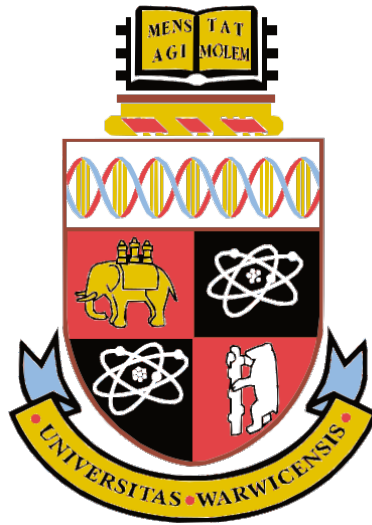




Large Gaps Between Primes

by

Simon Shackleton

MA4K8 Scholarly Report

Submitted to The University of Warwick

Mathematics Institute

April, 2026



Contents

1	Introduction	1
2	The Large Gaps Method	2
3	Probabilistic Estimates	3
3.1	Large Gaps Conjectures	4
4	Sieve Methods	6
4.1	Primes in Arithmetic Progressions	9
5	Erdős-Rankin Construction	10
6	Probabilistic Approach	13
6.1	GPY Method and Maynard-Tao Weights	14
6.2	Adapted Maynard-Tao Weights	16
7	Preliminary Sieve Estimates	17
8	Sieve Estimates	21
8.1	Setup	21
8.2	Calculations	21
9	Improvements	36
9.1	Outline	37
10	Acknowledgments	40

1 Introduction

“It is evident that the primes are randomly distributed but, unfortunately, we don’t know what ‘random’ means.” [13]

— R. C. Vaughan, (February, 1990)

Around 300BC Euclid proved that there are infinitely many primes. Once we know this, it is natural to try to understand their distribution. In particular, one may ask how large or how small we can guarantee the gaps between consecutive primes to be. In 1849 De Polignac conjectured that every even number is the gap between two consecutive primes infinitely often. This is, of course, a broader form of the famous twin prime conjecture. Similarly, Dickson conjectured in 1904 that there are infinitely many k -tuples for which each of $n + h_1, \dots, n + h_k$ are prime where $\mathcal{H} = \{h_1, \dots, h_k\}$ is an admissible set [14, Section 1]. We say that $\mathcal{H}$ is *admissible* if for all primes p there is some residue class modulo p which none of the h_i lie in.

Both of these problems are open, but headway was made in 2013 by Zhang [30] on De Polignac’s Conjecture and by Maynard [19] and Tao (unpublished) on Dickson’s Conjecture. The authors were able to show the existence of infinitely many prime gaps of bounded length.

The methods used by Maynard and Tao heavily influence the developments made on Large gaps between primes and shall be discussed for motivation. Throughout this project, we let $G(X)$ be the largest gap between consecutive primes of size at most X . More precisely,

$$G(X) := \max_{p_n \leq X} p_{n+1} - p_n \quad (1.1)$$

The prime number theorem tells us that the average gaps between primes in $[1, X]$ is asymptotically $\log X$. This, along with the pigeonhole principle, enables us to guarantee that $G(X) \geq (1 + o(1)) \log X$. The first major breakthrough in the large gaps problem was by Westzynthius [29], who showed that there are gaps between primes that are arbitrarily large relative to the average gap. In the 1930s, Erdős and Rankin showed that

$$G(X) \geq (c' + o(1)) \frac{\log X \log \log X \log \log \log X}{(\log \log X)^2} \quad (1.2)$$

for some $c' > 0$. In fact Erdős [6] showed (1.2) without the $\log \log \log \log X$ term and Rankin [25] later showed (1.2) with $c' = 1/3$. Erdős conjectured that c' can be taken arbitrarily large, famously offering a prize of \$5000 for showing this, the largest of the Erdős prizes [7]. This conjecture was resolved affirmatively by Ford, Green, Konyagin and Tao [9], and Maynard [21] independently.

Theorem 1.1. *We have that*

$$\limsup_{X \rightarrow \infty} \frac{G(X)}{\log X \log_2 X \log_4 X (\log_3 X)^{-2}} = \infty \quad (1.3)$$

Here, and throughout this project, $\log_v X$ denotes the v -fold logarithm.

The respective papers were released only a day apart, both giving distinct approaches. Ford, Green, Konyagin, Maynard and Tao [8] later collaborated to improve Theorem 1.1, proving the best result to date.

Theorem 1.2. *For any sufficiently large X , one has*

$$G(X) \gg \frac{\log X \log_2 X \log_4 X}{\log_3 X} \quad (1.4)$$

This project focuses on the method developed by Maynard to prove Theorem 1.1. We will first discuss the basic method for tackling large gaps between primes. Then we shall discuss some conjectures on large gaps and give a brief discussion of sieve methods, which will be the key tools in this approach. Then we will discuss the more key details of the approach taken by Erdős and Rankin and why Maynard's approach improves on previous attacks, before presenting the proof presented by Maynard. Finally, we will give an overview of the key ideas in the proof of Theorem 1.2, explaining how this improves on the method in [21].

2 The Large Gaps Method

We give the standard method of attack on the large gaps between primes problem. We follow [16] and [18]. First, we give an alternative proof that $G(X) \geq (1 + o(1)) \log X$. Consider the following list of numbers

$$P_n + 2, P_n + 3, \dots, P_n + n$$

Here, $P_n := \prod_{p \leq n} p$ denotes the n -th primorial. The above gives a sequence of $n - 1$ consecutive composite numbers. So we have constructed a gap between primes of at least size n . The prime number theorem tells us that $P_n = e^{n(1+o(1))}$. Choosing $n = (1 - \epsilon) \log X$ guaranties that $G(X) \geq (1 - \epsilon + o(1)) \log X$, and since ϵ was arbitrary, we have $G(X) \geq (1 + o(1)) \log X$.

The approach above was to directly construct some string of composite numbers. We can do this in a slightly different way. Consider the following set

$$\{2, 3, \dots, U\} \quad (2.1)$$

We note that every element of the above set is divisible by some $p \leq U$. In other words, each element m of our set in (2.1) satisfies $m \equiv 0 \pmod{p}$ for some prime $p \leq U$. This motivates the following definition

Definition 2.1. *A system of congruences $\{a_i \pmod{q_i}\}_{1 \leq i \leq k}$ covers a set of integers $\mathcal{N}$ if for every $n \in \mathcal{N}$ there is some $1 \leq i \leq k$ with $n \equiv a_i \pmod{q_i}$.*

So the set of congruences $\{0 \pmod{p}\}_{p \leq U}$ covers (2.1). Assume instead we have some more general system of congruences $\{a_p \pmod{p}\}_{p \leq x}$ that cover the set $\{1, 2, \dots, U\}$. Then the Chinese Remainder Theorem guarantees the existence of some $\gamma \in (x, x + \exp((1 + o(1))x)]$ satisfying $\gamma \equiv -a_p \pmod{p}$ for every prime $p \leq x$. So the string of numbers $\gamma + k$ for $k \in \{1, 2, \dots, U\}$ are all divisible by some $p \leq x$ whilst also being strictly large than x and hence must be composite. So now we have some prime $p_n \leq e^{x(1+o(1))}$ with $p_{n+1} - p_n \geq U$. If we then take $x = \log X$ we have proved that $G(X) \geq (1 + o(1))U$. So our goal to get a good lower bound for $G(X)$ is to show we can take U to be large, by choosing a suitable system of congruences covering $\{1, 2, \dots, U\}$.

However, this method is believed to be limited. Our approach above locates a string of integers divisible by some $p \leq x$ in the interval $(x, x + \exp((1 + o(1))x)]$. This is of course sufficient for the integers to be composite, but is really quite strong. In fact, this method gives us a lower bound of the maximal gap between integers coprime to P_x , denoted $j(P_x)$, and our method cannot go beyond this. Maier and Pomerance [18, Section 1] conjectured that $j(P_x) \leq x(\log x)^{2+o(1)}$ and so the limit of this method would therefore be to establish a bound of the form $G(X) \geq \log X(\log \log X)^{2+o(1)}$. We will see in Section 3.1 that this is expected to be far from the truth.

But this method does give us a foothold in the large gaps problem, and this helpful change of perspective has turned our question about finding long strings of consecutive composite numbers into a sieve theoretic covering problem. We can now borrow methods from sieve theory to tackle the problem. However, we postpone our discussion of sieve methods for now to discuss some conjectures on the size of $G(X)$.

3 Probabilistic Estimates

When trying to understand gaps between primes, it can be helpful to have some sort of heuristic model that indicates what we may expect to be the truth; gaining some understanding of how far away we may be from what seems to be true. Here, we follow [2], [13] and [16, Chapter 29]. A rather elementary way one can do this is via Cramér's model, given by a sequence of independent Bernoulli random variables $(X_n)_{n=1}^\infty$, which for $n \geq 3$ have probability distribution

$$\mathbb{P}(X_n = 1) = 1/\log n$$

and we set for completion $X_1 = 0$ and $X_2 = 1$ [2]. These random variables are intended to model the primes, as by the prime number theorem there are approximately $x/\log x$ primes up to x . That is, the primes have density approximately $1/\log x$ in $[1, x]$ and so we suspect that a number n has a probability of about $1/\log n$ of being prime and a probability of roughly $1 - 1/\log n$ of being composite. However, a problem one immediately notices is that Cramér's model is insensitive to arithmetic considerations. Indeed, this model suggests that there are infinitely many pairs of primes of the form $n, n + 1$, which is of course false. Noting this, we would somehow like these random variables to be conscious of congruence considerations modulo small primes. Granville [13], motivated by earlier work of Maier on irregularities in the distribution of primes [17], introduced a refinement of Cramér's model, which does exactly this. Let

$$\mathcal{N} := \{n > y^2 : (n, P_y) = 1\} \quad (3.1)$$

Here y is a parameter in $\mathbb{R}_{\geq 1}$ that we are free to choose. We call y the Cramér-Granville parameter. The sieve of Eratosthenes-Legendre, or a simple inclusion-exclusion argument, implies that $\mathcal{N}$ contains $\alpha := \prod_{p \leq y} (1 - 1/p)$ proportion of $\mathbb{N}$. So an integer $n \in \mathcal{N}$ should have a probability $\alpha^{-1}/\log n > 1/\log n$ of being prime. We now construct a new sequence of independent Bernoulli random variables $(Y_n)_{n=1}^\infty$, which for $n \geq 3$ have probability distribution

$$\mathbb{P}(Y_n = 1) = \alpha^{-1} \mathbf{1}_{n \in \mathcal{N}} / \log n \quad (3.2)$$

and again for completion, we take $Y_1 = 0$ and $Y_2 = 1$. We also take $\mathbf{1}_{n \in \mathcal{A}}$ to be the indicator function on some set $\mathcal{A}$ here and throughout this project. This sequence of random variables $(Y_n)_{n=1}^\infty$ is the Cramér-Granville model from [13]. Note that when we choose $y = 1$ this is exactly Cramér's model.

3.1 Large Gaps Conjectures

Conjecture 3.1 (Cramér's Conjecture, weak form).

$$G(X) = (\log X)^{2+o(1)} \quad (3.3)$$

We use Cramér's model to justify this, giving the argument in [2]. First we state the following well known lemma from Probability theory [1, Theorems 4.3, 4.4].

Lemma 3.2 (Borel-Cantelli). *Let $E_1, E_2, \dots$ be some events in a probability space. Let E be the event that infinitely many of the E_i s occur. Then*

1. *If $\sum_{i=1}^\infty \mathbb{P}(E_i) < \infty$, then $\mathbb{P}(E) = 0$.*
2. *If $\sum_{i=1}^\infty \mathbb{P}(E_i) = \infty$ and each of the E_i s are mutually independent, then $\mathbb{P}(E) = 1$.*

Cramér describes the locating primes like drawing balls from a sequence of urns $U_1, U_2, \dots$

with probability $1/\log n$ of drawing a white ball from urn U_n (n being prime) and $1 - 1/\log n$ of drawing a black ball (n being composite). Cramér then defines the random variables P_n to take the value k where k is the index of the k th Urn from which the n th white ball is drawn.

Cramér then considers the class C of all possible sequences $(P_n)_{n=1}^\infty$ and notes that the sequence of prime numbers $(p_n)_{n=1}^\infty$ is contained in this class. We then consider the order of the difference $P_{n+1} - P_n$. Let $c > 0$ be some constant and let E_m denote the event that all black balls are drawn from the urns U_{m+v} with $1 \leq v \leq c(\log m)^2$. The event in which the inequality

$$P_{m+1} - P_m > c(\log P_m)^2$$

is achieved infinitely often has the same probability as E_m occurring for an infinite number of m s. As each of the events are independent, we have

$$\mathbb{P}(E_m) = \prod_{i=1}^{c(\log m)^2} \left(1 - \frac{1}{\log(m+i)} \right) \asymp \frac{1}{m^c}$$

We first note that if $c > 1$ then we immediately see that $\sum_i \mathbb{P}(E_i) < \infty$ and hence by Lemma 3.2 we've that the inequality $P_{m+1} - P_m > c(\log m)^2$ is achieved for at most finitely many m with probability 1.

If instead $c < 1$ we construct non-overlapping intervals to ensure that the E_m s are independent. Let $m_1 = 2$ and $m_{j+1} = m_j + [c(\log j)^2] + 1$. Clearly we have $m_j \ll j(\log j)^2$ so $\sum_j \mathbb{P}(E_{m_j}) = \infty$. So we have by Lemma 3.2 that the inequality $P_{n+1} - P_n > c(\log n)^2$ is satisfied for infinitely many n with probability 1. Together, this shows that with probability 1 we have

$$\limsup_{n \rightarrow \infty} \frac{P_{n+1} - P_n}{(\log P_n)^2} = 1$$

This led Cramér to conjecture the following

Conjecture 3.3 (Cramér's Conjecture).

$$\limsup_{X \rightarrow \infty} \frac{G(X)}{(\log X)^2} = 1 \tag{3.4}$$

Of course, this would also lead one to posit (3.1). After such an observation, one may be interested if we can do any better than this. Perhaps we may hope for an asymptotic relation?

Conjecture 3.4 (Shanks).

$$G(X) \sim (\log X)^2 \tag{3.5}$$

Shanks [27] gives a similar heuristic probability argument to justify this.

We have already discussed how Cramér's model has some serious flaws, so we shall now apply the Cramér-Granville model to this situation, following [13] and [16, Chapter 29]. Taking E_k as the event that $Y_n = 0$ for all $n \in (k, k + c(\log k)^2]$, we have

$$\mathbb{P}(E_k) = \prod_{i=1}^{c(\log k)^2} \left(1 - \frac{\alpha^{-1} \mathbf{1}_{k+i \in \mathcal{N}}}{\log(k+i)} \right)$$

We now take k to be some multiple of P_y , where y is the Cramér-Granville parameter in (3.1). We note from Mertens' third estimate that we have $\alpha \sim e^{-\gamma}/\log y$, and for such a choice of k we have

$$\sum_{\substack{k < n \leq k + c(\log k)^2 \\ p \leq y \implies p \nmid n}} 1 \sim \pi(cy^2) \sim \frac{cy^2}{2 \log y} \quad (3.6)$$

This is because any $k+m$ in the range of the sum on the left hand side of (3.6) is divisible by $p \leq y$ if and only if $p|m$ and that $c(\log k)^2 \sim cy^2$ from the prime number theorem. So, for $c < 1$ the sum on the left hand side of (3.6) is equal to $\pi(cy^2)$ and for $c \geq 1$ it differs only by $|\{p_1 p_2 : y < p_1, p_2 \leq \sqrt{cy}(1 + o(1)) \text{ prime}\}| \asymp (\sqrt{c} - 1 + o(1))^2 y^2 / (\log y)^2$, which has negligible contribution. So we see that

$$\mathbb{P}(E_k) = \left(1 - \frac{e^{-\gamma} \log y}{y} \right)^{\frac{cy^2(1+o(1))}{2 \log y}} = k^{-ce^{\gamma}/2 + o(1)}$$

Reasoning now similarly to as we did for Cramér's conjecture, this leads us to conjecture

$$G(X) \geq (2e^{-\gamma} + o(1))(\log X)^2$$

This is quite a crushing blow for Cramér's and Shanks' conjectures, and in fact we do not know what the true value of $\limsup_{X \rightarrow \infty} G(X)$ should be.

4 Sieve Methods

We give a brief overview of the sieve and some results, as a background for the calculations to come. We follow [10, Chapters 1, 4, 5, 7] and [16, Chapters 18-21]. We consider a sequence of non-negative real numbers $\mathcal{A} = (a_n)_{n=1}^{\infty}$. We will have an abuse of notation where $\mathcal{A}$ will sometimes refer to $(a_n)_{n=1}^{\infty}$ and in other contexts refer to the support of $(a_n)_{n=1}^{\infty}$. In sieve theory, we are interested in the size of $\mathcal{A}$ when we remove elements with n satisfying some congruence conditions modulo a set of primes $\mathcal{P}$, more precisely, we hope to understand the sifting function

$$\mathcal{S}(\mathcal{A}, z) := \sum_{(n, \mathcal{P}(z))=1} a_n$$

where

$$\mathcal{P}(z) := \prod_{\substack{p \in \mathcal{P} \\ p < z}} p$$

The sifting function tells us the size of some very interesting sets. For example, if we choose $a_n = \mathbf{1}_{n \leq x}$, $\mathcal{P}$ to be the set of all primes and take $z = \sqrt{x}$, then the sifting function tells us the number of primes in the interval $[\sqrt{x}, x]$. If instead we choose $a_n = \mathbf{1}_{n=m(m+2)}$ for $m \leq x$, $\mathcal{P}$ to be the set of all primes and $z = \sqrt{x+2}$, then the sifting function gives us a count of the number of twin primes in $[\sqrt{x+2}, x]$. The sifting function can otherwise be written as

$$\mathcal{S}(\mathcal{A}, z) = \sum_{(n, \mathcal{P}(z))=1} a_n = \sum_n a_n \mathbf{1}_{(n, \mathcal{P}(z))=1}$$

It is standard to write the one indicator of $(n, \mathcal{P}(z))$ in convolution form

$$\sum_{d|(n, \mathcal{P}(z))} \mu(n) = \sum_{\substack{d|n \\ d|\mathcal{P}(z)}} \mu(d) = \mathbf{1}_{(n, \mathcal{P}(z))}$$

and making this substitution leads us to

$$\mathcal{S}(\mathcal{A}, z) = \sum_n a_n \sum_{d|(n, \mathcal{P}(z))} \mu(n) = \sum_{d|\mathcal{P}(z)} \mu(d) |\mathcal{A}_d| \quad (4.1)$$

where

$$|\mathcal{A}_d| := \sum_{n \equiv 0 \pmod{d}} a_n$$

Notice that this looks strikingly similar to simply performing an inclusion exclusion argument. For very small z an inclusion-exclusion argument would suffice as the error from approximating $|\mathcal{A}_d|$ should be small. However, our goal is to take z as large as possible (ideally $z = x^\theta$) and for such a large choice of z the contribution from our error terms when approximating $|\mathcal{A}_d|$ is large. We expect that every $|\mathcal{A}_d|$ can be written in the form

$$Xg(d) + r_d \quad (4.2)$$

Here, X is a smooth approximation to $|\mathcal{A}|$. $g(d)$ is called the density function and represents the proportion of $|\mathcal{A}|$ that we expect to be divisible by d and r_d is the error term. It is also standard to require some regularity in g that one expects from Mertens' 3rd estimate. Specifically, for $z > w \geq 2$ there exist constants $\kappa \geq 0$ and $K > 0$ such that g satisfies the following

$$\prod_{w \leq p < z} (1 + g(p))^{-1} \leq K \left(\frac{\log z}{\log w} \right)^\kappa \quad (4.3)$$

We are furthermore led to guess that $g(d)$ should be multiplicative, as we expect the events that a natural number n is divisible by two distinct primes to be independent. So, one may

guess that the right hand side of (4.1) can be written in Euler product form, becoming

$$\mathcal{S}(\mathcal{A}, z) = X \prod_{p \leq z} (1 - g(p)) + \sum_{d | \mathcal{P}(z)} \mu(d) r_d \quad (4.4)$$

One may then naively guess that the main contribution comes from this Euler product term. Unfortunately, it is difficult to show that any cancellation is occurring in the error sum in many scenarios, and in some cases it is not possible to show that significant cancellation is occurring at all! For example, take $a_n = \mathbf{1}_{n \leq x}$ and $z \leq \sqrt{x}$. The sifting function tells us the number of primes in $[\sqrt{x}, x]$ which by the prime number theorem is asymptotically $x/\log x$. In this case, we also have $g(p) = 1/p$ for each p . So plugging this into (4.4), Mertens' third theorem then gives us that the expected main term is asymptotically $2e^{-\gamma}x/\log x$. So we see that the remainder term must have the same order as the expected main term. An explanation for this is that for a natural number n , divisibility by distinct primes are not independent events. Consider, for example, three primes $p_1, p_2, p_3 \geq n^{1/3}$. It cannot be the case that n is divisible by all three primes simultaneously. So, our density function will not be multiplicative. Our assumption that g should be multiplicative is one of the main limitations of sieve methods.

How can we remedy this? Excluding some A_d may allow us to get a better grasp on the error term whilst still allowing us to bound the sifting function. More generally, in sieve methods one is interested in replacing the Möbius function on the right hand side of (4.1) with 'sieve weights' which behave like the Möbius function but reduce the error term. The first sieve weights were given by the Möbius function on a well chosen truncated support. Selberg instead introduced a sieve weight given by a smooth truncation of the Möbius function. The truncation being smooth is what may be attributed to it's success. We give the following result due to Selberg, in the form of [16, Theorem 21.1].

Theorem 4.1. *Let $\mathcal{A}$ and $\mathcal{P}$ satisfy $A_d = Xg(d) + r_d$. If $D \geq 1$ and $h(p) = g(p)/(1 - g(p))$ then*

$$\mathcal{S}(\mathcal{A}, \mathcal{P}) \leq \frac{X}{L} + \sum_{\substack{d \leq D \\ d | \mathcal{P}}} 3^{\omega(d)} r_d \quad \text{with } L = \sum_{\substack{m \leq \sqrt{D} \\ m | \mathcal{P}}} h(m) \quad (4.5)$$

We similarly have a fundamental lemma type result due to Selberg [10, Corollary 7.8].

Theorem 4.2 (Selberg Fundamental Lemma). *Let $g(d)$ be a density function satisfying (4.3). Let $k := \kappa + \log K$. For $\frac{\log D}{\log z} =: s \geq 2k + (2 + c)\sqrt{2k \log k}$ we have*

$$\mathcal{S}(\mathcal{A}, z) = \prod_{p | \mathcal{P}(z)} (1 - g(p)) \{1 + O(k^{-c})\} + 2\theta \sum_{\substack{d | \mathcal{P}(z) \\ d < D}} 3^{\omega(d)} |r_d|$$

where $|\theta| \leq 1$ and the implied constant is absolute.

These results give us a good understanding of the behaviour of the sifting function, as we hoped to achieve. Unfortunately, improving on these results is hindered by the parity problem of sieve theory; see [16, Chapter 21]. However, these results are sufficient for our purposes, and we will use these in some preliminary sieving arguments to set up the more complex calculations to follow. However, for these sieves to be of much use, we need to be able to bound our remainders well. To do this, we need to have a good understanding of the distribution of primes in arithmetic progressions.

4.1 Primes in Arithmetic Progressions

We follow [15, Chapter 3] and [16, Chapters 12, 18, 26]. One may guess that the primes in arithmetic progressions are distributed equally on average; for numbers $1 \leq a \leq q$ with $(a, q) = 1$ we may expect that

$$\pi(x; q, a) \sim \pi(x)/\phi(q) \quad (4.6)$$

In 1896 de la Vallée Poussin showed (4.6) using the ideas of Riemann and Dirichlet [3]. Following this, one becomes interested in the size of the moduli one can take in which the asymptotics still hold. In 1936 Walfisz, following the work of Siegel, and using the theory of Dirichlet L -series [28] proved the following

Theorem 4.3 (Siegel-Walfisz). *Let $D > 0$. There exists an absolute constant $c > 0$ such that if $1 \leq q \leq (\log x)^D$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$, then*

$$\pi(x; q, a) = \frac{\text{Li}(x)}{\phi(q)} + O_D(xe^{-c\sqrt{\log x}}) \quad (4.7)$$

This is in the form of [16, Theorem 12.1]. The Siegel-Walfisz theorem is a highly celebrated result. But how far away is this from what we expect to be true? Recall that the Generalized Riemann Hypothesis implies

$$\pi(x; q, a) = \frac{\text{Li}(x)}{\phi(q)} + O(\sqrt{x} \log x) \quad (x \geq 2, a \in (\mathbb{Z}/q\mathbb{Z})^*)$$

Indicating that (4.6) should still hold with $q \ll x^{1/2-\epsilon}$ for any $\epsilon > 0$. So the Siegel-Walfisz theorem seems far away from what we expect to be true. However, in the 1960s, from the development of the large sieve came one of the most celebrated results in mathematics, which we give in the form of [16, Theorem 18.9].

Theorem 4.4 (Bombieri-Vinogradov). *Fix $A \geq 0$. For $x \geq 2$ and $1 \leq Q \leq \sqrt{x}/(\log x)^{A+3}$, we have*

$$\sum_{q \leq Q} \max_{\substack{y \leq x \\ (a, q) = 1}} \left| \pi(y; q, a) - \frac{\text{Li}(y)}{\phi(q)} \right| \ll_A \frac{x}{(\log x)^{A+1}} \quad (4.8)$$

This theorem has proven to be invaluable in Sieve methods. We give another form of Theorem 4.4 which is more useful in certain arguments.

Corollary 4.4.1. *Let h be some positive integer and let*

$$E(x; q) = \max_{\substack{y \leq x \\ (a, q) = 1}} \left| \pi(y; q, a) - \frac{Li(y)}{\phi(q)} \right| \quad (4.9)$$

then given any positive constant U there exists a positive constant $C = C(U, h)$ such that

$$\sum_{d < x^{1/2}/\log^C x} \mu(d)^2 h^{\omega(d)} E(x; d) \ll_{U, h} \frac{x}{\log^U x} \quad (4.10)$$

This is a slightly weaker form of [15, Lemma 3.5].

Given $\theta > 0$ we say the primes have *level of distribution* θ if for all $A > 0$ (4.8) holds for $Q = x^\theta$. So the Bombieri-Vinogradov theorem tells us the primes have level of distribution θ for all $\theta < 1/2$. Considering this is exactly what the Riemann Hypothesis implies this is an outstanding result, because of this the Bombieri-Vinogradov theorem is often called the Generalised Riemann Hypothesis on average.

Can we do any better than this? Certainly no unconditional results have been able to improve on all $\theta < 1/2$. However it is believed that one can do considerably better.

Conjecture 4.5 (Elliott and Halberstam). *The primes have level of distribution θ for all $\theta < 1$.*

This conjecture [5] seems very far out of reach, but has been useful in demonstrating the limitations of the best known current methods in gaps between primes.

5 Erdős-Rankin Construction

Following section 2, we now want to come up with some explicit construction of a system of congruences with moduli being the primes up to x that covers the set $\mathbb{Z} \cap [1, U]$. All attacks on the large gaps problem follow the same set up, coming from the work of Erdős and Rankin. We give the set up as in [21]. Fix constants C_U, ϵ . We define the following parameters

$$y = \exp\left(\frac{(1 - \epsilon) \log x \log_3 x}{\log_2 x}\right), \quad z = \frac{x}{\log_2 x}, \quad U = \frac{C_U x \log y}{\log_2 x} \quad (5.1)$$

and we pick the following residue classes

$$a_p = 1 \text{ for primes } p \leq y \quad (5.2)$$

$$a_p = 0 \text{ for primes } y < p \leq z \quad (5.3)$$

for primes $\leq z$. Motivation for the choice of these parameters and residue classes comes from the anatomy of integers. Heuristically, we expect most integers to be divisible by some medium sized prime, so these congruence relations should cover most numbers up to x . It is standard to call the first sifting the sifting of integers in $[1, U]$ by the residue classes in (5.3). Here, sifting a set by some set of congruences means the removal of all elements that satisfy at least one of the congruence relations. This then leaves the first residual set. The second sifting is by the congruences in (5.2). Once these siftings have been performed, we are left with two sets

$$\mathcal{R}' = \{m \leq U : m \text{ is } y \text{ smooth}, (m-1, P_y) = 1\} \quad (5.4)$$

$$\mathcal{R} = \{mp \leq U : p > z, m \text{ is } y \text{ - smooth}, (mp-1, P_y) = 1\} \quad (5.5)$$

Here we say that an integer n is k -smooth if all prime divisors of n are less than or equal to k . We note that in $\mathcal{R}$, as $p > z$, the condition $(mp-1, P_y) = 1$ forces m to be even. We will use this throughout the argument. We now split $\mathcal{R}$ according to this integer m . For even m we let

$$\mathcal{R}_m = \{z < p \leq U/m : m \text{ is } y \text{ - smooth}, (mp-1, P_y) = 1\} \quad (5.6)$$

We are interested in understanding the sizes of $\mathcal{R}$ and $\mathcal{R}'$. We deviate from [21] and postpone the proof of the following estimates, which we will prove in section 7, to discuss how the method in [21] differs from [18], giving a motivation similar to that in [22]. We have

$$|\mathcal{R}'| \ll \frac{x}{(\log x)^{1+\epsilon}} \quad (5.7)$$

and

$$\sum_{U/z(\log_2 x)^2 \leq m < U/z} |\mathcal{R}_m| = o\left(\frac{C_U x}{\log x}\right), \quad \sum_{1 \leq m \leq U/z(\log_2 x)^2} |\mathcal{R}_m| = O\left(\frac{C_U x}{\log x}\right) \quad (5.8)$$

It therefore follows from the prime number theorem that we can trivially cover $\mathcal{R}'$ and $\mathcal{R}_m$ for $U/z(\log_2 x)^2 \leq m < U/z$ by choosing for each element of the respective sets an individual residue class for each prime in $(z, x/2)$.

The same may also be done for the whole of $\mathcal{R}' \cup \mathcal{R}$ by choosing for each element a corresponding residue class for a unique prime in $[x/2, x]$, enabling us to cover the whole of $\mathbb{Z} \cap [1, U]$ provided that C_U is sufficiently small. This would yield a result of the form (1.2). However, our goal is to now cover each $\mathcal{R}_m$ for $m \leq U/z(\log_2 x)^2$, where C_U can be taken arbitrarily large. To do this, we hope to show the existence of some residue class for a proportion of primes in $[x/2, x]$ containing many elements of $\mathcal{R}_m$. By doing this, we can increase the size of C_U while still covering all of $\mathbb{Z} \cap [1, U]$.

As noted in [18], if we were able to view $\mathcal{R}' \cup \mathcal{R}$ as a random set of residues (mod q)

where these are independent for each q in $(z, x]$ then we would expect to be able to sieve our approximately $(\log x)^{1+o(1)}$ members of $\mathcal{R}' \cup \mathcal{R}$ for a positive proportion of the q s, so we certainly expect to be able to do better than sift out a single element for each residue class. Some previous attempts to improve (1.2) took the approach of trying to find for some positive proportion of primes in $[x/2, x]$, a residue class that can sieve out two members of $\mathcal{R} \cup \mathcal{R}'$. Maier and Pomerance [18] succeeded in doing this by proving an averaging estimate for generalised twin primes in arithmetic progressions; that on average primes of a given gap in an arithmetic progression occur about as frequently as one would expect from the Cramér-Granville model, for example. Taking the moduli to be primes q in $[x/2, x]$ this succeeded in making the covering more efficient and improved the constant factor in (1.2) by showing that many pairs of members of $\mathcal{R}$ are congruent (mod q) for most q . Maynard in some sense takes a similar approach to this; however, he managed to show that the number of primes expected to lie in a given residue class, if the residue classes are chosen well, can be taken to be arbitrarily large.

We note that one of the key features that makes this particular situation difficult is that the modulus q is only slightly smaller than a given $p \in \mathcal{R}_m$. Attempting to find residue classes for q in $[x/2, x]$ containing many elements of $\mathcal{R}_m$ can be thought of somewhat differently. As we suggested above, if one could guarantee that for some arithmetic progression $(a + iq)_{i \leq k}$, that the progression captures many primes in the range of $\mathcal{R}_m$ we may be hopeful that many of these primes lie in $\mathcal{R}_m$. This would make a a good candidate for the chosen residue class, as if we were able to cover many elements of $\mathcal{R}_m$, we may be able to increase C_U and still cover the whole of $\mathcal{R}_m$. So we wish to guarantee the existence of some a such that many members of the progression $a + q, a + 2q, \dots, a + kq$ are prime. Motivated by the work on small gaps between primes, we believe that this can be done using the following theorem [22, Theorem 1].

Theorem 5.1. *Let $L_1, \dots, L_k$ be some linear integral functions of the form $L_i(n) = a_i n + b_i$. If for each prime p there exists some integer n_p where $\prod_{i=1}^k L_i(n_p)$ is coprime to p then there exists a constant $c > 0$ and a infinite number of integers n with at least $c \log k$ of the $L_1(n), \dots, L_k(n)$ prime.*

If we define $L_i(n) = n + h_i q$ for some suitable admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ then the existence of some residue class $a_q \pmod{q}$ that contains many primes is implied by the fact that many of the $n + h_i q$ are simultaneously prime for some n . So we will attempt to identify candidate residue classes in the same way we identified candidate n s for which many shifts of n are prime in the work on small gaps. Our goal now is to make this rigorous.

We find it helpful to attempt to cover each $\mathcal{R}_m$ with primes from disjoint intervals of $[x/2, x]$ that we denote $\mathcal{I}_m$, to avoid simple pathologies. Following the outlined approach, we claim we can show the following.

Proposition 5.2. Fix $\delta > 0$. For each $m \leq U/z(\log_2 x)^2$ and interval $\mathcal{I}_m \subset [x/2, x]$ of length at least $\delta|\mathcal{R}_m| \log x$ we have for $x > x_0(\delta, C_U)$ there is some choice of residue classes a_q for each $q \in \mathcal{I}_m$ such that each $p \in \mathcal{R}_m$ satisfies

$$p \equiv a_q \pmod{q} \quad (5.9)$$

for some $q \in \mathcal{I}_m$.

Remark. We note that for m in this range we have that $|\mathcal{R}_m| \gg x/(\log x)^2$, so $|\mathcal{I}_m| \gg \delta x / \log x$ guaranteeing $\sim |\mathcal{I}_m| / \log x$ primes in $\mathcal{I}_m$ by the prime number theorem.

Proof of Theorem 1.1 given Proposition 5.2. From (5.8) we have

$$\sum_{m < U/z(\log_2 x)^2} \delta|\mathcal{R}_m| \log x \ll \delta C_U x \quad (5.10)$$

and so we can choose sufficiently small δ relative to C_U such that we can pick each $\mathcal{I}_m \subset [x/2, x]$ non-overlapping for each even $m < U/z(\log_2 x)^2$. Hence by Proposition 5.2 we can guarantee that each $\mathcal{R}_m$ is covered by primes in $\mathcal{I}_m \subset [x/2, x]$. So by (5.7) and (5.8) we can cover all but a contribution of $o(x/\log x)$ of $\mathcal{R} \cup \mathcal{R}'$. The remaining uncovered elements of $\mathcal{R} \cup \mathcal{R}'$ can then be covered by individually picking residue classes of primes in $[z, x/2]$ for each element by noting $\pi(x/2) - \pi(z) \asymp x/\log x$. So we can choose residue classes of primes $p \leq x$ covering $[1, U]$ for each C_U given x sufficiently large, completing the proof. □

We note that we will instead prove an equivalent form of Proposition 5.2. We show that for a fixed choice of $\delta, \epsilon > 0$ and interval $\mathcal{I}_m \subset [x/2, x]$ of length at least $\delta|\mathcal{R}_m| \log x$ there exists a choice of residue classes such that one can cover all but $\epsilon|\mathcal{R}_m|$ members of $\mathcal{R}_m$ by prime elements of $\mathcal{I}_m$. Then one may append to $\mathcal{I}_m$ an interval of length $2\epsilon|\mathcal{R}_m| \log x$ and choose for each of the remaining $\epsilon|\mathcal{R}_m|$ elements a residue class for a single prime element of the appended interval, guaranteed by the prime number theorem. So we cover all of $\mathcal{R}_m$ by primes in an interval of length $(2\epsilon + \delta)|\mathcal{R}_m| \log x$. As ϵ and δ are arbitrary, this is equivalent to the statement of Proposition 5.2.

6 Probabilistic Approach

We outline the probabilistic method taken in [21]. For each even $m < U/z(\log_2 x)^2$ and prime $q \in \mathcal{I}_m$ we define a probability measure $\mu_{m,q}$ which gives the probability that we choose a particular residue class $a_q \pmod{q}$. Indeed, we want $\mu_{m,q}$ to be heavily concentrated on residue class in which many of the elements of $\mathcal{R}_m$ reside. Given such a probability measure we independently pick residue classes for each prime $q \in \mathcal{I}_m$. The

probability an individual $p \in \mathcal{R}_m$ does not lie in any of the chosen congruence classes is

$$\prod_{\substack{q \in \mathcal{I}_m \\ q \text{ prime}}} (1 - \mu_{m,q}(p)) = \exp \left(\sum_{\substack{q \in \mathcal{I}_m \\ q \text{ prime}}} \log(1 - \mu_{m,q}(p)) \right) \quad (6.1)$$

$$\leq \exp \left(- \sum_{\substack{q \in \mathcal{I}_m \\ q \text{ prime}}} \mu_{m,q}(p) \right) \quad (6.2)$$

So, if we can take the expected number of residue classes each prime $p \in \mathcal{R}_m$ lies in, let this be t , to be arbitrarily large, we will see that the expected number of elements of $\mathcal{R}_m$ that do not lie in any of these chosen congruence classes is at most $e^{-t}|\mathcal{R}_m|$. And so there must be some choice of residue classes which covers all but at most $\epsilon|\mathcal{R}_m|$ elements of $\mathcal{R}_m$ which by Proposition 5.2 is sufficient to prove Theorem 1.1.

6.1 GPY Method and Maynard-Tao Weights

We need to choose a probability measure $\mu_{m,q}$ that is concentrated on residue classes containing many elements of $\mathcal{R}_m$. We motivate our choice of measure, following [16, Chapter 28], [19] and [22]. Recall that the elements of $\mathcal{R}_m$ are simply primes satisfying some congruence conditions, so we want to identify residue classes containing many primes $p \leq U/m$ where we also have $(mp - 1, P_y) = 1$. We will see that ensuring small measure on residue classes unless similar congruences hold will be simple to enforce, so we now want to select some weight highly concentrated on primes in short intervals. To do this, the natural choice is the Maynard-Tao weights, which we now motivate.

The original problem concerning Maynard and Tao was to show the existence of bounded gaps between primes. The standard method of attack on this problem, developed by Goldston, Pintz and Yıldırım (abbreviated to GPY) is to for some admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ show that

$$\sum_{N \leq n \leq 2N} \left(\sum_{i=1}^k \chi_P(n + h_i) - \rho \right) w_n > 0 \quad (6.3)$$

where χ_P the indicator for the primes and w_n a choice of weight where $w_n \geq 0$ for every n . If we were able to show (6.3) then we would have for at least one $n \in [N, 2N]$ that

$$\sum_{i=1}^k \chi_P(n + h_i) - \rho > 0$$

So, at least $\lfloor \rho + 1 \rfloor$ of the $n + h_1, \dots, n + h_k$ are prime. So, our goal is now to find weights which should be concentrated on n for which many of the shifts $n + h_1, \dots, n + h_k$ are prime. It can be helpful to reformulate (6.3) in a more probabilistic light. The search for weights can equivalently be viewed as a search for a probability measure on $\mathbb{Z} \cap [N, 2N]$ defined

by

$$\mathbb{P}_{[N, 2N]}(m) := \frac{w_m}{\sum_{N \leq n \leq 2N} w_n} \quad (6.4)$$

With this in mind, showing (6.3) becomes equivalent to showing

$$\mathbb{E}_{[N, 2N]} \left[\sum_{i=1}^k \chi_P(n + a_i) \right] > \rho \quad (6.5)$$

So, what may be a good candidate for our probability measure? A constant or smooth, slowly changing function simply will not work due to the fact that the primes are sparse (have density zero) in the integers. Picking weights that depend on the prime values of $n + h_i$ i.e $w_n = \prod_{i=1}^k \chi_P(n + h_i)$ will fail as well, as understanding these weights is just as difficult as the question at hand; indeed, when viewing the w_n s as a probability measure as in (6.4), deducing the probability measure is well defined once we have scaled by the normalising constant would itself require us to guarantee the existence of prime k -tuples in $[N, 2N]$. A clear candidate for our weights is the Selberg sieve weights

$$w_n \propto \left(\sum_{\substack{d \mid \prod_{i=1}^k (n+a_i) \\ d \leq R}} \lambda_d \right)^2, \quad \lambda_d = \mu(d)(\log R/d)^k \quad (6.6)$$

As we discussed earlier, the Selberg sieve weights are so effective as they are a smooth truncation of the Möbius function, which is why their asymptotic sifting level with k (k being the dimension of the sieve) is better than the fundamental lemma sieve [10, Chapter 7]. Unfortunately, however, even assuming the Elliot-Halberstam conjecture this fails to show that bounded gaps between primes occur infinitely often. Goldston, Pintz, and Yıldırım [11] however, used a similar but more general choice of sieve weight, of the form

$$\lambda_d = \mu(d)F(\log R/d) \quad (6.7)$$

for a suitable, smooth function F to be optimised. GPY found that $F(x) = x^{k+l}$ was a suitable choice. The analysis of GPY showed that if we could take the level of distribution of the primes to be $\theta > 1/2$, then we could guarantee the existence of infinitely many bounded prime gaps. An annoyingly close result. Maynard and Tao's key contribution was the introduction of a yet again refined weight

$$w_n \propto \left(\sum_{\substack{d_1, \dots, d_k \\ d_i \mid n+a_i \ \forall i \\ \prod_{i=1}^k d_i \leq R}} \lambda_{d_1, \dots, d_k} \right)^2, \quad \lambda_{d_1, \dots, d_k} = \mu(d)F\left(\frac{\log d_1}{\log R}, \dots, \frac{\log d_k}{\log R}\right) \quad (6.8)$$

For a suitable smooth function $F : [0, \infty)^k \rightarrow \mathbb{R}$. This k -dimensional generalisation of the weights in (6.7) allows F to vary with each $n + h_i$ and forces each d_i to vary only with

$n + h_i$, making the weight more concentrated on n where many of $n + h_i$ are prime. The analysis of Maynard and Tao, using these weights, showed that there are infinitely many bounded gaps between primes, assuming only that the primes have level of distribution $\theta > 0$ [19]. Goldston and Yıldırım [12] considered similar weights to those in (6.8) but restricting to $d_i \leq R^{1/k}$, so the range of each d_i is restricted to account for the worst case scenario in which each of the d_i s are very close to $R^{1/k}$, to ensure that $\prod_{i=1}^k d_i \leq R$. This truncation on $\prod_{i=1}^k d_i$ by R is necessary to guarantee our main terms in our calculations are legitimately main terms, where our error terms will be bounded by the Bombieri-Vinogradov theorem. The Maynard-Tao weights improve on this, as the d_i s can be conscious of the size of one another. To ensure that $\prod_{i=1}^k d_i \leq R$ we now only need to require that each $d_j \leq R / \prod_{i \neq j} d_i$. In some sense, this increases the level of distribution of the sieve, allowing us to improve upon the work of GPY. This generalisation also allows one to exploit high dimensional phenomenon such as concentration of measure, which very roughly says a L^1 function depending on a large number of random variables takes values highly concentrated around it's center of mass. This relaxes the calculations for large k .

6.2 Adapted Maynard-Tao Weights

So, as the Maynard Tao weights identify n where many of the shifts are prime, we suspect that these should give good candidates for the residue classes (mod q) we choose. We adapt the weights in (6.8) to fit this particular situation. We follow closely [21, Section 4].

We choose an admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ where we take $h_i = p_{\pi(k)+i} P_w$ and w small enough such that $P_w = o(\log_2 x)$ (to ensure that each $n + h_i q$ is typically in the range of $\mathcal{R}_m$), for example a choice of $w = \log_4 x$ would suffice by the Prime Number Theorem. k is a constant that we take to be sufficiently large relative to ϵ, δ, C_U . In particular, w will be large compared with k . We take our probability measure to be

$$\mu_{m,q}(a) = \alpha_{m,q} \sum_{\substack{n \leq U/m \\ n \equiv a \pmod{q} \\ (n(mn-1), P_w)=1}} \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | n+h_i q \ \forall i}} \sum_{\substack{e_1, \dots, e_k \\ e_i | m(n+h_i q)-1 \ \forall i}} \lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \right)^2 \quad (6.9)$$

where $\alpha_{m,q}$ is a normalising constant to ensure $\sum_{a \pmod{q}} \mu_{m,q}(a) = 1$. We note the composition of $\lambda_{d_1, \dots, d_k, e_1, \dots, e_k} = \lambda_{d_1, \dots, d_k}^{(1)} \prod_{i=1}^k \lambda_{e_i}^{(2)}$ where $\lambda_{d_1, \dots, d_k}^{(1)}$ correspond to the weights in (6.8) which ensure that $\mu_{m,q}(a)$ can be large if there is some $n \equiv a \pmod{q}$ such that all of the $\{n + h_1 q, \dots, n + h_k q\}$ have no small prime factors. The $\lambda_{e_i}^{(2)}$ will correspond to a standard Selberg sieve and act to ensure that $\mu_{m,q}(a)$ is small unless $m(n + h_i q) - 1$ has no small prime factors less than y^ϵ .

We now give a heuristic argument for why we believe these weights may be successful in showing that the expected number of residue classes a prime $p \in \mathcal{R}_m$ lies in, $\sum_{q \in \mathcal{I}_m} \mu_{m,q}(p)$, may be taken arbitrarily large.

The application of the Maynard-Tao weights in the GPY method led to a proof of Theorem 5.1 and as discussed, did so by showing that if n in $[N, 2N]$ is chosen with probability proportional to the weights in (6.8) then $\mathbb{E}_{[N, 2N]} \left[\sum_{i=1}^k \chi_P(n + a_i) \right] \gg \log k$. This suggests that if we choose $\mu_{m,q}$ to mimic these weights (as we have in (6.9)) we may be able to guarantee that a multiple of $\log k$ elements of $\mathcal{R}_m$ lie in some residue class (mod q), where we can take k to be arbitrarily large as long as x is sufficiently large. Suppose that the primes covered are independent for each $q \in \mathcal{I}_m$, we may then expect that the primes p in $\mathcal{R}_m$ are covered by an approximately equal number of primes in $\mathcal{I}_m$. If that were the case, then we would expect each prime p to occupy approximately

$$\frac{|\mathcal{I}_m| \log k}{\log x |\mathcal{R}_m|} \quad (6.10)$$

many of the chosen residue classes as there are approximately $|\mathcal{I}_m|/\log x$ primes in $\mathcal{I}_m$ by our previous remark. Recalling that we took the length of $\mathcal{I}_m$ to be at least $\delta |\mathcal{R}_m| \log x$, then from (6.10) the expected number of residue classes we expect a prime p in $\mathcal{R}_m$ to lie in should be approximately some constant multiple of $\delta \log k$. So, if we can take k large enough, this should be larger than any constant, and our proof would be complete by the discussion in Section 6. We now aim to make this idea rigorous.

However, before we do this, we shall now give the proofs of our claimed sizes of $\mathcal{R}'$, $\mathcal{R}$ that we gave in Section 5.

7 Preliminary Sieve Estimates

We prove the claims on the sizes of $\mathcal{R}'$, $\mathcal{R}$ made in (5.7) and (5.8), following [18] and [21].

Lemma 7.1. *We have*

$$|\mathcal{R}'| \ll \frac{x}{(\log x)^{1+\epsilon}} \quad (7.1)$$

Proof. We give a proof similar to that presented in [18, Theorem 5.3]. First, consider $n \in \mathcal{R}'$ such that $n > x/(\log x)^{1+\epsilon}$ and n is divisible by some prime $p > y^{1/2}$. We hope to count the number of $n \in \mathcal{R}'$ that do not satisfy both of these conditions. For this, we need a count on smooth numbers. Following de Bruijn [4], we have the following count

$$\Psi(m, k) \sim m e^{-u \log u (1+o(1))} \quad (\log k > (\log m)^{2/3}) \quad (7.2)$$

where $\psi(m, k)$ denotes the number of k -smooth numbers less than or equal to m and $u := \log m / \log k$. So, the number of $n \in \mathcal{R}'$ that do not satisfy both of our conditions is

at most

$$\frac{x}{(\log x)^{1+\epsilon}} + \psi(U, y^{1/2}) = \frac{x}{(\log x)^{1+\epsilon}} + O\left(C_U \frac{x \log x \log_3 x}{\log_2 x} \exp\left(\frac{-2 \log_2 x (1 + o(1))}{(1 - \epsilon)}\right)\right) \quad (7.3)$$

$$= \frac{x}{(\log x)^{1+\epsilon}} + O\left(\frac{C_U x}{(\log x)^{1+2\epsilon}}\right) = \frac{x(1 + o(1))}{(\log x)^{1+\epsilon}} \quad (7.4)$$

So we consider only n of the form mp with $y \geq p > y^{1/2}$ and m satisfying

$$\frac{x}{y(\log x)^{1+\epsilon}} < m < \frac{U}{y^{1/2}}, \quad m \text{ is } y\text{-smooth} \quad (7.5)$$

For each m in this range, we now wish to estimate the number of $p \leq U/m$ with $(mp - 1, P_y) = 1$. To do this, one may appeal to [15, Theorem 4.2]. We instead opt to give an almost identical proof using Theorem 4.1. In this case, $\mathcal{A} = (mp - 1)_{p \leq U/m}$ and $g(p) = \mathbf{1}_{(m,p)=1}/(p - 1)$ for $p > 2$ and $g(2) = 1$ as m even. To bound the remainder, by taking $h = 3$ we have

$$\sum_{\substack{d \leq y^{1/5} \\ d \text{ square-free}}} 3^{\omega(d)} |r_d| \ll \frac{y^{1/2}}{(\log y)^3} \quad (7.6)$$

from Corollary 4.4.1. This will turn out to be negligible, noting $y^{1/2} \ll U/m$. So we may take $D = y^{1/5}$. We now bound L from below

$$L = \sum_{\substack{d \leq y^{1/10} \\ d \text{ square-free} \\ (m,d)=1}} \prod_{p|d} \frac{1/(p-1)}{1 - 1/(p-1)} \geq \sum_{\substack{d \leq y^{1/10} \\ d \text{ square-free} \\ (m,d)=1}} \frac{1}{d} \prod_{p|d} \sum_{k=0}^{\infty} \frac{1}{(p-1)^k} \quad (7.7)$$

$$\geq \sum_{\substack{d \leq y^{1/10} \\ (m,d)=1}} \frac{1}{d} \quad (7.8)$$

$$\gg \log y \prod_{p|m} \frac{p}{p+1} \quad (7.9)$$

where the last line follows from recalling that $\sum_{d \leq y^{1/10}} 1/d \gg \log y$ and that the sum of reciprocals of squares is convergent, so

$$\sum_{\substack{d \leq y^{1/10} \\ (m,d)=1}} \frac{1}{d} \prod_{p|m} \left(1 + \frac{1}{p}\right) \gg \log y \quad (7.10)$$

So Theorem 4.1 gives us that the count of such ps for a given m is

$$\ll \frac{U}{m \log(U/m) \log y} \prod_{p|m} \frac{p+1}{p} \ll \frac{U \log_2 x}{m(\log y)^2} \quad (7.11)$$

where we have used $\omega(m) \ll \log m$ and Mertens' third estimate, along with the bound $\log(U/m) \gg \log y$. Furthermore, we may use (7.2) to evaluate the number of m s that satisfy (7.5). Once we have this count, we may evaluate $\sum 1/m$ for m satisfying (7.5) using partial summation, which gives that this sum is

$$(\log x)^{-\epsilon(1-\epsilon)^{-1}+o(1)} \quad (7.12)$$

Hence

$$|\mathcal{R}'| \ll U \log_2 x (\log x)^{-\epsilon-\epsilon^2} (\log y)^{-2} + x (\log x)^{-1-\epsilon} \quad (7.13)$$

$$\ll x (\log x)^{-1-\epsilon} \quad (7.14)$$

□

We now consider the size of $\mathcal{R}$. To do this, we consider the size of each individual $\mathcal{R}_m$.

Lemma 7.2. *We have uniformly for $z + z/\log x \leq V \leq x(\log x)^2$ and $m \leq x$*

$$\begin{aligned} \#\{z < p \leq V \mid (mp-1, P_y) = 1\} = \\ \frac{V-z}{\log x} \left(\prod_{\substack{p \leq y \\ p \nmid m}} \frac{p-2}{p-1} \right) \{1 + O(\exp(-(\log_2 x)^{1/2}))\} \end{aligned}$$

and so in particular, for even $m \leq U(1 - 1/\log x)/z$ we have

$$|\mathcal{R}_m| \leq \frac{2e^{-\gamma}U(1+o(1))}{m(\log x)(\log y)} \left(\prod_{p>2} \frac{p(p-2)}{(p-1)^2} \right) \left(\prod_{p|m} \frac{p-1}{p-2} \right) \quad (7.15)$$

Remark. In [21] (7.15) is given with equality. However, we note that this would not be the case whenever $V = U/m = O(z)$, and instead the right hand side of (7.15) would be a strict bound. However, this does not affect the following calculations.

Proof of Lemma 7.2. We do not give the full details to avoid more somewhat unpleasant calculations. One arrives at the first estimate from using a fundamental lemma type sieve, such as in Theorem 4.2, along with the Bombieri-Vinogradov Theorem (particularly in the form of Corollary 4.4.1). The second estimate follows immediately from Mertens' third estimate and noting we can extend the Euler product product at the cost of a $o(1)$ term as it is absolutely convergent (since all terms in the product are of the form $1 + O(p^{-2})$).

□

Remark. Recall that the residue classes we picked for $p \leq y$ were $a_p = 1$ for each p . An alternative and perhaps more natural approach is to pick the a_p s 'greedily'; picking each one so that it contains the most integers not already in a occupied residue class. This

reduces the number of integers in $[1, U]$ not covered by the congruence classes chosen in (5.3) by at least a factor of $\log y$. This factor is recovered in the above lemma from the residue classes chosen in (5.2), so one can pick these classes for simplicity. However, we shall take this ‘greedy’ approach in Section 9.1.

Following Lemma 7.2 we now derive a bound for $|\mathcal{R}|$.

Lemma 7.3. *For $M \geq 2$ we have*

$$\sum_{U/zM \leq m < U/z} |\mathcal{R}_m| \ll \frac{U \log M}{\log x \log y} \quad (7.16)$$

and in particular

$$\sum_{U/z(\log_2 x)^2 \leq m < U/z} |\mathcal{R}_m| = o\left(\frac{C_U x}{\log x}\right), \quad \sum_{1 \leq m \leq U/z(\log_2 x)^2} |\mathcal{R}_m| = O\left(\frac{C_U x}{\log x}\right) \quad (7.17)$$

Proof. Define $w_1 = U/Mz$ and $w_2 = U(1 - 1/\log x)/z$. For $w_2 \leq m \leq U/z$ we cannot use Lemma 7.2 but we note the trivial bound $|\mathcal{R}_m| \ll z/\log x \ll U/(m \log x)$. So for $w_2 \leq m < U/z$ we gain a contribution of at most $O(U/(\log x)^2)$. We now consider $w_1 \leq m \leq w_2$. We have

$$\prod_{p|m, p>2} \frac{p-1}{p-2} \ll \prod_{p|m} \frac{p+1}{p} \leq \sum_{d|m} \frac{1}{d} \quad (7.18)$$

Remark. The final bound in (7.18) is given instead with equality in [21]. This however would only be true in the case where m is square-free, which need not be the case, so we instead give this with inequality.

With the first bound coming from noting $(n+1)/n$ decreasing and the second from simply expanding the product. This leads us to the bound

$$\sum_{w_1 \leq m < w_2} |\mathcal{R}_m| \ll \frac{U}{\log x \log y} \sum_{w_1 \leq m < w_2} \frac{1}{m} \sum_{d|m} \frac{1}{d} \quad (7.19)$$

and making the substitution $m = dm'$ we have

$$\sum_{w_1 \leq m < w_2} |\mathcal{R}_m| \ll \frac{U}{\log x \log y} \sum_{d \leq w_2} \frac{1}{d^2} \sum_{w_1/d \leq m' \leq w_2/d} \frac{1}{m'} \ll \frac{U(\log M + O(1))}{\log x \log y} \quad (7.20)$$

□

Now we have shown the sizes of $\mathcal{R}$ and $\mathcal{R}'$, we can proceed to proving Proposition 5.2.

8 Sieve Estimates

In this section, we give the proof of Proposition 5.2 as given by Maynard in [21]. We now define some notation that we will use throughout and give the precise form of the weights in (6.9).

8.1 Setup

Let

$$\omega_{m,q}(p) := |\{1 \leq n \leq p : n \text{ satisfies one of } n + h_i q \equiv 0 \pmod{p} \text{ or} \quad (8.1)$$

$$m(n + h_i q) \equiv 1 \pmod{p} \text{ for some } 1 \leq i \leq k\}| \quad (8.2)$$

and we extend $\omega_{m,q}$ to a completely multiplicative function over $\mathbb{N}$. We also define $\phi_{m,q}(p) = p - \omega_{m,q}(p)$ and similarly extend this to a completely multiplicative function. We define further

$$\mathfrak{S}_{m,q} = \prod_{p \leq y} \left(1 - \frac{\omega_{m,q}(p)}{p}\right) \left(1 - \frac{1}{p}\right)^{-2k} \quad (8.3)$$

note this product is non-zero as we've that $\omega_{m,q}(p) \leq 2$ for $p \leq w$ and $\omega_{m,q}(2) = 1$ as m is even.

We define $\lambda_{d_1, \dots, d_k, e_1, \dots, e_k}$ by

$$\lambda_{d_1, \dots, d_k, e_1, \dots, e_k} = \prod_{i=1}^k \mu(d_i) \mu(e_i) \sum_{j=1}^J \prod_{l=1}^k F_{l,j} \left(\frac{\log d_l}{\log x} \right) G \left(\frac{\log e_l}{\log y} \right) \quad (8.4)$$

for smooth non-negative functions $F_{l,j}, G : [0, \infty) \rightarrow \mathbb{R}$ not identically zero. The functions $F_{l,j}, G$ and the quantity J will be allowed to depend on k but independent of x, q . In particular this gives that $|\lambda_{d_1, \dots, d_k, e_1, \dots, e_k}| \ll_k 1$ which we exploit later. We also require for $j \in \{1, \dots, J\}$, we have

$$\sup \left\{ \sum_{i=1}^k u_i : F_{l,j}(u_i) \neq 0 \right\} \leq 1/10 \quad (8.5)$$

So in particular if we have that if $\lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \neq 0$ then $\prod_{i=1}^k d_i \leq x^{1/10}$. We also restrict G to be supported on $[0, 1]$. Finally we put

$$F(t_1, \dots, t_k) = \sum_{j=1}^J \prod_{l=1}^k F_{l,j}(t_l) \quad (8.6)$$

and we assume the $F_{l,j}$ are chosen such that F is symmetric.

8.2 Calculations

Our first goal is to asymptotically evaluate the normalising constant.

Lemma 8.1. *We have*

$$\alpha_{m,q}^{-1} = (1 + o_k(1)) \frac{U \mathfrak{S}_{m,q}}{m(\log x)^k (\log y)^k} I_k^{(1)}(F) I_k^{(2)}(G) \quad (8.7)$$

where $\mathfrak{S}_{m,q}$ given in (8.3) and

$$I_k^{(1)}(F) = \int \dots \int_{t_1, \dots, t_k \geq 0} F(t_1, \dots, t_k)^2 dt_1 \dots dt_k, \quad I_k^{(2)}(G) = \left(\int_{t=0}^{\infty} G'(t) dt \right)^k \quad (8.8)$$

Proof. As $\sum_{a \pmod{q}} \mu_{m,q}(a) = 1$, we are left to evaluate

$$\alpha_{m,q}^{-1} = \sum_{\substack{n \leq U/m \\ (n(mn-1), P_w)=1}} \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | n+h_i q \ \forall i}} \sum_{\substack{e_1, \dots, e_k \\ e_i | m(n+h_i q)-1 \ \forall i}} \lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \right)^2 \quad (8.9)$$

by expanding the square and reordering the sum this becomes

$$\sum_{\substack{d_1, \dots, d_k \\ d'_1, \dots, d'_k}} \sum_{\substack{e'_1, \dots, e'_k \\ e_1, \dots, e_k}} \lambda_{\mathbf{d}, \mathbf{e}} \lambda_{\mathbf{d}', \mathbf{e}'} \sum_{\substack{n \leq U/m \\ (n(mn-1), P_w)=1 \\ [d_i, d'_i] | n+h_i q \ \forall i \\ [e_i, e'_i] | m(n+h_i q)-1 \ \forall i}} 1 \quad (8.10)$$

Where we write $\lambda_{\mathbf{d}, \mathbf{e}} = \lambda_{d_1, \dots, d_k, e_1, \dots, e_k}$ and similarly for $\lambda_{\mathbf{d}', \mathbf{e}'}$ to ease notation. We wish to combine this into a sum over a single congruence class. We note that the inner sum is empty unless $d_1 d'_1, \dots, d_k d'_k, e_1 e'_1, \dots, e_k e'_k$ are coprime to P_w (if say $p | (d_i d'_i, P_w)$ then $p | n+h_i q$ and hence $p | n$ contradicting $(n, P_w) = 1$). We note also the inner sum empty unless the $d_1 d'_1, \dots, d_k d'_k$ pairwise coprime and the $e_1 e'_1, \dots, e_k e'_k$ pairwise coprime (say $p | (e_i e'_i, e_j e'_j)$ then $p | [e_i, e'_i] - [e_j, e'_j]$ and so $p | mq(h_i - h_j)$. Clearly $p \nmid m$ as if it did then $p | -1$ which clearly cannot happen. So we would have $p | q(h_i - h_j)$. So either $p | q$, however this contradicts q being prime as $p \leq y < q$ (recall that $y = x^{o(1)}$ and $q \in [x/2, x]$), or we have $p | (p_{\pi(k)+i} - p_{\pi(k)+j}) P_w$ contradicting $(e_i e'_i, P_w) = 1$. Finally we note if $p | (d_i d'_i, e_j e'_j)$ then $p | mq(h_j - h_i) - 1$ for all $1 \leq i, j \leq k$. Let us first list these conditions for ease of reference

A.1 $d_1 d'_1, \dots, d_k d'_k, e_1 e'_1, \dots, e_k e'_k$ are coprime to P_w

A.2 $d_1 d'_1, \dots, d_k d'_k$ pairwise coprime

A.3 $e_1 e'_1, \dots, e_k e'_k$ pairwise coprime

A.4 $p | (d_i d'_i, e_j e'_j)$ then $p | mq(h_j - h_i) - 1$ for all $1 \leq i, j \leq k$.

Remark. Our condition **A.4** is slightly different from that in [21] where this is instead given in the form $(d_i d'_i, e_j e'_j) | mq(h_j - h_i) - 1$ for all $1 \leq i, j \leq k$. This is a slight inaccuracy as this would force $p^2 | mq(h_j - h_i) - 1$ in the case when $p | d_i, d'_i, e_j, e'_j$ which need not be the case. However, the condition is treated as in **A.4** throughout [21] and therefore does not

affect the calculations.

If all of these conditions are satisfied then the set of congruences are consistent and we can combine the divisibility conditions using the chinese remainder theorem so that the sum runs over a individual residue class $(\text{mod } [d_1, d'_1, e_1, e'_1, \dots, d_k, d'_k, e_k, e'_k])$ and the $\phi_{m,q}(P_w)$ possible residue classes $(\text{mod } P_w)$ hence the sum runs over the $\phi_{m,q}(P_w)$ residue classes $(\text{mod } P_w[d_1, d'_1, e_1, e'_1, \dots, d_k, d'_k, e_k, e'_k])$. So (8.10) becomes

$$\sum'_{\substack{d_1, \dots, d_k \\ d'_1, \dots, d'_k}} \sum'_{\substack{e_1, \dots, e_k \\ e'_1, \dots, e'_k}} \lambda_{\mathbf{d}, \mathbf{e}} \lambda_{\mathbf{d}', \mathbf{e}'} \left(\frac{U \phi_{m,q}(P_w)}{m P_w[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']} + O(\phi_{m,q}(P_w)) \right) \quad (8.11)$$

where we have written $[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}'] = [d_1, d'_1, e_1, e'_1, \dots, d_k, d'_k, e_k, e'_k]$ and $\sum'$ for the conditions **A.1**, **A.2**, **A.3** and a slight reformulation of **A.4**, being $p|(d_i d'_i, e_j e'_j)$ then $p|(mq(h_j - h_i) - 1, P_y)$ for all $1 \leq i, j \leq k$. Note this is equivalent to **A.4** due to the support condition on G forces $e_j, e'_j \leq y$ for all $1 \leq j \leq k$. This will become useful later.

We first evaluate the error term. Writing $d = \prod_{i=1}^k d_i$ (and similarly d', e, e'), as noted earlier we must have $d, d' \leq x^{1/10}$ and $e, e' \leq y^k \leq x^\epsilon$. As $|\lambda_{d_1, \dots, d_k, e_1, \dots, e_k}| \ll_k 1$, We have

$$\phi_{m,q}(P_w) \sum'_{\substack{d_1, \dots, d_k \\ d'_1, \dots, d'_k}} \sum'_{\substack{e_1, \dots, e_k \\ e'_1, \dots, e'_k}} \lambda_{\mathbf{d}, \mathbf{e}} \lambda_{\mathbf{d}', \mathbf{e}'} \ll_k \log_2 x \sum_{d, d' \leq x^{1/10}} \tau_k(d) \tau_k(d') \sum_{e, e' \leq y^k} \tau_k(e) \tau_k(e') \quad (8.12)$$

$$\ll_{k, \epsilon} x^\epsilon x^{1/10} x^{1/10} x^\epsilon x^\epsilon \ll_k x^{1/2} \quad (8.13)$$

where we have used that for any integer $k \geq 2$ we have

$$\sum_{d \leq x} \tau_k(d) \sim \frac{x}{(k-1)!} (\log x)^{k-1} \quad (8.14)$$

See [23, Chapter 2]. So we can gather all our log terms simply into a x^ϵ term and we also use that we take w to be large relative to k to guarantee $y^k = x^{o_k(1)}$.

We Now focus our analysis on the main term.

Remark. Here, one may naturally be motivated by the analysis in [19], hoping to decouple and then use [19, Lemma 6.1]. However, our situation here is more complex. The contribution provided from the d, d' and e, e' terms that are coupled together by condition **A.4** is of the same order as the sum over the uncoupled terms. So we cannot evaluate this sum by simply evaluating over each individual summand independently, picking up only a negligible error. So we must take a different approach. One could, more classically, evaluate the sum using Peron's formula. However we opt to use Fourier analysis. We note however that these approaches are practically identical (noting the form of (8.17) for

example) and both approaches would use the same input (that being the behavior of zeta around it's simple pole at $s = 1$, which we will see as our calculations unfold).

Writing $\lambda_{\mathbf{d}, \mathbf{e}}, \lambda_{\mathbf{d}', \mathbf{e}'}$ in the form given in (8.4), the main term in (8.11) becomes

$$\frac{U\phi_{m,q}(P_w)}{mP_w} \sum_{j=1}^J \sum_{j'=1}^J \sum'_{d_1, \dots, d_k} \sum'_{\substack{e_1, \dots, e_k \\ d'_1, \dots, d'_k, e'_1, \dots, e'_k}} \frac{\prod_{l=1}^k \mu(d_l) \mu(d'_l) \mu(e_l) \mu(e'_l) H_{l,j,j'}(d_l, d'_l, e_l, e'_l)}{[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']} \quad (8.15)$$

where

$$H_{l,j,j'}(d_l, d'_l, e_l, e'_l) = F_{l,j}\left(\frac{\log d_l}{\log x}\right) G\left(\frac{\log e_l}{\log y}\right) F_{l,j'}\left(\frac{\log d'_l}{\log x}\right) G\left(\frac{\log e'_l}{\log y}\right) \quad (8.16)$$

The function $e^t F_{l,j}(t)$ can be extended to a smooth, compactly supported function on $\mathbb{R}$ and by the Fourier inversion formula has a Fourier expansion $\int_{\mathbb{R}} e^{-it\xi} f_{l,j}(\xi) d\xi$, where $f_{l,j}(\xi) = \int_{\mathbb{R}} e^{it\xi} F_{l,j}(t) dt$ the Fourier transform. From the smoothness and compact support of $e^t F_{l,j}(t)$ we can repeat integration by parts on the Fourier transform to get the bound $f_{l,j}(\xi) \ll_{k,A} (1 + |\xi|)^{-A}$ for any $A > 0$, hence $f_{l,j}(\xi)$ rapidly decreasing. We also have identical considerations for the Fourier transform g of $e^t G$. We gain the expressions

$$F_{l,j}\left(\frac{\log d_l}{\log x}\right) = \int_{\mathbb{R}} \frac{f_{l,j}(\xi_l)}{d_l^{\frac{1+i\xi_l}{\log x}}} d\xi_l, \quad G\left(\frac{\log e_l}{\log y}\right) = \int_{\mathbb{R}} \frac{g(\tau_l)}{e_l^{\frac{1+i\tau_l}{\log y}}} d\tau_l \quad (8.17)$$

so our sum over the d, d', e, e' variables in (8.15) becomes

$$\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left(\sum'_{d_1, \dots, d_k} \sum'_{\substack{e_1, \dots, e_k \\ d'_1, \dots, d'_k, e'_1, \dots, e'_k}} \frac{1}{[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']} \prod_{l=1}^k \frac{\mu(d_l) \mu(d'_l) \mu(e_l) \mu(e'_l)}{d_l^{\frac{1+i\xi_l}{\log x}} d'_l^{\frac{1+i\xi'_l}{\log x}} e_l^{\frac{1+i\tau_l}{\log y}} e'_l^{\frac{1+i\tau'_l}{\log y}}} \right) \quad (8.18)$$

$$\times \left(\prod_{l=1}^k f_{l,j}(\xi_l) f_{l,j'}(\xi'_l) g(\tau_l) g(\tau'_l) d\xi_l d\xi'_l d\tau_l d\tau'_l \right) \quad (8.19)$$

We are able to swap the order of summation and integration as the expression is absolutely convergent. We now focus our analysis to the first term in brackets in the integral. We note that as the summand and the restrictions on the sum **A.1-A.4** are multiplicative, we can rewrite the sum as a product of the form $\prod_p K_p$ where each K_p a function $K_p(\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_k, \tau_1, \dots, \tau_k, \tau'_1, \dots, \tau'_k)$. Our first observation is

$$K_p = \sum'_{\substack{d_1, \dots, d_k \\ d'_1, \dots, d'_k}} \sum'_{\substack{e_1, \dots, e_k \\ e'_1, \dots, e'_k}} \frac{1}{[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']} \prod_{l=1}^k \frac{\mu(d_l) \mu(d'_l) \mu(e_l) \mu(e'_l)}{d_l^{\frac{1+i\xi_l}{\log x}} d'_l^{\frac{1+i\xi'_l}{\log x}} e_l^{\frac{1+i\tau_l}{\log y}} e'_l^{\frac{1+i\tau'_l}{\log y}}} \quad (8.20)$$

$$= 1 + O_k(p^{-1-1/\log x}) \quad (8.21)$$

Which Maynard notes in [21] is $\ll (\log x)^{O_k(1)}$. To see this, we have

$$\prod_p K_p = 1 + O_k(p^{-1-1/\log x}) \quad (8.22)$$

$$\ll \prod_p \left(\frac{1}{1 - p^{-1-1/\log x}} \right)^{O_k(1)} = \zeta(1 + 1/\log x)^{O_k(1)} \ll_k (\log x)^{O_k(1)} \quad (8.23)$$

where we have used that for $0 < |s - 1| \leq 1$ we have

$$\zeta(s) = 1/(s - 1) + \gamma + O(|s - 1|) \quad (8.24)$$

where γ the Euler-Mascheroni constant. See [16, Chapter 5].

We note that from the fact that the $f_{i,j}(t) \ll (1 + |t|)^{-A}$ for all $A > 0$ and that we have identical bounds for g that we can restrict the integral in (8.18) to $|\xi_l|, |\xi'_l|, |\tau'_l|, |\tau_l| \leq \sqrt{\log x}$ as the contribution outside of this range gives for any $A > 0$ is

$$\ll_A (\log x)^{O_k(1)} \int \dots \int_{|\xi_l|, |\xi'_l|, |\tau'_l|, |\tau_l| > \sqrt{\log x} \ \forall l} (1 + |\xi_l|)^{-Ak} (1 + |\xi'_l|)^{-Ak} (1 + |\tau'_l|)^{-Ak} (1 + |\tau_l|)^{-Ak} \quad (8.25)$$

$$\ll_k (\log x)^{-2k} \quad (8.26)$$

by choosing sufficiently large A . To ease notation we let $s_l = (1 + i\xi_l)/\log x$, $r_l = (1 + i\tau_l)/\log y$ and similarly for s'_l, r'_l . We want to understand the structure of K_p . Due to its multiplicative properties we would like to write it in a form that exploits this. We first consider a prime $w < p \leq y$ such that $p \nmid \prod_{h, h' \in \mathcal{H}} m q(h - h') - 1$ or $p > y$. Note that for such a p our slight reformulation of condition **A.4** tells us that $(d_i d'_i, e_j e'_j) = 1$ for all $1 \leq i, j \leq k$ with d_i, d'_i, e_i, e'_i in the range of summation of K_p . So we deduce that

$$\begin{aligned} K_p &= 1 + \sum_{l=1}^k \frac{-1}{p^{1+s_l}} + \frac{-1}{p^{1+s'_l}} + \frac{1}{p^{1+s_l+s'_l}} + \frac{-1}{p^{1+e_l}} + \frac{-1}{p^{1+e'_l}} + \frac{1}{p^{1+e_l+e'_l}} \\ &= \prod_{l=1}^k \frac{(1 - p^{-1-s_l})(1 - p^{-1-s'_l})(1 - p^{-1-e_l})(1 - p^{-1-e'_l})}{(1 - p^{-1-s_l-s'_l})(1 - p^{-1-e_l-e'_l})} + O_k\left(\frac{1}{p^2}\right) \\ &= (1 + O_k(p^{-2})) \prod_{l=1}^k \frac{(1 - p^{-1-s_l})(1 - p^{-1-s'_l})(1 - p^{-1-e_l})(1 - p^{-1-e'_l})}{(1 - p^{-1-s_l-s'_l})(1 - p^{-1-e_l-e'_l})} \end{aligned} \quad (8.27)$$

From expanding out the brackets, and writing

$$\frac{1}{1 - p^{-1-s_l-s'_l}} = 1 + p^{-1-s_l-s'_l} + O(p^{-2}) \quad (8.28)$$

First, we note that (8.27) has terms similar to those in the Euler product of the zeta function. We will exploit this later.

If instead $w < p \leq y$ with $p \mid \prod_{h, h' \in \mathcal{H}} mq(h - h') - 1$ then we will have terms containing products of d_i, e_j if $p \mid mq(h_j - h_i) - 1$, and so we will have an extra term, compared with (8.27) of the form

$$(1 + O_k(p^{-2})) \prod_{l, j: p \mid mq(h_j - h_i) - 1} \left(1 + \sum_{\substack{\mathcal{T} \subseteq \{s_l, s'_l, e_j, e'_j\} \\ \mathcal{T} \cap \{s_l, s'_l\} \neq \emptyset \\ \mathcal{T} \cap \{e_j, e'_j\} \neq \emptyset}} (-1)^{\#\mathcal{T}} p^{-1 - \sum_{t \in \mathcal{T}} t} \right) \quad (8.29)$$

Recall, we have restricted $|s_i|, |e_j| < \sqrt{\log x} / \log y \leq 1 + o(1)$. We note that as, for example, $p^{-s_i} = 1 + O(\log p \sqrt{\log x} / \log y)$, then (8.29) becomes

$$(1 + O_k(p^{-2})) \prod_{l, j: p \mid mq(h_j - h_i) - 1} \left(1 + \frac{1}{p} + O\left(\frac{\log p \sqrt{\log x}}{p \log y}\right) \right) \quad (8.30)$$

$$= (1 + O_k(p^{-2})) \left(1 + \frac{1}{p} + O\left(\frac{\log p \sqrt{\log x}}{p \log y}\right) \right)^{\#\{l, j: p \mid mq(h_j - h_i) - 1\}} \quad (8.31)$$

$$= (1 + O_k(p^{-2})) \left(1 + \frac{\#\{l, j: p \mid mq(h_j - h_i) - 1\}}{p} + O_k\left(\frac{\log p \sqrt{\log x}}{p \log y}\right) \right) \quad (8.32)$$

$$= \left(1 + \frac{\#\{l, j: p \mid mq(h_j - h_i) - 1\}}{p} \right) \left(1 + O_k\left(\frac{1}{p^2} + \frac{\log p \sqrt{\log x}}{p \log y}\right) \right) \quad (8.33)$$

For p satisfying $w < p \leq y$ with $p \mid \prod_{h, h' \in \mathcal{H}} mq(h - h') - 1$ we note that all the $qh_i \pmod{p}$ are distinct. To see this, note that as $p > w$ we must have $p \nmid \prod_{h, h'} (h' - h)$ and as $p \leq y < q$, $p \nmid q$ as q is prime. So we see that

$$\omega_{m, q}(p) = 2k - \#\{l, j: p \mid mq(h_j - h_i) - 1\} \quad (8.34)$$

as clearly $n + h_i q, m(n + h_i q) - 1$ are distinct $\pmod{p}$ unless $p \mid mq(h_l - h_i) - 1$. So the first product in (8.33) becomes $1 - (\omega_{m, q}(p) - 2k)/p$. We also note that $\prod_{h, h' \in \mathcal{H}} mq(h - h') - 1 \ll x^{O_k(1)}$ (as clearly $m, q, (h' - h) \ll x$) and $\log y > (\log x)^{1-\epsilon}$. So now, in order to evaluate $\prod_p K_p$, we are finally left to understand

$$\prod_{\substack{p \mid \prod_{h, h'} (mq(h - h') - 1) \\ p > w}} \left(1 + O_K\left(\frac{1}{p^2} + \frac{\log p \sqrt{\log x}}{p \log y}\right) \right) \quad (8.35)$$

but we note that $1/n^2 + (\log n \sqrt{\log x} / n \log y)$ decreasing. So the largest contribution is from when p is as small as possible, i.e if $\prod_{h, h'} (mq(h - h') - 1)$ a primorial. In this case

then, we take the product over $p \ll_k \log x$. We can therefore bound (8.35) by

$$\exp \left(\sum_{w < p \ll_k \log x} \log \left(1 + O_k \left(\frac{1}{p^2} + \frac{\log p \sqrt{\log x}}{p \log y} \right) \right) \right) \quad (8.36)$$

$$= \exp \left(\sum_{w < p \ll_k \log x} O_k \left(\frac{1}{p^2} + \frac{\log p \sqrt{\log x}}{p \log y} \right) \right) \quad (8.37)$$

$$= \exp \left(O_k \left(w^{-1} + \frac{\log \log x}{(\log x)^{1/2-\epsilon}} \right) \right) \quad (8.38)$$

$$= 1 + o_k(1) \quad (8.39)$$

So we see that the contribution of (8.35) in $\prod K_p$ is negligible.

$K_p = 1$ whenever $p \leq w$. However, we noted earlier that the individual expressions for K_p with $p > w$ are reminiscent of the individual terms for the Euler product for zeta, so if we could take our product over all primes, we could write K_p in terms of zeta and some correction factor for our terms $p \leq w$. Namely,

$$\prod_{p \leq w} \prod_{l=1}^k \frac{(1 - p^{-1-s_l})(1 - p^{-1-s'_l})(1 - p^{-1-r_l})(1 - p^{-1-r'_l})}{(1 - p^{-1-s_l-s'_l})(1 - p^{-1-r_l-r'_l})} \quad (8.40)$$

We note that $\log w = o((\log x)^\epsilon)$, so in particular $\log p = o((\log x)^\epsilon)$ in this range and $s_l, s'_l, r_l, r'_l = o((\log x)^{-1/2+\epsilon})$ and as $p^{-s_l} = 1 + o((\log x)^{-1/2+2\epsilon})$, then $(1 - p^{-1-s_l}) = (1 - p^{-1})(1 + o((\log x)^{-1/2+2\epsilon}))$ and similarly for $p^{-s'_l}, p^{-r_l}, p^{-r'_l}, p^{-s_l-s'_l}, p^{-r_l-r'_l}$, so (8.40) becomes

$$\prod_{p \leq w} \prod_{l=1}^k (1 - p^{-1})^2 (1 + o(1)) = (1 + o_k(1)) \prod_{p \leq w} \left(1 - \frac{1}{p} \right)^{2k} \quad (8.41)$$

Combining all we have shown, we arrive at

$$\prod_{p > w} K_p = (1 + o_k(1)) \prod_{p \leq w} \left(1 + \frac{1}{p} \right)^{-2k} \prod_{w < p \leq y} \left(1 + \frac{2k - \omega_{m,q}(p)}{p} \right) \quad (8.42)$$

$$\times \prod_{l=1}^k \frac{\zeta \left(1 + \frac{2+i\xi_l+i\xi'_l}{\log x} \right) \zeta \left(1 + \frac{2+i\tau_l+i\tau'_l}{\log y} \right)}{\zeta \left(1 + \frac{1+i\xi_l}{\log x} \right) \zeta \left(1 + \frac{1+i\xi'_l}{\log x} \right) \zeta \left(1 + \frac{1+i\tau_l}{\log y} \right) \zeta \left(1 + \frac{1+i\tau'_l}{\log y} \right)} \quad (8.43)$$

noting simply that $\prod_{p > w} 1 + O_k(p^{-2}) = 1 + o_k(1)$. Now we can now exploit the well enough understood behavior of zeta close to it's simple pole at $s = 1$. Recalling (8.24) in particular, we know in the region where $|z| = o(1)$ we have the estimate $\zeta(1+z) = (1 + o(1))/z$. By

further recalling that $\log y \geq (\log x)^{1-\epsilon}$ we are left to evaluate

$$\int \dots \int (1 + o_k(1)) \prod_{l=1}^k \left(\frac{(1 + i\xi_l)(1 + i\xi'_l)(1 + i\tau_l)(1 + i\tau'_l)}{(2 + i\xi_l + i\xi'_l)(2 + i\tau_l + i\tau'_l)} \right) \quad (8.44)$$

$$\times f_{l,j}(\xi_l) f_{l,j'}(\xi'_l) g(\tau_l) g(\tau'_l) d\xi_l d\xi'_l d\tau_l d\tau'_l \quad (8.45)$$

with the integral over $|\xi_l|, |\xi'_l|, |\tau_l|, |\tau'_l| \leq \sqrt{\log x}$. All that is left is to show that for any f_1, f_2 from the $f_{l,j}, g$ we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(1 + i\xi)(1 + i\xi')}{2 + i\xi + i\xi'} f_1(\xi) f_2(\xi') d\xi d\xi' = \int_{t=0}^{\infty} F'_1(t) F'_2(t) dt \quad (8.46)$$

to see this, we recall that we had the expression

$$F(t) = \int_{\mathbb{R}} e^{-t(1+i\xi)} f(\xi) d\xi \quad (8.47)$$

Noting that we can differentiate under the integral sign due to the absolute convergence of the integral, we have

$$F'(t) = \int_{\mathbb{R}} \frac{-(1 + i\xi)f(\xi)}{e^{t(1+i\xi)}} d\xi \quad (8.48)$$

and then (8.46) follows from rearranging the integrals guaranteed by absolute convergence and Fubini's theorem.

Altogether, we get

$$\alpha_{m,q}^{-1} = \frac{(1 + o_k(1)) U \phi_{m,q}(P_w)}{m(\log x)^k (\log y)^k P_w} \prod_{p \leq w} \left(1 - \frac{1}{p}\right)^{-2k} \prod_{w < p \leq y} \left(1 - \frac{\omega_{m,q}(p) - 2k}{p}\right) \quad (8.49)$$

$$\times \left(\int_{t=0}^{\infty} G'(t)^2 \right)^k \sum_{j=1}^J \sum_{j'=1}^{J'} \prod_{l=1}^k \left(\int_{t=0}^{\infty} F'_{i,j}(t) F'_{i,j'}(t) dt \right) \quad (8.50)$$

$$= (1 + o_k(1)) \frac{U \mathfrak{S}_{m,q}}{m(\log x)^k (\log y)^k} I_k^{(1)}(F) I_k^{(2)}(G) \quad (8.51)$$

where we write

$$\left(1 - \frac{\omega_{m,q}(p) - 2k}{p}\right) = (1 + O_k(p^{-2})) \left(1 - \frac{\omega_{m,q}(p)}{p}\right) \left(1 - \frac{1}{p}\right)^{-2k} \quad (8.52)$$

and have used the fact that G and the $F_{l,j}$ are all non-negative (and not the zero function) to take the $o(1)$ errors as a factor at the front of the expression

□

Lemma 8.2. *Let $m < Uz^{-1}(\log_2 x)^{-2}$ be even and $p_0 \in \mathcal{R}_m$ with $h_k x < p_0 < U/m - h_k x$.*

Then

$$\sum_{q \in \mathcal{I}_m} \mu_{m,q}(p_0) \gg (1 + o_k(1)) \frac{k |\mathcal{I}_m| J_k^{(1)}(F) J_k^{(2)}(G)}{(\log x) |\mathcal{R}_m| I_k^{(1)}(F) I_k^{(2)}(G)} \quad (8.53)$$

where

$$J_k^{(1)}(F) = \int \dots \int_{t_1, \dots, t_{k-1} \geq 0} \left(\int_{t_k \geq 0} F(t_1, \dots, t_k) dt_k \right)^2 dt_1 \dots dt_k \quad (8.54)$$

$$J_k^{(2)}(G) = G(0)^2 \left(\int_{t=0}^{\infty} G'(t)^2 dt \right)^{k-1}. \quad (8.55)$$

Proof. Substituting in our definition for our probability measure $\mu_{m,q}$ we have

$$\sum_{q \in \mathcal{I}_m \text{ prime}} \mu_{m,q}(p_0) \quad (8.56)$$

$$= \sum_{q \in \mathcal{I}_m \text{ prime}} \alpha_{m,q} \sum_{\substack{n \leq U/m \\ n \equiv p_0 \pmod{q} \\ (n(mn-1), P_w)=1}} \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | n + h_i q}} \sum_{\substack{e_1, \dots, e_k \\ e_i | m(n + h_i q) - 1}} \lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \right)^2 \quad (8.57)$$

we note that as the inner terms are squared they are positive and so we can drop terms to get a lower bound, in particular we will only sum over n of the form $p_0 - hq$ where $h \in \mathcal{H}$. We note that for each $1 \leq i \leq k$ we have $(p_0 + (h_i - h)q, P_w) = (p_0, P_w) = 1$ and $(m(p_0 + (h_i - h)q) - 1, P_w) = (mp_0 - 1, P_w) = 1$ as $p_0 \in \mathcal{R}_m$ so all the terms $n = p_0 - hq$ appear in the sum (note $0 < n = p_0 - hq < U$ by assumption). So we get the bound

$$\sum_{q \in \mathcal{I}_m \text{ prime}} \mu_{m,q}(p_0) \quad (8.58)$$

$$\geq \sum_{h \in \mathcal{H}} \sum_{q \in \mathcal{I}_m \text{ prime}} \alpha_{m,q} \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | p_0 + (h_i - h)q}} \sum_{\substack{e_1, \dots, e_k \\ e_i | m(p_0 + (h_i - h)q) - 1}} \lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \right)^2 \quad (8.59)$$

We further split the sum over q into individual congruence classes mod P_w . This gives

$$\sum_{h \in \mathcal{H}} \sum_{\substack{w_0 \pmod{P_w} \\ (w_0, P_w)=1}} \sum_{\substack{q \in \mathcal{I}_m \text{ prime} \\ q \equiv w_0 \pmod{P_w}}} \alpha_{m,q} \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | p_0 + (h_i - h)q}} \sum_{\substack{e_1, \dots, e_k \\ e_i | m(p_0 + (h_i - h)q) - 1}} \lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \right)^2 \quad (8.60)$$

We can now use lemma 8.1 to write the normalising constant, however we hope to give this in a form somewhat less dependent on q . The dependence of $\alpha_{m,q}$ on q comes from the term $\mathfrak{S}_{m,q}$ which we hope to rewrite. First we note that for $p \leq w$ we have $\omega_{m,q}(p) = 1$ if $p|m$ and $\omega_{m,q}(p) = 2$ otherwise. We note further that for $w < p \leq y$ we have $\omega_{m,q}(p) = 2k$ when

$$p \nmid \prod_{h', h'' \in \mathcal{H}} (mq(h' - h'') - 1) \quad (8.61)$$

Recalling (8.3) and that m even then

$$\mathfrak{S}_{m,q}^{-1} = \prod_{p \leq y} \left(1 - \frac{\omega_{m,q}(p)}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^{2k} \quad (8.62)$$

$$\geq \mathfrak{S}_m \prod_{\substack{w < p \leq y \\ p \mid \prod_{h', h''} (mq(h' - h'') - 1)}} \left(1 - \frac{2k}{p}\right) \quad (8.63)$$

where we take

$$\mathfrak{S}_m = 2^{-(2k-1)} \left(\prod_{p \mid m, p > 2} \frac{p-2}{p-1} \right) \prod_{2 < p \leq w} \left(1 - \frac{2}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^{2k} \quad (8.64)$$

In [21] 8.63 has a factor of $(1 + o_k(1))$, however we note that this bound is, in fact, strict; making this factor ineffectual. To see this bound, note that here we have split the product into $p \leq w$ and $w < p \leq y$. In the latter case, we get the bound

$$\prod_{w < p \leq y} \left(1 - \frac{\omega_{m,q}(p)}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^{2k} \quad (8.65)$$

$$\geq \prod_{w < p \leq y} \left(1 - \frac{\omega_{m,q}(p)}{p}\right)^{-1} \left(1 - \frac{2k}{p}\right) \geq \prod_{\substack{w < p \leq y \\ p \mid \prod_{h', h''} (mq(h' - h'') - 1)}} \left(1 - \frac{2k}{p}\right) \quad (8.66)$$

We can truncate the product on the right hand side of (8.63) to $p \leq z_0 := \log x / \log_2 x$. To see this, we note that each individual term is smaller whenever p is small and as $\prod_{h', h''} (mq(h' - h'') - 1) \ll x^{O_k(1)}$ the error we pick up from this is at most

$$\prod_{z_0 \leq p \ll_k z_0} 1 - \frac{2k}{p} = 1 + o_k(1)$$

where we have used Merten's third estimate and the fact $\omega(x) \leq (1 + o(1)) \log x / \log_2 x$ from the prime number theorem.

Expanding the product in (8.63) gives us

$$\mathfrak{S}_{m,q}^{-1} \geq (1 + o_k(1)) \mathfrak{S}_m \sum_{\substack{a_{1,2}, \dots, a_{k,k-1} \mid P_{z_0}/P_w \\ a_{i,j} \mid mq(h_i - h_j) - 1}} \frac{(-2k)^{\omega(\mathbf{a})}}{[\mathbf{a}]} \quad (8.67)$$

where we write $\mathbf{a} = [a_{1,2}, \dots, a_{k,k-1}]$. We substitute this bound for $\mathfrak{S}_{m,q}^{-1}$ into lemma 8.1 and

substitute this into (8.60) giving

$$\sum_{q \in \mathcal{I}_m \text{ prime}} \mu_{m,q}(p_0) = \frac{(1 + o_k(1))m(\log x)^k(\log y)^k}{UI_k^{(1)}(F)I_k^{(2)}(G)} \quad (8.68)$$

$$\times \sum_{h \in \mathcal{H}} \sum_{\substack{w_0 \pmod{P_w} \\ (w_0, P_w) = 1}} \sum_{a_{1,2}, \dots, a_{k,k-1} | P_{z_0}/P_w} \frac{(-2k)^{\omega([\mathbf{a}]})}{[\mathbf{a}]} \quad (8.69)$$

$$\times \sum_{\substack{q \in \mathcal{I}_M \text{ prime} \\ p \equiv w_0 \pmod{P_w} \\ a_{i,j} | mq(h_i - h_j) - 1 \ \forall i \neq j}} \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | p_0 + (h_i - h)q}} \sum_{\substack{e_1, \dots, e_k \\ e_i | m(p_0 + (h_i - h)q) - 1}} \lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \right)^2 \quad (8.70)$$

we now focus on the sum over q , we concentrate on the sum over an individual $h \in \mathcal{H}$, we take h_k for convenience.

We first note that as $p_0 > xh_k$ and $(mp_0 - 1, P_y) = 1$ as $p_0 \in \mathcal{R}_m$ we can restrict the sum to $d_k = d'_k = e_k = e'_k = 1$ as there are no other divisors by considerations of the support of the F, G functions. Inserting these conditions into the sum, expanding the square and rearranging the sum over q becomes

$$\sum_{\substack{d_1, \dots, d_k \\ d'_1, \dots, d'_k \\ d_k = d'_k = 1}} \sum_{\substack{e_1, \dots, e_k \\ e'_1, \dots, e'_k \\ e_k = e'_k = 1}} \lambda_{\mathbf{d}, \mathbf{e}} \lambda_{\mathbf{d}', \mathbf{e}'} \sum_{\substack{q \in \mathcal{I}_m \text{ prime} \\ [d_i, d'_i] | p_0 + (h_i - h_k)q \ \forall i \\ [e_j, e'_j] | m(p_0 + (h_i - h_k)q) - 1 \ \forall j \\ q \equiv w_0 \pmod{P_w} \\ a_{i,j} | mq(h_i - h_j) - 1 \ \forall i \neq j}} 1 \quad (8.71)$$

one can verify that the inner sum will be empty unless the following conditions hold

- B.1** $d_k = d'_k = e_1 = e'_1 = 1$
- B.2** $d_1 d'_1, \dots, d_k d'_k, P_w$ pairwise coprime
- B.3** $e_1 e'_1, \dots, e_k e'_k, P_w$ pairwise coprime
- B.4** $a_{1,2}, \dots, a_{k,k-1}, m$ pairwise coprime
- B.5** $p | (d_i d'_i, e_j e'_j)$ then $p | mq(h_j - h_i) - 1$ for all $1 \leq i, j \leq k$.
- B.6** $(d_i d'_i, a_{j,l}) | (h_j - h_l)mp_0 + h_i - h_k$ for all $1 \leq i, j, l \leq k$
- B.7** $(e_i e'_i, a_{j,l}) | (h_j - h_l)(1 - p_0 m) - h_i + h_k$ for all $1 \leq i, j, l \leq k$

Here we reason analogously to deducing conditions **A.1-A.4**. Here **B.5** is of a slightly different form to that given in [21] (for the same reason condition **A.4** is of a different form to that in [21]) but again this ended up not affecting the calculations. If all these conditions are satisfied, then the conditions are consistent and we may apply the Chinese remainder theorem to take the inner sum to be over a single residue class modulo the least common

multiple of $d_1, \dots, d_k, e_1, \dots, e_k, d'_1, \dots, d'_k, e'_1, \dots, e'_k a_{1,2}, \dots, a_{k,k-1}, P_w$. We can also check that the residue class is coprime to the modulus by checking the residue classes in each original congruence is coprime to the modulus (for example, if $p|[e_j, e'_j]$ then clearly $(p, mp_0 - 1) = 1$ otherwise $p|m(h_j - h_k)p_0$ which is impossible as $p < y$. If $p|m$ then $p| - 1$ which of course cannot occur and finally $(h_i - h_k)$ is w -smooth, therefore $(m(h_j - h_k), p) = 1$. This forces the residue class to be coprime to p). So, the inner sum now becomes

$$\frac{\sum_{q \in \mathcal{I}_m \text{ prime}} 1}{\phi(P_w)\phi([\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}])} + O(E(x; P_w[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}])) \quad (8.72)$$

where $[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}]$ represents the least common multiple over the d, d', e, e', a variables and $E(x; q)$ is as in (4.9). We let $r = P_w[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}]$. Recall that $\prod_{i=1}^k d_i, \prod_{i=1}^k d'_i \ll x^{1/10}$ and $\prod_{i=1}^k e_i, \prod_{i=1}^k e'_i \ll x^{o_k(1)}$. Also as $a_{i,j}|p_{z_0} = x^{o(1)}$ we have $\prod_{i \neq j} a_{i,j} = x^{o_k(1)}$ so $r \ll x^{1/5+o_k(1)}$. We can use the trivial bound $E(x; q) \ll x/q$ and the Bombieri-Vinogradov theorem, to see the error term in (8.72) makes a contribution to (8.71) of order

$$\ll_k |\lambda_{\mathbf{d}, \mathbf{e}}| |\lambda_{\mathbf{d}', \mathbf{e}'}| \sum_{r \ll x^{1/5+\epsilon}} \tau_{k^2+4k}(r) E(x; r) \quad (8.73)$$

$$\ll_k \left(x \sum_{r \ll x^{1/5+\epsilon}} \frac{\tau_{k^2+4k}(r)^2}{r} \right)^{1/2} \left(\sum_{r \ll x^{1/5+\epsilon}} E(x; r) \right)^{1/2} \ll_k \frac{x}{(\log x)^{2k}} \quad (8.74)$$

from the Cauchy-Schwarz inequality. Here we have used the fact that $\tau_k^2 \ll \tau_{k^2}$ and that $\sum_{n \leq N} \tau_k(n) \ll N(\log N)^{k-1}$ for any integer $k \geq 2$ on the left most sum. This contribution will turn out to be negligible.

Recall from a previous remark that the number of primes in $\mathcal{I}_m$ is $\sim |\mathcal{I}_m|/\log x$. So, our main term in (8.71) simplifies to

$$\frac{(1 + o(1))|\mathcal{I}_m|}{\phi(P_w) \log x} \sum_{\substack{d_1, \dots, d_k \\ d'_1, \dots, d'_k}}^* \sum_{\substack{e_1, \dots, e_k \\ e'_1, \dots, e'_k}}^* \frac{\lambda_{\mathbf{d}, \mathbf{e}} \lambda_{\mathbf{d}', \mathbf{e}'}}{\phi([\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}])} \quad (8.75)$$

$$= \frac{(1 + o(1))|\mathcal{I}_m| G(0)^2}{\phi(P_w) \log x} \sum_{j=1}^J \sum_{j'=1}^J \sum_{\substack{d_1, \dots, d_{k-1} \\ d'_1, \dots, d'_{k-1}}}^{**} \sum_{\substack{e_1, \dots, e_{k-1} \\ e'_1, \dots, e'_{k-1}}}^{**} \quad (8.76)$$

$$\times \frac{\prod_{l=1}^{k-1} \mu(d_l) \mu(d'_l) \mu(e_l) \mu(e'_l) F_{l,j} \left(\frac{\log d_l}{\log x} \right) F_{l,j'} \left(\frac{\log d'_l}{\log x} \right) G \left(\frac{\log e_l}{\log x} \right) G \left(\frac{\log e'_l}{\log x} \right)}{\phi([\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}])} \quad (8.77)$$

from simply substituting the definitions of the weights. We have written $\sum^*$ to indicate the summation is restricted to the conditions in **B.1-B.7**, similarly we write $\sum^{**}$ to be the sum with constraints **B.2-B.7** as we have separated the contribution from d_k, d'_k, e_k, e'_k . We note this is of a form almost identical to that in (8.15) and so we can proceed with a essentially identical argument. The presence of the totient function in the denominator

affects K_p by a factor of $1 + O_k(p^{-2})$ from Taylor's theorem and so has negligible affect. The added presence of the $\mathbf{a}$ in the denominator means for $p|a_{i,j}$ we have $K_p \ll_k p^{-1}$, so the contribution to (8.71) when $a_{1,2}, \dots, a_{k,k-1} \neq 1, \dots, 1$ is

$$\ll_k \frac{\mathfrak{S}_{m,p_0,h_k}^{(2)} |\mathcal{I}_m|}{\phi(P_w)(\log x)^k (\log y)^{k-1}} \prod_{p|a_{i,j} \text{ for some } i,j} \frac{O_k(1)}{p} \quad (8.78)$$

where

$$\mathfrak{S}_{m,p_0,h}^{(2)} = \prod_{p \leq w} \left(1 - \frac{1}{p}\right)^{-(2k-2)} \prod_{w < p \leq y} \left(1 - \frac{\omega_{m,p_0,h}^{(2)}(p)}{p}\right) \left(1 - \frac{1}{p}\right)^{-(2k-2)} \quad (8.79)$$

$$\omega_{m,p_0,h}^{(2)}(p) = \#\{1 \leq n \leq p : p_0 + (h_i - h)n \equiv (\text{mod } p)\} \quad (8.80)$$

$$\text{or } m(p_0 + (h_i - h)n) \equiv (\text{mod } p) \text{ for some } i\}. \quad (8.81)$$

again by simply following the same reasoning as in the proof of Lemma 8.1. So (8.78) shows the contribution from $a_{1,2}, \dots, a_{k,k-1} \neq 1, \dots, 1$ is

$$\sum_{\substack{a_{1,2}, \dots, a_{k,k-1} | P_{z_0}/P_w \\ a_{1,2}, \dots, a_{k,k-1} \neq 1, \dots, 1}} \frac{(-2k)^{\omega(\mathbf{a})}}{[\mathbf{a}]} \quad (8.82)$$

$$\times \sum_{\substack{q \in \mathcal{I}_M \text{ prime} \\ p \equiv w_0 \pmod{P_w} \\ a_{i,j} | m q (h_i - h_j) - 1 \ \forall i \neq j}} \left(\sum_{d_1, \dots, d_k} \sum_{\substack{e_1, \dots, e_k \\ d_i | p_0 + (h_i - h)q \ e_i | m(p_0 + (h_i - h)q) - 1}} \lambda_{d_1, \dots, d_k, e_1, \dots, e_k} \right)^2 \quad (8.83)$$

$$\ll_k \frac{\mathfrak{S}_{m,p_0,h_k}^{(2)} |\mathcal{I}_m|}{\phi(P_w)(\log x)^k (\log y)^{k-1}} \left(\prod_{w < p \leq z_0} \left(1 + \frac{O_k(1)}{p^2}\right) - 1 \right) \quad (8.84)$$

$$= o_k \left(\frac{\mathfrak{S}_{m,p_0,h_k}^{(2)} |\mathcal{I}_m|}{\phi(P_w)(\log x)^k (\log y)^{k-1}} \right) \quad (8.85)$$

So we must have that the main contribution comes from when $a_{1,2}, \dots, a_{k,k-1} = 1, \dots, 1$. This case provides a contribution of

$$\frac{(1 + o_k(1)) \mathfrak{S}_{m,p_0,h_k}^{(2)} |\mathcal{I}_m| G(0)^2}{\phi(P_w)(\log x)^k (\log y)^{k-1}} \sum_{j=1}^J \sum_{j'=1}^J F_{k,j}(0) F_{k,j'}(0) \left(\int_{t=0}^{\infty} G'(t)^2 dt \right)^{k-1} \quad (8.86)$$

$$\times \prod_{l=1}^{k-1} \int_{t=0}^{\infty} F'_{l,j}(t) F'_{l,j'}(t) dt \quad (8.87)$$

$$= \frac{(1 + o_k(1)) \mathfrak{S}_{m,p_0,h_k}^{(2)} |\mathcal{I}_m| G(0)^2}{\phi(P_w)(\log x)^k (\log y)^{k-1}} J_k^{(1)}(F) J_k^{(2)}(G) \quad (8.88)$$

we note that we will obtain the same estimate as in (8.88) with $\mathfrak{S}_{m,p_0,h_k}^{(2)}$ replaced by

$\mathfrak{S}_{m,p_0,h}^{(2)}$ for a different $h \in \mathcal{H}$. This is due to our demand that F be symmetric forcing $J_k^{(1)}(F)$ to be independent of choice of $h \in \mathcal{H}$. If we now substitute (8.85) and (8.88) into the main term (8.68), we obtain

$$\sum_{q \in \mathcal{I}_m \text{ prime}} \mu_{m,q}(p_0) \geq \frac{(1 + o_k(1))m(\log y)J_k^{(1)}(F)J_k^{(2)}(G)|\mathcal{I}_m|}{UI_k^{(1)}(F)I_k^{(2)}(G)\phi(P_w)} \quad (8.89)$$

$$\times \sum_{h \in \mathcal{H}} \mathfrak{S}_m \mathfrak{S}_{m,p_0,h}^{(2)} \sum_{\substack{w_0 \pmod{P_w} \\ (w_0, P_w) = 1}} 1 \quad (8.90)$$

Clearly the inner sum is $\phi(P_w)$. For any $h \in \mathcal{H}$ we have $\omega_{m,p_0,h}^{(2)}(p) = 0$ for any $p \leq w$ (as $p|(h_i - h)$ but $p \nmid p_0$) and $\omega_{m,p_0,h}^{(2)}(p) \leq 2k - 2$ for all primes p . So by recalling the definitions of $\mathfrak{S}_m$ and $\mathfrak{S}_{m,p_0,h_k}^{(2)}$ from (8.64) and (8.79) we have

$$\mathfrak{S}_m \mathfrak{S}_{m,p_0,h_k}^{(2)} = 2^{-1} \prod_{p|m, p>2} \frac{p-2}{p-1} \prod_{2 < p < w} \frac{\left(1 - \frac{1}{p}\right)^2}{\left(1 - \frac{2}{p}\right)} \prod_{w < p \leq y} \frac{\left(1 - \frac{\omega_{m,p_0,h}^{(2)}(p)}{p}\right)}{\left(1 - \frac{1}{p}\right)^{2k-2}} \quad (8.91)$$

$$\gg_k \prod_{p|m, p>2} \frac{p-2}{p-1} \prod_{w < p \leq y} \left(1 - \frac{2k-2}{p}\right) \left(1 - \frac{1}{p}\right)^{-(2k-2)} \quad (8.92)$$

by noting $\prod_{p>2} (1 - 1/p)^2 / (1 - 2/p)$ a convergent Euler product. We can further bound (8.92) by

$$\gg_k \prod_{p|m, p>2} \frac{p-2}{p-1} \prod_{w < p \leq y} \left(1 - \frac{2k-2}{p}\right) \left(1 + \frac{2k-2}{p}\right) \quad (8.93)$$

As the product on the right-hand side of (8.93) also has each factor of the form $1 + O_k(p^{-2})$ this is also absolutely convergent. So we have

$$\mathfrak{S}_m \mathfrak{S}_{m,p_0,h_k}^{(2)} \gg_k (1 + o_k(1)) \prod_{p|m, p>2} \frac{p-2}{p-1} \quad (8.94)$$

Recalling the size of $|\mathcal{R}_m|$ given in Lemma 7.2 the result follows. □

We now make use of the key integral estimate in [19]

Lemma 8.3. *There exists a suitable choice of smooth functions F, G such that*

$$\frac{kJ_k^{(1)}J_k^{(2)}(G)}{I_k^{(1)}(F)I_k^{(2)}(G)} \gg \log k \quad (8.95)$$

Proof. We take $G(t)$ to be a smooth approximation to $1 - t$ supported on $t \in [0, 1]$ with the property $J_k^{(2)}(G)/I_k^{(2)}(G) \geq 1 - \epsilon$. The existence of such a G guaranteed since the set

of non-negative continuous functions supported on $[0, 1]$ is L^2 and L^1 dense in the set of non-negative continuous functions supported on $[0, 1]$ by the Stone-Weierstrass theorem. Let $F_k : [0, \infty)^k \rightarrow \mathbb{R}$ and $g : [0, \infty) \rightarrow \mathbb{R}$ be defined by

$$F_k(t_1, \dots, t_k) = \begin{cases} \prod_{i=1}^k g(kt_i) & \text{if } \sum_{i=1}^k t_i \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$g(t) = \begin{cases} 1/(1 + At) & t \in [0, T] \\ 0 & \text{otherwise} \end{cases}$$

where $A = \log k - 2 \log_2 k$ and $T = (e^A - 1)/A$.

Remark. Our choice of g was obtained by forming a lower bound for $kJ_k^{(1)}(F_k)/I_k^{(1)}(F_k)$ of roughly the form $(\int_{u=0}^{\infty} g(u)du)^2 / \int_{u=0}^{\infty} g(u)^2 du$ and then maximizing this lower bound subject to some constraints on g . Applying the Euler-Lagrange equation to do this, we find that the optimal g is of the given form.

We note that this is clearly a symmetric function defined on $\sum_{i=1}^k t_i \leq 1$. From [19, Section 7] we have $J_k^{(1)}(F_k)/I_k^{(1)}(F_k) \gg \log k/k$. We choose $F_{l,j}$ such that F is a smooth approximation to $F_k(10t_1, \dots, 10t_k)$ supported on $\{(t_1, \dots, t_k) \in [0, \infty)^k : \sum_{i=1}^k t_i \leq 1/10\}$ with

$$J_k^{(1)}(F)/I_k^{(1)}(F) \geq (1/10 - \epsilon) J_k^{(1)}(F_k)/I_k^{(1)}(F_k) \quad (8.96)$$

giving the result. We are able to choose such an approximation to F as the set of symmetric, non-negative linear combinations of direct products of smooth, compactly supported functions on $[0, 1]^k$ is L^2 - and L^1 - dense in the set of non-negative symmetric L^2 -integrable functions on $[0, 1]^k$.

□

Completing the Proof of Proposition 5.2. Since

$$|\mathcal{I}_m| = (1 + o(1))\delta|\mathcal{R}_m| \log x \quad (8.97)$$

From Lemmas 8.1 and 8.2 we have shown that any primes $p_0 \in \mathcal{R}_m$ satisfying $h_k x < p_0 < U/m - h_k x$ has that the expected number $\sum_q \mu_{m,q}(p_0)$ of times the residue class of p_0 is chosen is $\gg \delta \log x$. From Lemma 7.2 and recalling that $h_k = o_k(\log_2 x)$ from our choice of w , the number of primes $p_0 \in \mathcal{R}_m$ not satisfying $h_k x < p_0 < U/m - h_k x$ is $o_k(|\mathcal{R}_m|)$ for $m < U/z(\log_2 x)^2$.

So we have shown if we choose our residue classes randomly with probability according to $\mu_{m,q}$ then for all but $o_k(|\mathcal{R}_m|)$ primes $p_0 \in \mathcal{R}_m$ the expected number $\sum_q \mu_{m,q}(p_0)$ of times $p_0 \in \mathcal{R}_m$ is chosen is $\gg \delta \log k$ as we guessed in section 6. So if we choose k sufficiently

large relative to δ, ϵ we can guarantee that the expectation is larger than $\log \epsilon^{-1}$. So, as argued in Section 6, the expected number of $p_0 \in \mathcal{R}_m$ not in a chosen residue class must be less than $\epsilon |\mathcal{R}_m|$. So we can guarantee the existence of some choice of residue classes for which at most $\epsilon |\mathcal{R}_m|$ of the primes $p_0 \in \mathcal{R}_m$ are not chosen and so by the argument at the end of Section 5 we prove Proposition 5.2 and hence Theorem 1.1.

□

9 Improvements

As we stated in the introduction, Theorem 1.1 is not the best result on large gaps between primes to date. We will now give an informal discussion of the ideas in [8] that give us the best lower bound of $G(X)$, first following [22] for motivation. In the proof presented in [21] one may notice that there is the potential to allow k to grow slowly with x . In this precise argument, we see that taking k to be something like a fifth-fold logarithm of x may work, which would provide a quantitative improvement on (1.2). In fact Maynard acknowledges this in [21], however the problem we face is to do this one would need a version of Theorem 5.1 with uniformity in k , the number of linear equations, and uniformity in the coefficients of the linear functions as noted in [22, Section 5]; as in Theorem 5.1 our choice of c will ultimately depend on these factors. Such a variant of Theorem 5.1 was achieved in [20]. Using this improvement, one may introduce some technical modifications to the argument in [21] to even take k to be as large as $(\log x)^c$ for some small c . This gives a quantitative improvement of a factor of $\log_3 X / \log_4 X$ on (1.2). More precisely, we get a result of the form

$$G(X) \geq c'' \frac{\log X \log_2 X}{\log_3 X} \quad (9.1)$$

for some positive constant $c'' > 0$, as noted by Maynard in [22]. However, the main improvement made in [8] comes from the observation that the random approach to covering taken in Section 6 is rather inefficient, as we could have a lot of overlap in the covering by two residue classes independently chosen for two different primes. The authors of [8] adapt a hypergraph covering theorem of Pippenger and Spencer [24] to make the covering more efficient. This allows one to improve on (9.1) by an additional factor of $\log_4 X$ and gives an overall quantitative improvement of a factor of $\log_3 X$ on (1.2).

In fact, the use of graph theoretic covering techniques to guarantee more ideal sieving is not new, see [18, Section 4] for example. We will now give a brief overview of this argument, focusing on the covering optimisation element; as we have already demonstrated the success of our multi-dimensional sieve in this problem and perhaps the number of elements we expect to cover using our multi-dimensional sieve is more believable, once we have the result from [20].

9.1 Outline

We now direct our focus to the specific method from [8]. We again follow the program outlined in section 2: We attempt to cover the interval $[1, U]$ with the primes $p \leq x$ where we now take

$$U := c \frac{x \log x \log_3 x}{\log_2 x} \quad (9.2)$$

where c is some small fixed positive constant. In fact, we may instead choose to cover $[1, U] \setminus [1, x]$, noting that achieving this is equivalent to covering $[1, U]$ for our purposes and is slightly more convenient. However, we opt for a slightly different covering than in [21]. For primes in $[1, \log^{20} x]$ we choose the residue class $0 \pmod{p}$ and similarly for primes in $[y, x/2]$ we also choose $0 \pmod{p}$ where we now take $y = \exp(\log x \log_3 x / 4 \log_2 x)$.

By reasoning analogously to in Section 7 one would find that the survivors of this process are essentially the primes between x and U . First, let us set out some notation. Let

$$\mathcal{S} = \{s \text{ prime} : \log^{20} x < s \leq y\} \quad (9.3)$$

$$\mathcal{P} = \{p \text{ prime} : x/2 < p \leq x\} \quad (9.4)$$

$$\mathcal{Q} = \{q \text{ prime} : x < q \leq U\} \quad (9.5)$$

and for residue classes $\vec{a} = (a_s \pmod{s})_{s \in \mathcal{S}}$ and $\vec{b} = (a_p \pmod{p})_{p \in \mathcal{P}}$ we define the sifted sets

$$\mathcal{S}(\vec{a}) := \{n \in \mathbb{Z} : n \not\equiv a_s \pmod{s} \text{ for all } s \in \mathcal{S}\} \quad (9.6)$$

$$\mathcal{S}(\vec{b}) := \{n \in \mathbb{Z} : n \not\equiv b_p \pmod{p} \text{ for all } p \in \mathcal{P}\} \quad (9.7)$$

We reduce proving Theorem 1.2 to proving the following

Theorem 9.1. *Let x be sufficiently large and suppose that U satisfies 9.2. Then there exist vectors $\vec{a}$ and $\vec{b}$ such that*

$$|\mathcal{Q} \cap \mathcal{S}(\vec{a}) \cap \mathcal{S}(\vec{b})| \ll \frac{x}{\log x} \quad (9.8)$$

Again, reasoning analogously to Section 7 one would find that this is equivalent to proving Theorem 1.2.

To do this, we randomly choose and sieve out residue classes $\vec{a} = (a_s \pmod{s})_{s \in \mathcal{S}}$. The survivors of this are $\mathcal{Q} \cap \mathcal{S}(\vec{a})$. The cardinality of this set will typically have order $x \log_2 x / \log x$, from Metens' third estimate. So, our goal is now reduced to finding integers n_p for each prime between $x/2$ and x that cuts the set $\mathcal{Q} \cap \mathcal{S}(\vec{a})$ down to a set of cardinality $O(x / \log x)$ after sieving. We may allow the set of survivors to have size $O(x / \log x)$ as we can easily eliminate these by increasing x by some constant multiplicative factor. We will use our multi-dimensional sieve weights to achieve this.

This choice of covering differs from that in Section 5 in three ways. Firstly, we choose a random covering for ‘small’ primes (the primes in $\mathcal{S}$). Secondly, we have chosen a slightly different choice of y and finally we no longer have the equivalent of our parameter z . As discussed in a remark in Section 7 this random covering does not make too much of a difference, but does give us something else to optimise over, which is attractive. The different choice of y is chosen so that our preliminary sieve estimates performed to verify proving Theorem 9.1 is equivalent to proving Theorem 1.2, are legitimately of the desired order, specifically in our application of de Bruijn’s Theorem (7.2); as our U is now, of course, considerably large. We no longer need our parameter z as this was originally there to leave some left over primes to cover the $o(x/\log x)$ survivors from our sieving, however, as we have discussed in this set up, we may permit the survivor set to have cardinality $O(x/\log x)$, simply eliminating these by increasing x by some constant multiplicative factor, which we do not really care about as we do not keep track of the implied constant in (1.4).

If we naively believe that the residue classes we pick for each prime p between $x/2$ and x are largely non-overlapping in $\mathcal{Q} \cap \mathcal{S}(\vec{a})$, then to reduce the cardinality of $\mathcal{Q} \cap \mathcal{S}(\vec{a})$ to $O(x/\log x)$ after sieving we would need to guarantee the existence of some $\vec{b}$ such that each prime captures around $\log_2 x$ members of $\mathcal{Q} \cap \mathcal{S}(\vec{a})$. Following Section 6.2 we can guarantee the existence of many tuples $q + h_1 p, \dots, q + h_k p$ containing $\gg \log k$ primes. Following [20], one can take k to be as large as $\log^c x$ for a small absolute constant c , whilst still having uniformity in our estimate on the number of primes. This would therefore guarantee that each prime is eliminating $\gg \log_2 x$ elements of $\mathcal{Q} \cap \mathcal{S}(\vec{a})$. The problem we now face is the overlap of the residue classes. In [21] there was some inefficiency in choosing the residue classes randomly, as we expect some overlap. We now hope to adapt a covering theorem that allows us to choose these classes more efficiently.

To motivate this, consider for some index set $\mathcal{I}$ a collection of non-empty hypergraphs $(V, E_i)_{i \in \mathcal{I}}$ all on the same vertex set V . We hope to choose from each edge set E_i a e_i such that $\bigcup_{i \in \mathcal{I}} e_i$ covers as much of V as possible. This corresponds to our problem at hand by taking our vertex set to be $\mathcal{Q} \cap \mathcal{S}(\vec{a})$ and our edge sets to be E_p , the intersection of each $b_p \pmod{p}$ with $\mathcal{Q} \cap \mathcal{S}(\vec{a})$. We could first attempt to make this selection of edges randomly. If we do this, reasoning as we did in Section 6, the probability some vertex is left in $V \setminus \bigcup_{i \in \mathcal{I}} e_i$ would be

$$\prod_{i \in \mathcal{I}} (1 - \mathbb{P}(v \in e_i)) \leq \exp(-d_I(v)) \quad (9.9)$$

where we define the normalised degree of v to be

$$d_I(v) := \sum_{i \in \mathcal{I}} \mathbb{P}(v \in e_i) \quad (9.10)$$

To simplify the situation, let us naively assume some uniformity condition, so that $d_I(v) \approx d$ for each v . We would therefore expect the density of our sifted set $V \setminus \bigcup_{i \in \mathcal{I}} e_i$ to be $\exp(-d)$ as in our application the index set $\mathcal{I}$ is large and $d_I(v)$ is small. We are now interested in whether one can do better than this. An ideal scenario would be to modify the coupling we have between distinct edges e_i and e_j so that they are always disjoint. This would give us that the probability that an edge v is sieved out is

$$1 - \sum_{i \in \mathcal{I}} \mathbb{P}(v \in e_i) \quad (9.11)$$

So, again assuming an optimistic uniformity condition, this would indicate that the density of the sifted set would now be $\max(0, 1 - d)$ instead of $\exp(-d)$ and so d would now need only be 1 to almost entirely sieve out our vertex set.

Assuming this is possible, we set our vertex set V to be $\mathcal{Q} \cap \mathcal{S}(\vec{a})$. Returning to the context of Theorem 9.1, we suspect that $\mathcal{Q} \cap \mathcal{S}(\vec{a})$ will have size

$$\frac{U}{\log x} \prod_{s \in \mathcal{S}} \left(1 - \frac{1}{s}\right) \approx c \frac{x \log_2 x}{\log x} \quad (9.12)$$

so the set is roughly $c \log_2 x$ times larger than $\mathcal{P}$. Using our multidimensional sieve, we suspect to be able to guarantee many residue classes $b_p \pmod{p}$ containing roughly $\asymp \log_2 x$ elements of $\mathcal{Q} \cap \mathcal{S}(\vec{a})$ on average for most primes p in $\mathcal{P}$. Setting our index set $\mathcal{I} = \mathcal{P}$ and E_p to be the set of all such intersections $(b_p \pmod{p}) \cap V$ then guaranteeing $\mathcal{Q} \cap \mathcal{S}(\vec{a}) \cap \mathcal{S}(\vec{b})$ may be made small is the same as making $V \setminus \bigcup_{i \in \mathcal{I}} e_i$ small, for some suitable edges e_p drawn from E_p . By double counting, the expected density here should be

$$d \asymp \frac{\#\mathcal{P} \times \log_2 x}{\#V} \asymp \frac{1}{c} \quad (9.13)$$

So we would be able to sieve out $\mathcal{Q} \cap \mathcal{S}(\vec{a})$ almost entirely with optimal sieving. If we compare this to the random case, we expect the cardinality of $\mathcal{Q} \cap \mathcal{S}(\vec{a}) \cap \mathcal{S}(\vec{b})$ to be something like $\exp(-1/c) \times cx \log_2 x / \log x$. To ensure we can make this as small as necessary, we would need to make c just smaller than $1/\log_3 x$. This loss of $\log_3$ leads to the loss of $\log_4 X$ in (9.1) compared to Theorem 1.2.

So, this heuristic discussion motivates the desire for a more efficient covering, this is achieved in [8, Theorem 3]. However, it is not immediately obvious that we should be able to guarantee such disjoint coverings. We now describe the method in which one picks their congruence classes to ensure that this is possible.

As we stated earlier, the covering result in [8] is an adaptation of a covering theorem of Pippenger and Spencer which follows from previous work of Rödl [26]. Pippenger and Spencer very roughly make the following assumptions in order to prove their covering result:

1. (Uniform degrees) One has $d_I(v) \approx d$ for almost all v .
2. (Independence) The hypergraphs (V, E_i) do not depend on i .
3. (Small codegrees) The normalised codegrees $\sum_{i \in I} \mathbb{P}(v, w \in e_i)$ are small for $v \neq w$.
4. (Uniform covering) The edges e_i of E_i are of constant size.

One finds that in fact condition 2 may be removed as it is not necessary for the result to hold. In our context we can verify conditions 1 and 3, however it immediately seems that we will certainly not be able to establish a form of condition 4, as we will surely have a large fluctuation in the size of the sets $e_p = b_p \pmod{p}$, and it is the fact that there are choices of b_p making our e_p s exceptionally large that allows our covering to be successful. So, the authors of [8] adapt the method in [24] to fit the scenario where we have large fluctuations in the size of our edges.

The original argument of Pippenger and Spencer follows from the ‘Rödl nibble’ method from [26]. Roughly, this involves separating our indexing set $\mathcal{I}$ into a collection of subsets $\mathcal{I}_1, \dots, \mathcal{I}_m$. At each $\mathcal{I}_k$ we will perform a ‘nibble’. First, we select each e_i for $i \in \mathcal{I}_1$ uniformly and independently at random. Then for $\mathcal{I}_2$ we restrict our edges so that they are disjoint from the edges we have already picked in the first ‘nibble’, and then we pick our e_i s for $i \in \mathcal{I}_2$ independently at random under this new conditioned distribution. We then repeat this step for each $\mathcal{I}_k$ with $1 \leq k \leq m$ until we have exhausted our indexing set. This method certainly will guarantee better disjointness properties in our covering as m increases but makes it harder to analyse the probability of success.

As we noted, our e_i s have quite varying sizes, and unfortunately our nibbling processes is naturally biased towards discarding larger $b_p \pmod{p} \cap V$ compared with smaller such sets as certainly larger sets are more likely to have intersections with sets selected in our previous ‘nibbles’. One way to resolve this issue is to reweight the probability distribution on the e_i s during our nibble to attempt to eliminate this undesirable bias. Following this program, one can successfully guarantee the existence of a near disjoint covering which was exactly what we needed to prove our best result, Theorem 1.2.

10 Acknowledgments

The author would like to thank Adam Harper for suggesting the topic of large gaps between primes for this project and for all the helpful conversations, suggestions, and comments throughout.

Bibliography

References

- [1] P. Billingsley. *Probability and measure*. A Wiley-Interscience publication. Wiley, New York, 3. ed edition, 1995.
- [2] H. Cramér. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arithmetica*, 2(1):23–46, 1936.
- [3] C. De la Vallée Poussin. Sur la fonction $\zeta(s)$ de riemann et le nombre des nombres premiers inférieurs à une limite donnée. *Mémoires de l'Académie royale de Belgique*, 59(1):1–74, 1899.
- [4] N. G. deBruijn. On the number of positive integers $\leq x$ and free of prime factors $> y$. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen: Series A: Mathematical Sciences*, 54(1):50–60, 1951.
- [5] P. D. T. A. Elliott and H. Halberstam. A conjecture in prime number theory. *Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69)*, pages 59–72, 1970.
- [6] P. Erdős. On the difference of consecutive primes. *The Quarterly Journal of Mathematics*, 6(1):124–128, 1935.
- [7] P. Erdős. Some of my favourite unsolved problems. In *A Tribute to Paul Erdős (A. Baker, B. Bollobás, A. Hajnal, eds.)*, page 467–478. Cambridge University Press, 1990.
- [8] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao. Long gaps between primes. *Journal of the American Mathematical Society*, 31(1):65–105, 2018.
- [9] K. Ford, B. Green, S. Konyagin, and T. Tao. Large gaps between consecutive prime numbers. *Annals of Mathematics*, 183(3):935–974, 2016.
- [10] J. B. Friedlander and H. Iwaniec. *Opera de Cribro*. Colloquium Publications - American Mathematical Society, 2010.
- [11] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in tuples. I. *Ann. of Math. (2)*, 170(2):819–862, 2009.
- [12] D. A. Goldston and C. Y. Yıldırım. Higher correlations of divisor sums related to primes. III. Small gaps between primes. *Proc. Lond. Math. Soc. (3)*, 95(3):653–686, 2007.
- [13] A. Granville. Harald Cramér and the distribution of prime numbers. In *Proceedings of the symposium held in Stockholm, September 24–25, 1993*, pages 12–28. Scandinavian University Press, Oslo, 1995.

- [14] A. Granville. Primes in intervals of bounded length. *Bull. Amer. Math. Soc. (N.S.)*, 52(2):171–222, 2015.
- [15] H. Halberstam and H.-E. Richert. *Sieve Methods*. Academic Press, 1974.
- [16] D. Koukoulopoulos. *The Distribution of Prime Numbers*. Graduate Studies in Mathematics. American Mathematical Society, 2019.
- [17] H. Maier. Primes in short intervals. *Michigan Math. J.*, 32(2):221–225, 1985.
- [18] H. Maier and C. Pomerance. Unusually large gaps between consecutive primes. *Transactions of the American Mathematical Society*, 322(1):201–237, 1990.
- [19] J. Maynard. Small gaps between primes. *Annals of Mathematics*, 181(1):383–413, 2015.
- [20] J. Maynard. Dense clusters of primes in subsets. *Compositio Mathematica*, 152(7):1517–1554, 2016.
- [21] J. Maynard. Large gaps between primes. *Annals of Mathematics*, 183(3):915–933, 2016.
- [22] J. Maynard. Gaps between primes. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. II. Invited lectures*, pages 345–361. World Sci. Publ., Hackensack, NJ, 2018.
- [23] H. L. Montgomery and R. C. Vaughan. *Multiplicative Number Theory I: Classical Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, reprint edition, 2006.
- [24] N. Pippenger and J. Spencer. Asymptotic behavior of the chromatic index for hypergraphs. *Journal of Combinatorial Theory, Series A*, 51(1):24–42, 1989.
- [25] R. A. Rankin. The difference between consecutive prime numbers. *Journal of the London Mathematical Society*, s1-13(4):242–247, 10 1938.
- [26] V. Rödl. On a packing and covering problem. *Eur. J. Comb.*, 6(1):69–78, 1985.
- [27] D. Shanks. On maximal gaps between successive primes. *Mathematics of Computation*, 18(88):646–651, 1964.
- [28] A. Walfisz. Zur additiven zahlentheorie. ii. *Mathematische Zeitschrift*, 40(1):592–607, 1936.
- [29] E. Westzynthius. Über die verteilung der zahlen, die zu den n ersten primzahlen teilerfremd sind. *Commentat. Phys.-Math*, 25(5):1–37, 1931.
- [30] Y. Zhang. Bounded gaps between primes. *Annals of Mathematics*, 179(3):1121–1174, 2014.