifically seen in Lemma 3.2.

**1.5. Organization.** In Section 2, we introduce some estimates on smooth numbers and prove several results that are needed in later sections. In particular, several short sum estimates with restrictions on prime factorizations will be established. In Section 3, we prove results for random Euler products with the feature that the evaluation is on  $\alpha/2$  line instead of  $1/2$  line. Section 4 is devoted to the proof of Theorem 1.4 on a large range, where the saving comes from connecting the problem to critical GMC. The motivation for proving this result first is that it is quite straightforward, while also illustrating techniques used later on. In Section 5, we prove Theorem 1.3. This is the most involved part of the paper. Appendix A contains a proof of Lemma 2.5, which follows similarly to previous work of Hildebrand [19].

**Acknowledgement.** The authors would like to thank Adam Harper, Carl Schildkraut and K. Soundararajan for their interest in this paper and for helpful discussions. SH is supported by the Swinnerton-Dyer scholarship at the Warwick Mathematics Institute Centre for Doctoral Training. MWX is supported by a Simons Junior Fellowship from the Simons Foundation. Part of the work was done during several visits of MWX at Warwick Mathematics Institute, and the warm hospitality is greatly appreciated.

## 2. SMOOTH NUMBER RESULTS

In this section, we state some important results about smooth numbers that are used frequently throughout the paper. First of all, in Section 2.1, we state some estimates for the saddle point and its relation to the count of smooth numbers. We also prove a result, Lemma 2.4, that allows us to compare the count of  $y$ -smooth numbers to the count of  $y/d$ -smooth numbers. In Section 2.2, we prove some estimates for sums over integers that have restrictions on their prime factorization.

**2.1. Classical results.** We begin by stating three classical results on smooth numbers are frequently used throughout the paper.

**Lemma 2.1** (Saddle point estimate). *For  $(\log x)^2 \leq y \leq x$ ,  $u = \frac{\log x}{\log y}$  and  $\alpha(x, y)$  be the saddle point, we have*

$$\alpha(x, y) = 1 - \frac{\log(u \log(u+1))}{\log y} + O\left(\frac{1}{\log y}\right).$$

*In the reduced range  $(\log x)^2 \leq y \leq x^{\frac{1}{\log \log x}}$ , say, we have*

$$\alpha(x, y) = 1 - \frac{\log(u \log u)}{\log y} + O\left(\frac{\log \log u}{(\log y)(\log u)}\right).$$

*Proof.* These estimates follow from [20, Lemmas 1 and 2]. □

**Lemma 2.2** (Explicit smooth count). *Uniformly for  $x \geq y \geq 2$ , the number of  $y$  smooth numbers up to  $x$  satisfies*

$$\Psi(x, y) = \frac{x^\alpha \zeta(\alpha, y)}{\alpha \sqrt{2\pi(1 + (\log x)/y) \log x \log y}} \left(1 + O\left(\frac{1}{\log(u+1)} + \frac{1}{\log y}\right)\right),$$

where  $\alpha = \alpha(x, y)$  is the saddle point.

*Proof.* This is [20, Theorem 1]. □

**Lemma 2.3** (Smooth number comparison in  $x$ ). *For  $2 \leq y \leq x$ , and  $1 \leq d \leq x$ , we have the following relation*

$$\Psi(x/d, y) \ll \frac{1}{d^\alpha} \Psi(x, y),$$

where  $\alpha = \alpha(x, y)$  denotes the saddle-point corresponding to the  $y$ -smooth numbers less than  $x$ .

*Proof.* This is [3, Théorème 2.4(i)]. □

We now prove an analogue of the above result, where instead we dilate the smoothness parameter. Before stating this, we take  $\xi(u)$  to be the unique solution to the equation  $e^{\xi(u)} = 1 + u\xi(u)$  with  $\xi(1) = 0$ . This is a standard function that is frequently found in smooth number estimates, and it is straightforward to show that  $\xi(u) = \log(u \log u) + O(\frac{\log \log u}{\log u})$  (see [20, Lemma 1]).

**Lemma 2.4** (Smooth number comparison in  $y$ ). *Suppose that  $(\log x)^4 \leq y \leq x$  and that  $2 \leq d \leq y^{1/3}$ , say. We have*

$$\Psi(x, y/d) = \Psi(x, y) \exp\left((u - u_d)\xi(u) + O\left(\frac{u(\log d)^2}{(\log y)^2}\right) + O(E(u))\right),$$

where  $u_d := \frac{\log x}{\log y/d}$  and  $u = \frac{\log x}{\log y}$ . For any fixed  $\varepsilon > 0$ , the error term  $E(u)$  satisfies  $E(u) \ll_\varepsilon 1$  uniformly in the range  $e^{(\log \log x)^{5/3+\varepsilon}} \leq y \leq x$  and  $E(u) \ll_\varepsilon u \exp(-(\log u)^{3/5-\varepsilon})$  uniformly on the range  $(\log x)^4 \leq y \leq e^{(\log \log x)^{5/3+\varepsilon}}$ .

*Remark.* Assuming the Riemann Hypothesis, one can use the work of Hildebrand [18] to obtain  $E(u) \ll 1$  on the entire range.

*Proof.* It follows from the work of Hildebrand [19, pg 291] (see the last equation there) that we have the unconditional bound

$$\Psi(x, y/d) = x\rho(u_d) \exp\left(O(E(u))\right),$$

uniformly for  $y$  in this range. Note that we can replace occurrences of  $u_d$  in the exponential with  $u$ , since these quantities are of the same order of magnitude. Taking  $d = 1$  and comparing, we have

$$\Psi(x, y/d) = \Psi(x, y) \frac{\rho(u_d)}{\rho(u)} \exp\left(O(E(u))\right).$$

We now make use of estimates for the Dickman function, specifically using the fact that  $\rho(u - v) = \rho(u)e^{v\xi(u) + O(v^2/(u+v^2))}$  for  $0 \leq v \leq u$ . This estimate is [25, Equation (5.114),

III.5], which, as mentioned there, follows by combining [4, Lemma 6.1] and [25, Equation (5.63), III.5]). Applying this estimate, we find that

$$\rho(u) = \rho(u_d) \exp\left((u_d - u)\xi(u_d) + O\left(\frac{u(\log d)^2}{(\log y)^2}\right)\right).$$

We now use the fact from [20, Lemma 1] that  $\xi(v) = \log(v \log v) + O(\frac{\log \log v}{\log v})$ . By the defining equation and the asymptotic formula, one can readily find that  $\xi'(u) \ll \frac{1}{u}$ . Therefore, by the mean value theorem and the estimate for  $\xi(u)$ , we have  $\xi(u_d) - \xi(u) \ll \frac{\log d}{\log y}$ . We deduce that

$$\Psi(x, y/d) = \Psi(x, y) \exp\left((u - u_d)\xi(u) + O\left(\frac{u(\log d)^2}{(\log y)^2}\right) + O(E(u))\right),$$

as required.  $\square$

**2.2. Short sums over prime restricted integers.** In various parts of our paper, it will be necessary to estimate short sums over integers whose prime factors all lie in a given range. For this task, we employ the following lemmas:

**Lemma 2.5** (Restricted sum over many primes). *Fix any small  $\varepsilon > 0$  and let  $x \geq 2$  be a parameter tending to infinity. Suppose  $y, h, \delta$  are parameters such that  $e^{(\log \log x)^{5/3+\varepsilon}} \leq y \leq x^{1/\delta}$ ,  $x/y^{1/3} \leq h \leq x/2$ , and  $\frac{1}{\log y} \leq \delta \leq \frac{1}{10}$ . Uniformly in these parameters, we have*

$$\frac{1}{h} \sum_{\substack{x < n \leq x+h \\ p|n \Rightarrow p \in (y^\delta, y]}} 1 \ll \frac{1}{x} \cdot \frac{\Psi(x, y)}{\delta \log y}.$$

The proof of this will follow similarly to Hildebrand [19, Theorem 3], and we prove it in Appendix A). Note that the lower bound on  $y$  comes from the current best known error term in the prime number theorem. In the case of long intervals, this sum is asymptotically evaluated in [3, Théorème 2.1] and [29, Theorem 1]. Our result should incur no loss when  $\delta \ll \frac{1}{\log u}$ , seeing as then the smooth and rough conditions behave roughly as though they are independent of one another. One can also find more general sums of this form in [5].

**Lemma 2.6** (Restricted sum over large primes). *Suppose that  $2 \leq h \leq t \leq x$ , and that  $(\log x)^3 \leq y \leq x^{\frac{1}{\log \log x}}$ . Suppose that  $0 \leq d \leq C$ , for some large constant  $C$  (for example,  $C = e^{100}$  will suffice). Then we have*

$$\sum_{\substack{t \leq n \leq t+h \\ p|n \Rightarrow p \in (y/d, y]}} 1 \ll h^{\alpha(x^2, y)} e^{\frac{3du \log u}{\log y}} \log x,$$

where the implied constant is uniform in all parameters.

*Remark.* It may be surprising on first glance that we can obtain a power of  $\alpha(x^2, y)$  in this result, which is smaller than  $\alpha(x, y)$ . This is possible since our sum is only over integers consisting of primes on the same scale as  $y$ . In fact, one could prove a far more

flexible result with a more general parameter in place of  $\alpha(x^2, y)$ . However, motivated by Section 1.3, we make this choice in advance and stick with it for simplicity.

*Proof.* Suppose that  $h < y$ . Applying the trivial bound and using Lemma 2.1, we have

$$\sum_{\substack{t \leq n \leq t+h \\ p|n \Rightarrow p \in (y/d, y]}} 1 \leq h \leq h^{\alpha(x^2, y)} y^{1-\alpha(x^2, y)} \ll h^{\alpha(x^2, y)} (u \log u) \leq h^{\alpha(x^2, y)} \log x,$$

so the lemma holds for  $h < y$ . The case where  $h \geq y$  follows similarly to [10, Smooth Numbers Result 3]. First, we decompose the sum. Letting  $P(n)$  denote the largest prime factor of  $n$ , we have

$$\sum_{\substack{t \leq n \leq t+h \\ p|n \Rightarrow p \in (y/d, y]}} 1 = \sum_{\substack{h/y < m \leq h \\ m/P(m) \leq h/y \\ p|m \Rightarrow p \in (y/d, y]}} \sum_{\substack{\frac{t}{m} < \ell \leq \frac{t+h}{m} \\ p|\ell \Rightarrow p \in [P(m), y]}} 1 \leq h \sum_{\substack{h/y < m \leq h \\ p|m \Rightarrow p \in (y/d, y]}} \frac{1}{m}.$$

The first equality comes from the decomposition  $n = m\ell$  in a suitable way. Breaking this up into dyadic intervals, we find that

$$\sum_{\substack{t \leq n \leq t+h \\ p|n \Rightarrow p \in (y/d, y]}} 1 \leq \sum_{j=0}^{\log y} e^j \sum_{\substack{\frac{h}{e^{j+1}} < m \leq \frac{h}{e^j} \\ p|m \Rightarrow p \in (y/d, y]}} 1 \leq \sum_{j=0}^{\log y} e^j \sum_{\substack{m \leq \frac{h}{e^j} \\ p|m \Rightarrow p \in (y/d, y]}} 1.$$

Note that our bound is independent of  $t$ . Applying Rankin's trick, for any  $\beta > 0$ , we have

$$\sum_{\substack{m \leq \frac{h}{e^j} \\ p|m \Rightarrow p \in (y/d, y]}} 1 \leq \left(\frac{h}{e^j}\right)^\beta \sum_{\substack{m \leq \frac{h}{e^j} \\ p|m \Rightarrow p \in (y/d, y]}} \frac{1}{m^\beta} \leq \left(\frac{h}{e^j}\right)^\beta \prod_{y/d < p \leq y} \left(1 - \frac{1}{p^\beta}\right)^{-1}.$$

Inserting this estimate into the previous inequality and taking  $\beta = \alpha(x^2, y)$ , we find that

$$\sum_{\substack{t \leq n \leq t+h \\ p|n \Rightarrow p \in (y/d, y]}} 1 \leq h^{\alpha(x^2, y)} \prod_{y/d < p \leq y} \left(1 - \frac{1}{p^{\alpha(x^2, y)}}\right)^{-1} \sum_{j=0}^{\log y} e^{j(1-\alpha(x^2, y))}.$$

Applying Lemma 2.1, the geometric series is  $\ll u \log y = \log x$ , so it remains to evaluate the product over primes. Again, using Lemma 2.1, we find that  $\alpha(x^2, y) \geq 3/5$  for  $y \geq (\log x)^3$ , and Taylor expanding, we have

$$\prod_{y/d < p \leq y} \left(1 - \frac{1}{p^{\alpha(x^2, y)}}\right)^{-1} \ll \exp\left(\sum_{y/d < p \leq y} \frac{1}{p^{\alpha(x^2, y)}}\right),$$

Finally, we have

$$\sum_{y/d < p \leq y} \frac{1}{p^{\alpha(x^2, y)}} \leq \frac{d}{y^{\alpha(x^2, y)}} \sum_{p \leq y} 1 \leq \frac{6dy^{1-\alpha(x^2, y)}}{5 \log y} \leq \frac{3u \log u}{\log y},$$

when  $x$  is sufficiently large. Here we have used the fact that  $y^{1-\alpha(x^2, y)} \leq \frac{5}{2}u \log u$  for  $x$  sufficiently large, which follows from Lemma 2.1. Putting all of this together, we have shown that

$$\sum_{\substack{t \leq n \leq t+h \\ p|n \Rightarrow p \in (y/d, y]}} 1 \ll h^{\alpha(x^2, y)} e^{\frac{3du \log u}{\log y}} \log x,$$

as required.  $\square$

### 3. EULER PRODUCT RESULTS

In this section, we collect useful results on (random) Euler products. The main results of interest, Lemma 3.2 and Theorem 3.4, give estimates for random Euler products which will later be key to the proofs of Theorems 1.4, 5.1 and 5.2. The main feature of these results (different from most other work in the area) is that we are working on the line  $\operatorname{Re}(s) = \alpha/2$  instead of the line  $\operatorname{Re}(s) = 1/2$ , due to the smoothness condition and saddle point feature.

We begin by introducing the version of Plancherel's identity that will be used, now ubiquitous in the study of random multiplicative functions.

**Lemma 3.1** (Multiplicative Plancherel's identity). *Let  $(a_n)_{n=1}^\infty$  be a sequence of complex numbers, and let  $A(s) = \sum_{n=1}^\infty \frac{a_n}{n^s}$  denote the corresponding Dirichlet series, and  $\sigma_c$  the abscissa of convergence. Then for any  $\sigma > \max\{0, \sigma_c\}$ , we have*

$$\int_0^\infty \frac{|\sum_{1 \leq n \leq x} a_n|^2}{x^{1+2\sigma}} dx = \frac{1}{2\pi} \int_{-\infty}^\infty \left| \frac{A(\sigma + it)}{\sigma + it} \right|^2 dt.$$

*Proof.* This is [22, Equation (5.26)].  $\square$

The next lemma allows us to obtain cancellation in random Euler product integrals on the line  $\operatorname{Re}(s) = \alpha/2$ , provided that the truncation length is small enough in terms of  $\alpha$ .

**Lemma 3.2** (Euler product close to half line). *Let  $c = 2e^{-2}$ . For  $4/5 \leq \alpha < 1$ , we define  $z = e^{\frac{c}{1-\alpha}}$ , so that  $F_z(s) = \prod_{p \leq e^{\frac{c}{1-\alpha}}} (1 - \frac{f(p)}{p^s})^{-1}$ . Then we have*

$$\mathbb{E} \left( \int_{-1/2}^{1/2} \left| F_z \left( \frac{\alpha}{2} + it \right) \right|^2 dt \right)^{2/3} \ll \left( \frac{1}{(1-\alpha) \sqrt{\log \frac{1}{1-\alpha}}} \right)^{2/3}.$$

*Proof.* Let  $x \geq 10$ , and define  $F_{x,k}(s) := \prod_{p \leq x e^{-(k+1)}} (1 - \frac{f(p)}{p^s})^{-1}$ . We then have

$$\mathbb{E} \left( \frac{e^k (1-q) \sqrt{\log \log x}}{\log x} \int_{-1/2}^{1/2} \left| F_{x,k} \left( 1/2 - \frac{k}{\log x} + it \right) \right|^2 dt \right)^q \ll 1,$$

uniformly for all  $0 \leq k \leq \lfloor \log \log x \rfloor$  and  $2/3 \leq q \leq 1 - \frac{1}{\sqrt{\log \log x}}$ . This is proved in [12] in the section titled: ‘‘Proof of the upper bound in Theorem 1, assuming Key Propositions



1 and 2". To obtain the lemma, we apply this result with  $k = 1$ ,  $x = e^{\frac{1}{1-\alpha}}$ , and  $q = 2/3$ . Since

$$\mathbb{E} \left( \int_{-1/2}^{1/2} \left| F_z \left( \frac{\alpha}{2} + it \right) \right|^2 dt \right)^{2/3} = \mathbb{E} \left( \int_{-1/2}^{1/2} \left| F_{x,1} \left( 1/2 - \frac{1}{\log x} + it \right) \right|^2 dt \right)^{2/3},$$

the result follows.  $\square$

We now state a general expectation result that will help us to prove Theorem 3.4. Again, we care about the case that the Euler product is not necessarily evaluated on the critical line.

**Lemma 3.3** (Expectation evaluation). *If  $f$  is a Steinhaus random multiplicative function, then uniformly for any  $k \in \mathbb{N}$ , any  $\beta \geq 3/4$ , any real  $\alpha_1, \dots, \alpha_k$ , any real  $100(1 + (\sum_{j=1}^k |\alpha_j|)^{8/3}) \leq x \leq y$  and  $t_1, \dots, t_k$ , we have*

$$\begin{aligned} & \mathbb{E} \prod_{j=1}^k \prod_{x < p \leq y} \left| 1 - \frac{f(p)}{p^{\beta/2 + it_k}} \right|^{-2\alpha_j} \\ &= \exp \left( \sum_{x < p \leq y} \frac{\sum_{j=1}^k \alpha_j^2 + 2 \sum_{1 \leq j < l \leq k} \alpha_j \alpha_l \cos((t_l - t_j) \log p)}{p^\beta} + O \left( \frac{M}{x^{1/8} \log x} \right) \right), \end{aligned}$$

where  $M := \max \{ \sum_{j=1}^k |\alpha_j|, (\sum_{j=1}^k |\alpha_j|)^3 \}$ . In particular, for any  $|\alpha| \leq 100$ , say, we have

$$\mathbb{E} |F_y(\beta/2 + it)|^{2\alpha} \asymp \zeta(\beta, y)^{\alpha^2},$$

where the implied constant is uniform for  $\alpha$  in this range. Furthermore, if  $|t - t'| \leq \frac{1}{\delta \log y}$  for  $\delta \geq 1$ , then we have

$$\mathbb{E} \left| \frac{F_y(\beta/2 + it)}{F_y(\beta/2 + it')} \right|^{2\alpha} \leq e^{O(1+|\alpha|^{11/3})} \zeta(\beta, y)^{\frac{\alpha^2}{\delta^2}}, \quad (3.1)$$

where the implied constant in the  $e^{O(1+|\alpha|^{11/3})}$  term is uniform in  $\alpha$ .

*Proof.* The first part of the statement follows identically to [13, Euler Product Result 1]. We then note that

$$\mathbb{E} |F_y(\beta/2 + it)|^{2\alpha} \asymp \mathbb{E} \prod_{M^8 < p \leq y} \left| 1 - \frac{f(p)}{p^{\beta/2 + it}} \right|^{-2\alpha}.$$

Applying the first part of the lemma with  $k = 1$ , and using the fact that

$$\begin{aligned} \zeta(\beta, y) &= \exp \left( - \sum_{p \leq y} \log \left( 1 - \frac{1}{p^\beta} \right) \right) \\ &= \exp \left( \sum_{p \leq y} \frac{1}{p^\beta} + O \left( \sum_{p \leq y} \frac{1}{p^{2\beta}} \right) \right) \asymp \exp \left( \sum_{p \leq y} \frac{1}{p^\beta} \right), \end{aligned}$$

for  $\beta \geq 3/4$ , we deduce that  $\mathbb{E}|F_y(\beta/2+it)|^{2\alpha} \asymp \zeta(\beta, y)^{\alpha^2}$ . For the last part of the lemma, we begin by noting that

$$\mathbb{E} \left| \frac{F_y(\beta/2+it)}{F_y(\beta/2+it')} \right|^{2\alpha} \leq e^{O(1+|\alpha|^{11/3})} \mathbb{E} \prod_{x < p \leq y} \left| 1 + \frac{f(p)}{p^{\beta/2+it}} \right|^{-2\alpha} \left| 1 + \frac{f(p)}{p^{\beta/2+it'}} \right|^{2\alpha},$$

for  $x = 100(1+2|\alpha|)^{8/3}$ , say. Applying the first part of the lemma, we find that

$$\mathbb{E} \left| \frac{F_y(\beta/2+it)}{F_y(\beta/2+it')} \right|^{2\alpha} \leq e^{O(1+|\alpha|^{11/3})} \exp \left( \sum_{x < p \leq y} \frac{\alpha^2(1 - \cos((t-t') \log p))}{p^\beta} \right).$$

Note that the “big Oh” term in the last exponential factor can be engulfed into the  $e^{O(1+|\alpha|^{11/3})}$  term. Using the facts that  $|\cos x - 1| \leq |x|^2$  and  $|t - t'| \leq \frac{1}{\delta \log y}$ , we have

$$\mathbb{E} \left| \frac{F_y(\beta/2+it)}{F_y(\beta/2+it')} \right|^{2\alpha} \leq e^{O(1+|\alpha|^{11/3})} \exp \left( \sum_{p \leq y} \frac{\alpha^2}{\delta^2 p^\beta} \right).$$

Finally, it follows similarly that  $\exp \left( k \sum_{p \leq y} \frac{1}{p^\beta} \right) = e^{O(k)} \zeta(\beta, y)^k$ , where the implied constant is uniform in  $k$ , so applying this with  $k = \frac{\alpha^2}{\delta^2}$  completes the proof.  $\square$

We are now ready to state and prove our main Euler product estimate.

**Theorem 3.4** (Main Euler product result). *Suppose that  $3/4 \leq \beta \leq 1$ . For any  $1/2 \leq q \leq 1$ , we have uniformly*

$$\zeta(\beta, y)^{q^2} \ll \mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2+it)}{\beta/2+it} \right|^2 dt \right)^q \ll (\log y)^{1-q} \zeta(\beta, y)^{q^2} \exp \left( O(\log \zeta(\beta, y))^{8/11} \right).$$

The lower bound holds uniformly on the larger range  $0 \leq q \leq 1$ .

*Remark.* This lower bound can likely be improved to include a factor of  $(\log y)^{1-q}$ , but we content ourselves with this weaker bound, since we only include it to illustrate that our theorem is close to being sharp.

*Proof.* The proof of the lower bound is relatively straightforward. For any  $0 < q \leq 1$ , it holds that

$$\begin{aligned} \mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2+it)}{\beta/2+it} \right|^2 dt \right)^q &\gg \mathbb{E} \left( \int_{-1/2}^{1/2} |F_y(\beta/2+it)|^2 dt \right)^q \\ &\geq \int_{-1/2}^{1/2} \mathbb{E} |F_y(\beta/2+it)|^{2q} dt, \end{aligned}$$

where the second line follows from applying Hölder's inequality to the integral. An application of Lemma 3.3 gives that

$$\mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2+it)}{\beta/2+it} \right|^2 dt \right)^q \gg \zeta(\beta, y)^{q^2},$$

as required. For the upper bound, we begin by splitting the integral and applying sub-additivity of  $x \rightarrow x^q$  ( $0 < q < 1$ ), so we obtain

$$\begin{aligned} \mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q &\ll \mathbb{E} \left( \int_{|t| \leq \zeta(\beta, y)} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q \\ &\quad + \mathbb{E} \left( \int_{|t| > \zeta(\beta, y)} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q. \end{aligned} \quad (3.2)$$

Applying Hölder's inequality followed by Lemma 3.3, the second term satisfies

$$\ll \left( \int_{|t| > \zeta(\beta, y)} \mathbb{E} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q \ll 1.$$

In light of our lower bound, this contributes negligibly to (3.2), so we have

$$\begin{aligned} \mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q &\ll \mathbb{E} \left( \int_{|t| \leq \zeta(\beta, y)} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q \\ &\ll \mathbb{E} \left( \sum_{|n| \leq \zeta(\beta, y)} \frac{1}{(|n| + 1)^2} \int_{n-1/2}^{n+1/2} |F_y(\beta/2 + it)|^2 dt \right)^q \\ &\ll \sum_{|n| \leq \zeta(\beta, y)} \frac{1}{(|n| + 1)^{2q}} \mathbb{E} \left( \int_{n-1/2}^{n+1/2} |F_y(\beta/2 + it)|^2 dt \right)^q, \end{aligned}$$

where the last line follows from sub-additivity of  $x \rightarrow x^q$ . Seeing as the law of  $(f(p))_{p \text{ prime}}$  is the same as that  $(f(p)p^{it})_{p \text{ prime}}$  for any fixed  $t$ , we have

$$\begin{aligned} \mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q &\ll \sum_{|n| \leq \zeta(\beta, y)} \frac{1}{(|n| + 1)^{2q}} \mathbb{E} \left( \int_{-1/2}^{1/2} |F_y(\beta/2 + it)|^2 dt \right)^q \\ &\ll \log(\zeta(\beta, y)) \mathbb{E} \left( \int_{-1/2}^{1/2} |F_y(\beta/2 + it)|^2 dt \right)^q, \end{aligned}$$

uniformly for any  $1/2 \leq q \leq 1$ . Note that if  $q > 1/2$  is fixed, the factor of  $\log \zeta(\beta, y)$  can be removed, since in that case the sum over  $n$  is convergent. It remains to estimate

$$\mathbb{E} \left( \int_{-1/2}^{1/2} |F_y(\beta/2 + it)|^2 dt \right)^q.$$

As is motivated in Section 1.3, we want to discretize the interval  $[-\frac{1}{2}, \frac{1}{2}]$  further into short sub-intervals (of length roughly  $\frac{1}{\log y}$ ), and then apply sub-additivity. We first introduce a parameter,  $\delta > 0$ , which will be chosen to depend on  $\beta$ . Then (3.1) is certainly at most

$$\frac{1}{(\delta \log y)^q} \mathbb{E} \left( \sum_{|j| \leq \delta \log y} \left| F_y \left( \beta/2 + i \frac{j}{\delta \log y} \right) \right|^2 (\delta \log y) \int_{\frac{j}{\delta \log y}}^{\frac{j+1}{\delta \log y}} \left| \frac{F_y(\beta/2 + it)}{F_y(\beta/2 + i \frac{j}{\delta \log y})} \right|^2 dt \right)^q.$$

The factor of  $\frac{1}{(\delta \log y)^q}$  has been introduced to normalise the integral. For  $q \in [1/2, 1]$ , the function  $x \rightarrow x^q$  is sub-additive, so we can move this sum over  $j$  to the outside, giving

the upper bound

$$\frac{1}{(\delta \log y)^q} \sum_{|j| \leq \delta \log y} \mathbb{E} \left[ \left| F_y \left( \beta/2 + i \frac{j}{\delta \log y} \right) \right|^{2q} \left( (\delta \log y) \int_{\frac{j}{\delta \log y}}^{\frac{j+1}{\delta \log y}} \left| \frac{F_y(\beta/2 + it)}{F_y(\beta/2 + i \frac{j}{\delta \log y})} \right|^2 dt \right)^q \right].$$

We now apply Hölder's inequality with an exponent  $1 + \varepsilon$  to the expectation, where  $0 < \varepsilon < 1$  that will be chosen later. Overall, this gives

$$\begin{aligned} \mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q &\leq \frac{\log \zeta(\beta, y)}{(\delta \log y)^q} \sum_{|j| \leq \delta \log y} \left( \mathbb{E} \left| F_y \left( \beta/2 + i \frac{j}{\delta \log y} \right) \right|^{2q(1+\varepsilon)} \right)^{\frac{1}{1+\varepsilon}} \\ &\quad \times \left[ \mathbb{E} \left( (\delta \log y) \int_{\frac{j}{\delta \log y}}^{\frac{j+1}{\delta \log y}} \left| \frac{F_y(\beta/2 + it)}{F_y(\beta/2 + i \frac{j}{\delta \log y})} \right|^2 dt \right)^{q(1+\frac{1}{\varepsilon})} \right]^{\frac{\varepsilon}{1+\varepsilon}}. \end{aligned}$$

By Lemma 3.3, the first expectation is  $\ll \zeta(\beta, y)^{q^2(1+\varepsilon)^2} \ll \zeta(\beta, y)^{q^2(1+3\varepsilon)}$ , so we have the upper bound

$$\ll \frac{\log \zeta(\beta, y)}{(\delta \log y)^q} \sum_{|j| \leq \delta \log y} \zeta(\beta, y)^{q^2(1+3\varepsilon)} \left[ \mathbb{E} \left( (\delta \log y) \int_{\frac{j}{\delta \log y}}^{\frac{j+1}{\delta \log y}} \left| \frac{F_y(\beta/2 + it)}{F_y(\beta/2 + i \frac{j}{\delta \log y})} \right|^2 dt \right)^{q(1+\frac{1}{\varepsilon})} \right]^{\frac{\varepsilon}{1+\varepsilon}}.$$

Applying Hölder's inequality with exponent  $q(1 + \frac{1}{\varepsilon})$  to the integral, this is

$$\ll \frac{\log \zeta(\beta, y)}{(\delta \log y)^q} \sum_{|j| \leq \delta \log y} \zeta(\beta, y)^{q^2(1+3\varepsilon)} \left( (\delta \log y) \int_{\frac{j}{\delta \log y}}^{\frac{j+1}{\delta \log y}} \mathbb{E} \left| \frac{F_y(\beta/2 + it)}{F_y(\beta/2 + i \frac{j}{\delta \log y})} \right|^{2q(1+\frac{1}{\varepsilon})} dt \right)^{\frac{\varepsilon}{1+\varepsilon}}.$$

Applying Lemma 3.3, the expectation in our integral is  $\leq e^{O(1+1/\varepsilon^{11/3})} \zeta(\beta, y)^{\frac{q^2}{\delta}(1+\frac{1}{\varepsilon})^2}$ . We deduce that

$$\mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q \ll (\log \zeta(\beta, y)) \delta^{1-q} (\log y)^{1-q} e^{O(1/\varepsilon^{8/3})} \zeta(\beta, y)^{q^2(1+3\varepsilon+\frac{1}{\delta^2}+\frac{1}{\delta^2\varepsilon})}.$$

Taking  $\delta = (\log \zeta(\beta, y))^{1/2}$  and  $\varepsilon = (\log \zeta(\beta, y))^{-3/11}$ , we have

$$\mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_y(\beta/2 + it)}{\beta/2 + it} \right|^2 dt \right)^q \ll (\log y)^{1-q} e^{O((\log \zeta(\beta, y))^{8/11})} \zeta(\beta, y)^{q^2},$$

which completes the proof.  $\square$

#### 4. LARGE SMOOTHNESS PARAMETER: MULTIPLICATIVE CHAOS REGIME

In this section, we prove Theorem 1.4. As mentioned in section 1.4, the proof essentially follows that of Harper [12], but in our case, we only condition on relatively few primes. Since the remaining random Euler product behaves as though it is on the half-line, we are in a similar setting to Harper [12], and we can apply Lemma 3.2 to obtain a non-trivial saving. The additional input needed for this strategy is an estimate for the count of smooth numbers in short intervals without small prime factors, which is provided by Lemma 2.5.

The theorem will follow by iteratively applying the following proposition:

**Proposition 4.1.** *Let  $v > 2$  be a large constant so that  $\alpha(x, y) \leq 1 - \frac{10}{\log y}$  whenever  $y \leq x^{1/v}$  (it can be verified using Lemma 2.1 that such a constant exists), and let  $c = 2e^{-2}$ . For any  $e^{(\log \log x)^2} \leq y \leq x^{1/v}$ , we have*

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| \ll \frac{\Psi(x, y)^{1/2}}{\left( \log \frac{1}{1-\alpha(x, y)} \right)^{1/4}},$$

where the implied constant is uniform in  $y$ .

*Remark.* We reiterate that the constant  $c = 2e^{-2}$  is unimportant, and is just chosen to allow for straightforward application of Lemma 3.2 later on.

*Proof of Theorem 1.4, assuming Proposition 4.1.* We begin by noting that, in the case where  $u = \frac{\log x}{\log y} \leq v$  (i.e. when  $x^{1/v} \leq y \leq x$ ), the theorem follows identically to the work of Harper<sup>2</sup> [12] (which is not surprising as  $y \geq x^{1/v}$  is now very close to  $x$ , and so the smoothness restriction is negligible). Now suppose that  $y \leq x^{1/v}$ . By the triangle inequality, we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq \mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| + \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|. \quad (4.1)$$

If  $u \geq \log \log \log x$ , then the first term can easily be shown to dominate. Specifically, in this range of  $u$ , we apply Proposition 4.1 to handle the first term and Cauchy–Schwarz inequality to handle the second term (together with a mean square calculation), giving

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq \frac{\Psi(x, y)^{1/2}}{\left( \log \frac{1}{1-\alpha(x, y)} \right)^{1/4}} + \Psi(x, e^{\frac{c}{1-\alpha}})^{1/2}.$$

It follows readily from Lemma 2.1 that  $e^{\frac{c}{1-\alpha}} \leq y^{1/2}$ , say, and a short calculation using Lemma 2.2 allows one to deduce that

$$\Psi(x, \sqrt{y}) \ll \Psi(x, y) e^{-u \log u / 3} \ll \frac{\Psi(x, y)}{\left( \log \frac{1}{1-\alpha(x, y)} \right)^{1/2}},$$

and the result follows immediately.

It remains to handle the case where  $v \leq u \leq \log \log \log x$ . The main idea is to iteratively apply the decomposition (4.1) and then use Proposition 4.1 to evaluate the resulting

---

<sup>2</sup>For example, one can follow the proof of Proposition 1 in [12, Section 2.4], just removing a finite number of  $k$  in the initial splitting of the sum. Since this will only decrease the overall size, the same upper bound follows.

absolute first moment. Define  $y_0 = y$ , and  $y_{i+1} = e^{\frac{c}{1-\alpha(x, y_i)}}$  for  $i \geq 0$ . Iterating (4.1)  $\ell$ -times, we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq \sum_{k=0}^{\ell-1} \mathbb{E} \left| \sum_{\substack{n \leq x \\ y_{k+1} < P(n) \leq y_k}} f(n) \right| + \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y_\ell}} f(n) \right|.$$

We apply Proposition 4.1 to the first  $\ell$  terms and Cauchy–Schwarz inequality to the latter term, giving

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq \sum_{k=0}^{\ell-1} \frac{\Psi(x, y_k)^{1/2}}{\left( \log \frac{1}{1-\alpha(x, y_k)} \right)^{1/4}} + \Psi(x, y_\ell).$$

It follows from our conditions on  $y$  and Lemma 2.1 that  $y_1 = e^{\frac{c}{1-\alpha(x, y_0)}} \leq \sqrt{y_0} = \sqrt{y}$ , and similarly we have  $y_k \leq y^{1/2^k}$ . Taking  $\ell = \lfloor \log \log \log x \rfloor$ , say, we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq \frac{1}{(\log \log x)^{1/4}} \sum_{k=0}^{\ell-1} \Psi(x, y^{1/2^k})^{1/2} + \Psi(x, y^{1/2^\ell}).$$

We now make use of the relation between  $\Psi(x, y)$  and the Dickman function,  $\rho$ . It follows from [25, Theorem III.5.8] that  $\Psi(x, y^{1/2^k}) \asymp x\rho(2^k u)$  uniformly for  $k$  in our range of summation. Applying the estimate  $\frac{\rho(2u)}{\rho(u)} \leq \frac{1}{2}$  that holds for all  $u \geq 1$  (since  $\rho$  is rapidly decreasing, this ratio is maximised at  $u = 1$ ), we deduce that  $\Psi(x, y^{1/2^k}) \ll \frac{x}{2^k} \rho(u) \ll \frac{x}{2^k} \Psi(x, y)$ . Finally, since  $\Psi(x, y^{1/2^\ell}) \ll \Psi(x, x^{\frac{1}{\log \log x}}) \ll \Psi(x, y) / \log \log x$ , say, we deduce that

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq \frac{\Psi(x, y)^{1/2}}{(\log \log x)^{1/4}},$$

which agrees with Theorem 1.4 on this range of  $u$ . This completes the proof.  $\square$

We now proceed with the proof of Proposition 4.1.

*Proof of Proposition 4.1.* Recall that  $c = 2e^{-2}$ . We have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| = \mathbb{E} \left| \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} f(m) \sum_{\substack{n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|.$$

Let  $\mathbb{E}_\alpha$  denote the expectation conditioned on primes  $p \leq e^{\frac{c}{1-\alpha}}$ . By the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| &= \mathbb{E} \mathbb{E}_\alpha \left| \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} f(m) \sum_{\substack{n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right| \\ &\leq \mathbb{E} \left( \mathbb{E}_\alpha \left| \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} f(m) \sum_{\substack{n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \right)^{1/2} \\ &= \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \left| \sum_{\substack{n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \right)^{1/2}. \end{aligned}$$

We wish to replace the sum over  $m$  by an integral, which will in turn allow us to apply Plancherel’s identity. First, we introduce a “dummy integral” over  $t$ , writing

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| &\leq \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \left| \sum_{\substack{n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \right)^{1/2} \\ &= \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 dt \right)^{1/2}. \end{aligned}$$

The first term on the right-hand is

$$\begin{aligned} &\ll \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{n \leq x/t \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 dt \right)^{1/2} + \\ &\quad \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{x/t < n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 dt \right)^{1/2}. \end{aligned}$$

We show that the second term is small. By the Cauchy–Schwarz inequality, we can bound it from above by

$$\left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \sum_{\substack{x/t < n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} 1 dt \right)^{1/2} \ll \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \sum_{\substack{x/m(1+\frac{1}{X}) < n \leq x/m \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} 1 \right)^{1/2}.$$

We make the choice  $X = \log \log x$ . Change the order of summation, and apply Lemma 2.3 and Lemma 2.5 to get that (Noting carefully that the conditions for Lemma 2.5 are



satisfied in our application below)

$$\begin{aligned} & \ll \left( \sum_{\substack{n \leq x \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} \sum_{\substack{x/n(1+\frac{1}{X}) < m \leq x/n \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} 1 \right)^{1/2} \ll \left( (1-\alpha) \sum_{\substack{n \leq x \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} \frac{1}{X} \Psi(x/n, y) \right)^{1/2} \\ & \ll \left( \frac{\Psi(x, y)(1-\alpha)}{X} \sum_{\substack{n \leq x \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} \frac{1}{n^\alpha} \right)^{1/2} \ll \frac{\Psi(x, y)^{1/2}}{X^{1/2}} = \sqrt{\frac{\Psi(x, y)}{\log \log x}}. \end{aligned}$$

At this stage, we have

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| & \ll \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{n \leq x/t \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 dt \right)^{1/2} + \sqrt{\frac{\Psi(x, y)}{\log \log x}} \\ & = \mathbb{E} \left( \int_{e^{\frac{c}{1-\alpha}}}^x \left| \sum_{\substack{n \leq x/t \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \frac{X}{m} dt \right)^{1/2} + \sqrt{\frac{\Psi(x, y)}{\log \log x}}. \end{aligned}$$

Applying Lemma 2.5, we have

$$\sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (e^{\frac{c}{1-\alpha}}, y]}} \frac{X}{m} \ll (1-\alpha) \frac{\Psi(t, y)}{t},$$

uniformly for  $t \geq y$ . In the remaining range  $e^{\frac{c}{1-\alpha}} \leq t < y$ , the estimate follows using a simple sieve argument (see, for example, [22, Theorem 3.6]), sieving out all primes  $p \leq (e^{\frac{c}{1-\alpha}})^{1/10}$ , say. Making the change of variables  $t = x/z$ , we have

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| & \ll (1-\alpha)^{1/2} \mathbb{E} \left( \int_1^{x/e^{\frac{1}{1-\alpha}}} \left| \sum_{\substack{n \leq z \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \Psi(x/z, y) \frac{dz}{z} \right)^{1/2} \\ & \quad + \sqrt{\frac{\Psi(x, y)}{\log \log x}}. \end{aligned}$$

Applying Lemma 2.3, we have  $\Psi(x/z, y) \ll \Psi(x, y)/z^\alpha$ , and completing the integral, we obtain

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| & \ll (\Psi(x, y)(1-\alpha))^{1/2} \mathbb{E} \left( \int_1^\infty \left| \sum_{\substack{n \leq z \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \frac{dz}{z^{1+\alpha}} \right)^{1/2} \\ & \quad + \sqrt{\frac{\Psi(x, y)}{\log \log x}}. \end{aligned} \tag{4.2}$$

The second term is negligible for our upper bound. We next manipulate the first term so that it can be bounded by using Lemma 3.2, and for this purpose, we let  $z = e^{\frac{c}{1-\alpha}}$ , so that  $F_z(s) = \prod_{p \leq e^{\frac{c}{1-\alpha}}} (1 - \frac{f(p)}{p^s})^{-1}$ . Applying Plancherel's identity, Lemma 3.1, followed by Hölder's inequality, we obtain

$$\begin{aligned} \mathbb{E} \left( \int_1^\infty \left| \sum_{\substack{n \leq z \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \frac{dz}{z^{1+\alpha}} \right)^{1/2} &= \mathbb{E} \left( \int_{-\infty}^\infty \left| \frac{F_z(\frac{\alpha}{2} + it)}{\frac{\alpha}{2} + it} \right|^2 dt \right)^{1/2} \\ &\leq \left( \mathbb{E} \left( \int_{-\infty}^\infty \left| \frac{F_z(\frac{\alpha}{2} + it)}{\frac{\alpha}{2} + it} \right|^2 dt \right)^{2/3} \right)^{3/4}. \end{aligned}$$

The above expectation satisfies

$$\begin{aligned} \mathbb{E} \left( \int_{-\infty}^\infty \left| \frac{F_z(\frac{\alpha}{2} + it)}{\frac{\alpha}{2} + it} \right|^2 dt \right)^{2/3} &\ll \mathbb{E} \left( \sum_{n \in \mathbb{Z}} \frac{1}{|n|^2 + 1} \int_{n-1/2}^{n+1/2} \left| F_z\left(\frac{\alpha}{2} + it\right) \right|^2 dt \right)^{2/3} \\ &\ll \sum_{n \in \mathbb{Z}} \frac{1}{n^{4/3} + 1} \mathbb{E} \left( \int_{n-1/2}^{n+1/2} \left| F_z\left(\frac{\alpha}{2} + it\right) \right|^2 dt \right)^{2/3} \\ &\ll \mathbb{E} \left( \int_{-1/2}^{1/2} \left| F_z\left(\frac{\alpha}{2} + it\right) \right|^2 dt \right)^{2/3}. \end{aligned}$$

where to obtain the last line, we have used translation invariance in law. Finally, by Lemma 3.2 with  $q = 2/3$ , we have

$$\mathbb{E} \left( \int_{-1/2}^{1/2} \left| F_z\left(\frac{\alpha}{2} + it\right) \right|^2 dt \right)^{2/3} \ll \left( \frac{1}{(1-\alpha)\sqrt{\log \frac{1}{1-\alpha}}} \right)^{2/3},$$

so overall this gives

$$\mathbb{E} \left( \int_1^\infty \left| \sum_{\substack{n \leq z \\ P(n) \leq e^{\frac{c}{1-\alpha}}}} f(n) \right|^2 \frac{dz}{z^{1+\alpha}} \right)^{1/2} \ll \frac{1}{(1-\alpha)^{1/2} (\log \frac{1}{1-\alpha})^{1/4}}.$$

Inserting this into (4.2), we deduce that

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ e^{\frac{c}{1-\alpha}} < P(n) \leq y}} f(n) \right| \ll \frac{\Psi(x, y)^{1/2}}{\left( \log \frac{1}{1-\alpha(x, y)} \right)^{1/4}},$$

which completes the proof of Proposition 4.1.  $\square$

## 5. MODERATELY SMOOTH NUMBERS: PROOF OF THEOREM 1.3

**5.1. Separating into two ranges of smoothness parameter.** In this section, we prove Theorem 1.3, which we split into two cases depending on the size of  $y$ . When  $y$  is somewhat “large” compared to  $x$ , say  $y \gg e^{(\log x)^{1/2+o(1)}}$ , the proof will follow roughly the same structure as the proof of Proposition 4.1, first conditioning on some portion

of the primes to obtain a mean square of partial sums. In this case, we condition on all  $(f(p))_{p \leq y/2}$ . Note that if we condition on fewer primes (and apply Cauchy–Schwarz inequality to more  $f(p)$ ), then we lose significantly. This is because we are working far from the  $1/2$ -line, so large primes contribute more to the Euler product, and it is the expectation of the *random* Euler product integrals that ultimately gives us a larger saving, as described in Section 1.3. On this range of larger  $y$ , we will prove the following result:

**Theorem 5.1.** *For  $e^{\sqrt{10 \log x}} \leq y \leq x^{\frac{1}{\log \log x}}$ , we have*

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \Psi(x^2, y)^{1/4} \exp\left(O(u^{8/11})\right) (\log x)^{5/8} (\log y)^{13/8},$$

where all implied constants are uniform for  $y$  in this range.

We will prove this in Section 5.2. The result will be a direct consequence of Proposition 5.3, which proves a similar result over  $n \leq x$  that have  $P(n) \in (y/2, y]$ .

Unfortunately, when  $y$  is smaller than roughly  $e^{(\log x)^{1/2+o(1)}}$ , the proof strategy used in Proposition 5.3 breaks down, which is a byproduct of our saving being too large relative to  $y$ . We are able to circumvent these issues and prove the following:

**Theorem 5.2.** *For  $(\log x)^{20} \leq y \leq e^{\sqrt{10 \log x}}$ , we have*

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \Psi(x^2, y)^{1/4} \exp\left(\frac{10u \log u}{\log y} + O(u^{8/11})\right),$$

where all implied constants are uniform for  $y$  in this range.

This result is proved in Section 5.3. We will explain why we need to handle these ranges differently, but first, we show how one derives Theorem 1.3 from these results.

*Proof of Theorem 1.3, assuming Theorems 5.1 and 5.2.* It follows immediately from these results that

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \Psi(x^2, y)^{1/4} \exp\left(\frac{10u \log u}{\log y} + O(u^{8/11})\right) (\log x)^{5/8} (\log y)^{13/8}, \quad (5.1)$$

for  $(\log x)^{20} \leq y \leq x^{\frac{1}{\log \log x}}$ . We will see that this result only gives a non-trivial saving in a smaller range that includes  $(\log x)^{30} \leq y \leq x^{\frac{1}{7 \log \log x}}$ , and we will also prove the statement (1.3). First note that Lemma 2.2 implies that

$$\Psi(x^2, y) \asymp \Psi(x, y)^2 x^{2(\alpha(x^2, y) - \alpha(x, y))} \frac{\zeta(\alpha(x^2, y), y)}{\zeta(\alpha(x, y), y)^2} \sqrt{\log x \log y},$$

for  $(\log x)^{20} \leq y \leq x^{\frac{1}{\log \log x}}$ . In the same range, it follows from Lemma 2.1 that

$$x^{2(\alpha(x^2, y) - \alpha(x, y))} = \exp(-2u \log 2(1 + o(1))),$$

where the  $o(1)$  term goes to zero uniformly. Note that for  $(\log x)^3 \leq y \leq x$ , we have

$$\log \zeta(\alpha(x, y), y) = \sum_{p \leq y} \frac{1}{p^{\alpha(x, y)}} + O(1).$$

We can evaluate this using [22, Lemma 7.4] to find that

$$\log \zeta(\alpha, y) = \frac{y^{1-\alpha}}{(1-\alpha) \log y} + \log\left(\frac{1}{1-\alpha}\right) + O\left(\frac{y^{1-\alpha}}{(1-\alpha)^2 (\log y)^2}\right).$$

Applying Lemma 2.1 and evaluating, we obtain

$$\log \zeta(\alpha, y) = u + \log\left(\frac{1}{1-\alpha}\right) + O\left(\frac{u \log \log u}{\log u}\right), \quad (5.2)$$

uniformly for  $(\log x)^{20} \leq y \leq x^{\frac{1}{\log \log x}}$ . Note that the first term dominates whenever  $\log \log x = o(u)$ . We then have

$$\Psi(x^2, y) \asymp \Psi(x, y)^2 \exp(-2u \log 2(1 + o(1))) \left(\frac{(1 - \alpha(x, y))^2}{1 - \alpha(x^2, y)}\right) \sqrt{\log x \log y},$$

and applying Lemma 2.1, we have  $\left(\frac{(1 - \alpha(x, y))^2}{1 - \alpha(x^2, y)}\right) \sqrt{\log x \log y} \asymp u \log u$ , so we find that

$$\Psi(x^2, y) \asymp \Psi(x, y)^2 \exp(-2u \log 2(1 + o(1))), \quad (5.3)$$

uniformly for  $(\log x)^{20} \leq y \leq x^{\frac{1}{\log \log x}}$ . Inserting this estimate into equation (5.1), it follows that

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \sqrt{\Psi(x, y) \exp\left(-u \left(\log 2 - \frac{20 \log u}{\log y} + o(1)\right)\right) (\log x)^{5/4} (\log y)^{13/4}}.$$

uniformly for  $(\log x)^{20} \leq y \leq x^{\frac{1}{\log \log x}}$ . Now, in the range  $e^{(\log \log x)^{1+\varepsilon}} \leq y \leq x^{\frac{1}{(\log \log x)^{1+\varepsilon}}}$ , say, we have  $\frac{\log u}{\log y} = o(1)$  and  $\log \log x = o(u)$ , where the  $o(1)$  terms goes to zero uniformly for fixed  $\varepsilon > 0$ . Therefore, equation (1.3) holds on this range of  $y$ , as required. Furthermore, a simple calculation shows that this result gives better than squareroot cancellation for  $y \geq (\log x)^C$ , so long as  $C > \frac{20}{\log 2} \approx 28.854$ , and we also require that  $u > \frac{9}{2 \log 2} \log \log x \approx 6.492 \log \log x$  to overcome the  $(\log x)^{5/4} (\log y)^{13/4}$  term. The range  $(\log x)^{30} \leq y \leq x^{\frac{1}{7 \log \log x}}$  suffices, and we always have a saving of at least  $e^{cu}$  for some fixed  $c > 0$  on this range.  $\square$

We now briefly elaborate on the reason that our proof strategy cannot cover the full range of  $y$ . Similarly to in Proposition 4.1, in the proof of Proposition 5.3, we have a “smoothing step” where we work to approximate an outer sum with an integral. This step, which is done at around equation (5.4) in the proof of Proposition 5.3, introduces a dichotomy that is mediated by some parameter  $X$  of our choosing. Specifically, the smoothing step:

- (i) Introduces an error roughly of size  $\sqrt{\Psi(x, y)/X}$ . In order for this term not to dominate the main term, we need to take  $X \gg e^{cu}$ , for some  $c > \log 2$ , say.
- (ii) Leads us to estimate sums of the form

$$\sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m},$$

for  $t \in (y/2, x]$ , which we need to bound when evaluating the integrand in our main term (in Proposition 5.3, this step is performed in equation (5.6)). To handle the case when  $t$  is small, we need  $y/X \gg 1$  if we want to obtain a meaningful estimate.

In order to satisfy these conditions simultaneously, we need to have  $y \gg e^{cu}$ , say, which roughly corresponds to  $y \geq e^{(\log x)^{1/2+o(1)}}$ . To circumvent this problem and handle the case when  $y$  is small, we perform a technical splitting of our original sum (see equation (5.7)), which relies on the fact that, even though the part of  $n$  coming from primes  $p \in (y/2, y]$  (say) could be small, for the vast majority of  $n \leq x$  with  $P(n) \leq y$ , there will be a large factor of  $n$  divisible only by primes on some (potentially different) scale. This idea allows us to essentially boost the lower range of  $t$  in the short sum above and circumvent this problem.

**5.2. Proof of Theorem 5.1.** As mentioned above, the theorem will be a straightforward consequence of the following proposition:

**Proposition 5.3.** *For  $e^{\sqrt{2 \log x}} \leq y \leq x^{\frac{1}{\log \log x}}$ , we have*

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \in (y/2, y]}} f(n) \right| \ll \Psi(x^2, y)^{1/4} e^{O(u^{8/11})} (\log x)^{5/8} (\log y)^{5/8}.$$

*Proof of Theorem 5.1, assuming Proposition 5.3.* We will take advantage of the fact that the density of smooth numbers decreases rapidly as the smoothness parameter decreases. By Hölder's inequality, we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq \frac{y}{2^{\lceil \frac{1}{2} \log_2 y \rceil}}}} f(n) \right| + \sum_{j=0}^{\lceil \frac{1}{2} \log_2 y \rceil} \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \in (\frac{y}{2^{j+1}}, \frac{y}{2^j}]}} f(n) \right|,$$

where, for the course of this short proof,  $\log_2$  denotes the logarithm of base 2. Applying the Cauchy–Schwarz inequality, the first term on the right-hand side is

$$\leq \left( \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq \frac{y}{2^{\lceil \frac{1}{2} \log_2 y \rceil}}}} f(n) \right|^2 \right)^{1/2} \leq (\Psi(x, \sqrt{y}))^{1/2}.$$

One can readily check<sup>3</sup> that this is far smaller than the main term in Theorem 5.1. For the remaining expectations, note carefully that we can apply Proposition 5.3 on this range of smoothness parameter, and that  $\frac{\log x}{\log y/2^j}$  varies only by a constant over this range, so the “big Oh” term in the exponent is uniformly bounded by  $O(u^{8/11})$  with  $u = \frac{\log x}{\log y}$ . Applying this proposition and using the fact that the bound there is certainly increasing in the smoothness parameter, we deduce that

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \Psi(x^2, y)^{1/4} e^{O(u^{8/11})} (\log x)^{5/8} (\log y)^{13/8},$$

completing the proof.  $\square$

We proceed with the proof of Proposition 5.3.

*Proof of Proposition 5.3.* Letting  $\tilde{\mathbb{E}}$  denote the expectation conditioned on the values of  $f(p)$  for primes  $p \leq y/2$ , we have

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \in (y/2, y]}} f(n) \right| &= \mathbb{E} \tilde{\mathbb{E}} \left| \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} f(m) \sum_{\substack{n \leq x/m \\ P(n) \leq y/2}} f(n) \right| \\ &\leq \mathbb{E} \left( \tilde{\mathbb{E}} \left| \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} f(m) \sum_{\substack{n \leq x/m \\ P(n) \leq y/2}} f(n) \right|^2 \right)^{1/2} \\ &= \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} \left| \sum_{\substack{n \leq x/m \\ P(n) \leq y/2}} f(n) \right|^2 \right)^{1/2}, \end{aligned}$$

where the second line follows by applying Cauchy–Schwarz to the conditional expectation. We now introduce a dummy integral as was done in the proof of Proposition 4.1. This gives

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \in (y/2, y]}} f(n) \right| \leq \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{n \leq x/m \\ P(n) \leq y/2}} f(n) \right|^2 dt \right)^{1/2},$$

---

<sup>3</sup>One way to do this would be to trivially note that  $\Psi(x, \sqrt{y}) \leq \Psi(x, y^{2/3})$ , and then to apply Lemma 2.4 followed by equation (5.3).

where  $X$  is some parameter chosen later. Splitting the innermost sum at  $n = x/t$ , we have

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \in (y/2, y]}} f(n) \right| &\ll \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{n \leq x/t \\ P(n) \leq y/2}} f(n) \right|^2 dt \right)^{1/2} \\ &\quad + \mathbb{E} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{x/t < n \leq x/m \\ P(n) \leq y/2}} f(n) \right|^2 dt \right)^{1/2}. \end{aligned} \quad (5.4)$$

We will show that the second term is small when  $X$  is chosen to be sufficiently large. Specifically, by applying the Cauchy–Schwarz inequality to the expectation, the second term is smaller than

$$\begin{aligned} \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \sum_{\substack{x/t < n \leq x/m \\ P(n) \leq y/2}} 1 dt \right)^{1/2} &\leq \left( \sum_{\substack{1 < m \leq x \\ p|m \Rightarrow p \in (y/2, y]}} \sum_{\substack{x/m(1+\frac{1}{X}) < n \leq x/m \\ P(n) \leq y/2}} 1 \right)^{1/2} \\ &\leq \left( \sum_{\substack{n \leq x \\ P(n) \leq y/2}} \sum_{\substack{x/n(1+\frac{1}{X}) < m \leq x/n \\ p|m \Rightarrow p \in (y/2, y]}} 1 \right)^{1/2}. \end{aligned}$$

To handle this term, we first remove the contribution from the largest  $n$ . Note that we may impose the upper bound  $n \leq 2x/y$  on the above sum due to the prime factor restriction on the  $m$ . We take  $X = e^{0.9u}$  and recall that  $y \geq e^{\sqrt{2}\log x}$ . Straightforward estimates show that we certainly have  $y \geq 2X$ , so that the innermost sum over  $m$  will always be of length  $\geq 1$ . Therefore, the contribution from  $n > x/yX$  to the above is

$$\sum_{\substack{x/yX < n \leq 2x/y \\ P(n) \leq y/2}} \sum_{\substack{x/n(1+\frac{1}{X}) < m \leq x/n \\ p|m \Rightarrow p \in (y/2, y]}} 1 \ll \sum_{\substack{x/yX < n \leq 2x/y \\ P(n) \leq y/2}} \frac{x}{nX} \ll \sum_{i=0}^{\log X} \sum_{\substack{2x/ye^{i+1} < n \leq 2x/ye^i \\ P(n) \leq y/2}} \frac{x}{nX}.$$

Invoking Lemma 2.3 and the fact that  $e^i \leq X \leq y$ , we deduce that the above is

$$\ll \sum_{i=0}^{\log X} \frac{ye^i}{X} \sum_{\substack{2x/ye^{i+1} < n \leq 2x/ye^i \\ P(n) \leq y/2}} 1 \ll \Psi(x, y) \sum_{i=0}^{\log X} \frac{y^{2(1-\alpha(x, y))}}{X} \ll \Psi(x, y) \frac{u^4}{X},$$

where, to obtain the last estimate, we have applied Lemma 2.1. It follows that the second term in (5.4) is

$$\ll \left( \sum_{\substack{n \leq x/yX \\ P(n) \leq y/2}} \sum_{\substack{x/n(1+\frac{1}{X}) < m \leq x/n \\ p|m \Rightarrow p \in (y/2, y]}} 1 \right)^{1/2} + \sqrt{\frac{\Psi(x, y)}{e^{0.8u}}},$$

say. To estimate the first term above, note that the innermost sum over  $m$  is of length  $\geq \frac{x}{nX} \geq y$ , so Lemma 2.6 is applicable. Note that we may replace the  $\alpha(x^2, y)$  appearing



there with  $\alpha(x, y)$  since the latter is slightly bigger. This will help us to compare the error to  $\Psi(x, y)$ . We then find that the first term in the previous display is

$$\ll \left(\frac{x}{X}\right)^{\alpha(x, y)/2} (\log x) e^{O(\log u)} \left( \sum_{\substack{n \leq x/yX \\ P(n) \leq y/2}} \frac{1}{n^{\alpha(x, y)}} \right)^{1/2} \ll \frac{x^{\alpha(x, y)/2} \zeta(\alpha, y)^{1/2} (\log x) e^{O(\log u)}}{X^{\alpha(x, y)/2}}.$$

To obtain this estimate, in our application of Lemma 2.6, we have used the crude estimate  $e^{\frac{6u \log u}{\log y}} \leq e^{O(\log u)}$  uniformly for  $y \geq e^{\sqrt{2 \log x}}$ . Using Lemma 2.2 followed by Lemma 2.1 and the fact that  $X = e^{0.9u}$ , we find that the above is

$$\ll \frac{\Psi(x, y)^{1/2}}{X^{\alpha(x, y)/2}} (\log x)^{3/2} (\log y)^{1/2} e^{O(\log u)} \ll \left( \frac{\Psi(x, y) (\log x)^3 (\log y)}{e^{0.8u}} \right)^{1/2}.$$

Putting this all together, equation (5.4) becomes

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \in (y/2, y]}} f(n) \right| &\ll \mathbb{E} \left( \left| \int_{y/2}^x \sum_{\substack{n \leq x/t \\ P(n) \leq y/2}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} dt \right)^{1/2} \\ &\quad + \left( \frac{\Psi(x, y) (\log x)^3 (\log y)}{e^{0.8u}} \right)^{1/2}. \end{aligned} \quad (5.5)$$

We first show that the second term on the right-hand side will be negligible compared to our desired bound. By (5.3) and Lemma 2.1, we have

$$\Psi(x, y)^{1/2} \ll \Psi(x^2, y)^{1/4} \exp\left(\frac{u \log 2}{2} (1 + o(1))\right) (\log x)^{-1/8} (\log y)^{-5/8},$$

where the  $o(1)$  term goes to zero uniformly for  $e^{\sqrt{2 \log x}} \leq y \leq x^{\frac{1}{\log \log x}}$ . Hence

$$\begin{aligned} \left( \frac{\Psi(x, y) (\log x)^3 (\log y)}{e^{0.8u}} \right)^{1/2} &\ll \Psi(x^2, y)^{1/4} e^{-0.1u} (\log x)^{5/4} \\ &= \Psi(x^2, y)^{1/4} e^{-0.1u} u^{5/8} (\log x)^{5/8} (\log y)^{5/8}. \end{aligned}$$

The right-hand side is  $\ll \Psi(x^2, y)^{1/4} (\log x)^{5/8} (\log y)^{5/8}$ , so is acceptable in view of Proposition 5.3. By (5.5), to complete the proof, it remains to show that

$$\mathbb{E} \left( \left| \int_{y/2}^x \sum_{\substack{n \leq x/t \\ P(n) \leq y/2}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} dt \right)^{1/2} \ll \Psi(x^2, y)^{1/4} e^{O(u^{8/11})} (\log x)^{5/8} (\log y)^{5/8}.$$

By Lemma 2.6, we have

$$\sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} \ll \left( \frac{t}{X} \right)^{\alpha(x^2, y) - 1} e^{\frac{6u \log u}{\log y}} (\log x) \quad (5.6)$$

on our range of integration. Applying this bound, we have

$$\begin{aligned} \mathbb{E} \left( \int_{y/2}^x \left| \sum_{\substack{n \leq x/t \\ P(n) \leq y/2}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} dt \right)^{1/2} &\ll \\ e^{\frac{3u \log u}{2 \log y}} X^{\frac{1-\alpha(x^2, y)}{2}} (\log x)^{1/2} \mathbb{E} \left( \int_{y/2}^x \left| \sum_{\substack{n \leq x/t \\ P(n) \leq y/2}} f(n) \right|^2 t^{\alpha(x^2, y)-1} dt \right)^{1/2}. \end{aligned}$$

Using Lemma 2.1 and the fact that  $X = e^{0.9u}$ , we have  $X^{\frac{1-\alpha(x^2, y)}{2}} \ll e^{\frac{u \log u}{2 \log y}}$ . Then, performing the change of variables  $t = x/z$ , we obtain an upper bound of size

$$x^{\alpha(x^2, y)/2} e^{\frac{7u \log u}{2 \log y}} (\log x)^{1/2} \mathbb{E} \left( \int_1^{2x/y} \left| \sum_{\substack{n \leq z \\ P(n) \leq y/2}} f(n) \right|^2 \frac{dz}{z^{1+\alpha(x^2, y)}} \right)^{1/2}.$$

Completing the integral and applying Plancherel's identity, Lemma 3.1, we have

$$\begin{aligned} \mathbb{E} \left( \int_{y/2}^x \left| \sum_{\substack{n \leq x/t \\ P(n) \leq y/2}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} dt \right)^{1/2} &\ll \\ x^{\alpha(x^2, y)/2} e^{\frac{7u \log u}{2 \log y}} (\log x)^{1/2} \mathbb{E} \left( \int_{\mathbb{R}} \left| \frac{F_{y/2}(\alpha(x^2, y)/2 + it)}{\alpha(x^2, y)/2 + it} \right|^2 dt \right)^{1/2}. \end{aligned}$$

We now apply Theorem 3.4 to deduce that the expectation is

$$\ll x^{\alpha(x^2, y)/2} \zeta(\alpha(x^2, y), y)^{1/4} e^{\frac{7u \log u}{2 \log y}} (\log y)^{1/2} (\log x)^{1/2} e^{O((\log \zeta(\alpha(x^2, y), y))^{8/11})},$$

and seeing as we have  $\log \zeta(\alpha(x^2, y), y) \ll u$  for  $y \leq x^{\frac{1}{\log \log x}}$  (see equation (5.2)) and  $\frac{u \log u}{\log y} \ll u^{8/11}$  uniformly for  $y \geq e^{\sqrt{2 \log x}}$ , this is

$$\ll x^{\alpha(x^2, y)/2} \zeta(\alpha(x^2, y), y)^{1/4} e^{O(u^{8/11})} (\log y)^{1/2} (\log x)^{1/2}.$$

Finally, comparing this to Lemma 2.2, we deduce that

$$\mathbb{E} \left( \int_{y/2}^x \left| \sum_{\substack{n \leq x/t \\ P(n) \leq y/2}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) \leq m \leq t \\ p|m \Rightarrow p \in (y/2, y]}} \frac{X}{m} dt \right)^{1/2} \ll \Psi(x^2, y)^{1/4} e^{O(u^{8/11})} (\log y)^{5/8} (\log x)^{5/8},$$

as desired. This completes the proof of Proposition 5.3.  $\square$

**5.3. Proof of Theorem 5.2.** We now handle the case where  $y \leq e^{\sqrt{10 \log x}}$ . As mentioned, we begin by splitting up the original sum depending on the contribution from primes on different scales.

*Proof of Theorem 5.2.* Suppose that  $(\log x)^{20} \leq y \leq e^{\sqrt{10 \log x}}$ . Let  $Z, M \geq 2$  be some parameters specified shortly, and let  $\mathcal{K} = \lceil \frac{\log y}{10 \log M} \rceil$ . For any  $y$ -smooth integer  $n$ , we decompose it as

$$n = m_1 \dots m_{\mathcal{K}} s, \quad m_i := \prod_{\substack{p \in (y/M^i, y/M^{i-1}] \\ p^\alpha \parallel n}} p^\alpha \text{ for } 1 \leq i \leq \mathcal{K},$$

so that  $m_i$  consists of the prime factors that lie in  $(y/M^i, y/M^{i-1}]$  (if there are no such prime factors we have  $m_i = 1$ ), and  $s$  denotes the part consisting of primes below  $y/M^{\mathcal{K}} \approx y^{9/10}$ . To fix ideas, we take  $Z = e^u$  and  $M = 5e^{10(\log y)/u}$ . Note that this choice of  $M$  is always bounded, and is approximately equal to 5 unless  $y$  is very close to the upper range  $e^{\sqrt{10 \log x}}$  (we find this technical adjustment useful, for example, in the estimate of  $S_3$  below). Given a  $y$ -smooth number  $n$  of this form, we will split up our sum depending on a parameter  $k$ , which will be the smallest integer  $1 \leq k \leq \mathcal{K}$  such that  $m_k > Z$ . Integers where no such  $k$  exists are handled separately, and, crucially, since there are relatively few  $y$ -smooth numbers with a large  $y^{9/10}$ -smooth part, the contribution from the case where  $m_i \leq Z$  for all  $1 \leq i \leq \mathcal{K}$  will be small. For the remaining terms, if  $Z$  is sufficiently large, then we will be able to efficiently approximate the sum over  $m_k$  by an integral, as discussed at the end of Section 5.1. Using this splitting and applying the triangle inequality, we have

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \leq S_1 + S_2 + S_3, \quad (5.7)$$

where

$$\begin{aligned} S_1 &:= \mathbb{E} \left| \sum_{\substack{Z < m \leq x \\ p|m \Rightarrow p \in (\frac{y}{M}, y]}} f(m) \sum_{\substack{n \leq \frac{x}{m} \\ P(n) \leq \frac{y}{M}}} f(n) \right|, \\ S_2 &:= \mathbb{E} \left| \sum_{k=2}^{\mathcal{K}} \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} f(m_1 \dots m_k) \sum_{\substack{n \leq \frac{x}{m_1 \dots m_k} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|, \\ S_3 &:= \mathbb{E} \left| \sum'_{\substack{m_1, \dots, m_{\mathcal{K}} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} f(m_1 \dots m_{\mathcal{K}}) \sum_{\substack{n \leq \frac{x}{m_1 \dots m_{\mathcal{K}}} \\ P(n) \leq \frac{y}{M^{\mathcal{K}}}}} f(n) \right|. \end{aligned}$$

Here  $\sum'$  denotes that the sum should include the cases where  $m_i = 1$ . We have separated the  $k = 1$  term, which corresponds to  $S_1$ , and we will see that this gives the dominant contribution to (5.7).

We begin by bounding the final term,  $S_3$ , by exploiting the fact that there are few terms in the sum. By the Cauchy–Schwarz inequality and Lemma 2.3, we have

$$S_3 \leq \left( \sum'_{\substack{m_1, \dots, m_K \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^i-1}]}} \sum_{\substack{n \leq \frac{x}{m_1 \dots m_K} \\ P(n) \leq \frac{y}{M^K}}} 1 \right)^{1/2} \ll \left( \sum'_{\substack{m_1, \dots, m_K \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^i-1}]}} \frac{\Psi(x, y^{9/10})}{m_1^{\alpha(x, y^{9/10})} \dots m_K^{\alpha(x, y^{9/10})}} \right)^{1/2}.$$

Removing the condition on the prime factors, we find that

$$S_3 \ll \Psi(x, y^{9/10})^{1/2} Z^{\mathcal{K}(1-\alpha(x, y^{9/10}))/2} (\log Z)^{\mathcal{K}/2}$$

Here we recall that  $Z = e^u$  and  $M = 5e^{10(\log y)/u}$ . Using Lemma 2.1, we deduce that  $1 - \alpha(x, y^{9/10}) \leq \frac{6 \log u}{5 \log y}$  when  $x$  is sufficiently large. Using also the fact that  $\mathcal{K} = \lceil \frac{\log y}{10 \log M} \rceil$ , we have the crude bound

$$Z^{\mathcal{K}(1-\alpha(x, y^{9/10}))/2} (\log Z)^{\mathcal{K}/2} \ll \exp\left(\frac{u \log u}{25} + \frac{u \log u}{100}\right) \ll \exp\left(\frac{u \log u}{20}\right),$$

say. Note carefully that to bound the  $(\log Z)^{\mathcal{K}/2}$  term, we have made vital use of the fact that  $M$  becomes large when  $y \approx e^u$ . Seeing as  $\Psi(x, y) = xe^{-u \log u(1+o(1))}$  uniformly for  $(\log x)^2 \leq y \leq x^{\frac{1}{\log \log x}}$  (this follows immediately from Lemmas 2.1 and 2.2, in addition to a simple estimate for  $\zeta(\alpha, y)$  as in (5.2)), from which it follows immediately that  $\Psi(x, y^{9/10}) = xe^{-\frac{10}{9}u \log u(1+o(1))}$  uniformly for  $(\log x)^{20} \leq y \leq e^{\sqrt{10} \log x}$ . We deduce that

$$S_3 \ll x^{1/2} \exp\left(\left(\frac{1}{2} - \frac{1}{200}\right)u \log u(1+o(1))\right) \ll \sqrt{\frac{\Psi(x, y)}{e^{100u}}},$$

say. We will see later that this term contributes a negligible amount compared to the main term in Theorem 5.2.

We now proceed with bounding the first two terms in (5.7). Let  $\mathbb{E}^{(k)}$  denote the expectation conditioned on  $(f(p))_{p \leq y/M^k}$ . Applying the triangle inequality, followed by the Cauchy–Schwarz inequality to the conditional expectation, it follows that

$$\begin{aligned} S_2 &\leq \sum_{k=2}^{\mathcal{K}} \mathbb{E} \left| \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^i-1}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^k-1}]}} f(m_1 \dots m_k) \sum_{\substack{n \leq \frac{x}{m_1 \dots m_k} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right| \\ &\leq \sum_{k=2}^{\mathcal{K}} \mathbb{E} \left( \mathbb{E}^{(k)} \left| \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^i-1}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^k-1}]}} f(m_1 \dots m_k) \sum_{\substack{n \leq \frac{x}{m_1 \dots m_k} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 \right)^{1/2} \\ &= \sum_{k=2}^{\mathcal{K}} \mathbb{E} \left( \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^i-1}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^k-1}]}} \left| \sum_{\substack{n \leq \frac{x}{m_1 \dots m_k} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 \right)^{1/2}. \end{aligned}$$

Similarly, conditioning only on  $(f(p))_{p \leq y/M}$  and applying Cauchy–Schwarz, we have

$$S_1 \leq \mathbb{E} \left( \sum_{\substack{Z < m \leq x \\ p|m \Rightarrow p \in (\frac{y}{M}, y]}} \left| \sum_{\substack{n \leq \frac{x}{m} \\ P(n) \leq \frac{y}{M}}} f(n) \right|^2 \right)^{1/2}.$$

Let  $X = e^{0.9u}$  (as in the proof of Proposition 5.3), and recall that  $Z = e^u$ . Note that  $Z$  is larger than  $X$ , which is vital in allowing us to control sums of length  $Z/X$ . Combining the above bounds and introducing a dummy integral, we deduce from (5.7) that

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| &\ll \mathbb{E} \left( \sum_{\substack{Z < m \leq x \\ p|m \Rightarrow p \in (\frac{y}{M}, y]}} \frac{X}{m} \int_m^{m(1+\frac{1}{X})} \left| \sum_{\substack{n \leq \frac{x}{m} \\ P(n) \leq \frac{y}{M}}} f(n) \right|^2 dt \right)^{1/2} + \sqrt{\frac{\Psi(x, y)}{e^{100u}}} \\ &+ \sum_{k=2}^{\mathcal{K}} \mathbb{E} \left( \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} \frac{X}{m_k} \int_{m_k}^{m_k(1+\frac{1}{X})} \left| \sum_{\substack{n \leq \frac{x}{m_1 \dots m_k} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 dt \right)^{1/2} \quad (5.8) \end{aligned}$$

We now bound the last term; the first term being handled similarly. Splitting the sum at  $n = \frac{x}{m_1 \dots m_{k-1} t}$ , the last term on the right-hand side is

$$\begin{aligned} &\ll \sum_{k=2}^{\mathcal{K}} \mathbb{E} \left( \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} \frac{X}{m_k} \int_{m_k}^{m_k(1+\frac{1}{X})} \left| \sum_{\substack{n \leq \frac{x}{m_1 \dots m_{k-1} t} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 dt \right)^{1/2} \\ &+ \sum_{k=2}^{\mathcal{K}} \mathbb{E} \left( \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} \frac{X}{m_k} \int_{m_k}^{m_k(1+\frac{1}{X})} \left| \sum_{\substack{\frac{x}{m_1 \dots m_{k-1} t} < n \leq \frac{x}{m_1 \dots m_k} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 dt \right)^{1/2}. \end{aligned}$$

As in the previous proofs, the first term will give the dominant contribution, and we will first bound the second term. Applying the Cauchy–Schwarz inequality, the second term is

$$\ll \sum_{k=2}^{\mathcal{K}} \left( \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \sum_{\substack{Z < m_k \leq \frac{x}{m_1 \dots m_{k-1}} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} \sum_{\substack{\frac{x}{m_1 \dots m_k(1+1/X)} < n \leq \frac{x}{m_1 \dots m_k} \\ P(n) \leq \frac{y}{M^k}}} 1 \right)^{1/2}.$$

Exchanging the order of summation, this is

$$\ll \sum_{k=2}^{\mathcal{K}} \left( \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \sum_{\substack{n \leq \frac{x}{m_1 \dots m_{k-1} Z} \\ P(n) \leq \frac{y}{M^k}}} \sum_{\substack{\frac{x}{m_1 \dots m_{k-1} n(1+1/X)} < m_k \leq \frac{x}{m_1 \dots m_{k-1} n} \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} 1 \right)^{1/2},$$

and applying Lemma 2.6 (noting that we can replace  $\alpha(x^2, y)$  there by  $\alpha(x, y)$  for an upper bound), this is

$$\ll \sum_{k=2}^{\mathcal{K}} \left( \frac{x^{\alpha(x,y)} e^{\frac{20u \log u}{\log y} + O(\log u)}}{X^{\alpha(x,y)}} \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \frac{1}{m_1^{\alpha(x,y)} \dots m_{k-1}^{\alpha(x,y)}} \sum_{\substack{n \leq \frac{x}{m_1 \dots m_{k-1}} \\ P(n) \leq \frac{y}{M^k}}} \frac{1}{n^{\alpha(x,y)}} \right)^{1/2}.$$

Here we have used the fact that  $\frac{3Mu \log u}{\log y / M^{k-1}} \leq \frac{20u \log u}{\log y} + O(\log u)$  for  $y \leq e^{\sqrt{10 \log x}}$ , uniformly in  $k$ , to bound the exponential term in Lemma 2.6. We have also engulfed the  $\log x$  term from that lemma into the  $e^{O(\log u)}$  term. Removing the size restrictions on each of the sums, this is

$$\ll \frac{x^{\alpha(x,y)/2} \zeta(\alpha(x,y), y)^{1/2}}{X^{\alpha(x,y)/2}} \exp\left(\frac{10u \log u}{\log y} + O(\log u)\right) \ll \sqrt{\frac{\Psi(x, y)}{e^{0.7u}}} \exp\left(\frac{10u \log u}{\log y}\right),$$

in view of Lemma 2.2. Here we have used the fact that  $X = e^{0.9u}$  and  $\alpha(x, y) \geq \frac{9}{10}(1+o(1))$  for  $(\log x)^{20} \leq y$ . The first term in (5.8) can be handled similarly, instead splitting the sum at  $x/t$ , and the analogous error there can be seen to satisfy the same bound. This leads to

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| &\ll \mathbb{E} \left( \int_Z^x \left| \sum_{\substack{n \leq \frac{x}{t} \\ P(n) \leq \frac{y}{M}}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) < m \leq t \\ p|m \Rightarrow p \in (\frac{y}{M}, y]}} \frac{X}{m} \right)^{1/2} + \sqrt{\frac{\Psi(x, y)}{e^{0.7u}}} \exp\left(\frac{10u \log u}{\log y}\right) \\ &+ \sum_{k=2}^{\mathcal{K}} \mathbb{E} \left( \sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \int_Z^{\frac{x(1+1/X)}{m_1 \dots m_{k-1}}} \left| \sum_{\substack{n \leq \frac{x}{m_1 \dots m_{k-1} t} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 \sum_{\substack{t/(1+\frac{1}{X}) < m_k \leq t \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} \frac{X}{m_k} dt \right)^{1/2}, \end{aligned}$$

where we have exchanged the order of summation and integration. Again, applying Lemma 2.6 we have

$$\sum_{\substack{t/(1+\frac{1}{X}) < m_k \leq t \\ p|m_k \Rightarrow p \in (\frac{y}{M^k}, \frac{y}{M^{k-1}}]}} \frac{X}{m_k} \ll \left( \frac{t}{X} \right)^{\alpha(x^2, \frac{y}{M^{k-1}}) - 1} \exp\left(\frac{18u \log u}{\log y} + O(\log u)\right),$$

which also holds for  $k = 1$ . Here we have used the fact that  $\frac{3Mu \log u}{\log y / M^{k-1}} \leq \frac{18u \log u}{\log y} + O(\log u)$  uniformly in  $k$ , and also that  $\log x \leq e^{O(\log u)}$  uniformly for  $(\log x)^{20} \leq y \leq e^{\sqrt{10 \log x}}$ . In the exponent of  $t/X$ , it is important that we maintain the saddle point corresponding to the smoothness parameter  $y/M^{k-1}$  so that our errors do not compound. We deduce that

$$\mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| \ll \tilde{S}_1 + \tilde{S}_2 + \sqrt{\frac{\Psi(x, y)}{e^{0.7u}}} \exp\left(\frac{10u \log u}{\log y} + O(\log u)\right), \quad (5.9)$$

where

$$\tilde{S}_1 := X^{\frac{1}{2}(1-\alpha(x^2, y))} \exp\left(\frac{9u \log u}{\log y} + O(\log u)\right) \mathbb{E}\left(\int_Z^x \left| \sum_{\substack{n \leq \frac{x}{t} \\ P(n) \leq \frac{y}{M}}} f(n) \right|^2 t^{\alpha(x^2, y)-1}\right)^{1/2} \quad (5.10)$$

and

$$\begin{aligned} \tilde{S}_2 &:= \exp\left(\frac{9u \log u}{\log y} + O(\log u)\right) \sum_{k=2}^K X^{\frac{1}{2}(1-\alpha(x^2, \frac{y}{M^{k-1}}))} \\ &\quad \times \mathbb{E}\left(\sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \int_Z^{\frac{x(1+\frac{1}{X})}{m_1 \dots m_{k-1}}} \left| \sum_{\substack{n \leq \frac{x}{m_1 \dots m_{k-1} t} \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 t^{\alpha(x^2, \frac{y}{M^{k-1}})-1} dt\right)^{1/2}. \end{aligned} \quad (5.11)$$

We first focus on bounding the expectation that appears in  $\tilde{S}_2$ . Performing the change of variables  $z = \frac{x}{m_1 \dots m_{k-1} t}$ , the expectation there is equal to

$$\mathbb{E}\left(\sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \frac{x^{\alpha(x^2, \frac{y}{M^{k-1}})}}{(m_1 \dots m_{k-1})^{\alpha(x^2, \frac{y}{M^{k-1}})}} \int_1^{\frac{x}{m_1 \dots m_{k-1} Z}} \left| \sum_{\substack{n \leq z \\ P(n) \leq \frac{y}{M^k}}} f(n) \right|^2 \frac{dz}{z^{1+\alpha(x^2, \frac{y}{M^{k-1}})}}\right)^{1/2}.$$

Completing the integral, and applying Plancherel's identity (Lemma 3.1), this is

$$\leq \mathbb{E}\left(\sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \frac{x^{\alpha(x^2, \frac{y}{M^{k-1}})}}{(m_1 \dots m_{k-1})^{\alpha(x^2, \frac{y}{M^{k-1}})}} \int_{\mathbb{R}} \left| \frac{F_{y/M^k}(\alpha(x^2, \frac{y}{M^{k-1}}) + it)}{\alpha(x^2, \frac{y}{M^{k-1}}) + it} \right|^2 dt\right)^{1/2}.$$

Note that the integral term depends on  $k$  but does not depend on any of the  $m_i$ , so we can now evaluate the summation over  $m_1, \dots, m_{k-1}$ . Dropping the prime factor conditions and breaking into dyadic intervals, we have

$$\sum'_{\substack{m_1, \dots, m_{k-1} \leq Z \\ p|m_i \Rightarrow p \in (\frac{y}{M^i}, \frac{y}{M^{i-1}}]}} \frac{1}{(m_1 \dots m_{k-1})^{\alpha(x^2, \frac{y}{M^{k-1}})}} \ll (\log Z)^{k-1} Z^{(k-1)(1-\alpha(x^2, \frac{y}{2^{k-1}}))}.$$

Therefore the expectation in (5.11) is at most

$$x^{\alpha(x^2, \frac{y}{M^{k-1}})/2} (\log Z)^{\frac{k-1}{2}} Z^{\frac{k-1}{2}(1-\alpha(x^2, \frac{y}{M^{k-1}}))} \mathbb{E}\left(\int_{\mathbb{R}} \left| \frac{F_{y/M^k}(\alpha(x^2, \frac{y}{M^{k-1}}) + it)}{\alpha(x^2, \frac{y}{M^{k-1}}) + it} \right|^2 dt\right)^{1/2}.$$

Applying Theorem 3.4 and increasing the number of primes in the product, this is

$$\ll x^{\alpha(x^2, \frac{y}{M^{k-1}})/2} \zeta(\alpha(x^2, \frac{y}{M^{k-1}}), y/M^{k-1})^{1/4} \exp(O(u^{8/11})) (\log Z)^{\frac{k-1}{2}} Z^{\frac{k-1}{2}(1-\alpha(x^2, \frac{y}{M^{k-1}}))}.$$

By Lemma 2.2 (or equation (1.6)), we have

$$x^{\alpha(x^2, \frac{y}{M^{k-1}})/2} \zeta(\alpha(x^2, \frac{y}{M^{k-1}}), y/M^{k-1})^{1/4} \ll \Psi(x^2, y/M^{k-1})^{1/4} (\log x \log y)^{1/8},$$



so we find that the expectation in (5.11) is

$$\ll \Psi(x^2, y/M^{k-1})^{1/4} \exp\left(O(u^{8/11})\right) (\log Z)^{\frac{k-1}{2}} Z^{\frac{k-1}{2}(1-\alpha(x^2, \frac{y}{M^{k-1}}))}, \quad (5.12)$$

where the powers of  $\log x$  and  $\log y$  have been engulfed into the  $\exp(O(u^{8/11}))$  term. Finally, seeing as  $Z = e^u$ , and using Lemma 2.1 to deduce that  $1 - \alpha(x^2, \frac{y}{M^{k-1}}) \leq \frac{6 \log u}{5 \log y}$  when  $x$  is large, we have

$$(\log Z)^{\frac{k-1}{2}} Z^{\frac{k-1}{2}(1-\alpha(x^2, \frac{y}{M^{k-1}}))} \ll \exp\left(\frac{(k-1) \log u}{2} \left(1 + \frac{6u}{5 \log y}\right)\right)$$

Applying Lemma 2.4 and using the facts that  $\frac{1}{1-x} - 1 \geq x$  (when considering  $u - u_d$ ) and  $\xi(u) = \log u(1 + o(1))$  uniformly for  $u$  in this range, we have

$$\Psi(x^2, y/M^{k-1})^{1/4} \ll \Psi(x^2, y)^{1/4} \exp\left(-\frac{(k-1)u \log u \log M}{2 \log y} (1 + o(1)) + O(E(u))\right),$$

where  $E(u)$  is as stated in the lemma and the  $o(1)$  term goes to zero uniformly for  $y$  in the considered range. Using the fact that  $M = 5e^{10(\log y)/u}$  and applying these estimates to (5.12), we have that the expectation in (5.11) is

$$\ll \Psi(x^2, y)^{1/4} \exp\left(O(u^{8/11})\right).$$

Inserting this back into (5.11), we have

$$\tilde{S}_2 \ll \Psi(x^2, y)^{1/4} \exp\left(\frac{9u \log u}{\log y} + O(u^{8/11})\right) \sum_{k=2}^{\mathcal{K}} X^{\frac{1}{2}(1-\alpha(x^2, \frac{y}{M^{k-1}}))},$$

Seeing as  $X = e^{0.9u}$  and by Lemma 2.1, we have  $X^{\frac{1}{2}(1-\alpha(x^2, y/M^{k-1}))} \leq \exp\left(\frac{3u \log u}{4 \log y}\right)$  when  $x$  is sufficiently large. We certainly then have

$$\tilde{S}_2 \ll \Psi(x^2, y)^{1/4} \exp\left(\frac{10u \log u}{\log y} + O(u^{8/11})\right),$$

which is the bound in our theorem. We are left with estimating  $\tilde{S}_1$ . The proof is done in the same way. First, completing the integral, performing the change of variables  $z = x/t$ , and applying Plancherel's identity (Lemma 3.1), we obtain

$$\mathbb{E}\left(\int_Z^x \left| \sum_{\substack{n \leq \frac{x}{t} \\ P(n) \leq \frac{y}{M}}} f(n) \right|^2 t^{\alpha(x^2, y)-1} dt\right)^{1/2} \leq \mathbb{E}\left(\int_{\mathbb{R}} \left| \frac{F_{y/M}(\alpha(x^2, y) + it)}{\alpha(x^2, y) + it} \right|^2\right)^{1/2}.$$

By Theorem 3.4, and extending the product over primes, we have

$$\mathbb{E}\left(\int_Z^x \left| \sum_{\substack{n \leq \frac{x}{t} \\ P(n) \leq \frac{y}{M}}} f(n) \right|^2 t^{\alpha(x^2, y)-1} dt\right)^{1/2} \ll x^{\alpha(x^2, y)/2} \zeta(\alpha(x^2, y), y)^{1/4} \exp\left(O(u^{8/11})\right),$$

where we have again made use of (5.2) to obtain the bound  $\log \zeta(\alpha, y) \ll u$ . By Lemma 2.2, we then have, similarly to before

$$\mathbb{E} \left( \int_Z^x \left| \sum_{\substack{n \leq \frac{x}{t} \\ P(n) \leq \frac{y}{M}}} f(n) \right|^2 t^{\alpha(x^2, y) - 1} dt \right)^{1/2} \ll \Psi(x^2, y)^{1/4} \exp \left( O(u^{8/11}) \right).$$

Inserting this estimate into (5.10), we have

$$\tilde{S}_1 \ll \Psi(x^2, y)^{1/4} \exp \left( \frac{9u \log u}{\log y} + O(u^{8/11}) \right) X^{\frac{1}{2}(1 - \alpha(x^2, y))},$$

and similarly to before, we deduce that

$$\tilde{S}_1 \ll \Psi(x^2, y)^{1/4} \exp \left( \frac{10u \log u}{\log y} + O(u^{8/11}) \right).$$

Inserting these estimates into (5.9), we see that

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{n \leq x \\ P(n) \leq y}} f(n) \right| &\ll \Psi(x^2, y)^{1/4} \exp \left( \frac{10u \log u}{\log y} + O(u^{8/11}) \right) \\ &\quad + \sqrt{\frac{\Psi(x, y)}{e^{0.7u}}} \exp \left( \frac{10u \log u}{\log y} + O(\log u) \right). \end{aligned}$$

Finally, it follows from equation (5.3) that  $\sqrt{\Psi(x, y)e^{-0.7u}} \ll \Psi(x^2, y)^{1/4}$  uniformly for  $(\log x)^{20} \leq y \leq e^{\sqrt{10 \log x}}$  (vitally, this uses the fact that  $\log 2 \leq 0.7$ ). Therefore, the first term dominates, completing the proof of Theorem 5.2.  $\square$

## APPENDIX A. PROOF OF LEMMA 2.5

*Proof of Lemma 2.5.* As mentioned, this lemma is a natural extension of [19, Theorem 3]. To begin, we take  $y \geq 2$ ,  $1 \leq u \leq \exp((\log y)^{3/5 - \varepsilon})$ ,  $y^{-1/3} \leq h \leq 1/2$  and let  $\frac{1}{\log y} \leq \delta \leq \frac{100}{\log u}$ . Note carefully that in our proof, mimicking Hildebrand [19], we redefine  $h$  to be the relative length of the sum (i.e. here  $h$  plays the role of  $h/x$  in the statement of Lemma 2.5). These conditions therefore correspond to the case where  $e^{(\log \log x)^{5/3 + \varepsilon}} \leq y \leq x$ ,  $\frac{1}{\log y} \leq \delta \leq \frac{100}{\log u}$  and  $x/y^{1/3} \leq h \leq x/2$  in the lemma.

We make use of the standard notation

$$\Psi_q(x, y) = \#\{n \leq x : p \mid n \Rightarrow (p, q) = 1 \text{ and } p \leq y\},$$

and in our proof we fix  $q = \prod_{p \leq y^\delta} p$ . For any  $z \geq y$ , we have

$$\int_1^z \frac{\Psi_q(z, y) - \Psi_q(t, y)}{t} dt = \sum_{\substack{n \leq z \\ p \mid n \Rightarrow y^\delta < p \leq y}} \log n,$$

and upon writing  $\log n = \sum_{d|n} \Lambda(d)$ , it follows that

$$\Psi_q(z, y) \log z = \sum_{\substack{p^m \leq z \\ y^\delta < p \leq y}} \Psi_q(z/p^m, y) \log p + \int_1^z \frac{\Psi_q(t, y)}{t} dt.$$

Applying the formula above twice and taking the difference gives

$$\begin{aligned} \Psi_q(y^u, y) - \Psi_q(y^u(1-h), y) &= \frac{1}{u \log y} \left( \int_{y^u(1-h)}^{y^u} \frac{\Psi_q(t, y)}{t} dt - \Psi_q(y^u(1-h), y) \log \left( \frac{1}{1-h} \right) \right. \\ &\quad \left. + \sum_{\substack{p^m \leq y^u \\ y^\delta < p \leq y}} \left( \Psi_q(y^u/p^m, y) - \Psi_q(y^u(1-h)/p^m, y) \right) \log p \right). \end{aligned}$$

Rewriting the term  $\Psi_q(y^u(1-h), y) \log \left( \frac{1}{1-h} \right)$  into the integral leads to

$$\begin{aligned} \Psi_q(y^u, y) - \Psi_q(y^u(1-h), y) &= \frac{1}{u \log y} \left( \int_{y^u(1-h)}^{y^u} \frac{\Psi_q(t, y) - \Psi_q(y^u(1-h), y)}{t} dt \right. \\ &\quad \left. + \sum_{\substack{p^m \leq y^u \\ y^\delta < p \leq y}} \left( \Psi_q(y^u/p^m, y) - \Psi_q(y^u(1-h)/p^m, y) \right) \log p \right). \quad (\text{A.1}) \end{aligned}$$

We define  $\Delta_h(\delta, y, u) \geq 0$  to satisfy

$$\Psi_q(y^u, y) - \Psi_q(y^u(1-h), y) = \Delta_h(\delta, y, u) \frac{hy^u \rho(u)}{\delta \log y},$$

and we set

$$\Delta_h^*(\delta, y, u) := \sup_{1/2 \leq u' \leq u} \Delta_h(\delta, y, u').$$

We will show that uniformly for  $e^{(\log \log x)^{5/3+\varepsilon}} \leq y \leq x, y^{-1/3} \leq h \leq 1/2, \frac{1}{\log y} \leq \delta \leq \frac{100}{\log u}$ :

$$\Delta_h^*(\delta, y, u) \ll 1. \quad (\text{A.2})$$

If this is proved, then one can see that by using  $\Psi(x, y) \gg x\rho(u)$  in our range of parameters ([17]), then the Lemma is proved.

Starting from now, we focus on proving (A.2). We first consider the integral term in (A.1). We have

$$\begin{aligned} \int_{y^u(1-h)}^{y^u} \frac{\Psi_q(t, y) - \Psi_q(y^u(1-h), y)}{t} dt &\leq \frac{hy^u \rho(u) \Delta_h(\delta, y, u)}{\delta \log y} \log \left( \frac{1}{1-h} \right) \\ &\ll \frac{hy^u \rho(u) \Delta_h^*(\delta, y, u)}{\delta \log y}, \end{aligned}$$

and the implied constant is absolute. We turn our attention to

$$\begin{aligned} \sum_{\substack{y < p^m \leq y^u \\ y^\delta < p \leq y}} \left( \Psi_q(y^u/p^m, y) - \Psi_q(y^u(1-h)/p^m, y) \right) \log p \\ = \frac{1}{\delta \log y} \sum_{\substack{y < p^m \leq y^u \\ y^\delta < p \leq y}} \Delta_h \left( \delta, y, u - \frac{\log p^m}{\log y} \right) \frac{hy^u \log p}{p^m} \rho \left( u - \frac{\log p^m}{\log y} \right). \end{aligned}$$

We apply [19, Lemma 3], which states that

$$\sum_{\substack{y < p^m \leq y^u \\ p \leq y}} \frac{\log p}{p^m} \rho \left( u - \frac{\log p^m}{\log y} \right) \ll \rho(u),$$

uniformly for  $y \geq 2$  and  $u \leq y^{1/4}$ , so we deduce that

$$\sum_{\substack{y < p^m \leq y^u \\ y^\delta < p \leq y}} \left( \Psi_q(y^u/p^m, y) - \Psi_q(y^u(1-h)/p^m, y) \right) \log p \ll \frac{hy^u \rho(u) \Delta_h^*(\delta, y, u)}{\delta \log y}.$$

Inserting our estimates into (A.1) and rewriting the left hand side, we deduce that  $\Delta_h(\delta, y, u) =$

$$\frac{1}{\rho(u)u \log y} \sum_{\substack{p^m \leq y \\ y^\delta < p \leq y}} \Delta_h \left( \delta, y, u - \frac{\log p^m}{\log y} \right) \rho \left( u - \frac{\log p^m}{\log y} \right) \frac{\log p}{p^m} + O \left( \frac{\Delta_h^*(\delta, y, u)}{u \log y} \right),$$

uniformly for  $y \geq 2$  and  $u \leq y^{1/4}$ . We deduce that  $\Delta_h(\delta, y, u)$  is smaller than or equal to

$$\frac{1}{u \rho(u) \log y} \sum_{p^m \leq y} \Delta_h \left( \delta, y, u - \frac{\log p^m}{\log y} \right) \rho \left( u - \frac{\log p^m}{\log y} \right) \frac{\log p}{p^m} + O \left( \frac{\Delta_h^*(\delta, y, u)}{u \log y} \right).$$

Applying [19, Lemma 4], which states that for every fixed  $\varepsilon > 0$  and uniformly for  $y \geq 2, u \geq 1$  and  $0 \leq \theta \leq 1$  the following holds

$$\begin{aligned} \sum_{p^m \leq y^\theta} \rho \left( u - \frac{\log p^m}{\log y} \right) \frac{\log p}{p^m} &= (\log y) \int_{u-\theta}^u \rho(t) dt \\ &+ O_\varepsilon \left( \rho(u) (1 + u \log^2(u+1) \exp(-(\log y)^{3/5-\varepsilon})) \right), \end{aligned}$$

to find that, for  $u \leq \exp((\log y)^{3/5-\varepsilon})$ , we have

$$\begin{aligned} \Delta_h(\delta, y, u) &\leq \frac{1}{u \rho(u)} \left[ \Delta_h^*(\delta, y, u) \int_{u-1/2}^u \rho(t) dt + \Delta_h^*(\delta, y, u-1/2) \int_{u-1}^{u-1/2} \rho(t) dt \right] \\ &\quad + O \left( \frac{\Delta_h^*(\delta, y, u)}{u \log y} \right). \end{aligned}$$

Seeing as  $\int_{u-1}^u \rho(t) dt = u\rho(u)$  (see [19, Lemma 1]), we have

$$\begin{aligned} 1 &= \frac{1}{u\rho(u)} \int_{u-1/2}^u \rho(t) dt + \frac{1}{u\rho(u)} \int_{u-1}^{u-1/2} \rho(t) dt \\ &=: \alpha(u) + (1 - \alpha(u)). \end{aligned}$$

Hence

$$\Delta_h(\delta, y, u) \leq \left( \Delta_h^*(\delta, y, u)\alpha(u) + \Delta_h^*(\delta, y, u - 1/2)(1 - \alpha(u)) \right) + O\left(\frac{\Delta_h^*(\delta, y, u)}{u \log y}\right).$$

As is noted in the proof of Theorem 1 in [19], we have  $\alpha(u) \leq 1/2$  (this follows from [19, Lemma 1]). We deduce that the quantity

$$\begin{aligned} \frac{1}{2} \left( \Delta^*(\delta, y, u) + \Delta^*\left(\delta, y, u - \frac{1}{2}\right) \right) - \left( \alpha(u)\Delta^*(\delta, y, u) + (1 - \alpha(u))\Delta^*\left(\delta, y, u - \frac{1}{2}\right) \right) \\ = \left( \frac{1}{2} - \alpha(u) \right) \left( \Delta^*(\delta, y, u) - \Delta^*\left(\delta, y, u - \frac{1}{2}\right) \right) \end{aligned}$$

is nonnegative. Therefore,

$$\Delta_h(\delta, y, u) \leq \frac{1}{2} \left( \Delta_h^*(\delta, y, u) + \Delta_h^*(\delta, y, u - 1/2) \right) + O\left(\frac{\Delta_h^*(\delta, y, u)}{u \log y}\right).$$

By taking the sup of  $\Delta_h(\delta, y, u')$  over  $1/2u'u$  and using the above bound, we find that there exists some constant  $C > 0$ ,

$$\Delta_h^*(\delta, y, u) \leq \left( 1 + C \frac{1}{u \log y} \right) \Delta_h^*(\delta, y, u - 1/2).$$

Iterating this inequality, we deduce that

$$\Delta_h^*(\delta, y, u) \leq e^{O(\frac{\log u}{\log y})} \Delta_h^*(\delta, y, 1). \quad (\text{A.3})$$

Recall that for  $c \in [1/2, 1]$ ,

$$\Delta_h(\delta, y, c) = \frac{\delta \log y}{hy^c \rho(c)} \left( \Psi_q(y^c, y) - \Psi_q(y^c(1 - h), y) \right).$$

Finally, seeing as

$$\Psi_q(y^c, y) - \Psi_q(y^c(1 - h), y) = \sum_{\substack{y^c(1-h) < n \leq y^c \\ p|n \Rightarrow p \in (y^\delta, y]}} 1 \leq \sum_{\substack{y^c(1-h) < n \leq y^c \\ p|n \Rightarrow p > y^\delta}} 1,$$

and it follows from a simple sieve argument (for example, by [22, Theorem 3.6]) that

$$\sum_{\substack{y^c(1-h) < n \leq y^c \\ p|n \Rightarrow p > y^\delta}} 1 \ll \frac{hy^c}{\delta \log y},$$

uniformly for  $c \in [1/2, 1]$ ,  $h \geq y^{-1/3}$ , and  $\frac{1}{\log y} \leq \delta \leq \frac{1}{10}$ . We therefore have

$$\Delta_h(\delta, y, c) \ll 1,$$

uniformly for  $c \in [1/2, 1]$ . Combining this with (A.3), we deduce that

$$\Delta_h^*(\delta, y, u) \leq e^{O(\frac{\log u}{\log y})} \ll 1,$$

which follows from the fact that  $y \geq e^{(\log \log x)^{5/3+\varepsilon}}$ . This completes the proof of (A.2) and the lemma is proved. □

## REFERENCES

- [1] Rachid Caich. Random multiplicative functions and typical size of character in short intervals. 2024. Preprint available at [arxiv.org/abs/2402.06426](https://arxiv.org/abs/2402.06426).
- [2] Christine Chang. Hybrid statistics of a random model of zeta over intervals of varying length. 2024. Preprint available at [arxiv.org/abs/2404.08575](https://arxiv.org/abs/2404.08575).
- [3] Régis de la Bretèche and Gérald Tenenbaum. Propriétés statistiques des entiers friables. *Ramanujan J.*, 9(1-2):139–202, 2005.
- [4] E. Fouvry and G. Tenenbaum. Répartition statistique des entiers sans grand facteur premier dans les progressions arithmétiques. *Proc. London Math. Soc. (3)*, 72(3):481–514, 1996.
- [5] John B. Friedlander. Integers free from large and small primes. *Proc. London Math. Soc. (3)*, 33(3):565–576, 1976.
- [6] Ofir Gorodetsky and Mo Dick Wong. On the limiting distribution of sums of random multiplicative functions. 2025. Preprint available at [arxiv.org/abs/2508.12956](https://arxiv.org/abs/2508.12956).
- [7] Andrew Granville. Smooth numbers: computational number theory and beyond. In *Algorithmic number theory: lattices, number fields, curves and cryptography*, volume 44 of *Math. Sci. Res. Inst. Publ.*, pages 267–323. Cambridge Univ. Press, Cambridge, 2008.
- [8] Seth Hardy. The distribution of partial sums of random multiplicative functions with a large prime factor. 2025. Preprint available at [arxiv.org/abs/2503.06256](https://arxiv.org/abs/2503.06256).
- [9] A. J. Harper. Sharp conditional bounds for moments of the Riemann zeta function. 2013. Preprint available at [arxiv.org/abs/1305.4618](https://arxiv.org/abs/1305.4618).
- [10] A. J. Harper. Minor arcs, mean values, and restriction theory for exponential sums over smooth numbers. *Compos. Math.*, 152(6):1121–1158, 2016.
- [11] A. J. Harper. Moments of random multiplicative functions, II: High moments. *Algebra Number Theory*, 13(10):2277–2321, 2019.
- [12] A. J. Harper. Moments of random multiplicative functions, I: Low moments, better than squareroot cancellation, and critical multiplicative chaos. *Forum Math. Pi*, 8:e1, 95, 2020.
- [13] A. J. Harper. Almost sure large fluctuations of random multiplicative functions. *Int. Math. Res. Not. IMRN*, pages 2095–2138, 2023.
- [14] A. J. Harper. The typical size of character and zeta sums is  $o(\sqrt{x})$ . 2023. Preprint available at [arxiv.org/abs/2301.04390](https://arxiv.org/abs/2301.04390).
- [15] A. J. Harper. Moments of random multiplicative functions, III: A short review. 2024. Preprint available at [arxiv.org/abs/2410.11523](https://arxiv.org/abs/2410.11523).
- [16] H. Helson. Hankel forms. *Studia Math.*, 198(1):79–84, 2010.
- [17] Douglas Hensley. The number of positive integers  $\leq x$  and free of prime factors  $> y$ . *Journal of Number Theory*, 21:286–298, 1985.
- [18] Adolf Hildebrand. Integers free of large prime factors and the Riemann hypothesis. *Mathematika*, 31(2):258–271, 1984.
- [19] Adolf Hildebrand. On the number of positive integers  $\leq x$  and free of prime factors  $> y$ . *J. Number Theory*, 22(3):289–307, 1986.

- [20] Adolf Hildebrand and Gérald Tenenbaum. On integers free of large prime factors. *Trans. Amer. Math. Soc.*, 296(1):265–290, 1986.
- [21] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math. (2)*, 183(3):1015–1056, 2016.
- [22] H. L. Montgomery and R. C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2007.
- [23] K. Soundararajan. Smooth numbers in short intervals, 2010. Preprint available at [arxiv.org/abs/1009.1591](https://arxiv.org/abs/1009.1591).
- [24] Kannan Soundararajan. Moments of the Riemann zeta function. *Annals of Mathematics*, 170(2):981–993, 2009.
- [25] Gérald Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 46 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, french edition, 1995.
- [26] Victor Y. Wang and Max Wenqiang Xu. Harper’s beyond square-root conjecture. *International Mathematics Research Notices*, 2025:rnf279, 2025.
- [27] Aurel Wintner. Random factorizations and Riemann’s hypothesis. *Duke Math. J.*, 11:267–275, 1944.
- [28] Max Wenqiang Xu. Better than square-root cancellation for random multiplicative functions. *Trans. Amer. Math. Soc. Ser. B*, 11:482–507, 2024.
- [29] Ti Zuo Xuan. Integers with no large prime factors. *Acta Arith.*, 69(4):303–327, 1995.
- [30] Khalid Younis. Asymptotics for smooth numbers in short intervals, 2024. Preprint available at [arxiv.org/abs/2409.05761](https://arxiv.org/abs/2409.05761).

MATHEMATICS INSTITUTE, ZEEMAN BUILDING, UNIVERSITY OF WARWICK, COVENTRY CV4 7AL, ENGLAND

*Email address:* `seth.hardy@warwick.ac.uk`

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, 251 MERCER STREET, NEW YORK 10012, USA

*Email address:* `maxxu1729@gmail.com`