LONG GAPS BETWEEN PRIMES

KEVIN FORD, BEN GREEN, SERGEI KONYAGIN, JAMES MAYNARD, AND TERENCE TAO

ABSTRACT. Let p_n denote the n-th prime. We prove that

$$\max_{p_n \leqslant X} (p_{n+1} - p_n) \gg \frac{\log X \log\log X \log\log\log\log X}{\log\log\log X}$$

for sufficiently large X, improving upon recent bounds of the first three and fifth authors and of the fourth author. Our main new ingredient is a generalization of a hypergraph covering theorem of Pippenger and Spencer, proven using the Rödl nibble method.

CONTENTS

1. INTRODUCTION

Let p_n denote the n^{th} prime, and let

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

It is clear from the prime number theorem that

$$G(X) \geqslant (1 + o(1)) \log X,$$

as the *average* gap between the prime numbers which are $\leqslant X$ is $\sim \log X$. In 1931, Westzynthius [46] proved that infinitely often, the gap between consecutive prime numbers can be an arbitrarily large multiple of the average gap, that is, $G(X)/\log X \to \infty$ as $X \to \infty$, improving upon prior results of Backlund [2] and Brauer-Zeitz [5]. Moreover, he proved the quantitative bound[1]

$$G(X) \gg \frac{\log X \log_3 X}{\log_4 X}.$$

[1]As usual in the subject, $\log_2 x = \log\log x$, $\log_3 x = \log\log\log x$, and so on. The conventions for asymptotic notation such as $\ll$ and $o()$ will be defined in Section 2.

In 1935 Erdős [11] sharpened this to

$$G(X) \gg \frac{\log X \log_2 X}{(\log_3 X)^2}$$

and in 1938 Rankin [40] made a subsequent improvement

$$G(X) \geqslant (c + o(1))\frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2}$$

with $c = \frac{1}{3}$. The constant c was increased several times: to $\frac{1}{2}e^\gamma$ by Schönhage [43], then to $c = e^\gamma$ by Rankin [41], to $c = 1.31256e^\gamma$ by Maier and Pomerance [32] and, most recently, to $c = 2e^\gamma$ by Pintz [36].

Recently, in two independent papers [13, 35], the authors showed that c could be taken to be arbitrarily large, answering in the affirmative a long-standing conjecture of Erdős [12]. The methods of proof in [13] and [35] both introduced estimates on primes in very short arithmetic progressions, but these differed in some key aspects. The arguments in [13] used recent results [22, 21, 23] on the number of solutions to linear equations in primes, whereas the arguments in [35] instead relied on multidimensional prime-detecting sieves introduced in [33]. Our main theorem is the following quantitative improvement.

Theorem 1 (Large prime gaps)**.** *For any sufficiently large X, one has*

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

The implied constant is effective.

Our overall approach combines ideas from the two previous papers [13, 35]. There are two key ingredients which allow us to obtain the quantitative improvement. Firstly, we incorporate a uniform version of the multidimensional sieve approach as worked out in [34], which gives a quantitative improvement to the underlying estimates about primes. Secondly, we prove a generalization of a hypergraph covering theorem of Pippenger and Spencer [37], which allows for an essentially optimal means of translating such estimates into a result on large gaps between primes. It is this covering theorem which is the key new ingredient in our work, and may be of independent interest.

All approaches which obtain quantitative improvements beyond the average bound $G(X) \gg \log X$ have used a sieving argument which is conjectured to be unable to produce a result stronger than $G(X) \gg \log X(\log_2 X)^{2+o(1)}$. Moreover, in light of the essentially optimal bounds in our covering theorem for this problem and the current limitations of the multidimensional sieve estimates, Theorem 1 appears to be the strongest result one can hope to prove without improvements towards the Hardy-Littlewood prime k-tuples conjecture, or a radically new approach.

In a sequel [15] to this paper, a subset of the authors will extend the above theorem to also cover chains of consecutive large gaps between primes, by combining the methods in this paper with the Maier matrix method. In view of this, we have written some of the key propositions in this paper in slightly more generality than is strictly necessary to prove Theorem 1, as the more general versions of these results will be useful in the sequel [15].

The results and methods of this paper have also subsequently been applied by Maier and Rassias [31] (extending the method of the first author, Heath-Brown and the third author [14]) to obtain large prime gaps (of the order of that in Theorem 1) that contain a perfect k^{th} power of a prime for a fixed k, and by Baker and Freiberg [3] to obtain lower bounds on the density of limit points of prime gaps normalized by any function that grows slightly more slowly than the one in Theorem 1. We refer the interested reader to these papers for further details.

1.1. **Historical background.** Based on a probabilistic model of primes, Cramér [8] conjectured that

$$\limsup_{X\to\infty} \frac{G(X)}{\log^2 X} = 1.$$

Granville [20] offered a refinement of Cramér's model and has conjectured that the $\limsup$ above is in fact at least $2e^{-\gamma} = 1.1229\ldots$. These conjectures are well beyond the reach of our methods. Cramér's model also predicts that the normalized prime gaps $\frac{p_{n+1}-p_n}{\log p_n}$ should have exponential distribution, that is, $p_{n+1} - p_n \geqslant C \log p_n$ for about $e^{-C}\pi(X)$ primes $\leqslant X$, for any fixed $C > 0$. Numerical evidence from prime calculations up to $4 \cdot 10^{18}$ [44] matches this prediction quite closely, with the exception of values of C close to $\log X$, for which there is very little data available. In fact, $\max_{X\leqslant 4\cdot 10^{18}} G(X)/\log^2 X \approx 0.9206$, slightly below the predictions of Cramér and Granville.

Unconditional upper bounds for $G(X)$ are far from the conjectured truth, the best being $G(X) \ll X^{0.525}$ and due to Baker, Harman and Pintz [4]. Even the Riemann Hypothesis only[2] furnishes the bound $G(X) \ll X^{1/2}\log X$ [7].

All works on lower bounds for $G(X)$ have followed a similar overall plan of attack: show that there are at least $G(X)$ consecutive integers in $(X/2, X]$, each of which has a "very small" prime factor. To describe the results, we make the following definition.

Definition 1. *Let x be a positive integer. Define $Y(x)$ to be the largest integer y for which one may select residue classes $a_p \bmod p$, one for each prime $p \leqslant x$, which together "sieve out" (cover) the whole interval $[y] = \{1, \ldots, y\}$. Equivalently, $Y(x)$ is the largest integer m so that there are m consecutive integers, each with a factor in common with $P(x)$.*

The relation between this function Y and gaps between primes is encoded in the following simple lemma.

Lemma 1.1. *Write $P(x)$ for the product of the primes less than or equal to x. Then*

$$G(P(x) + x) \geqslant Y(x).$$

Proof. Set $y = Y(x)$, and select residue classes $a_p \bmod p$, one for each prime $p \leqslant x$, which cover $[y]$. By the Chinese remainder theorem there is some m, $x < m \leqslant x + P(x)$, with $m \equiv -a_p \pmod p$ for all primes $p \leqslant x$. We claim that all of the numbers $m + 1, \ldots, m + y$ are composite, which means that there is a gap of length y amongst the primes less than $m + y$, thereby concluding the proof of the lemma. To prove the claim, suppose that $1 \leqslant t \leqslant y$. Then there is some p such that $t \equiv a_p \pmod p$, and hence $m + t \equiv -a_p + a_p \equiv 0 \pmod p$, and thus p divides $m + t$. Since $m + t > m > x \geqslant p$, $m + t$ is indeed composite. □

By the prime number theorem we have $P(x) = e^{(1+o(1))x}$. Thus the bound of Lemma 1.1 implies that

$$G(X) \geqslant Y\big((1 + o(1)) \log X\big)$$

as $X \to \infty$. In particular, Theorem 1 is a consequence of the bound

$$Y(x) \gg \frac{x \log x \log_3 x}{\log_2 x}, \tag{1.1}$$

which we will establish later in this paper. This improves on the bound $Y(x) \gg \frac{x \log x \log_3 x}{\log_2^2 x}$ obtained by Rankin [40].

[2]Some slight improvements are available if one also assumes some form of the pair correlation conjecture; see [26].

The function Y is intimately related to *Jacobsthal's function* j. If n is a positive integer then $j(n)$ is defined to be the maximal gap between integers coprime to n. In particular $j(P(x))$ is the maximal gap between numbers free of prime factors $\leqslant x$, or equivalently 1 plus the longest string of consecutive integers, each divisible by some prime $p \leqslant x$. The Chinese remainder theorem construction given in the proof of Lemma 1.1 in fact proves that

$$Y(x) = j(P(x)) - 1. \tag{1.2}$$

This observation, together with results in the literature, gives upper bounds for Y. The best upper bound known is $Y(x) \ll x^2$, which comes from Iwaniec's work [28] on Jacobsthal's function. It is conjectured by Maier and Pomerance that in fact $Y(x) \ll x(\log x)^{2+o(1)}$. This places a serious (albeit conjectural) upper bound on how large gaps between primes we can hope to find via lower bounds for $Y(x)$: a bound in the region of $G(X) \gtrapprox \log X(\log\log X)^{2+o(1)}$, far from Cramér's conjecture, appears to be the absolute limit of such an approach.

The lower bound on certain values of Jacobsthal's function provided by (1.1), (1.2) can be inserted directly into [39, Theorem 1] to obtain a lower bound for the maximum over l of $p(k,l)$, the least prime in the arithmetic progression $l \bmod k$, in the case when the modulus k has no small prime factors. We have

Corollary 1. *For any natural number k, let $M(k)$ denote the maximum value of $p(k,l)$ over all l coprime to k. Suppose that k has no prime factors less than or equal to x for some $x \leqslant \log k$. Then, if x is sufficiently large (in order to make $\log_2 x$, $\log_3 x$ positive), one has the lower bound*

$$M(k) \gg k\frac{x\log x\log_3 x}{\log_2 x}.$$

Proof. Apply [36, Theorem 1] with $m = P(x)$ if $x \leqslant \frac{1}{2}\log k$ and with $m = P(\frac{1}{2}\log k)$ if $\frac{1}{2}\log k < x \leqslant \log k$. □

In view of [39, Theorem 3], one may also expect to be able to prove a lower bound of the form

$$M(k) \gg \phi(k)\frac{\log k\log_2 k\log_4 k}{\log_3 k} \tag{1.3}$$

for a set of natural numbers k of density 1, but we were unable to find a quick way to establish this from the results in this paper [3].

1.2. **Method of proof.** Our methods here are a combination of those in our previous papers [13, 35], which are in turn based in part on arguments in earlier papers, particularly those of Rankin [40] and Maier-Pomerance [32]. In order to make the lower bound in Theorem 1 as efficient as possible, we combine these ideas with a generalization of some arguments of Pippenger and Spencer [37].

As noted above, to prove Theorem 1, it suffices to sieve out an interval $[y]$ by residue classes $a_p \bmod p$ for each prime $p \leqslant x$, where $y \asymp \frac{x\log x\log_3 x}{\log_2 x}$. Actually, it is permissible to have $O(\frac{x}{\log x})$ survivors in $[y]$ that are not sieved out by these residue classes, since one can easily eliminate such survivors by increasing x by a constant multiplicative factor. Also, for minor technical reasons, it is convenient to sieve out $[y]\backslash[x]$ rather than $[y]$.

Following [13], we will sieve out $[y]\backslash[x]$ by the residue classes $0 \bmod p$ both for very small primes p ($p \leqslant \log^{20} x$) and medium primes p (between $z := x^{\log_3 x/(4\log_2 x)}$ and $x/2$). The survivors of this process

[3] Inequality (1.3) has recently been established by Li, Pratt and Shakan [30] for every positive integer k except those with more than $\exp\{(1/2-\varepsilon)\log_2 k\log_4 k/\log_3 k\}$ prime factors, $\varepsilon > 0$ fixed.

are essentially the set $\mathcal{Q}$ of primes between x and y. After this initial sieving, the next stage will be to randomly sieve out residue classes $\tilde{\mathbf{a}} = (\mathbf{a}_s \bmod s)_{s\in\mathcal{S}}$ for small primes s (between $\log^{20} x$ and z). (This approach differs slightly from the approach taken in [35] and earlier papers, in which a fixed residue class is used for all small (and very small) primes instead.) This cuts down the set of primes $\mathcal{Q}$ to a smaller set $\mathcal{Q} \cap S(\tilde{\mathbf{a}})$, whose cardinality is typically on the order of $\frac{x}{\log x} \log_2 x$. The remaining task is then to select integers n_p for each prime p between $x/2$ and x, such that the residue classes $n_p \bmod p$ cut down $\mathcal{Q} \cap S(\tilde{\mathbf{a}})$ to a set of survivors of size $O(\frac{x}{\log x})$.

Assuming optimistically that one can ensure that the different residue classes $n_p \bmod p$ largely remove disjoint sets from $\mathcal{Q} \cap S(\tilde{\mathbf{a}})$, we are led to the need to select the integers n_p so that each $n_p \bmod p$ contains about $\log_2 x$ of the primes in $\mathcal{Q} \cap S(\tilde{\mathbf{a}})$. In [13], the approach taken was to use recent results on linear equations in primes [21, 22, 23] to locate arithmetic progressions $q, q + r!p, \ldots, q + (r-1)r!p$ consisting entirely of primes for some suitable r, and then to take $n_p = q$. Unfortunately, due to various sources of ineffectivity in the known results on linear equations in primes, this method only works when r is fixed or growing extremely slowly in x, whereas here we would need to take r of the order of $\log_2 x$. To get around this difficulty, we use instead the methods from [35], which are based on the multidimensional sieve methods introduced in [33] to obtain bounded intervals with many primes. A routine modification of these methods gives tuples $q + h_1 p, \ldots, q + h_k p$ which contain $\gg \log k$ primes, for suitable large k; in fact, by using the calculations in [34], one can take k as large as $\log^c x$ for some small absolute constant c (e.g. $c = 1/5$), so that the residue class $q \bmod p$ is guaranteed to capture $\gg \log_2 x$ primes in $\mathcal{Q}$.

There is however a difficulty due to the overlap between the residue classes $n_p \bmod p$. In both of the previous papers [13, 35], the residue classes were selected randomly and independently of each other, but this led to a slight inefficiency in the sieving: with each residue class $n_p \bmod p$ containing approximately $\log_2 x$ primes, probabilistic heuristics suggest that one would have needed the original survivor set $\mathcal{Q} \cap S(\tilde{\mathbf{a}})$ to have size about $\frac{x}{\log x} \frac{\log_2 x}{\log_3 x}$ rather than $\frac{x}{\log x} \log_2 x$ if one is to arrive at $O(\frac{x}{\log x})$ after the final sieving process. This ultimately leads to the bound

$$G(X) \gg \frac{\log X \log_2 X}{\log_3 X}, \tag{1.4}$$

as worked out in in unpublished work of the fourth author - an additional loss of $\log_4 x$ compared to Theorem 1.

To avoid this difficulty, we use some ideas from the literature on efficient hypergraph covering. Of particular relevance is the work of Pippenger and Spencer [37] in which it is shown that whenever one has a large hypergraph $G = (V, E)$ which is uniform both in the sense of edges $e \in E$ having constant cardinality, and also in the sense of the degrees $\#\{e \in E : v \in e\}$ being close to constant in v, one can efficiently cover most of V by almost disjoint edges in E. Unfortunately, the results in [37] are not directly applicable for a number of technical reasons, the most serious of which is that the analogous hypergraph in this case (in which the vertices are the sifted set $\mathcal{Q} \cap S(\tilde{\mathbf{a}})$ and the edges are sets of the form $\{q \in \mathcal{Q} \cap S(\tilde{\mathbf{a}}) : q \equiv n_p \pmod p\}$ for various n_p, p) does not have edges of constant cardinality. However, by modifying the "Rödl nibble" or "semi-random" method used to prove the Pippenger-Spencer theorem, we are able to obtain a generalization of that theorem in which the edges are permitted to have variable cardinality. This generalization is purely combinatorial in nature and may be of independent interest beyond the application here to large prime gaps.

We will make a series of reductions to prove Theorem 1. To aid the reader, we summarize the chain of implications below, indicating in which Section each implication or Theorem is proven (above or below),

and in which Section one may find a statement of each Theorem (in parentheses).

$$\begin{array}{ccccccc} & & & \overset{\S5}{} & & & \\ & & & \text{Thm 3 } (\S4.2) & & & \\ & & & \Downarrow & & & \\ & & & \text{Cor 4 } (\S4.3) & & & \\ \text{Thm 5 } (\S6) & \underset{\S7,8}{\Longrightarrow} & \text{Thm 4 } (\S4) & \underset{\S4,5}{\overset{\Downarrow}{\Longrightarrow}} & \text{Thm 2 } (\S3) & \underset{\S3}{\Longrightarrow} & \text{Thm 1} \end{array}$$

The deduction of Theorem 1 from Theorem 2 is easy, and codifies the reduction of the problem to that of finding residue classes for primes in $S \cup (x/2, x]$ which cover all the primes in $\mathcal{Q}$ with $O(x/\log x)$ exceptions. Theorem 5, proved using the sieve methods from [34], postulates the existence of a weight function with certain average properties. It implies the existence of residue classes $n_p \bmod p$ for primes $p \in (x/2, x]$, each containing many primes of Q, and moreover that each prime $q \in \mathcal{Q} \cap S(\tilde{\mathbf{a}})$ is covered by about the same number of these congruence classes $n_p \bmod p$. These properties are quantified in Theorem 4. Showing that Theorem 4 implies Theorem 2, i.e. that there exist choices for n_p which *efficiently* cover most of the primes in $q \in \mathcal{Q} \cap S(\tilde{\mathbf{a}})$, is accomplished with our new hypergraph covering tool. The fundamental result is Theorem 3, which is written in a very general form and is consequently rather long to state. Corollary 4 is a somewhat shorter version specialized for our purposes.

For ease of reading, we have endeavored to separate the combinatorial arguments of Section 4, 5 and 6 from the number theoretic arguments of Section 7 and 8. Indeed, a reader only interested in our hypergraph covering result Theorem 3 can read Section 4 and 5 as a standalone paper. A reader only interested in the number theoretic part of Theorem 1 can just read Section 7 and 8 provided they are willing to assume the reduction of Theorem 2 to Theorem 5. The deduction of Theorem 2 from the purely combinatorial Corollary 4 and the purely number theoretic Theorem 5 is performed in the second half of Section 4 and in Section 6, and does not require reading the more specialized Section 5, 7 or 8.

1.3. **Acknowledgments.** The research of SK was partially performed while he was visiting KF at the University of Illinois at Urbana–Champaign. Research of KF and SK was also carried out in part at the University of Chicago. KF and SK are thankful to Prof. Wilhelm Schlag for hosting these visits. KF also thanks the hospitality of the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences.

Also, the research of the SK and TT was partially performed while SK was visiting the Institute for Pure and Applied Mathematics (IPAM) at UCLA, which is supported by the National Science Foundation.

The research of JM was conducted partly while he was a CRM-ISM postdoctoral fellow at the Université de Montréal, and partly while he was a Fellow by Examination at Magdalen College, Oxford. He also thanks Andrew Granville and Daniel Fiorilli for many useful comments and suggestions. TT also thanks Noga Alon and Van Vu for help with the references.

KF was supported by NSF grant DMS-1201442. BG was supported by ERC Starting Grant 279438, *Approximate algebraic structure*, and a Simons Investigator grant. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1266164.

2. Notational conventions

In most of the paper, x will denote an asymptotic parameter going to infinity, with many quantities allowed to depend on x. The symbol $o(1)$ will stand for a quantity tending to zero as $x \to \infty$. The same convention applies to the asymptotic notation $X \sim Y$, which means $X = (1 + o(1))Y$. We use $X = O(Y)$, $X \ll Y$, and $Y \gg X$ to denote the claim that there is a constant $C > 0$ such that $|X| \leqslant CY$ throughout the domain

of the quantity X. We adopt the convention that C is independent of any parameter unless such dependence is indicated, e.g. by subscript such as $\ll_k$. In all of our estimates here, the constant C will be effective (we will not rely on ineffective results such as Siegel's theorem). If we can take the implied constant C to equal 1, we write $f = O_{\leqslant}(g)$ instead. Thus for instance

$$X = (1 + O_{\leqslant}(\varepsilon))Y$$

is synonymous with

$$(1-\varepsilon)Y \leqslant X \leqslant (1+\varepsilon)Y.$$

Finally, we use $X \asymp Y$ synonymously with $X \ll Y \ll X$.

When summing or taking products over the symbol p, it is understood that p is restricted to be prime.

Given a modulus q and an integer n, we use $n \bmod q$ to denote the congruence class of n in $\mathbb{Z}/q\mathbb{Z}$.

Given a set A, we use 1_A to denote its indicator function, thus $1_A(x)$ is equal to 1 when $x \in A$ and zero otherwise. Similarly, if E is an event or statement, we use 1_E to denote the indicator, equal to 1 when E is true and 0 otherwise. Thus for instance $1_A(x)$ is synonymous with $1_{x\in A}$.

We use $\#A$ to denote the cardinality of A, and for any positive real z, we let $[z] := \{n \in \mathbf{N} : 1 \leqslant n \leqslant z\}$ denote the set of natural numbers up to z.

Our arguments will rely heavily on the probabilistic method. Our random variables will mostly be discrete (in the sense that they take at most countably many values), although we will occasionally use some continuous random variables (e.g. independent real numbers sampled uniformly from the unit interval $[0,1]$). As such, the usual measure-theoretic caveats such as "absolutely integrable", "measurable", or "almost surely" can be largely ignored by the reader in the discussion below. We will use boldface symbols such as $\mathbf{X}$ or $\mathbf{a}$ to denote random variables (and non-boldface symbols such as X or a to denote deterministic counterparts of these variables). Vector-valued random variables will be denoted in arrowed boldface, e.g. $\vec{\mathbf{a}} = (\mathbf{a}_s)_{s\in\mathcal{S}}$ might denote a random tuple of random variables $\mathbf{a}_s$ indexed by some index set $\mathcal{S}$.

We write $\mathbb{P}$ for probability, and $\mathbb{E}$ for expectation. If $\mathbf{X}$ takes at most countably many values, we define the *essential range* of $\mathbf{X}$ to be the set of all X such that $\mathbb{P}(\mathbf{X} = X)$ is non-zero, thus $\mathbf{X}$ almost surely takes values in its essential range. We also employ the following conditional expectation notation. If E is an event of non-zero probability, we write

$$\mathbb{P}(F|E) := \frac{\mathbb{P}(F \wedge E)}{\mathbb{P}(E)}$$

for any event F, and

$$\mathbb{E}(\mathbf{X}|E) := \frac{\mathbb{E}(\mathbf{X}1_E)}{\mathbb{P}(E)}$$

for any (absolutely integrable) real-valued random variable $\mathbf{X}$. If $\mathbf{Y}$ is another random variable taking at most countably many values, we define the conditional probability $\mathbb{P}(F|\mathbf{Y})$ to be the random variable that equals $\mathbb{P}(F|\mathbf{Y} = Y)$ on the event $\mathbf{Y} = Y$ for each Y in the essential range of $\mathbf{Y}$, and similarly define the conditional expectation $\mathbb{E}(\mathbf{X}|\mathbf{Y})$ to be the random variable that equals $\mathbb{E}(\mathbf{X}|\mathbf{Y} = Y)$ on the event $\mathbf{Y} = Y$. We observe the idempotency property

$$\mathbb{E}(\mathbb{E}(\mathbf{X}|\mathbf{Y})) = \mathbb{E}\mathbf{X} \tag{2.1}$$

whenever $\mathbf{X}$ is absolutely integrable and $\mathbf{Y}$ takes at most countably many values.

We will make frequent use of the basic inequalities of Markov

$$\mathbb{P}(X \geqslant \lambda) \leqslant \frac{\mu}{\lambda}, \qquad \mu = \mathbb{E}X > 0, \lambda > 0, \tag{2.2}$$

and Chebyshev

$$\mathbb{P}\left(|X-\mu| \geqslant \lambda\sqrt{\mathbb{E}|X-\mu|^2}\right) \leqslant \frac{1}{\lambda^2}, \qquad \lambda > 0, \mu = \mathbb{E}X \in \mathbb{R}, \mathbb{E}|X-\mu|^2 > 0. \tag{2.3}$$

The latter implies, when the variance $\mathbb{E}|X-\mu|^2$ is small, that a random variable is highly concentrated.

Lemma 2.1. *Suppose that for some $A > 0$ and $0 < \varepsilon < 1$ we have*

$$\mu = \mathbb{E}X = A(1 + O_{\leqslant}(\varepsilon)), \quad \mathbb{E}X^2 = A^2(1 + O_{\leqslant}(\varepsilon)).$$

Then, for any $\delta > \varepsilon$ we have

$$\mathbb{P}(|X - A| \geqslant \delta A) \leqslant \frac{4\varepsilon}{(\delta-\varepsilon)^2}.$$

Proof. We first derive an upper bound on the variance

$$\mathbb{E}|X-\mu|^2 = \mathbb{E}X^2 - \mu^2 = A^2 O_{\leqslant}(\varepsilon + 2\varepsilon + \varepsilon^2) \leqslant 4\varepsilon A^2.$$

Then, using (2.3), we obtain

$$\begin{aligned}\mathbb{P}(|X-A| \geqslant \delta A) &\leqslant \mathbb{P}\left(|X-\mu| \geqslant (\delta-\varepsilon)A\right) \\ &\leqslant \mathbb{P}\left(|X-\mu| \geqslant \frac{\delta-\varepsilon}{2\sqrt{\varepsilon}}\sqrt{\mathbb{E}|X-\mu|^2}\right) \leqslant \frac{4\varepsilon}{(\delta-\varepsilon)^2}.\end{aligned}$$

□

We also require Hoeffding's inequality (see e.g. [16, Theorem 7.20]).

Lemma 2.2. *Let $X_1, \dots, X_m$ be independent random variables with $\mathbb{E}X_i = 0$ and $|X_i| \leqslant B_i$ almost surely for each i. Then, for any real $t > 0$,*

$$\mathbb{P}\left(|X_1 + \cdots + X_m| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{2(B_1^2 + \cdots + B_m^2)}\right).$$

3. Sieving a set of primes

We begin by using a variant of the Westzynthius-Erdős-Rankin method to reduce this problem to a problem of sieving a set $\mathcal{Q}$ of *primes* in $[y]\backslash[x]$, rather than integers in $[y]\backslash[x]$.

Given a large real number x, define

$$y := cx\frac{\log x \log_3 x}{\log_2 x}, \tag{3.1}$$

where c is a certain (small) fixed positive constant. Also let

$$z := x^{\log_3 x/(4\log_2 x)}, \tag{3.2}$$

and introduce the three disjoint sets of primes

$$\mathcal{S} := \{s \text{ prime} : \log^{20} x < s \leqslant z\}, \tag{3.3}$$

$$\mathcal{P} := \{p \text{ prime} : x/2 < p \leqslant x\}, \tag{3.4}$$

$$\mathcal{Q} := \{q \text{ prime} : x < q \leqslant y\}. \tag{3.5}$$

For residue classes $\vec{a} = (a_s \bmod s)_{s\in\mathcal{S}}$ and $\vec{b} = (b_p \bmod p)_{p\in\mathcal{P}}$, define the sifted sets

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

and likewise

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

We then have

Theorem 2 (Sieving primes)**.** *Let x be sufficiently large and suppose that y obeys* (3.1)*. Then there are vectors $\vec{a} = (a_s \bmod s)_{s\in\mathcal{S}}$ and $\vec{b} = (b_p \bmod p)_{p\in\mathcal{P}}$, such that*

$$\#(\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{b})) \ll \frac{x}{\log x}. \tag{3.6}$$

We prove Theorem 2 in subsequent sections. Here, we show how this theorem implies (1.1), and hence Theorem 1.

Let $\vec{a}$ and $\vec{b}$ be as in Theorem 2. We extend the tuple $\vec{a}$ to a tuple $(a_p)_{p\leqslant x}$ of congruence classes $a_p \bmod p$ for all primes $p \leqslant x$ by setting $a_p := b_p$ for $p \in \mathcal{P}$ and $a_p := 0$ for $p \notin \mathcal{S} \cup \mathcal{P}$, and consider the sifted set

$$\mathcal{T} := \{n \in [y]\backslash[x] : n \not\equiv a_p \pmod p \text{ for all } p \leqslant x\}.$$

The elements of $\mathcal{T}$, by construction, are not divisible by any prime in $(0, \log^{20} x]$ or in $(z, x/2]$. Thus, each element must either be a z-smooth number (i.e., a number with all prime factors at most z), or must consist of a prime greater than $x/2$, possibly multiplied by some additional primes that are all at least $\log^{20} x$. However, from (3.1) we know that $y = o(x \log x)$. Thus, we see that an element of $\mathcal{T}$ is either a z-smooth number or a prime in $\mathcal{Q}$. In the second case, the element lies in $\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{b})$. Conversely, every element of $\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{b})$ lies in $\mathcal{T}$. Thus, $\mathcal{T}$ only differs from $\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{b})$ by a set $\mathcal{R}$ consisting of z-smooth numbers in $[y]$.

To estimate $\#\mathcal{R}$, let

$$u := \frac{\log y}{\log z},$$

so from (3.1), (3.2) one has $u \sim 4\frac{\log_2 x}{\log_3 x}$. By standard counts for smooth numbers (e.g. de Bruijn's theorem [6]) and (3.1), we thus have

$$\#\mathcal{R} \ll y e^{-u\log u + O(u\log\log(u+2))} = \frac{y}{\log^{4+o(1)} x} = o\left(\frac{x}{\log x}\right).$$

Thus, we find that $\#\mathcal{T} \ll x/\log x$.

Next, let C be a sufficiently large constant such that $\#\mathcal{T}$ is less than the number of primes in $(x, Cx]$. By matching each of these surviving elements to a distinct prime in $(x, Cx]$ and choosing congruence classes appropriately, we thus find congruence classes $a_p \bmod p$ for $p \leqslant Cx$ which cover *all* of the integers in $(x, y]$. In the language of Definition 1, we thus have

$$Y(Cx) \geqslant y - x + 1,$$

and (1.1) follows from (3.1).

Remark 1. One can replace the appeal to de Bruijn's theorem here by the simpler bounds of Rankin [40, Lemma II], if one makes the very minor change of increasing the 4 in the denominator of (3.2) to 5, and also makes similar numerical changes to later parts of the argument.

It remains to establish Theorem 2. This is the objective of the remaining sections of the paper.

4. Efficient hypergraph covering

In the previous section we reduced matters to obtaining residue classes $\vec{a}$, $\vec{b}$ such that the sifted set $\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{b})$ is small. In this section we use a hypergraph covering theorem, generalizing a result of Pippenger and Spencer [37], to reduce the task to that of finding residue classes $\vec{b}$ that have large intersection with $\mathcal{Q} \cap S(\vec{a})$.

4.1. **Heuristic discussion.** Consider the following general combinatorial problem. Let $(V, E_i)_{i \in I}$ be a collection of (non-empty) hypergraphs on a fixed finite vertex set V indexed by some finite index set I. In other words, V and I are finite sets, and for each $i \in I$, E_i is a (non-empty) collection of subsets of V. The problem is then to select a single edge e_i from each set E_i in such a way that the union $\bigcup_{i \in I} e_i$ covers as much of the vertex set V as possible. (In the context considered in [37], one considers choosing many edges from a single hypergraph (V, E), which in our context would correspond to the special case when (V, E_i) was independent of i.) One should think of the set $V \setminus \bigcup_{i \in I} e_i$ as a sifted version of V, with each e_i representing one step of the sieve.

One simple way to make this selection is a random one: one chooses a random edge $\mathbf{e}_i$ uniformly at random from E_i, independently in i. In that case, the probability that a given vertex $v \in V$ survives the sifting (that is, it avoids the random union $\bigcup_{i \in I} \mathbf{e}_i$) is equal to

$$\prod_{i \in I} (1 - \mathbb{P}(v \in \mathbf{e}_i)).$$

In applications, the index set I is large and the probabilities $\mathbb{P}(v \in \mathbf{e}_i)$ are small, in which case the above expression may be well approximated by

$$\exp(-d_I(v))$$

where we define the *normalized degree* $d_I(v)$ of v to be the quantity

$$d_I(v) := \sum_{i \in I} \mathbb{P}(v \in \mathbf{e}_i).$$

If we make the informal uniformity assumption

(i) One has $d_I(v) \approx d$ for all (or almost all) vertices v,

we thus expect the sifted set $V \setminus \bigcup_{i \in I} \mathbf{e}_i$ to have density approximately $\exp(-d)$.

Can one do better than this? Choosing the $\mathbf{e}_i$ independently is somewhat inefficient because it allows different random edges $\mathbf{e}_i, \mathbf{e}_j$ to collide with each other. If we could somehow modify the coupling between the $\mathbf{e}_i$ so that they were always disjoint, then the probability that a given vertex $v \in V$ survives the sieve would now become

$$1 - \sum_{i \in I} \mathbb{P}(v \in \mathbf{e}_i) = 1 - d_I(v).$$

This suggests that one could in principle lower the density of the sifted set from $\exp(-d)$ to $1 - d$ (or $\max(1-d, 0)$, since the density clearly cannot be negative), and in particular to sift out V almost completely as soon as d exceeds 1.

Suppose for the moment that such an optimal level of sieve efficiency is possible, and return briefly to consideration of Theorem 2. We set the vertex set V equal to $\mathcal{Q} \cap S(\vec{a})$ for some suitable choice of $\vec{a}$. If we set

$$y := cx \frac{\log x \log_3 x}{\log_2 x}$$

for some small $c > 0$ (in accordance with (3.1)), then standard probabilistic heuristics (together with Mertens' theorem and (3.1), (3.3)) suggest that V should have cardinality about

$$\frac{y}{\log x} \times \prod_{s \in \mathcal{S}} \left(1 - \frac{1}{s}\right) \approx c \frac{x}{\log x} \log_2 x,$$

so in particular this set is roughly $c \log_2 x$ times larger than $\mathcal{P}$. In later sections, we will use the multidimensional sieve from [35], [34] to locate for most primes p in $\mathcal{P}$, a large number of residue classes $b_p \bmod p$ that each intersect $\mathcal{Q} \cap S(\vec{a})$ in roughly $\asymp \log_2 x$ elements on the average. If we let E_p be the set of all such intersections $(b_p \bmod p) \cap V$, then the task of making $\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{b})$ small is essentially the same as making the sifted set $V \backslash \bigcup_{p \in \mathcal{P}} e_p$ small, for some suitable edges e_p drawn from E_p. By double counting, the expected density d here should be roughly

$$d \asymp \frac{\#\mathcal{P} \times \log_2 x}{\#V} \asymp \frac{1}{c},$$

and so one should be able to sieve out $\mathcal{Q} \cap S(\vec{a})$ more or less completely once c is small enough if one had optimal sieving. In contrast, if one used independent sieving, one would expect the cardinality of $\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{b})$ to be something like $\exp(-1/c) \times c \frac{x}{\log x} \log_2 x$, which would only be acceptable if c was slightly smaller than $\frac{1}{\log_3 x}$. This loss of $\log_3 x$ ultimately leads to the loss of $\log_4 X$ in (1.4) as compared against Theorem 1.

It is thus desirable to obtain a general combinatorial tool for achieving near-optimal sieve efficiency for various collections $(V, E_i)_{i \in I}$ of hypergraphs. The result of Pippenger and Spencer [37] (extending previous results of Rödl [42] and Frankl and Rödl [17], as well as unpublished work of Pippenger) asserts, very roughly speaking, that one can almost attain this optimal efficiency under some further assumptions beyond (i), which we state informally as follows:

(ii) The hypergraphs (V, E_i) do not depend on i.
(iii) The *normalized codegrees* $\sum_{i \in I} \mathbb{P}(v, w \in \mathbf{e}_i)$ for $v \neq w$ are small.
(iv) The edges e_i of E_i are of *constant* size, thus there is a k such that $\#e_i = k$ for all i and all $e_i \in E_i$.

The argument is based on the *Rödl nibble* from [42], which is a variant of the *semi-random method* from [1]. Roughly speaking, the idea is to break up the index set I into smaller pieces $I_1, \ldots, I_m$. For the first I_1, we perform a "nibble" by selecting the $\mathbf{e}_i$ for $i \in I_1$ uniformly and independently. For the next nibble at I_2, we restrict (or condition) the $\mathbf{e}_i$ for $i \in I_2$ to avoid the edges arising in the first nibble, and *then* select $\mathbf{e}_i$ for $i \in I_2$ independently at random using this conditioned distribution. We continue performing nibbles at $I_3, \ldots, I_m$ (restricting the edges at each nibble to be disjoint from the edges of previous nibbles) until the index set I is exhausted. Intuitively, this procedure enjoys better disjointness properties than the completely independent selection scheme, but it is harder to analyze the probability of success. To achieve the latter task, Pippenger and Spencer rely heavily on the four hypotheses (i)-(iv).

In our context, hypothesis (iii) is easily satisfied, and (i) can also be established. Hypothesis (ii) is not satisfied (the E_p vary in p), but it turns out that the argument of Pippenger and Spencer can easily be written in such a way that this hypothesis may be discarded. But it is the failure of hypothesis (iv) which is the most severe difficulty: the size of the sets $e_p = (b_p \bmod p) \cap V$ can fluctuate quite widely for different choices of p or b_p. This creates an undesirable bias in the iterative nibbling process: with each nibble, larger edges $\mathbf{e}_i$ have a reduced chance of survival compared with smaller edges, simply because they have more elements that could potentially intersect previous nibbles. Given that one expects the larger edges to be the most useful for the purposes of efficient sieving, this bias is a significant problem. One could try to rectify

the issue by partitioning the edge sets E_i depending on the cardinality of the edges, and working on one partition at a time, but this seriously impacts hypothesis (i) in a manner that we were not able to handle.

Our resolution to this problem is to modify the iterative step of the nibbling process by *reweighting* the probability distribution of the $\mathbf{e}_i$ at each step to cancel out the bias incurred by conditioning an edge $\mathbf{e}_i$ to be disjoint from previous nibbles. It turns out that there is a natural choice of reweighting for this task even when the normalized degrees $d_I(v)$ vary in v. As a consequence, we can obtain a version of the Pippenger-Spencer theorem in which hypothesis (ii) is essentially eliminated and (i), (iv) significantly weakened, leaving only (iii) as the main hypothesis. We remark that a somewhat similar relaxation of hypotheses (i)-(iv) was obtained by Kahn in [29], although the statement in [29] is not exactly in a form convenient for our applications here.

4.2. **Statement of covering theorem.** We now rigorously state the hypergraph covering theorem that we will use. In order to apply this theorem for our application, we will need a probabilistic formulation of this theorem which does not, at first glance, bear much resemblance to the combinatorial formulation appearing in [37]; we will discuss the connections between these formulations shortly. We will also phrase the theorem in a completely quantitative fashion, avoiding the use of asymptotic notation; this will be convenient for the purposes of proving the theorem via induction (on the number m of "nibbles").

Theorem 3 (Probabilistic covering)**.** *There exists a constant $C_0 \geqslant 1$ such that the following holds. Let $D, r, A \geqslant 1$, $0 < \kappa \leqslant 1/2$, and let $m \geqslant 0$ be an integer. Let $\delta > 0$ satisfy the smallness bound*

$$\delta \leqslant \left(\frac{\kappa^A}{C_0 \exp(AD)}\right)^{10^{m+2}}. \tag{4.1}$$

Let $I_1, \dots, I_m$ be disjoint finite non-empty sets, and let V be a finite set. For each $1 \leqslant j \leqslant m$ and $i \in I_j$, let $\mathbf{e}_i$ be a random finite subset of V. Assume the following:

- *(Edges not too large) With probability 1, we have for all $j = 1, \dots, m$ and all $i \in I_j$*

$$\#\mathbf{e}_i \leqslant r; \tag{4.2}$$

- *(Each sieve step is sparse) For all $j = 1, \dots, m$, $i \in I_j$ and $v \in V$,*

$$\mathbb{P}(v \in \mathbf{e}_i) \leqslant \frac{\delta}{(\#I_j)^{1/2}}; \tag{4.3}$$

- *(Very small codegrees) For every $j = 1, \dots, m$, and distinct $v_1, v_2 \in V$,*

$$\sum_{i \in I_j} \mathbb{P}(v_1, v_2 \in \mathbf{e}_i) \leqslant \delta \tag{4.4}$$

- *(Degree bound) If for every $v \in V$ and $j = 1, \dots, m$ we introduce the normalized degrees*

$$d_{I_j}(v) := \sum_{i \in I_j} \mathbb{P}(v \in \mathbf{e}_i) \tag{4.5}$$

and then recursively define the quantities $P_j(v)$ for $j = 0, \dots, m$ and $v \in V$ by setting

$$P_0(v) := 1 \tag{4.6}$$

and

$$P_{j+1}(v) := P_j(v) \exp(-d_{I_{j+1}}(v)/P_j(v)) \tag{4.7}$$

for $j = 0, \dots, m-1$ and $v \in V$, then we have

$$d_{I_j}(v) \leqslant DP_{j-1}(v) \qquad (1 \leqslant j \leqslant m, v \in V) \tag{4.8}$$

and

$$P_j(v) \geqslant \kappa \qquad (0 \leqslant j \leqslant m, v \in V). \tag{4.9}$$

Then we can find random variables $\mathbf{e}'_i$ for each $i \in \bigcup_{j=1}^m I_j$ with the following properties:

(a) *For each $i \in \bigcup_{j=1}^m I_j$, the essential support of $\mathbf{e}'_i$ is contained in the essential support of $\mathbf{e}_i$, union the empty set singleton $\{\emptyset\}$. In other words, almost surely $\mathbf{e}'_i$ is either empty, or is a set that $\mathbf{e}_i$ also attains with positive probability.*

(b) *For any $0 \leqslant J \leqslant m$ and any finite subset e of V with $\#e \leqslant A - 2rJ$, one has*

$$\mathbb{P}\left(e \subset V \backslash \bigcup_{j=1}^{J} \bigcup_{i \in I_j} \mathbf{e}'_i\right) = \left(1 + O_{\leqslant}(\delta^{1/10^{J+1}})\right) P_J(e) \tag{4.10}$$

where

$$P_j(e) := \prod_{v \in e} P_j(v). \tag{4.11}$$

We prove this theorem in Section 5. It is likely that the smallness condition (4.1) can be relaxed, for instance by modifying the techniques from [45]. However, this would not lead to any significant improvement in the final bound on $G(X)$ in Theorem 1, as in our application the condition (4.1) is already satisfied with some room to spare. The parameter r does not appear explicitly in the smallness requirement (4.1), but is implicit in that requirement since the conclusion is trivially true unless $2r < A$.

One may deduce special cases of this theorem which are close to the original hypergraph covering lemma of Pippenger and Spencer. These were included in an earlier draft of this paper (as Corollaries 2 and 3), and will now be described in a future, separate paper.

4.3. **Applying the covering theorem.** We now specialize Theorem 3 to a situation relevant for the application to large prime gaps, given by Corollary 4. Very roughly, this states that if we have a large collection $\{\mathbf{e}_p : p \in \mathcal{P}'\}$ of random subsets of a set $\mathcal{Q}'$, then there is a realization of these random variables which covers almost all of $\mathcal{Q}'$, provided the subsets are suitably 'uniform' and all elements of $\mathcal{Q}$ are covered more than once on average. Clearly we cannot hope to cover $\mathcal{Q}$ unless almost all elements are covered at least once on average, and similarly if the subsets are too highly correlated then one can easily produce examples where $\mathcal{Q}'$ will not be covered. Thus some form of both of these assumptions is necessary. For our application to prime gaps, the random sets $\mathbf{e}_p$ will be all elements of $\mathcal{Q}'$ in a randomly chosen residue class $\bmod p$ (which will be produced by the multidimensional sieve in the later sections.)

Corollary 4. *Let $x \to \infty$. Let $\mathcal{P}'$, $\mathcal{Q}'$ be sets with $\#\mathcal{P}' \leqslant x$ and $(\log_2 x)^3 < \#\mathcal{Q}' \leqslant x^{100}$. For each $p \in \mathcal{P}'$, let $\mathbf{e}_p$ be a random subset of $\mathcal{Q}'$ satisfying the size bound*

$$\#\mathbf{e}_p \leqslant r = O\left(\frac{\log x \log_3 x}{\log_2^2 x}\right) \qquad (p \in \mathcal{P}'). \tag{4.12}$$

Assume the following:

- *(Sparsity) For all $p \in \mathcal{P}'$ and $q \in \mathcal{Q}'$,*

$$\mathbb{P}(q \in \mathbf{e}_p) \leqslant x^{-1/2-1/10}. \tag{4.13}$$

- *(Small codegrees) For any distinct $q_1, q_2 \in \mathcal{Q}'$,*

$$\sum_{p\in\mathcal{P}'} \mathbb{P}(q_1, q_2 \in \mathbf{e}_p) \leqslant x^{-1/20}. \tag{4.14}$$

- *(Elements covered more than once in expectation) For all but at most $\frac{1}{(\log_2 x)^2}\#\mathcal{Q}'$ elements $q \in \mathcal{Q}'$, we have*

$$\sum_{p\in\mathcal{P}'} \mathbb{P}(q \in \mathbf{e}_p) = C + O_{\leqslant}\left(\frac{1}{(\log_2 x)^2}\right) \tag{4.15}$$

for some quantity C, independent of q, satisfying

$$\frac{5}{4}\log 5 \leqslant C \ll 1. \tag{4.16}$$

Then for any positive integer m with

$$m \leqslant \frac{\log_3 x}{\log 5}, \tag{4.17}$$

we can find random sets $\mathbf{e}'_p \subseteq \mathcal{Q}'$ for each $p \in \mathcal{P}'$ such that $\mathbf{e}'_p$ is either empty or a subset of $\mathcal{Q}'$ which $\mathbf{e}_p$ attains with positive probability, and that

$$\#\{q \in \mathcal{Q}' : q \notin \mathbf{e}'_p \text{ for all } p \in \mathcal{P}'\} \sim 5^{-m}\#\mathcal{Q}'$$

with probability $1 - o(1)$. More generally, for any $\mathcal{Q}'' \subset \mathcal{Q}'$ with cardinality at least $(\#\mathcal{Q}')/\sqrt{\log_2 x}$, one has

$$\#\{q \in \mathcal{Q}'' : q \notin \mathbf{e}'_p \text{ for all } p \in \mathcal{P}'\} \sim 5^{-m}\#\mathcal{Q}''$$

with probability $1 - o(1)$. The decay rates in the $o(1)$ and $\sim$ notation are uniform in $\mathcal{P}'$, $\mathcal{Q}'$, $\mathcal{Q}''$.

Remarks. For the arguments in this paper, we only need the case $\mathcal{Q}'' = \mathcal{Q}'$, but the more general situation $\mathcal{Q}'' \subset \mathcal{Q}'$ will be of use in the sequel [15] of this paper when we consider chains of large gaps.

From (4.13) and (4.15), it follows that $\#\mathcal{P}' \gg x^{1/2+1/10}$.

Proof. It suffices to establish the claim for x sufficiently large, as the claim is trivial for bounded x. The number of exceptional elements q of $\mathcal{Q}'$ that fail (4.15) is $o(5^{-m}\#\mathcal{Q}'')$, thanks to (4.17). Thus we may discard these elements from $\mathcal{Q}'$ and assume that (4.15) holds for *all* $q \in \mathcal{Q}'$, and deduce the conclusions of the corollary with the modified set $\mathcal{Q}'$.

By (4.16), we may find disjoint intervals $\mathscr{I}_1, \dots, \mathscr{I}_m$ in $[0, 1]$ with length

$$|\mathscr{I}_j| = \frac{5^{1-j}\log 5}{C} \tag{4.18}$$

for $j = 1, \dots, m$. Let $\vec{\mathbf{t}} = (\mathbf{t}_p)_{p\in\mathcal{P}'}$ be a tuple of elements $\mathbf{t}_p$ of $[0, 1]$ drawn uniformly and independently at random for each $p \in \mathcal{P}'$ (independently of the $\mathbf{e}_p$), and define the random sets

$$I_j = I_j(\vec{\mathbf{t}}) := \{p \in \mathcal{P}' : \mathbf{t}_p \in \mathscr{I}_j\}$$

for $j = 1, \dots, m$. These sets are clearly disjoint.

We will verify (for a suitable choice of $\vec{\mathbf{t}}$) the hypotheses of Theorem 3 with the indicated sets I_j and random variables $\mathbf{e}_p$, and with suitable choices of parameters $D, r, A \geqslant 1$ and $0 < \kappa \leqslant 1/2$, and $V = \mathcal{Q}'$.

Set

$$\delta := x^{-1/20} \tag{4.19}$$

and observe from (4.13) and $\#\mathcal{P}' \leqslant x$ that one has

$$\mathbb{P}(q \in \mathbf{e}_p) \leqslant \frac{\delta}{(\#I_j)^{1/2}} \tag{4.20}$$

for all $j = 1, \dots, m$, $p \in I_j$, and $q \in \mathcal{Q}'$. Clearly the small codegree condition (4.14) implies that

$$\sum_{p \in I_j} \mathbb{P}(q_1, q_2 \in \mathbf{e}_p) \leqslant \delta \qquad (1 \leqslant j \leqslant m). \tag{4.21}$$

Let $q \in \mathcal{Q}'$, $1 \leqslant j \leqslant m$ and consider the independent random variables $(\mathbf{X}_p^{(q,j)}(\vec{\mathbf{t}}))_{p \in \mathcal{P}'}$, where

$$\mathbf{X}_p^{(q,j)}(\vec{\mathbf{t}}) = \begin{cases} \mathbb{P}(q \in \mathbf{e}_p) & \text{if } p \in I_j \\ 0 & \text{otherwise.} \end{cases}$$

By (4.15), (4.16) and (4.18), for every j and every $q \in \mathcal{Q}'$,

$$\sum_{p \in \mathcal{P}'} \mathbb{E}\mathbf{X}_p^{(q,j)}(\vec{\mathbf{t}}) = \sum_{p \in \mathcal{P}'} \mathbb{P}(q \in \mathbf{e}_p)\mathbb{P}(p \in I_j(\vec{\mathbf{t}})) = |\mathscr{I}_j| \sum_{p \in \mathcal{P}'} \mathbb{P}(q \in \mathbf{e}_p) = 5^{1-j} \log 5 + O_{\leqslant}\left(\frac{4/5}{(\log_2 x)^2}\right).$$

By (4.13), we have $|\mathbf{X}_p^{(q,j)}(\vec{\mathbf{t}})| \leqslant x^{-1/2-1/10}$ for all p, and hence by Hoeffding's inequality (Lemma 2.2),

$$\begin{aligned} \mathbb{P}\left(\Big|\sum_{p \in \mathcal{P}'} (\mathbf{X}_p^{(q,j)}(\vec{\mathbf{t}}) - \mathbb{E}\mathbf{X}_p^{(q,j)}(\vec{\mathbf{t}}))\Big| \geqslant \frac{1}{(\log_2 x)^2}\right) &\leqslant 2\exp\left\{-\frac{(\log_2 x)^{-4}}{2x^{-1-1/5}\#I_j}\right\} \\ &\leqslant 2\exp\left\{-\frac{x^{1/5}}{(\log_2 x)^4}\right\} \ll \frac{1}{x^{200}}. \end{aligned}$$

By the upper bound on $\#\mathcal{Q}'$, there is a deterministic choice $\vec{t}$ of $\vec{\mathbf{t}}$ (and hence $I_1, \dots, I_m$) such that for *every* $q \in \mathcal{Q}'$ and *every* $j = 1, \dots, m$, we have

$$\Big|\sum_{p \in \mathcal{P}'} (\mathbf{X}_p^{(q,j)}(\vec{t}) - \mathbb{E}\mathbf{X}_p^{(q,j)}(\vec{\mathbf{t}}))\Big| < \frac{1}{(\log_2 x)^2}.$$

We fix this choice $\vec{t}$ (so that the I_j are now deterministic), and we conclude that

$$\sum_{p \in \mathcal{P}'} \mathbf{X}_p^{(q,j)}(\vec{t}) = \sum_{p \in I_j} \mathbb{P}(q \in \mathbf{e}_p) = 5^{1-j} \log 5 + O_{\leqslant}\left(\frac{2}{(\log_2 x)^2}\right) \tag{4.22}$$

uniformly for all $j = 1, \dots, m$, and all $q \in \mathcal{Q}'$.

Inserting (4.22) into the definition (4.5) of $d_{I_j}(q)$ and using the bound (4.17) on m, we now have

$$d_{I_j}(q) = (1 + O_{\leqslant}(2/\log_2 x))5^{-j+1} \log 5$$

for all $q \in \mathcal{Q}'$ and $1 \leqslant j \leqslant m$. A routine induction using (4.6), (4.7) then shows (for x sufficiently large) that

$$P_j(q) = (1 + O_{\leqslant}(4^j/\log_2 x))5^{-j} \quad (0 \leqslant j \leqslant m),$$

and hence that

$$P_j(q) = 5^{-j}(1 + O_{\leqslant}((\log_2 x)^{-\nu})) \qquad (0 \leqslant j \leqslant m), \tag{4.23}$$

where $\nu = \log(5/4)/\log 5$. In particular we have

$$d_{I_j}(q) \leqslant DP_{j-1}(q) \qquad (1 \leqslant j \leqslant m)$$

for some $D = O(1)$, and

$$P_j(q) \geqslant \kappa \qquad (1 \leqslant j \leqslant m),$$

where

$$\kappa \gg 5^{-m}.$$

We now set

$$A := 2rm + 2.$$

By our bounds on m (4.17) and r (4.12),

$$A \ll \frac{\log x \log_3^2 x}{\log_2^2 x},$$

and so

$$\frac{\kappa^A}{C_0 \exp(AD)} \geqslant \exp\left(-O\left(\frac{\log x \log_3^3 x}{\log_2^2 x}\right)\right).$$

By (4.17) and (4.19), we see that

$$\delta^{1/10^{m+2}} \leqslant \exp\left(-\frac{\log x}{2000(\log_2 x)^{\log 10/\log 5}}\right),$$

and so (4.1) is satisfied x is large enough (note that $\log 10/\log 5 < 2$). Thus all the hypotheses of Theorem 3 have been verified for this choice of parameters (note that A, κ and D are independent of $\mathcal{P}'$, $\mathcal{Q}'$).

Applying Theorem 3 (with $V = \mathcal{Q}'$) and using (4.23), one thus obtains random variables $\mathbf{e}'_p$ for $p \in \bigcup_{j=1}^m I_j$ whose essential range is contained in the essential range of $\mathbf{e}_p$ together with $\emptyset$, such that

$$\mathbb{P}\left(q \notin \bigcup_{j=1}^m \bigcup_{p \in I_j} \mathbf{e}'_p\right) = 5^{-m}\left(1 + O((\log_2 x)^{-\nu})\right) \tag{4.24}$$

for all $q \in \mathcal{Q}'$, and

$$\mathbb{P}\left(q_1, q_2 \notin \bigcup_{j=1}^m \bigcup_{p \in I_j} \mathbf{e}'_p\right) = 5^{-2m}\left(1 + O((\log_2 x)^{-\nu})\right) \tag{4.25}$$

for all distinct $q_1, q_2 \in \mathcal{Q}'$.

Set $\mathbf{e}'_p = \emptyset$ for $p \in \mathcal{P}' \backslash \bigcup_{j=1}^m I_j$. Let $\mathcal{Q}''$ be as in the corollary, and consider the random variable

$$\mathbf{Y} := \#\{q \in \mathcal{Q}'' : q \notin \mathbf{e}'_p \text{ for all } p \in \mathcal{P}'\} = \sum_{q \in \mathcal{Q}''} 1_{q \notin \bigcup_{j=1}^m \bigcup_{p \in I_j} \mathbf{e}'_p}.$$

Using (4.24) and (4.25), we obtain

$$\mathbb{E}\mathbf{Y} = 5^{-m}\left(1 + O((\log_2 x)^{-\nu})\right) \#\mathcal{Q}''$$

and

$$\mathbb{E}\mathbf{Y}^2 = 5^{-2m}\left(1 + O((\log_2 x)^{-\nu})\right) (\#\mathcal{Q}'')^2 + O(5^{-m}\#\mathcal{Q}'') = 5^{-2m}\left(1 + O((\log_2 x)^{-\nu})\right) (\#\mathcal{Q}'')^2,$$

(here we use (4.17) and the mild bound $\#\mathcal{Q}'' > (\log_2 x)^2$), and so from Lemma 2.1 we have

$$\mathbf{Y} \sim 5^{-m}\#\mathcal{Q}''$$

with probability $1 - o(1)$, as required. □

In view of the above corollary, we may now reduce Theorem 2 to the following claim.

Theorem 4 (Random construction)**.** *Let x be a sufficiently large real number and define y by* (3.1)*. Then there is a quantity C with*

$$C \asymp \frac{1}{c} \tag{4.26}$$

with the implied constants independent of c, a tuple of positive integers $(h_1, \dots, h_r)$ with $r \leqslant \sqrt{\log x}$, and some way to choose random vectors $\vec{\mathbf{a}} = (\mathbf{a}_s \bmod s)_{s\in\mathcal{S}}$ and $\vec{\mathbf{n}} = (\mathbf{n}_p)_{p\in\mathcal{P}}$ of congruence classes $\mathbf{a}_s \bmod s$ and integers $\mathbf{n}_p$ respectively, obeying the following:

- *For every $\vec{a}$ in the essential range of $\vec{\mathbf{a}}$, one has*

$$\mathbb{P}(q \in \mathbf{e}_p(\vec{a})|\vec{\mathbf{a}} = \vec{a}) \leqslant x^{-1/2-1/10} \quad (p \in \mathcal{P}), \tag{4.27}$$

  where $\mathbf{e}_p(\vec{a}) := \{\mathbf{n}_p + h_i p : 1 \leqslant i \leqslant r\} \cap \mathcal{Q} \cap S(\vec{a})$.
- *With probability $1 - o(1)$ we have that*

$$\#(\mathcal{Q} \cap S(\vec{\mathbf{a}})) \sim 80c\frac{x}{\log x}\log_2 x. \tag{4.28}$$

- *Call an element $\vec{a}$ in the essential range of $\vec{\mathbf{a}}$* good *if, for all but at most $\frac{x}{\log x \log_2 x}$ elements $q \in \mathcal{Q} \cap S(\vec{a})$, one has*

$$\sum_{p\in\mathcal{P}} \mathbb{P}(q \in \mathbf{e}_p(\vec{a})|\vec{\mathbf{a}} = \vec{a}) = C + O_{\leqslant}\left(\frac{1}{(\log_2 x)^2}\right). \tag{4.29}$$

  Then $\vec{\mathbf{a}}$ is good with probability $1 - o(1)$.

We now show why Theorem 4 implies Theorem 2. By (4.26), we may choose $0 < c < 1/2$ small enough so that (4.16) holds. Take

$$m = \left\lfloor \frac{\log_3 x}{\log 5} \right\rfloor.$$

Now let $\vec{\mathbf{a}}$ and $\vec{\mathbf{n}}$ be the random vectors guaranteed by Theorem 4. Suppose that we are in the probability $1 - o(1)$ event that $\vec{\mathbf{a}}$ takes a value $\vec{a}$ which is good and such that (4.28) holds. Fix some $\vec{a}$ within this event. We may apply Corollary 4 with $\mathcal{P}' = \mathcal{P}$ and $\mathcal{Q}' = \mathcal{Q} \cap S(\vec{a})$ for the random variables $\mathbf{n}_p$ conditioned to $\vec{\mathbf{a}} = \vec{a}$. A few hypotheses of the Corollary must be verified. First, (4.15) follows from (4.29). The small codegree condition (4.14) is also quickly checked. Indeed, for distinct $q_1, q_2 \in \mathcal{Q}'$, if $q_1, q_2 \in \mathbf{e}_p(\vec{a})$ then $p|q_1 - q_2$. But $q_1 - q_2$ is a nonzero integer of size at most $x \log x$, and is thus divisible by at most one prime $p_0 \in \mathcal{P}'$. Hence

$$\sum_{p\in\mathcal{P}'} \mathbb{P}(q_1, q_2 \in \mathbf{e}_p(\vec{a})) = \mathbb{P}(q_1, q_2 \in \mathbf{e}_{p_0}(\vec{a})) \leqslant x^{-1/2-1/10},$$

the sum on the left side being zero if p_0 doesn't exist. By Corollary 4, there exist random variables $\mathbf{e}'_p(\vec{a})$, whose essential range is contained in the essential range of $\mathbf{e}_p(\vec{a})$ together with $\emptyset$, and satisfying

$$\{q \in \mathcal{Q} \cap S(\vec{a}) : q \notin \mathbf{e}'_p(\vec{a}) \text{ for all } p \in \mathcal{P}\} \sim 5^{-m}\#(\mathcal{Q} \cap S(\vec{a})) \ll \frac{x}{\log x}$$

with probability $1-o(1)$, where we have used (4.28). Since $\mathbf{e}'_p(\vec{a}) = \{\mathbf{n}'_p + h_i p : 1 \leqslant i \leqslant r\} \cap \mathcal{Q} \cap S(\vec{a})$ for some random integer $\mathbf{n}'_p$, it follows that

$$\{q \in \mathcal{Q} \cap S(\vec{a}) : q \not\equiv \mathbf{n}'_p \pmod p \text{ for all } p \in \mathcal{P}\} \ll \frac{x}{\log x}$$

with probability $1-o(1)$. Taking a specific $\vec{\mathbf{n}}' = \vec{n}'$ for which this relation holds and setting $b_p = n'_p$ for all p concludes the proof of the claim (3.6) and establishes Theorem 2.

It remains to establish Theorem 4. This will be achieved in later sections.

5. Proof of the covering theorem

We now prove Theorem 3. Let C_0 be a sufficiently large absolute constant.

We induct on m. The case $m = 0$ is vacuous, so suppose that $m \geqslant 1$ and that the claim has already been proven for $m-1$. Let $D, r, A, \kappa, \delta, I_j, \mathbf{e}_i, V$ be as in the theorem. By the induction hypothesis, we can already find random variables $\mathbf{e}'_i$ for $i \in \bigcup_{j=1}^{m-1} I_j$ obeying the conclusions (a), (b) of the theorem for $m-1$. In particular, we may form the partially sifted set

$$\mathbf{W} := V \backslash \bigcup_{j=1}^{m-1} \bigcup_{i \in I_j} \mathbf{e}'_i,$$

and we have

$$\mathbb{P}(e \subset \mathbf{W}) = (1 + O_{\leqslant}(\delta^{1/10^m}))P_{m-1}(e) \tag{5.1}$$

whenever $e \subset V$ has cardinality $\#e \leqslant A - 2r(m-1)$.

Our task is then to construct random variables $\mathbf{e}'_i$ for $i \in I_m$, possibly coupled with existing random variables such as $\mathbf{W}$, whose essential range is contained in that of $\mathbf{e}_i$ together with the empty set, and such that

$$\mathbb{P}\left(e \subset \mathbf{W} \backslash \bigcup_{i \in I_m} \mathbf{e}'_i\right) = \left(1 + O_{\leqslant}(\delta^{1/10^{m+1}})\right) P_m(e) \tag{5.2}$$

for all finite subsets e of V with $\#e \leqslant A - 2rm$. Note that we may assume that $A > 2rm$, as the claim (4.10) is trivial otherwise. In particular we have

$$A - 2r(m-1) > 2r. \tag{5.3}$$

From (4.9), (4.11) we note that

$$P_j(\tilde{e}) \geqslant \kappa^{\#\tilde{e}} \tag{5.4}$$

whenever $j = 1, \dots, m$ and all $\tilde{e} \subset V$. In particular, by (5.4) and (4.2), whenever $\tilde{e}_i$ is in the essential range of $\mathbf{e}_i$, we have

$$P_j(\tilde{e}_i) \geqslant \kappa^r. \tag{5.5}$$

For future reference, we observe that from (5.3) and (4.1), we have

$$r\kappa^{-r} \leqslant A\kappa^{-r} \leqslant A^2\kappa^{-2r} \leqslant A^2 D\kappa^{-A} \leqslant \delta^{-1/10^{m+2}}. \tag{5.6}$$

For each $i \in I_m$, and every W in the essential range of $\mathbf{W}$, define the normalization factor

$$X_i(W) := \mathbb{E}\left(\frac{1_{\mathbf{e}_i \subset W}}{P_{m-1}(\mathbf{e}_i)}\right) = \sum_{\tilde{e}_i \subset W} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)}{P_{m-1}(\tilde{e}_i)}. \tag{5.7}$$

We will see shortly, and this is crucial to our argument, that $X_i(\mathbf{W})$ concentrates to 1. With this in mind, we let $F_i = F_i(\mathbf{W})$ be the event that

$$|X_i(\mathbf{W}) - 1| \leqslant \delta^{\frac{1}{3\times 10^m}}. \tag{5.8}$$

Very small values of $X_i(W)$, in particular sets W with $X_i(W) = 0$, are problematic for us and must be avoided. Fortunately, this occurs with very small probability.

We now define the random variables $\mathbf{e}'_i$ for $i \in I_m$. If $F_i(\mathbf{W})$ fails, we set $\mathbf{e}'_i = \emptyset$. Otherwise, if $F_i(\mathbf{W})$ holds, then after conditioning on a fixed value W of $\mathbf{W}$, we choose $\mathbf{e}'_i$ from the essential range of $\mathbf{e}_i$ using the conditional probability distribution

$$\mathbb{P}(\mathbf{e}'_i = \tilde{e}_i | \mathbf{W} = W) := \frac{1_{\tilde{e}_i \subset W}}{X_i(W)} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)}{P_{m-1}(\tilde{e}_i)} \tag{5.9}$$

for all $\tilde{e}_i$ in the essential range of $\mathbf{e}_i$, and also require that the $\mathbf{e}'_i$ are conditionally jointly independent for $i \in I_m$ on each event $\mathbf{W} = W$. Note from (5.7) that (5.9) defines a probability distribution, and so the $\mathbf{e}'_i$ are well defined as random variables. Informally, $\mathbf{e}'_i$ is $\mathbf{e}_i$ conditioned to the event $\mathbf{e}_i \subset W$, and then reweighted by $P_{m-1}(\mathbf{e}_i)$ to compensate for the bias caused by this conditioning.

Lemma 5.1. *We have*

$$\mathbb{P}(F_i(\mathbf{W})) = 1 - O(\delta^{\frac{1}{3\times 10^m}}).$$

Proof. By Lemma 2.1, it suffices to show that

$$\mathbb{E}X_i(\mathbf{W}) = 1 + O(\delta^{\frac{1}{10^m}}) \tag{5.10}$$

and

$$\mathbb{E}(X_i(\mathbf{W})^2) = 1 + O(\delta^{\frac{1}{10^m}}). \tag{5.11}$$

We begin with (5.10). Let $\tilde{e}_i$ be in the essential range of $\mathbf{e}_i$. From (4.2) and (5.3) we have

$$\#\tilde{e}_i \leqslant r \leqslant A - 2r(m-1)$$

and thus by (5.7) and (5.1), we have

$$\begin{aligned}\mathbb{E}X_i(\mathbf{W}) &= \sum_W \mathbb{P}(\mathbf{W} = W) \sum_{\tilde{e}_i \subset W} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)}{P_{m-1}(\tilde{e}_i)} \\ &= \sum_{\tilde{e}_i} \mathbb{P}(\mathbf{e}_i = \tilde{e}_i) \frac{\mathbb{P}(\tilde{e}_i \subset \mathbf{W})}{P_{m-1}(\tilde{e}_i)} = 1 + O_{\leqslant}(\delta^{\frac{1}{10^m}}).\end{aligned}$$

Now we show (5.11). Let $\tilde{e}_i$ and $\hat{e}_i$ be in the essential range of $\mathbf{e}_i$. From (4.2), (5.3) we have

$$\#\tilde{e}_i \cup \hat{e}_i \leqslant A - 2r(m-1)$$

and from (4.11) we have

$$\frac{P_{m-1}(\tilde{e}_i \cup \hat{e}_i)}{P_{m-1}(\tilde{e}_i) P_{m-1}(\hat{e}_i)} = \frac{1}{P_{m-1}(\tilde{e}_i \cap \hat{e}_i)}$$

and thus by (5.7) and (5.1) we have

$$\mathbb{E}(X_i(\mathbf{W})^2) = \sum_{\tilde{e}_i,\hat{e}_i} \mathbb{P}(\mathbf{e}_i = \tilde{e}_i)\mathbb{P}(\mathbf{e}_i = \hat{e}_i)\frac{\mathbb{P}(\tilde{e}_i \cup \hat{e}_i \subset \mathbf{W})}{P_{m-1}(\tilde{e}_i)P_{m-1}(\hat{e}_i)}$$
$$= \Big(1 + O_{\leqslant}(\delta^{\frac{1}{10^m}})\Big) \sum_{\tilde{e}_i,\hat{e}_i} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)\mathbb{P}(\mathbf{e}_i = \hat{e}_i)}{P_{m-1}(\tilde{e}_i \cap \hat{e}_i)}.$$

The denominator $P_{m-1}(\tilde{e}_i \cap \hat{e}_i)$ is 1 if $\tilde{e}_i \cap \hat{e}_i = \emptyset$, and is at least κ^r otherwise, thanks to (5.5). Thus, by (4.2), (4.3) and a union bound,

$$\sum_{\tilde{e}_i,\hat{e}_i} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)\mathbb{P}(\mathbf{e}_i = \hat{e}_i)}{P_{m-1}(\tilde{e}_i \cap \hat{e}_i)} = 1 + O\left(\kappa^{-r} \sum_{\tilde{e}_i} \mathbb{P}(\mathbf{e}_i = \tilde{e}_i) \sum_{v \in \tilde{e}_i} \mathbb{P}(v \in \mathbf{e}_i)\right) = 1 + O(r\delta\kappa^{-r}),$$

and the claim (5.11) follows from (5.6). □

It remains to verify (5.2). Let e be a fixed subset of V with

$$\#e \leqslant A - 2rm. \tag{5.12}$$

For any W in the essential range of $\mathbf{W}$, let $Y(W)$ denote the quantity

$$Y(W) := \mathbb{P}\left(e \subset W \backslash \bigcup_{i \in I_m} \mathbf{e}'_i | \mathbf{W} = W\right).$$

From (4.7), (4.11), (2.1), our task is now to show that

$$\mathbb{E}Y(\mathbf{W}) = \Big(1 + O_{\leqslant}(\delta^{1/10^{m+1}})\Big)\, P_{m-1}(e) \exp\left(-\sum_{v \in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right).$$

Clearly $Y(\mathbf{W})$ is only non-zero when $e \subset \mathbf{W}$. From (5.1) we have

$$\mathbb{P}(e \subset \mathbf{W}) = (1 + O_{\leqslant}(\delta^{1/10^m}))P_{m-1}(e), \tag{5.13}$$

so it will suffice to show that

$$\mathbb{E}(Y(\mathbf{W})|e \subset \mathbf{W}) = \Big(1 + O(\delta^{\frac{1}{9\times 10^m}})\Big) \exp\left(-\sum_{v \in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right).$$

From (4.8), (5.12) and (4.1), we have

$$\exp\left(-\sum_{v \in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right) \geqslant \exp(-AD) \geqslant \delta^{1/10^{m+2}},$$

so it suffices to show that

$$\mathbb{E}(Y(\mathbf{W})|e \subset \mathbf{W}) = \Big(1 + O(\delta^{\frac{1}{9\times 10^m}})\Big) \exp\left(-\sum_{v \in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right) + O(\delta^{\frac{1}{8\times 10^m}}). \tag{5.14}$$

Suppose that W is in the essential range of $\mathbf{W}$ with $e \subset W$. As the $\mathbf{e}'_i$, $i \in I_m$, are jointly conditionally independent on the event $\mathbf{W} = W$, we may factor $Y(W)$ as

$$Y(W) = \prod_{i \in I_m} (1 - \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W)).$$

Since $\mathbf{e}'_i = \emptyset$ if $F_i(W)$ fails, we may write

$$Y(W) = \prod_{i \in I_m} (1 - 1_{F_i(W)} \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W)).$$

Now suppose that $i \in I_m$ and that W is such that $F_i(W)$ holds. From the union bound we have

$$\mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) \leqslant \sum_{v \in e} \mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W} = W).$$

From (5.9), (5.8), and (5.5), we have

$$\mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W} = W) = \sum_{\tilde{e}_i : v \in \tilde{e}_i} \mathbb{P}(\mathbf{e}'_i = \tilde{e}_i | \mathbf{W} = W) \ll \kappa^{-r} \mathbb{P}(v \in \mathbf{e}_i),$$

and hence by (4.3), (5.12)

$$\mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) \ll A\kappa^{-r}\delta/(\#I_m)^{1/2}.$$

From Taylor's expansion, we then have

$$1 - 1_{F_i(W)} \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) = \exp\left(-1_{F_i(W)} \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) + O((A\kappa^{-r}\delta)^2/\#I_m)\right).$$

From (5.6), we have $(A\kappa^{-r}\delta)^2 = O(\delta^{\frac{1}{9\times 10^m}})$, and so

$$Y(W) = (1 + O(\delta^{\frac{1}{9\times 10^m}})) \exp\left(-1_{F_i(W)} \sum_{i \in I_m} \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W)\right).$$

Next, we apply inclusion-exclusion to write

$$\mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) = \sum_{v \in e} \mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W} = W) - O\left(\sum_{v,w \in e : v \neq w} \mathbb{P}(v, w \in \mathbf{e}'_i | \mathbf{W} = W)\right).$$

The error term is handled by summing (5.9) over all $\tilde{e}_i$ with $v, w \in \tilde{e}_i$, and using (5.8) and (5.5). For distinct $v, w \in e$, we have

$$\mathbb{P}(v, w \in \mathbf{e}'_i | \mathbf{W} = W) = \sum_{\tilde{e}_i : v,w \in \tilde{e}_i} \mathbb{P}(\mathbf{e}'_i = \tilde{e}_i | \mathbf{W} = W) \ll \kappa^{-r} \sum_{\tilde{e}_i : v,w \in \tilde{e}_i} \mathbb{P}(\mathbf{e}_i = \tilde{e}_i) \ll \kappa^{-r} \mathbb{P}(v, w \in \mathbf{e}_i).$$

Hence by (4.4), (5.12)

$$\sum_{i \in I_m} \sum_{\substack{v,w \in e \\ v \neq w}} \mathbb{P}(v, w \in \mathbf{e}'_i | \mathbf{W} = W) \ll \kappa^{-r} A^2 \max_{\substack{v,w \in e \\ v \neq w}} \sum_{i \in I_m} \mathbb{P}(v, w \in \mathbf{e}_i) \ll A^2 \kappa^{-r} \delta.$$

From (5.6), we have $A^2\kappa^{-r}\delta = O(\delta^{\frac{1}{9\times 10^m}})$, and so

$$Y(W) = (1 + O(\delta^{\frac{1}{9\times 10^m}})) \exp\left(-1_{F_i(W)} \sum_{v \in e} \sum_{i \in I_m} \mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W} = W)\right).$$

Also we trivially have $0 \leqslant Y(W) \leqslant 1$. Thus, to prove (5.14), it suffices to show that

$$\sum_{v \in e} \sum_{i \in I_m} 1_{F_i(\mathbf{W})} \mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W}) = \sum_{v \in e} \frac{d_{I_m}(v)}{P_{m-1}(v)} + O(\delta^{\frac{1}{9\times 10^m}})$$

with probability $1 - O(\delta^{\frac{1}{8\times 10^m}})$, conditionally on the event that $e \subset \mathbf{W}$. From (5.12), (5.6), and the union bound, it thus suffices to show that for each $v \in e$, one has

$$\sum_{i \in I_m} 1_{F_i(\mathbf{W})} \mathbb{P}(v \in \mathbf{e}_i' | \mathbf{W}) = \frac{d_{I_m}(v)}{P_{m-1}(v)} + O(\delta^{\frac{1}{8\times 10^m}}) \tag{5.15}$$

with probability $1 - O(\delta^{\frac{1}{7\times 10^m}})$, conditionally on the event that $e \subset \mathbf{W}$.

We have

$$1_{F_i(\mathbf{W})} \mathbb{P}(v \in \mathbf{e}_i' | \mathbf{W}) = \frac{1_{F_i(\mathbf{W})}}{X_i(\mathbf{W})} \sum_{\tilde{e}_i : v \in \tilde{e}_i} 1_{\tilde{e}_i \subset \mathbf{W}} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)}{P_{m-1}(\tilde{e}_i)} \tag{5.16}$$

and, by (5.8),

$$\frac{1_{F_i(\mathbf{W})}}{X_i(\mathbf{W})} = 1 + O((1 - 1_{F_i(\mathbf{W})}) + \delta^{\frac{1}{3\times 10^m}}). \tag{5.17}$$

Upon inserting (5.16) and (5.17) into (5.15), the left side of (5.15) breaks into two pieces, a "main term" and an "error term".

Let us first estimate the error

$$\sum_{i \in I_m} O\left(1 - 1_{F_i(\mathbf{W})} + \delta^{\frac{1}{3\times 10^m}}\right) \sum_{\tilde{e}_i : v \in \tilde{e}_i} 1_{\tilde{e}_i \subset \mathbf{W}} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)}{P_{m-1}(\tilde{e}_i)}.$$

By (5.5) and (4.5), we may bound this by

$$O(\kappa^{-r}) \sum_{i \in I_m} (1 - 1_{F_i(\mathbf{W})} + \delta^{\frac{1}{3\times 10^m}})) \mathbb{P}(v \in \mathbf{e}_i) = O(\kappa^{-r}) d_{I_m}(v)(1 - 1_{F_i(\mathbf{W})} + \delta^{\frac{1}{3\times 10^m}}).$$

By Lemma 5.1, the unconditional expectation of this random variable is

$$O\left(\kappa^{-r} \delta^{\frac{1}{3\times 10^m}} d_{I_m}(v)\right).$$

Thus, by (5.13), the conditional expectation of this random variable to the event $e \subset \mathbf{W}$ is

$$\ll \kappa^{-r} \delta^{\frac{1}{3\times 10^m}} \frac{d_{I_m}(v)}{P_{m-1}(e)} \ll \kappa^{-A} \delta^{\frac{1}{3\times 10^m}}.$$

Here we used (4.8), (5.5) and (5.3) to obtain the second bound. By (5.6), this can be bounded by

$$O(\delta^{\frac{2}{7\times 10^m}}).$$

Thus, by Markov's inequality, this error is $O(\delta^{\frac{1}{7\times 10^m}})$ with probability $1 - O(\delta^{\frac{1}{7\times 10^m}})$, conditionally on $e \subset \mathbf{W}$. By the triangle inequality, it thus suffices to show that the main term satisfies

$$\sum_{i \in I_m} \sum_{\tilde{e}_i : v \in \tilde{e}_i} 1_{\tilde{e}_i \subset \mathbf{W}} \frac{\mathbb{P}(\mathbf{e}_i = \tilde{e}_i)}{P_{m-1}(\tilde{e}_i)} = \frac{d_{I_m}(v)}{P_{m-1}(v)} + O(\delta^{\frac{1}{8\times 10^m}})$$

with probability $1 - O(\delta^{\frac{1}{7\times 10^m}})$, conditionally on $e \subset \mathbf{W}$.

By a conditional version of Lemma 2.1 (replacing $\mathbb{E}X$ and $\mathbb{E}X^2$ with $\mathbb{E}(X|E)$ and $\mathbb{E}(X^2|E)$, respectively), together with (4.8), (4.1), it suffices to show that

$$\mathbb{E}\left(\sum_{i\in I_m}\sum_{\tilde{e}_i:v\in\tilde{e}_i} 1_{\tilde{e}_i\subset\mathbf{W}}\frac{\mathbb{P}(\mathbf{e}_i=\tilde{e}_i)}{P_{m-1}(\tilde{e}_i)}\Big|e\subset\mathbf{W}\right) = \frac{d_{I_m}(v)}{P_{m-1}(v)} + O(\delta^{\frac{1}{2\times 10^m}}) \tag{5.18}$$

and

$$\mathbb{E}\left(\sum_{i,i'\in I_m}\sum_{\substack{\tilde{e}_i:v\in\tilde{e}_i\\ \hat{e}_i:v\in\hat{e}_i}} 1_{\tilde{e}_i\subset\mathbf{W}}\frac{\mathbb{P}(\mathbf{e}_i=\tilde{e}_i)}{P_{m-1}(\tilde{e}_i)}1_{\hat{e}_i\subset\mathbf{W}}\frac{\mathbb{P}(\mathbf{e}_{i'}=\hat{e}_i)}{P_{m-1}(\hat{e}_i)}\Big|e\subset\mathbf{W}\right) = \left(\frac{d_{I_m}(v)}{P_{m-1}(v)}\right)^2 + O(\delta^{\frac{1}{2\times 10^m}}). \tag{5.19}$$

We begin with (5.18). For any given $i\in I_m$, we have from (5.1), (5.3) that

$$\frac{\mathbb{P}(e\cup\tilde{e}_i\subset\mathbf{W})}{\mathbb{P}(e\subset\mathbf{W})} = (1+O(\delta^{1/10^m}))\frac{P_{m-1}(e\cup\tilde{e}_i)}{P_{m-1}(e)}.$$

By (4.11), we can rewrite

$$\frac{P_{m-1}(e\cup\tilde{e}_i)}{P_{m-1}(\tilde{e}_i)P_{m-1}(e)} = \frac{1}{P_{m-1}(v)P_{m-1}(\tilde{e}_i\cap e\backslash\{v\})}.$$

By (2.1), we may thus write the left-hand side of (5.18) as

$$\sum_{i\in I_m}\sum_{\tilde{e}_i:v\in\tilde{e}_i}\frac{\mathbb{P}(\mathbf{e}_i=\tilde{e}_i)}{P_{m-1}(\tilde{e}_i)}\frac{\mathbb{P}(e\cup\tilde{e}_i\subset\mathbf{W})}{\mathbb{P}(e\subset\mathbf{W})} = \frac{1+O(\delta^{1/10^m})}{P_{m-1}(v)}\sum_{i\in I_m}\sum_{\tilde{e}_i:v\in\tilde{e}_i}\frac{\mathbb{P}(\mathbf{e}_i=\tilde{e}_i)}{P_{m-1}(\tilde{e}_i\cap e\backslash\{v\})}.$$

As in the proof of Lemma 5.1, $P_{m-1}(\tilde{e}_i\cap e\backslash\{v\})$ equals 1 unless $\tilde{e}_i$ and $e\backslash\{v\}$ have a common element, in which case it is $\geqslant\kappa^r$ by (5.5). Thus

$$\frac{1}{P_{m-1}(\tilde{e}_i\cap e\backslash\{v\})} = 1+O\Big(\kappa^{-r}\sum_{w\in e\backslash\{v\}}1_{w\in\tilde{e}_i}\Big).$$

From (4.5) one has

$$\sum_{i\in I_m}\sum_{\tilde{e}_i:v\in\tilde{e}_i}\mathbb{P}(v\in\mathbf{e}_i) = d_{I_m}(v),$$

and from (4.4) one has

$$\sum_{i\in I_m}\mathbb{P}(v,w\in\mathbf{e}_i)\leqslant\delta$$

for all $w\neq v$. Therefore, by (5.12), the left side of (5.18) is

$$\frac{1+O(\delta^{1/10^m})}{P_{m-1}(v)}\left(d_{I_m}(v)+O(A\delta\kappa^{-r})\right).$$

The claim now follows from (5.6) and (4.8).

Now we prove (5.19). For any $i,i'\in I_m$, we have from (5.1), (5.3) that

$$\frac{\mathbb{P}(\tilde{e}_i\cup\hat{e}_i\cup e\subset\mathbf{W})}{\mathbb{P}(e\subset\mathbf{W})} = (1+O(\delta^{1/10^m}))\frac{P_{m-1}(\tilde{e}_i\cup\hat{e}_i\cup e)}{P_{m-1}(e)},$$

so we are reduced (after applying (4.8), (5.6)) to showing that

$$\sum_{i,i' \in I_m} \sum_{\substack{\tilde{e}_i : v \in \tilde{e}_i \\ \hat{e}_i : v \in \hat{e}_i}} \mathbb{P}(\mathbf{e}_i = \tilde{e}_i)\mathbb{P}(\mathbf{e}_{i'} = \hat{e}_i) \frac{P_{m-1}(v)^2 P_{m-1}(\tilde{e}_i \cup \hat{e}_i \cup e)}{P_{m-1}(\tilde{e}_i) P_{m-1}(\hat{e}_i) P_{m-1}(e)} = d_{I_m}(v)^2 + O(\delta^{\frac{1}{10^m}}).$$

The quantity $\frac{P_{m-1}(v)^2 P_{m-1}(\tilde{e}_i \cup \hat{e}_i \cup e)}{P_{m-1}(\tilde{e}_i) P_{m-1}(\hat{e}_i) P_{m-1}(e)}$ is equal to 1 when the intersection of any two of $\tilde{e}_i, \hat{e}_i$ and e is $\{v\}$, and is $O(\kappa^{-2r})$ otherwise thanks to (5.5). Hence we may estimate this ratio by

$$1 + O\left(\kappa^{-2r} \sum_{w \in e \setminus \{v\}} (1_{w \in \tilde{e}_i} + 1_{w \in \hat{e}_i})\right) + O\left(\kappa^{-2r} \sum_{w \in \tilde{e}_i \setminus \{v\}} 1_{w \in \hat{e}_i}\right).$$

From (4.5) one has

$$\sum_{i,i' \in I_m} \mathbb{P}(v \in \mathbf{e}_i)\mathbb{P}(v \in \mathbf{e}_{i'}) = d_{I_m}(v)^2,$$

so from (5.6) it suffices to show that

$$\sum_{i,i' \in I_m} \sum_{w \in e \setminus \{v\}} \mathbb{P}\left(v \in \mathbf{e}_i, v \in \mathbf{e}_{i'}, w \in \mathbf{e}_i\right) \leqslant DA\delta, \tag{5.20}$$

$$\sum_{i,i' \in I_m} \sum_{w \in e \setminus \{v\}} \mathbb{P}\left(v \in \mathbf{e}_i, v \in \mathbf{e}_{i'}, w \in \mathbf{e}_{i'}\right) \leqslant DA\delta, \tag{5.21}$$

$$\sum_{i,i' \in I_m} \mathbb{E}\left[1_{v \in \mathbf{e}_i, v \in \mathbf{e}_{i'}} \left(\#(\mathbf{e}_i \cap \mathbf{e}_{i'}) - 1\right)\right] \leqslant Dr\delta. \tag{5.22}$$

For (5.20), we use (4.5) to write the left-hand side as

$$d_{I_m}(v) \sum_{w \in e \setminus \{v\}} \sum_{i \in I_m} \mathbb{P}(v, w \in \mathbf{e}_i),$$

which by (4.8), (5.12), (4.4) is bounded by $DA\delta$, as desired. Similarly for (5.21). For (5.22), we take expectations in $\mathbf{e}_{i'}$ first using (2.1), (4.4) to upper bound the left-hand side of (5.22) by

$$\sum_{i \in I_m} \mathbb{E}\left(1_{v \in \mathbf{e}_i} \sum_{w \in \mathbf{e}_i \setminus \{v\}} \delta\right),$$

which by (4.2), (4.5), (4.8) is bounded by $Dr\delta$, as desired. This proves (5.19), which implies (5.15) and in turn (5.14). The proof of Theorem 3 is now complete.

6. USING A SIEVE WEIGHT

If r is a natural number, an *admissible r-tuple* is a tuple $(h_1, \dots, h_r)$ of distinct integers $h_1, \dots, h_r$ that do not cover all residue classes modulo p, for any prime p. For instance, the tuple $(p_{\pi(r)+1}, \dots, p_{\pi(r)+r})$ consisting of the first r primes larger than r is an admissible r-tuple.

We will establish Theorem 4 by a probabilistic argument involving a certain weight function, the details of which may be found in the following.

Theorem 5 (Existence of good sieve weight)**.** *Let x be a sufficiently large real number and let y be defined by* (3.1)*. Let $\mathcal{P}, \mathcal{Q}$ be defined by* (3.4)*,* (3.5)*. Let r be a positive integer with*

$$r_0 \leqslant r \leqslant \log^{1/5} x \tag{6.1}$$

for some sufficiently large absolute constant r_0, and let $(h_1, \dots, h_r)$ be an admissible r-tuple contained in $[2r^2]$. Then one can find a positive quantity

$$\tau \geqslant x^{-o(1)} \tag{6.2}$$

and a positive quantity $u = u(r)$ depending only on r with

$$u \asymp \log r \tag{6.3}$$

and a non-negative function $w : \mathcal{P} \times \mathbb{Z} \to \mathbb{R}^+$ supported on $\mathcal{P} \times (\mathbb{Z} \cap [-y, y])$ with the following properties:

- *Uniformly for every $p \in \mathcal{P}$, one has*

$$\sum_{n \in \mathbb{Z}} w(p, n) = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \tau \frac{y}{\log^r x}. \tag{6.4}$$

- *Uniformly for every $q \in \mathcal{Q}$ and $i = 1, \dots, r$, one has*

$$\sum_{p \in \mathcal{P}} w(p, q - h_i p) = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \tau \frac{u}{r} \frac{x}{2 \log^r x}. \tag{6.5}$$

- *Uniformly for every $h = O(y/x)$ that is not equal to any of the h_i, one has*

$$\sum_{q \in \mathcal{Q}} \sum_{p \in \mathcal{P}} w(p, q - hp) = O\left(\frac{1}{\log_2^{10} x} \tau \frac{x}{\log^r x} \frac{y}{\log x}\right). \tag{6.6}$$

- *Uniformly for all $p \in \mathcal{P}$ and $n \in \mathbb{Z}$,*

$$w(p, n) = O(x^{1/3 + o(1)}). \tag{6.7}$$

Remark 2. One should think of $w(p, n)$ as being a smoothed out indicator function for the event that $n + h_1 p, \dots, n + h_r p$ are all almost primes in $[y]$. As essentially discovered in [33], by choosing the smoothing correctly, one can ensure that approximately $\log r$ of the elements of this tuple $n + h_1 p, \dots, n + h_r p$ are genuinely prime rather than almost prime, when weighted by $w(p, n)$; this explains the presence of the bounds (6.3). The estimate (6.6) is not, strictly speaking, needed for our current argument; however, it is easily obtained by our methods, and will be of use in a followup work [15] to this paper in which the analogue of Theorem 1 for chains of large gaps is established.

The proof of this theorem will rely on the estimates for multidimensional prime-detecting sieves established by the fourth author in [34], and will be the focus of subsequent sections. In this section, we show how Theorem 5 implies Theorem 4.

Let $x, c, y, z, \mathcal{S}, \mathcal{P}, \mathcal{Q}$ be as in Theorem 4. We set r to be the maximum value permitted by Theorem 5, namely

$$r := \lfloor \log^{1/5} x \rfloor \tag{6.8}$$

and let $(h_1, \dots, h_r)$ be the admissible r-tuple consisting of the first r primes larger than r, thus $h_i = p_{\pi(r)+i}$ for $i = 1, \dots, r$. From the prime number theorem we have $h_i = O(r \log r)$ for $i = 1, \dots, r$, and so we have $h_i \in [2r^2]$ for $i = 1, \dots, r$ if x is large enough (there are many other choices possible, e.g.

$(h_1, \dots, h_r) = (1^2, 3^2, \dots, (2r-1)^2)$). We now invoke Theorem 5 to obtain quantities τ, u and a weight $w : \mathcal{P} \times \mathbb{Z} \to \mathbb{R}^+$ with the stated properties.

For each $p \in \mathcal{P}$, let $\tilde{\mathbf{n}}_p$ denote the random integer with probability density

$$\mathbb{P}(\tilde{\mathbf{n}}_p = n) := \frac{w(p,n)}{\sum_{n' \in \mathbb{Z}} w(p,n')}$$

for all $n \in \mathbb{Z}$ (we will not need to impose any independence conditions on the $\tilde{\mathbf{n}}_p$). From (6.4), (6.5) we have

$$\sum_{p \in \mathcal{P}} \mathbb{P}(q = \tilde{\mathbf{n}}_p + h_i p) = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{u}{r} \frac{x}{2y} \qquad (q \in \mathcal{Q}, 1 \leqslant i \leqslant r). \tag{6.9}$$

Also, from (6.4), (6.7), (6.2) and (3.1), one has

$$\mathbb{P}(\tilde{\mathbf{n}}_p = n) \ll x^{-1/2-1/6+o(1)} \tag{6.10}$$

for all $p \in \mathcal{P}$ and $n \in \mathbb{Z}$.

We choose the random vector $\vec{\mathbf{a}} := (\mathbf{a}_s \bmod s)_{s \in \mathcal{S}}$ by selecting each $\mathbf{a}_s \bmod s$ uniformly at random from $\mathbb{Z}/s\mathbb{Z}$, independently in s and independently of the $\tilde{\mathbf{n}}_p$. The resulting sifted set $S(\vec{\mathbf{a}})$ is a random periodic subset of $\mathbb{Z}$ with density

$$\sigma := \prod_{s \in \mathcal{S}} \left(1 - \frac{1}{s}\right).$$

From the prime number theorem (with sufficiently strong error term), (3.2) and (3.3),

$$\sigma = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{\log(\log^{20} x)}{\log z} = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{80 \log_2 x}{\log x \log_3 x / \log_2 x},$$

so in particular we see from (3.1) that

$$\sigma y = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) 80 c x \log_2 x. \tag{6.11}$$

We also see from (6.8) that

$$\sigma^r = x^{o(1)}. \tag{6.12}$$

We have a useful correlation bound:

Lemma 6.1. *Let $t \leqslant \log x$ be a natural number, and let $n_1, \dots, n_t$ be distinct integers with $|n_i| \leqslant x^2$ for each i. Then one has*

$$\mathbb{P}(n_1, \dots, n_t \in S(\vec{\mathbf{a}})) = \left(1 + O\left(\frac{1}{\log^{16} x}\right)\right) \sigma^t.$$

Proof. For each $s \in \mathcal{S}$, the integers $n_1, \dots, n_t$ occupy t distinct residue classes modulo s, unless s divides one of $n_i - n_j$ for $1 \leqslant i < j \leqslant t$. Since $s \geqslant \log^{20} x$ and $|n_i - n_j| \leqslant 2x^2$, the latter possibility occurs at most $O(t^2 \log x) = O(\log^3 x)$ times. Thus the probability that $\mathbf{a}_s \bmod s$ avoids all of the $n_1, \dots, n_t$ is

equal to $1-\frac{t}{s}$ except for $O(\log^3 x)$ values of s, where it is instead $(1+O(\frac{1}{\log^{19} x}))(1-\frac{t}{s})$. Thus,

$$\begin{aligned}\mathbb{P}(n_1,\dots,n_t \in S(\vec{\mathbf{a}})) &= \left(1+O\left(\frac{1}{\log^{19} x}\right)\right)^{O(\log^3 x)} \prod_{s\in\mathcal{S}}\left(1-\left(\frac{t}{s}\right)\right)\\ &= \left(1+O\left(\frac{1}{\log^{16} x}\right)\right)\sigma^t \prod_{s\in\mathcal{S}}\left(1+O\left(\frac{t^2}{s^2}\right)\right)\\ &= \left(1+O\left(\frac{1}{\log^{16} x}\right)\right)\sigma^t.\end{aligned}$$

□

Among other things, this gives the claim (4.28):

Corollary 5. *With probability* $1-o(1)$*, we have*

$$\#(\mathcal{Q}\cap S(\vec{\mathbf{a}})) \sim \sigma\frac{y}{\log x} \sim 80c\frac{x}{\log x}\log_2 x. \tag{6.13}$$

Proof. From Lemma 6.1, we have

$$\mathbb{E}\#(\mathcal{Q}\cap S(\vec{\mathbf{a}})) = \left(1+O\left(\frac{1}{\log^{16} x}\right)\right)\sigma\#\mathcal{Q}$$

and

$$\mathbb{E}\#\big((\mathcal{Q}\cap S(\vec{\mathbf{a}}))\big)^2 = \left(1+O\left(\frac{1}{\log^{16} x}\right)\right)(\sigma\#\mathcal{Q}+\sigma^2(\#\mathcal{Q})(\#\mathcal{Q}-1)),$$

and so by the prime number theorem we see that the random variable $\#\mathcal{Q}\cap S(\vec{\mathbf{a}})$ has mean $(1+o(\frac{1}{\log_2 x}))\sigma\frac{y}{\log x}$ and variance $O\left(\frac{1}{\log^{16} x}(\sigma\frac{y}{\log x})^2\right)$. The claim then follows from Lemma 2.1 (with plenty of room to spare). □

For each $p\in\mathcal{P}$, we consider the quantity

$$X_p(\vec{a}) := \mathbb{P}(\tilde{\mathbf{n}}_p + h_i p \in S(\vec{a}) \text{ for all } i=1,\dots,r), \tag{6.14}$$

and let $\mathcal{P}(\vec{a})$ denote the set of all the primes $p\in\mathcal{P}$ such that

$$X_p(\vec{a}) = \left(1+O_{\leqslant}\left(\frac{1}{\log^3 x}\right)\right)\sigma^r. \tag{6.15}$$

In light of Lemma 6.1, we expect most primes in $\mathcal{P}$ to lie in $\mathcal{P}(\vec{a})$, and this will be confirmed below (Lemma 6.3). We now define the random variables $\mathbf{n}_p$ as follows. Suppose we are in the event $\vec{\mathbf{a}}=\vec{a}$ for some $\vec{a}$ in the range of $\vec{\mathbf{a}}$. If $p\in\mathcal{P}\backslash\mathcal{P}(\vec{a})$, we set $\mathbf{n}_p=0$. Otherwise, if $p\in\mathcal{P}(\vec{a})$, we define $\mathbf{n}_p$ to be the random integer with conditional probability distribution

$$\mathbb{P}(\mathbf{n}_p = n|\vec{\mathbf{a}}=\vec{a}) := \frac{Z_p(\vec{a};n)}{X_p(\vec{a})}, \quad Z_p(\vec{a};n) = 1_{n+h_j p\in S(\vec{a}) \text{ for } j=1,\dots,r}\mathbb{P}(\tilde{\mathbf{n}}_p=n), \tag{6.16}$$

with the $\mathbf{n}_p$ ($p\in\mathcal{P}(\vec{a})$) jointly independent, conditionally on the event $\vec{\mathbf{a}}=\vec{a}$. From (6.14) we see that these random variables are well defined.

The first claim (4.27) of Theorem 4 now follows immediately from (6.10), (6.16) and (6.15), and so we are left to establish the final two assertions.

Lemma 6.2. *With probability $1-o(1)$, we have*

$$\sigma^{-r}\sum_{i=1}^{r}\sum_{p\in\mathcal{P}(\vec{\mathbf{a}})}Z_p(\vec{\mathbf{a}};q-h_ip)=\left(1+O\left(\frac{1}{\log_2^3 x}\right)\right)\frac{u}{\sigma}\frac{x}{2y} \tag{6.17}$$

for all but at most $\frac{x}{2\log x\log_2 x}$ of the primes $q\in\mathcal{Q}\cap S(\vec{\mathbf{a}})$.

Let $\vec{a}$ be good (recall the definition from Theorem 4) and $q\in\mathcal{Q}\cap S(\vec{a})$. Substituting definition (6.16) into the left hand side of of (6.17), using (6.15), and observing that $q=\mathbf{n}_p+h_ip$ is only possible if $p\in\mathcal{P}(\vec{\mathbf{a}})$ (since $\mathbf{n}_p=0$ for $p\in\mathcal{P}\backslash\mathcal{P}(\vec{a})$), we find that

$$\begin{aligned}\sigma^{-r}\sum_{i=1}^{r}\sum_{p\in\mathcal{P}(\vec{a})}Z_p(\vec{a};q-h_ip)&=\sigma^{-r}\sum_{i=1}^{r}\sum_{p\in\mathcal{P}(\vec{a})}X_p(\vec{a})\mathbb{P}(\mathbf{n}_p=q-h_ip|\vec{\mathbf{a}}=\vec{a})\\&=\left(1+O\left(\frac{1}{\log^3 x}\right)\right)\sum_{i=1}^{r}\sum_{p\in\mathcal{P}(\vec{a})}\mathbb{P}(\mathbf{n}_p=q-h_ip|\vec{\mathbf{a}}=\vec{a})\\&=\left(1+O\left(\frac{1}{\log^3 x}\right)\right)\sum_{p\in\mathcal{P}}\mathbb{P}(q\in\mathbf{e}_p(\vec{a})|\vec{\mathbf{a}}=\vec{a}),\end{aligned}$$

where $\mathbf{e}_p(\vec{a})=\{\mathbf{n}_p+h_ip:1\leqslant i\leqslant r\}\cap\mathcal{Q}\cap S(\vec{a})$ is as defined in Theorem 4. Relation (4.29) (that is, $\vec{\mathbf{a}}$ is good with probability $1-o(1)$) follows upon noting that by (6.8), (6.3) and (6.11),

$$C:=\frac{u}{\sigma}\,\frac{x}{2y}\asymp\frac{1}{c}.$$

Before proving Lemma 6.2, we first confirm that $\mathcal{P}\backslash\mathcal{P}(\vec{\mathbf{a}})$ is small with high probability.

Lemma 6.3. *With probability $1-O(1/\log^3 x)$, $\mathcal{P}(\vec{\mathbf{a}})$ contains all but $O(\frac{1}{\log^3 x}\frac{x}{\log x})$ of the primes $p\in\mathcal{P}$. In particular, $\mathbb{E}\#\mathcal{P}(\vec{\mathbf{a}})=\#\mathcal{P}(1+O(1/\log^3 x))$.*

Proof. By linearity of expectation and Markov's inequality, it suffices to show that for each $p\in\mathcal{P}$, we have $p\in\mathcal{P}(\vec{\mathbf{a}})$ with probability $1-O(\frac{1}{\log^6 x})$. By Lemma 2.1, it suffices to show that

$$\mathbb{E}X_p(\vec{\mathbf{a}})=\mathbb{P}(\tilde{\mathbf{n}}_p+h_ip\in S(\vec{\mathbf{a}})\text{ for all }i=1,\dots,r)=\left(1+O\left(\frac{1}{\log^{12}x}\right)\right)\sigma^r \tag{6.18}$$

and

$$\mathbb{E}X_p(\vec{\mathbf{a}})^2=\mathbb{P}(\tilde{\mathbf{n}}_p^{(1)}+h_ip,\tilde{\mathbf{n}}_p^{(2)}+h_ip\in S(\vec{\mathbf{a}})\text{ for all }i=1,\dots,r)=\left(1+O\left(\frac{1}{\log^{12}x}\right)\right)\sigma^{2r}, \tag{6.19}$$

where $\tilde{\mathbf{n}}_p^{(1)},\tilde{\mathbf{n}}_p^{(2)}$ are independent copies of $\tilde{\mathbf{n}}_p$ that are also independent of $\vec{\mathbf{a}}$.

The claim (6.18) follows from Lemma 6.1 (performing the conditional expectation over $\tilde{\mathbf{n}}_p$ first). A similar application of Lemma 6.1 allows one to write the left-hand side of (6.19) as

$$\left(1+O\left(\frac{1}{\log^{16}x}\right)\right)\mathbb{E}\sigma^{\#\{\tilde{\mathbf{n}}_p^{(l)}+h_ip:i=1,\dots,r;l=1,2\}}.$$

From (6.10) we see that the quantity $\#\{\tilde{\mathbf{n}}_p^{(l)}+h_ip:i=1,\dots,r;l=1,2\}$ is equal to $2r$ with probability $1-O(x^{-1/2-1/6+o(1)})$, and is less than $2r$ otherwise. The claim now follows from (6.12). □

Proof of Lemma 6.2. We first show that replacing $\mathcal{P}(\vec{\mathbf{a}})$ with $\mathcal{P}$ has negligible effect on the sum, with probability $1-o(1)$. Fix i and substitute $n = q - h_i p$. By Markov's inequality, it suffices to show that

$$\mathbb{E}\sum_n \sigma^{-r} \sum_{p\in\mathcal{P}\setminus\mathcal{P}(\vec{\mathbf{a}})} Z_p(\vec{\mathbf{a}};n) = o\left(\frac{u}{\sigma}\frac{x}{2y}\,\frac{1}{r}\frac{1}{\log_2^3 x}\,\frac{x}{\log x\log_2 x}\right). \tag{6.20}$$

By Lemma 6.1, we have

$$\begin{aligned}\mathbb{E}\sum_n \sigma^{-r}\sum_{p\in\mathcal{P}} Z_p(\vec{\mathbf{a}};n) &= \sigma^{-r}\sum_{p\in\mathcal{P}}\sum_n \mathbb{P}(\tilde{\mathbf{n}}_p = n)\mathbb{P}(n+h_j p\in S(\vec{\mathbf{a}}) \text{ for } j=1,\dots,r)\\ &= \left(1+O\left(\frac{1}{\log^{16} x}\right)\right)\#\mathcal{P}.\end{aligned}$$

Next, by (6.15) and Lemma 6.3 we have

$$\begin{aligned}\mathbb{E}\sum_n \sigma^{-r}\sum_{p\in\mathcal{P}(\vec{\mathbf{a}})} Z_p(\vec{\mathbf{a}};n) &= \sigma^{-r}\sum_{\vec{a}}\mathbb{P}(\vec{\mathbf{a}}=\vec{a})\sum_{p\in\mathcal{P}(\vec{a})} X_p(\vec{a})\\ &= \left(1+O\left(\frac{1}{\log^3 x}\right)\right)\,\mathbb{E}\,\#\mathcal{P}(\vec{\mathbf{a}}) = \left(1+O\left(\frac{1}{\log^3 x}\right)\right)\#\mathcal{P};\end{aligned}$$

subtracting, we conclude that the left-hand side of (6.20) is $O(\#\mathcal{P}/\log^3 x) = O(x/\log^4 x)$. The claim then follows from (3.1) and (6.1).

By (6.20), it suffices to show that with probability $1-o(1)$, for all but at most $\frac{x}{2\log x\log_2 x}$ primes $q\in\mathcal{Q}\cap S(\vec{\mathbf{a}})$, one has

$$\sum_{i=1}^r\sum_{p\in\mathcal{P}} Z_p(\vec{\mathbf{a}};q-h_i p) = \left(1+O_{\leqslant}\left(\frac{1}{\log_2^3 x}\right)\right)\sigma^{r-1}u\frac{x}{2y}. \tag{6.21}$$

Call a prime $q\in\mathcal{Q}$ *bad* if $q\in\mathcal{Q}\cap S(\vec{\mathbf{a}})$ but (6.21) fails. Using Lemma 6.1 and (6.9), we have

$$\begin{aligned}\mathbb{E}\Big[\sum_{q\in\mathcal{Q}\cap S(\vec{\mathbf{a}})}\sum_{i=1}^r\sum_{p\in\mathcal{P}} Z_p(\vec{\mathbf{a}};q-h_i p)\Big] &= \sum_{q,i,p}\mathbb{P}(q+(h_j-h_i)p\in S(\vec{\mathbf{a}}) \text{ for all } j=1,\dots,r)\mathbb{P}(\tilde{\mathbf{n}}_p = q-h_i p)\\ &= \left(1+O\left(\frac{1}{\log_2^{10} x}\right)\right)\frac{\sigma y}{\log x}\,\sigma^{r-1}u\,\frac{x}{2y}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\Big[\sum_{q\in\mathcal{Q}\cap S(\vec{\mathbf{a}})}\Big(\sum_{i=1}^r\sum_{p\in\mathcal{P}} Z_p(\vec{\mathbf{a}};q-h_i p)\Big)^2\Big] &= \sum_{\substack{p_1,p_2,q\\ i_1,i_2}}\mathbb{P}(q+(h_j-h_{i_\ell})p_\ell\in S(\vec{\mathbf{a}}) \text{ for } j=1,\dots,r;\ell=1,2)\\ &\qquad\times\mathbb{P}(\tilde{\mathbf{n}}^{(1)}_{p_1} = q-h_{i_1}p_1)\mathbb{P}(\tilde{\mathbf{n}}^{(2)}_{p_2} = q-h_{i_2}p_2)\\ &= \left(1+O\left(\frac{1}{\log_2^{10} x}\right)\right)\frac{\sigma y}{\log x}\,\left(\sigma^{r-1}u\,\frac{x}{2y}\right)^2,\end{aligned}$$

where $(\tilde{\mathbf{n}}^{(1)}_{p_1})_{p_1\in\mathcal{P}}$ and $(\tilde{\mathbf{n}}^{(2)}_{p_2})_{p_2\in\mathcal{P}}$ are independent copies of $(\tilde{\mathbf{n}}_p)_{p\in\mathcal{P}}$ over $\vec{\mathbf{a}}$. In the last step we used the fact that the terms with $p_1=p_2$ contribute negligibly.

By Lemma 2.1 it follows that the number of bad q is $\ll \frac{\sigma y}{\log x}\frac{1}{\log_2^3 x} \ll \frac{x}{\log x \log_2^2 x}$ with probability $1 - O(1/\log_2 x)$. This concludes the proof. □

It remains to establish Theorem 5. This is the objective of the remaining sections of the paper.

7. Multidimensional sieve estimates

We now recall a technical multidimensional sieve estimate from [34] (a minor variant of [34, Proposition 6.1]). In this section we will follow the notation from [34], which is a little different from that in the rest of this paper, with the exception that we will take the set denoted $\mathcal{P}$ in that paper to be equal to the set $\mathscr{P}$ of all primes from the outset.

A *linear form* will be a function $L : \mathbb{Z} \to \mathbb{Z}$ of the form $L(n) = l_1 n + l_2$ with integer coefficients l_1, l_2 and $l_1 \neq 0$. Let $\mathcal{A}$ be a set of integers. Given a linear form $L(n) = l_1 n + l_2$, we define the sets

$$\begin{aligned}\mathcal{A}(x) &:= \{n \in \mathcal{A} : x \leqslant n \leqslant 2x\},\\ \mathcal{A}(x; q, a) &:= \{n \in \mathcal{A}(x) : n \equiv a \pmod q\},\\ \mathscr{P}_{L,\mathcal{A}}(x) &:= L(\mathcal{A}(x)) \cap \mathscr{P},\\ \mathscr{P}_{L,\mathcal{A}}(x; q, a) &:= L(\mathcal{A}(x; q, a)) \cap \mathscr{P},\end{aligned}$$

for any $x > 0$ and congruence class $a \bmod q$, and define the quantity

$$\varphi_L(q) := \varphi(|l_1|q)/\varphi(|l_1|),$$

where φ is the Euler totient function. We recall the standard bounds

$$X \geqslant \varphi(X) \gg \frac{X}{\log_2 X} \tag{7.1}$$

since $\varphi(X)/X$ is smallest when X is composed only of primes $\ll \log X$. Thanks to this bound, most factors of the form $\frac{X}{\varphi(X)}$ appearing below become relatively harmless, and we recommend that they may be ignored for a first reading.

A finite set $\mathcal{L} = \{L_1, \dots, L_k\}$ of linear forms is said to be *admissible* if $\prod_{i=1}^k L_i(n)$ has no fixed prime divisor; that is, for every prime p there exists an integer n_p such that $\prod_{i=1}^k L_i(n_p)$ is not divisible by p.

Definition 2. [34] *Let x be a large quantity, let $\mathcal{A}$ be a set of integers, $\mathcal{L} = \{L_1, \dots, L_k\}$ a finite set of linear forms, and B a natural number. We allow $\mathcal{A}, \mathcal{L}, k, B$ to vary with x. Let $0 < \theta < 1$ be a quantity independent of x. Let $\mathcal{L}'$ be a subset of $\mathcal{L}$. We say that the tuple $(\mathcal{A}, \mathcal{L}, \mathscr{P}, B, x, \theta)$* obeys Hypothesis 1 at *$\mathcal{L}'$ if we have the following three estimates:*

(1) *($\mathcal{A}(x)$ is well-distributed in arithmetic progressions) We have*

$$\sum_{q \leqslant x^\theta} \max_a \left| \#\mathcal{A}(x; q, a) - \frac{\#\mathcal{A}(x)}{q} \right| \ll \frac{\#\mathcal{A}(x)}{\log^{100k^2} x}.$$

(2) *($\mathscr{P}_{L,\mathcal{A}}(x)$ is well-distributed in arithmetic progressions) For any $L \in \mathcal{L}'$, we have*

$$\sum_{q \leqslant x^\theta;\, (q,B)=1} \max_{a:(L(a),q)=1} \left| \#\mathscr{P}_{L,\mathcal{A}}(x; q, a) - \frac{\#\mathscr{P}_{L,\mathcal{A}}(x)}{\varphi_L(q)} \right| \ll \frac{\#\mathscr{P}_{L,\mathcal{A}}(x)}{\log^{100k^2} x}.$$

(3) *($\mathcal{A}(x)$ not too concentrated) For any $q < x^\theta$ and $a \in \mathbb{Z}$ we have*

$$\#\mathcal{A}(x;q,a) \ll \frac{\#\mathcal{A}(x)}{q}.$$

In [34] this definition was only given in the case $\mathcal{L}' = \mathcal{L}$, but we will need the (mild) generalization to the case in which $\mathcal{L}'$ is a (possibly empty) subset of $\mathcal{L}$.

As is common in analytic number theory, we will have to address the possibility of an exceptional Siegel zero. As we want to keep all our estimates effective, we will not rely on Siegel's theorem or its consequences (such as the Bombieri-Vinogradov theorem). Instead, we will rely on the Landau-Page theorem, which we now recall. Throughout, χ denotes a Dirichlet character.

Lemma 7.1 (Landau-Page theorem)**.** *Let $Q \geqslant 100$. Suppose that $L(s,\chi) = 0$ for some primitive character χ of modulus at most Q, and some $s = \sigma + it$. Then either*

$$1 - \sigma \gg \frac{1}{\log(Q(1+|t|))},$$

or else $t = 0$ and χ is a quadratic character χ_Q, which is unique. Furthermore, if χ_Q exists, then its conductor q_Q is square-free apart from a factor of at most 8, and obeys the lower bound

$$q_Q \gg \frac{\log^2 Q}{\log_2^4 Q}.$$

Proof. See e.g. [9, Chapter 14]. The final estimate follows from the bound $1 - \beta \gg q^{-1/2}\log^{-2} q$ for a real zero β of $L(s,\chi)$ with χ of modulus q, which can also be found in [9, Chapter 14]. □

We can then eliminate the exceptional character by deleting at most one prime factor of q_Q - an idea used previously by Hildebrand and Maier [27].

Corollary 6. *Let $Q \geqslant 100$. Then there exists a quantity B_Q which is either equal to 1 or is a prime of size*

$$B_Q \gg \log_2 Q$$

with the property that

$$1 - \sigma \gg \frac{1}{\log(Q(1+|t|))}$$

whenever $L(\sigma + it, \chi) = 0$ and χ is a character of modulus at most Q and coprime to B_Q.

Proof. If the exceptional character χ_Q from Lemma 7.1 does not exist, then take $B_Q := 1$; otherwise we take B_Q to be the largest prime factor of q_Q. As q_Q is square-free apart from a factor of at most 8, we have $\log q_Q \ll B_Q$ by the prime number theorem, and the claim follows. □

We will only need the above definition in the following special case:

Lemma 7.2. *Let x be a large quantity. Then there exists a natural number $B \leqslant x$, which is either 1 or a prime, such that the following holds. Let $\mathcal{A} := \mathbb{Z}$, let $\theta := 1/3$, and let $\mathcal{L} = \{L_1, \dots, L_k\}$ be a finite set of linear forms $L_i(n) = a_i n + b_i$ (which may depend on x) with $k \leqslant \log^{1/5} x$, $1 \leqslant |a_i| \leqslant \log x$, and $|b_i| \leqslant x\log^2 x$. Let $x \leqslant y \leqslant x\log^2 x$, and let $\mathcal{L}'$ be a subset of $\mathcal{L}$ such that L_i is non-negative on $[y, 2y]$ and a_i is coprime to B for all $L_i \in \mathcal{L}'$. Then $(\mathcal{A}, \mathcal{L}, \mathscr{P}, B, y, \theta)$ obeys Hypothesis 1 at $\mathcal{L}'$ with absolute implied constants (i.e. the bounds in Hypothesis 1 are uniform over all such choices of $\mathcal{L}$ and y).*

Proof. Parts (1) and (3) of Hypothesis 1 are easy; the only difficult verification is (2). We apply Corollary 6 with $Q := \exp(c_1\sqrt{\log x})$ for some small absolute constant c_1 to obtain a quantity $B := B_Q$ with the stated properties. By the Landau-Page theorem (see [9, Chapter 20]), we have that if c_1 is sufficiently small then we have the effective bound

$$\phi(q)^{-1}\sideset{}{^*}\sum_{\chi} |\psi(z,\chi)| \ll x\exp(-3c\sqrt{\log x}) \tag{7.2}$$

for all $1 < q < \exp(2c\sqrt{\log x})$ with $(q,B) = 1$ and all $z \leqslant x\log^4 x$. Here the summation is over all primitive $\chi \bmod q$ and $\psi(z,\chi) = \sum_{n\leqslant z}\chi(n)\Lambda(n)$. Following a standard proof of the Bombieri-Vinogradov Theorem (see [9, Chapter 28], for example), we have (for a suitable constant $c > 0$)

$$\sum_{\substack{q<x^{1/2-\epsilon}\\(q,B)=1}} \sup_{\substack{(a,q)=1\\ z\leqslant x\log^4 x}} \left|\pi(z;q,a) - \frac{\pi(z)}{\phi(q)}\right| \ll x\exp(-c\sqrt{\log x}) + \log x \sum_{\substack{q<\exp(2c\sqrt{\log x})\\(q,B)=1}} \sideset{}{^*}\sum_{\chi} \sup_{z\leqslant x\log^4 x} \frac{|\psi(z,\chi)|}{\phi(q)}. \tag{7.3}$$

Combining these two statements and using the triangle inequality gives the bound required for (2). □

We now recall the construction of sieve weights from [34, Section 7]. On first reading we recommend the reader not pay too much attention to the details; the key point is the existence of a weight $w(n)$ which will establish Theorem 5. The reason it is necessary to know the construction is the technical issue that the weights $w(n)$ depend on a given admissible set of linear forms, and we require that the final estimates obtained are essentially uniform over similar admissible sets.

Let $W := \prod_{p\leqslant 2k^2;\, p\nmid B} p$. For each prime p not dividing B, let $r_{p,1}(\mathcal{L}) < \cdots < r_{p,\omega_{\mathcal{L}}(p)}(\mathcal{L})$ be the elements n of $[p]$ for which $p|\prod_{i=1}^k L_i(n)$. If p is also coprime to W, then for each $1 \leqslant a \leqslant \omega_{\mathcal{L}}(p)$, let $j_{p,a} = j_{p,a}(\mathcal{L})$ denote the least element of $[k]$ such that $p|L_{j_{p,a}}(r_{p,a}(\mathcal{L}))$.

Let $\mathcal{D}_k(\mathcal{L})$ denote the set

$$\begin{aligned}\mathcal{D}_k(\mathcal{L}) := \{(d_1,\dots,d_k) \in \mathbf{N}^k : \mu^2(d_1\dots d_k) = 1; (d_1\dots d_k, WB) = 1;\\ (d_j,p) = 1 \text{ whenever } p\nmid BW \text{ and } j \neq j_{p,1},\dots,j_{p,\omega_{\mathcal{L}}(p)}\}.\end{aligned}$$

Define the singular series

$$\mathfrak{S}(\mathcal{L}) := \prod_{p\nmid B}\left(1-\frac{\omega_{\mathcal{L}}(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k},$$

and

$$\mathfrak{S}_{WB}(\mathcal{L}) := \prod_{p\nmid WB}\left(1-\frac{\omega_{\mathcal{L}}(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k},$$

the function

$$\varphi_{\omega_{\mathcal{L}}}(d) := \prod_{p|d}(p-\omega_{\mathcal{L}}(p)),$$

and let R be a quantity of size

$$x^{\theta/10} \leqslant R \leqslant x^{\theta/3}.$$

Let $F : \mathscr{R}^k \to \mathscr{R}$ be a smooth function supported on the simplex

$$\mathcal{R}_k = \{(t_1,\dots,t_k) \in \mathbb{R}_+^k : t_1 + \cdots + t_k \leqslant 1\}.$$

For any $(r_1,\dots,r_k)\in\mathcal{D}_k(\mathcal{L})$ define

$$y_{(r_1,\dots,r_k)}(\mathcal{L}) := \frac{1_{\mathcal{D}_k(\mathcal{L})}(r_1,\dots,r_k)W^kB^k}{\varphi(WB)^k}\mathfrak{S}_{WB}(\mathcal{L})F\left(\frac{\log r_1}{\log R},\dots,\frac{\log r_k}{\log R}\right).$$

For any $(d_1,\dots,d_k)\in\mathcal{D}_k(\mathcal{L})$, define

$$\lambda_{(d_1,\dots,d_k)}(\mathcal{L}) := \mu(d_1\dots d_k)d_1\dots d_k \sum_{d_i|r_i \text{ for } i=1,\dots,k} \frac{y_{(r_1,\dots,r_k)}(\mathcal{L})}{\varphi_{\omega_{\mathcal{L}}}(r_1\dots r_k)},$$

and then define the function $w=w_{k,\mathcal{L},B,R}:\mathbb{Z}\to\mathbb{R}^+$ by

$$w(n) := \left(\sum_{d_1,\dots,d_k: d_i|L_i(n) \text{ for all } i} \lambda_{(d_1,\dots,d_k)}(\mathcal{L})\right)^2. \tag{7.4}$$

We note that the restriction of the support of F to $\mathcal{R}_k$ means that $\lambda_{(d_1,\dots,d_k)}(\mathcal{L})$ and $y_{(r_1,\dots,r_k)}$ are supported on the set

$$\mathcal{S}_k(\mathcal{L}) = \mathcal{D}_k(\mathcal{L})\cap\{(d_1,\dots,d_k): \prod_{i=1}^k d_i \leqslant R\}.$$

We then have the following result, a slightly modified form of Proposition 6.1 from [34]:

Theorem 6. *Fix $\theta,\alpha>0$. Then there exists a constant C depending only on θ,α such that the following holds. Suppose that $(\mathcal{A},\mathcal{L},\mathscr{P},B,x,\theta)$ obeys Hypothesis 1 at some subset $\mathcal{L}'$ of $\mathcal{L}$. Write $k:=\#\mathcal{L}$, and suppose that $x\geqslant C$, $B\leqslant x^\alpha$, and $C\leqslant k\leqslant \log^{1/5}x$. Moreover, assume that the coefficients a_i,b_i of the linear forms $L_i(n)=a_in+b_i$ in $\mathcal{L}$ obey the size bound $|a_i|,|b_i|\leqslant x^\alpha$ for all $i=1,\dots,k$. Then there exists a smooth function $F:\mathbb{R}^k\to\mathbb{R}$ depending only on k and supported on the simplex $\mathcal{R}_k$, and quantities I_k,J_k depending only on k with*

$$I_k \gg (2k\log k)^{-k}$$

and

$$J_k \asymp \frac{\log k}{k} I_k \tag{7.5}$$

such that, for $w(n)$ given in terms of F as above, the following assertions hold uniformly for $x^{\theta/10}\leqslant R\leqslant x^{\theta/3}$.

- *We have*

$$\sum_{n\in\mathcal{A}(x)} w(n) = \left(1+O\left(\frac{1}{\log^{1/10}x}\right)\right)\frac{B^k}{\varphi(B)^k}\mathfrak{S}(\mathcal{L})\#\mathcal{A}(x)(\log R)^kI_k. \tag{7.6}$$

- *For any linear form $L(n)=a_Ln+b_L$ in $\mathcal{L}'$ with a_L coprime to B and $L(n)>R$ on $[x,2x]$, we have*

$$\begin{aligned}\sum_{n\in\mathcal{A}(x)} 1_{\mathscr{P}}(L(n))w(n) &= \left(1+O\left(\frac{1}{\log^{1/10}x}\right)\right)\frac{\phi(|a_L|)}{|a_L|}\frac{B^{k-1}}{\varphi(B)^{k-1}}\mathfrak{S}(\mathcal{L})\#\mathscr{P}_{L,\mathcal{A}}(x)(\log R)^{k+1}J_k \\ &\quad + O\left(\frac{B^k}{\varphi(B)^k}\mathfrak{S}(\mathcal{L})\#\mathcal{A}(x)(\log R)^{k-1}I_k\right).\end{aligned} \tag{7.7}$$

- *Let $L(n) = a_0 n + b_0$ be a linear form such that the discriminant*

$$\Delta_L := |a_0| \prod_{j=1}^{k} |a_0 b_j - a_j b_0|$$

is non-zero (in particular L is not in $\mathcal{L}$). Then

$$\sum_{n \in \mathcal{A}(x)} 1_{\mathscr{P} \cap [x^{\theta/10}, +\infty)}(L(n)) w(n) \ll \frac{\Delta_L}{\varphi(\Delta_L)} \frac{B^k}{\varphi(B)^k} \mathfrak{S}(\mathcal{L}) \# \mathcal{A}(x) (\log R)^{k-1} I_k. \tag{7.8}$$

- *We have the crude upper bound*

$$w(n) \ll x^{2\theta/3 + o(1)} \tag{7.9}$$

for all $n \in \mathbb{Z}$.

Here all implied constants depend only on θ, α and the implied constants in the bounds of Hypothesis 1.

Proof. The first estimate (7.6) is given by [34, Proposition 9.1], (7.7) follows from [34, Proposition 9.2] in the case $(a_L, B) = 1$, (7.8) is given by [34, Proposition 9.4] (taking $\xi := \theta/10$ and $D := 1$), and the final statement (7.9) is given by part (iii) of [34, Lemma 8.5]. The bounds for J_k and I_k are given by [34, Lemma 8.6]. □

We remark that the estimate (7.8) is only needed here to establish the estimate (6.6) which is not, strictly speaking, necessary for the results of this paper, but will be useful in a subsequent work [15] based on this paper.

8. Verification of sieve estimates

We can now prove Theorem 5. Let $x, y, r, h_1, \dots, h_r$ be as in that theorem.

We set

$$\begin{aligned} \mathcal{A} &:= \mathbb{Z}, \\ \alpha &:= 2, \\ \theta &:= 1/3, \\ k &:= r, \\ R &:= (x/4)^{\theta/3}, \end{aligned}$$

and let $B = x^{o(1)}$ be the quantity from Lemma 7.2.

We define the function $w : \mathcal{P} \times \mathbb{Z} \to \mathbb{R}^+$ by setting

$$w(p, n) := 1_{[-y,y]}(n) w_{k, \mathcal{L}_p, B, R}(n)$$

for $p \in \mathcal{P}$ and $n \in \mathbb{Z}$, where $\mathcal{L}_p$ is the (ordered) collection of linear forms $n \mapsto n + h_i p$ for $i = 1, \dots, r$, and $w_{k, \mathcal{L}_p, B, R}$ was defined in (7.4). Note that the admissibility of the r-tuple $(h_1, \dots, h_r)$ implies the admissibility of the linear forms $n \mapsto n + h_i p$.

A key point is that many of the important components of $w_{k, \mathcal{L}_p, B, R}$ are essentially uniform in p. Indeed, for any prime s, the polynomial $\prod_{i=1}^{k} (n + h_i p)$ is divisible by s only at the residue classes $-h_i p \bmod s$. From this we see that

$$\omega_{\mathcal{L}_p}(s) = \#\{h_i \pmod s\} \text{ whenever } s \neq p.$$

In particular, $\omega_{\mathcal{L}_p}(s)$ is independent of p as long as s is distinct from p, so

$$\mathfrak{S}(\mathcal{L}_p) = \left(1 + O\left(\frac{k}{x}\right)\right)\mathfrak{S}, \tag{8.1}$$
$$\mathfrak{S}_{BW}(\mathcal{L}_p) = \left(1 + O\left(\frac{k}{x}\right)\right)\mathfrak{S}_{BW},$$

for some $\mathfrak{S}$, $\mathfrak{S}_{BW}$ independent of p, with the error terms uniform in p. Moreover, if $s \nmid WB$ then $s > 2k^2$, so all the h_i are distinct $\mathrm{mod} s$ (since the h_i are less than $2k^2$). Therefore, if $s \nmid pWB$ we have $\omega_{\mathcal{L}_p}(s) = k$ and

$$\{j_{s,1}(\mathcal{L}_p), \dots, j_{s,\omega(s)}(\mathcal{L}_p)\} = \{1, \dots, k\}.$$

Since all $p \in \mathcal{P}$ are at least $x/2 > R$, we have $s \neq p$ whenever $s \leqslant R$. From this we see that $\mathcal{D}_k(\mathcal{L}_p) \cap \{(d_1, \dots, d_k) : \prod_{i=1}^k d_i \leqslant R\}$ is independent of p, and so we have

$$\lambda_{(d_1,\dots,d_k)}(\mathcal{L}_p) = \frac{\mathfrak{S}(\mathcal{L}_p)}{\mathfrak{S}}\lambda_{(d_1,\dots,d_k)} = \left(1 + O\left(\frac{k}{x}\right)\right)\lambda_{(d_1,\dots,d_k)},$$

for some $\lambda_{(d_1,\dots,d_k)}$ independent of p, and where the error term is independent of $d_1, \dots, d_k$.

It is clear that w is non-negative and supported on $\mathcal{P} \times [-y, y]$, and from (7.9) we have (6.7). We set

$$\tau := 2\frac{B^k}{\varphi(B)^k}\mathfrak{S}(\log R)^k(\log x)^k I_k \tag{8.2}$$

and

$$u := \frac{\varphi(B)}{B}\frac{\log R}{\log x}\frac{kJ_k}{2I_k}. \tag{8.3}$$

Since B is either 1 or prime, we have

$$\frac{\varphi(B)}{B} \asymp 1,$$

and from the definition of R we also have

$$\frac{\log R}{\log x} \asymp 1. \tag{8.4}$$

From (7.5) we thus obtain (6.3). From [34, Lemma 8.1(i)] we have

$$\mathfrak{S} \geqslant x^{-o(1)},$$

and from [34, Lemma 8.6] we have

$$I_k = x^{o(1)},$$

and so we have the lower bound (6.2). (In fact, we also have a matching upper bound $\tau \leqslant x^{o(1)}$, but we will not need this.)

It remains to verify the estimates (6.4), (6.5) and (6.6). We begin with (6.4). Let p be an element of $\mathcal{P}$. We shift the n variable by $3\lfloor y \rfloor$ and rewrite

$$\sum_{n\in\mathbb{Z}} w(p,n) = \sum_{n\in\mathcal{A}(2\lfloor y\rfloor)} w_{k,\mathcal{L}_p-3\lfloor y\rfloor,B,R}(n)$$

where $\mathcal{L}_p - 3\lfloor y\rfloor$ denotes the set of linear forms $n \mapsto n + h_i p - 3\lfloor y\rfloor$ for $i = 1, \dots, k$. This set of linear forms remains admissible, and

$$\mathfrak{S}(\mathcal{L}_p - 3\lfloor y\rfloor) = \mathfrak{S}(\mathcal{L}_p) = \left(1 + O\left(\frac{k}{x}\right)\right)\mathfrak{S}.$$

The claim (6.4) now follows from (8.2) and the first conclusion (7.6) of Theorem 6 (with x replaced by $2\lfloor y\rfloor$, $\mathcal{L}' = \emptyset$, and $\mathcal{L} = \mathcal{L}_p - 3\lfloor y\rfloor$), using Lemma 7.2 to obtain Hypothesis 1.

Now we prove (6.5). Fix $q \in \mathcal{Q}$ and $i \in \{1, \dots, k\}$. We introduce the set $\tilde{\mathcal{L}}_{q,i}$ of linear forms $\tilde{L}_{q,i,1}, \dots, \tilde{L}_{q,i,k}$, where

$$\tilde{L}_{q,i,i}(n) := n$$

and

$$\tilde{L}_{q,i,j}(n) := q + (h_j - h_i)n \qquad (1 \leqslant j \leqslant k, j \neq i)$$

We claim that this set of linear forms is admissible. Indeed, for any prime $s \neq q$, the solutions of

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

are $n \equiv 0$ and $n \equiv -q(h_j - h_i)^{-1} \pmod{s}$ for $h_j \not\equiv h_i \pmod{s}$, the number of which is equal to $\#\{h_j \pmod{s}\}$. Thus,

$$\mathfrak{S}(\tilde{\mathcal{L}}_{q,i}) = \left(1 + O\left(\frac{k}{x}\right)\right)\mathfrak{S},$$
$$\mathfrak{S}_{BW}(\tilde{\mathcal{L}}_{q,i}) = \left(1 + O\left(\frac{k}{x}\right)\right)\mathfrak{S}_{BW},$$

as before. Again, for $s \nmid WB$ we have that the h_i are distinct $\pmod{s}$, and so if $s < R$ and $s \nmid WB$ we have $\omega_{\tilde{\mathcal{L}}_{q,i}}(s) = k$ and

$$\{j_{s,1}(\tilde{\mathcal{L}}_{q,i}), \dots, j_{s,\omega(s)}(\tilde{\mathcal{L}}_{q,i})\} = \{1, \dots, k\}.$$

In particular, $\mathcal{D}_k(\tilde{\mathcal{L}}_{q,i}) \cap \{(d_1, \dots, d_k) : \prod_{i=1}^k d_i \leqslant R\}$ is independent of q, i and so

$$\lambda_{(d_1,\dots,d_k)}(\tilde{\mathcal{L}}_{q,i}) = \left(1 + O\left(\frac{k}{x}\right)\right)\lambda_{(d_1,\dots,d_k)},$$

where again the $O(\frac{k}{x})$ error is independent of $d_1, \dots, d_k$. From this, since $q - h_i p$ takes values in $[-y, y]$, we have that

$$w_{k,\tilde{\mathcal{L}}_{q,i},B,R}(p) = \left(1 + O\left(\frac{k}{x}\right)\right) w_{k,\mathcal{L}_p,B,R}(q - h_i p)$$

whenever $p \in \mathcal{P}$ (note that the d_i summation variable implicit on both sides of this equation is necessarily equal to 1). Thus, recalling that $\mathcal{P} = \mathscr{P} \cap (x/2, x]$, we can write the left-hand side of (6.5) as

$$\left(1 + O\left(\frac{k}{x}\right)\right) \sum_{n\in\mathcal{A}(x/2)} 1_{\mathscr{P}}(\tilde{L}_{q,i,i}(n)) w_{k,\tilde{\mathcal{L}}_{q,i},B,R}(n).$$

Applying the second conclusion (7.7) of Theorem 6 (with x replaced by $x/2$, $\mathcal{L}' = \{\tilde{L}_{q,i,i}\}$, and $\mathcal{L} = \tilde{\mathcal{L}}_{q,i}$) and using Lemma 7.2 to obtain Hypothesis 1, this expression becomes

$$\left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{B^{k-1}}{\varphi(B)^{k-1}} \mathfrak{S} \# \mathscr{P}_{\tilde{L}_{q,i,i},\mathcal{A}}(x/2) (\log R)^{k+1} J_k$$
$$+ O\left(\frac{B^k}{\varphi(B)^k} \mathfrak{S} \# \mathcal{A}(x/2) (\log R)^{k-1} I_k\right).$$

Clearly $\#\mathcal{A}(x/2) = O(x)$, and from the prime number theorem one has

$$\# \mathscr{P}_{\tilde{L}_{q,i,i},\mathcal{A}}(x/2) = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{x}{2 \log x}.$$

for any fixed $C > 0$. Using (8.2), (8.3), we can thus write the left-hand side of (6.5) as

$$\left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{u}{k} \tau \frac{x}{2 \log^k x} + O\left(\frac{1}{\log R} \tau \frac{x}{\log^k x}\right).$$

From (6.1), (6.3), the second error term may be absorbed into the first, and (6.5) follows.

Finally, we prove (6.6). Fix $h = O(y/x)$ not equal to any of the h_i, and fix $p \in \mathcal{P}$. By the prime number theorem, it suffices to show that

$$\sum_{q \in \mathcal{Q}} w(p, q - hp) \ll \frac{1}{\log_2^{10} x} \tau \frac{y}{\log^k x}.$$

By construction, the left-hand side is the same as

$$\sum_{x - hp < n \leqslant y - hp} 1_{\mathscr{P}}(n + hp) w_{k,\mathcal{L}_p,B,R}(n)$$

which we can shift as

$$\sum_{n \in \mathcal{A}(\lfloor y \rfloor - \lfloor x \rfloor)} 1_{\mathscr{P} \cap [x^{\theta/10}, +\infty)}(n - \lfloor y \rfloor + 2\lfloor x \rfloor) w_{k,\mathcal{L}_p - \lfloor y \rfloor + 2\lfloor x \rfloor - hp,B,R}(n).$$

Applying (7.8), we may then bound this by

$$\ll \frac{\Delta}{\varphi(\Delta)} \frac{B^k}{\varphi(B)^k} \mathfrak{S}(\mathcal{L}_p - \lfloor y \rfloor + 2\lfloor x \rfloor - hp) y (\log R)^{k-1} I_k = \frac{\Delta}{\varphi(\Delta)} \frac{B^k}{\varphi(B)^k} \mathfrak{S}(\mathcal{L}_p) y (\log R)^{k-1} I_k$$

where

$$\Delta := \prod_{i=1}^{k} |hp - h_i p|.$$

Applying (8.1), (8.2), we may simplify the above upper bound as

$$\ll \frac{\Delta}{\varphi(\Delta)} \frac{y}{(\log R)(\log x)^k} \tau.$$

Now $h - h_i = O(y/x) = O(\log x)$ for each i, hence $\Delta \leqslant (O(x \log x))^k$, and it follows from (7.1), (8.4) and (6.1) that

$$\frac{\Delta}{\varphi(\Delta)} \ll \log_2 \Delta \ll \log_2 x \ll \frac{\log R}{\log_2^{10} x}.$$

This concludes the proof of Theorem 5, and hence Theorem 1.

REFERENCES

[1] M. Ajtai, J. Komlós, E. Szemerédi, *A dense infinite Sidon sequence*, European J. Combin. **2** (1981), no. 1, 1–11.

[2] R. J. Backlund, *Über die Differenzen zwischen den Zahlen, die zu den ersten n Primzahlen teilerfremd sind*, Commentationes in honorem E. L. Lindelöf. Annales Acad. Sci. Fenn. **32** (1929), Nr. 2, 1–9.

[3] R. C. Baker, T. Freiberg, *Limit points and long gaps between primes*, Quart. J. Math. Oxford **67** (2) (2016), 233–260.

[4] R. C. Baker, G. Harman and J. Pintz, *The difference between consecutive primes. II.*, Proc. London Math. Soc. (3) **83** (2001), no. 3, 532–562.

[5] A. Brauer, H. Zeitz, *Über eine zahlentheoretische Behauptung von Legendre*, Sitzungsberichte Berliner Math. Ges. **29** (1930), 116–125.

[6] N. G. de Bruijn, *On the number of positive integers $\leqslant x$ and free of prime factors $> y$.* Nederl. Acad. Wetensch. Proc. Ser. A. **54** (1951) 50–60.

[7] H. Cramér, *Some theorems concerning prime numbers*, Ark. Mat. Astr. Fys. **15** (1920), 1–33.

[8] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers,* Acta Arith. **2** (1936), 396–403.

[9] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics vol. 74, Springer-Verlag, New York, 2000.

[10] L. E. Dickson, History of the theory of numbers, vol. III, Carnegie Inst. of Washington, Washington, DC 1919, 1920, 1923.

[11] P. Erdős, *On the difference of consecutive primes*, Quart. J. Math. Oxford Ser. **6** (1935), 124–128.

[12] P. Erdős, *Some of my favourite unsolved problems*, in A Tribute to Paul Erdős (A. Baker, B. Bollobás, A. Hajnal, eds.), Cambridge Univ. Press, 1990, pp. 467–478.

[13] K. Ford. B. Green, S. Konyagin, T. Tao, *Large gaps between consecutive prime numbers*, Annals of Math. **183** (2016), 935–974.

[14] K. Ford. D. R. Heath-Brown, S. Konyagin, *Large gaps between consecutive prime numbers containing perfect powers*, Analytic Number Theory, in honor of Helmut Maier's 60th birthday, C. Pomerance and M. Th. Rassias (eds.), Springer-Verlag, 2015, pp. 83–92.

[15] K. Ford, J. Maynard, T. Tao, *Chains of large gaps between primes*, to appear in the book "Irregularities in the distribution of prime numbers", dedicated to Helmut Maier.

[16] S. Foucart and H. Rouhut, *A mathematical introduction to compressive sensing*, Birkhaüser, 2013.

[17] P. Frankl, V. Rödl, *Near perfect coverings in graphs and hypergraphs*, European J. Combin. **6** (1985), no. 4, 317–326.

[18] J. Friedlander, H. Iwaniec, Opera de cribro. American Mathematical Society Colloquium Publications, 57. American Mathematical Society, Providence, RI, 2010.

[19] P. X. Gallagher, *A large sieve density estimate near $\sigma = 1$*, Invent. Math. **11** (1970), 329–339.

[20] A. Granville, *Harald Cramér and the distribution of prime numbers*, Scandanavian Actuarial J. (1995), 12–28.

[21] B. J. Green and T. C. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Annals of Math. **175** (2012), no. 2, 465–540.

[22] B. J. Green and T. C. Tao, *Linear equations in primes,* Annals of Math. **171** (2010), no. 3, 1753–1850.

[23] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Annals of Math. **176** (2012), 1231–1372.

[24] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[25] G. H. Hardy, J. E. Littlewood, *Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes*, Acta Math. **44** (1923), 1–70.

[26] D. R. Heath-Brown, *Gaps between primes, and the pair correlation of zeros of the zeta function*, Acta Arith. **41** (1982), no. 1, 85–99.

[27] A. Hildebrand, H. Maier, *Irregularities in the distribution of primes in short intervals*, J. Reine Angew. Math. **397** (1989), 162–193.

[28] H. Iwaniec, *On the problem of Jacobsthal*, Demonstratio Math. **11** (1978), 225–231.

[29] J. Kahn, *A linear programming perspective on the Frankl-Rödl-Pippenger theorem*, Random Structures Algorithms **8** (1996), no. 2, 149–157

[30] J. Li, K. Pratt, and G. Shakan, *A lower bound for the least prime in an arithmetic progression*, peprint, `arXiv:1607.02543`

[31] H. Maier and M. Th. Rassias, *Large gaps between consecutive prime numbers containing perfect k-th powers of prime numbers*. J. Functional Anal., to appear. `arXiv:1512.03936`.

[32] H. Maier and C. Pomerance, *Unusually large gaps between consecutive primes*. Trans. Amer. Math. Soc. **322** (1990), no. 1, 201–237.
[33] J. Maynard, *Small gaps between primes*, Annals of Math. (2) **181** (2015), no. 1, 383–413.
[34] J. Maynard, *Dense clusters of primes in subsets*, Compos. Math. **152** (2016), no. 7, 1517–1554.
[35] J. Maynard, *Large gaps between primes*, Annals of Math. **183** (2016), 915–933.
[36] J. Pintz, *Very large gaps between consecutive primes.* J. Number Theory **63** (1997), no. 2, 286–301.
[37] N. Pippenger, J. Spencer, *Asymptotic behavior of the chromatic index for hypergraphs*, J. Combin. Theory Ser. A **51** (1989), no. 1, 24–42.
[38] D. H. J. Polymath, *Variants of the Selberg sieve, and bounded gaps between primes*, Research in the Mathematical Sciences 1:12 (2014).
[39] C. Pomerance, *A note on the least prime in an arithmetic progression*, J. Number Theory **12** (1980), no. 2, 218–223.
[40] R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.
[41] R. A. Rankin, *The difference between consecutive prime numbers. V,* Proc. Edinburgh Math. Soc. (2) 13 (1962/63), 331–332.
[42] V. Rödl, *On a packing and covering problem*, European J. Combin. **6** (1985), no. 1, 69–78.
[43] A. Schönhage, *Eine Bemerkung zur Konstruktion grosser Primzahllücken*, Arch. Math. **14** (1963), 29–30.
[44] T. Oliveira e Silva, S. Herzog, S. Pardi, *Empirical verification of the even Goldbach conjecture and computation of prime gaps up to* 4×10^{18}, Math. Comp. **83** (2014), 2033–2060.
[45] V. Vu, *New bounds on nearly perfect matchings in hypergraphs: higher codegrees do help*, Random Structures Algorithms **17** (2000), no. 1, 29–63.
[46] E. Westzynthius, *Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind,* Commentationes Physico–Mathematicae, Societas Scientarium Fennica, Helsingfors **5**, no. 25, (1931) 1–37.

DEPARTMENT OF MATHEMATICS, 1409 WEST GREEN STREET, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801, USA
E-mail address: ford@math.uiuc.edu

MATHEMATICAL INSTITUTE, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, ENGLAND
E-mail address: ben.green@maths.ox.ac.uk

STEKLOV MATHEMATICAL INSTITUTE, 8 GUBKIN STREET, MOSCOW, 119991, RUSSIA
E-mail address: konyagin@mi.ras.ru

MATHEMATICAL INSTITUTE, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, ENGLAND
E-mail address: james.alexander.maynard@gmail.com

DEPARTMENT OF MATHEMATICS, UCLA, 405 HILGARD AVE, LOS ANGELES CA 90095, USA
E-mail address: tao@math.ucla.edu