An explicit version of Chen’s theorem assuming the Generalized
Riemann Hypothesis

Matteo Bordignon∗ Valeriia Starichkova

November 17, 2022

Abstract

We prove that assuming the Generalized Riemann Hypothesis every even integer larger than
exp(\exp(15.85)) can be written as the sum of a prime number and a number that has at most two
prime factors.

1 Introduction

In [1], Johnston and the authors of this paper recently built upon unpublished work of Yamada [20] to
prove an effective and explicit variant of Chen’s theorem that holds for a limited set of natural numbers.
Namely, [1, Corollary 4] states the following.

Theorem 1. Every even integer \( N > \exp(\exp(34.5)) \) can be represented as the sum of a prime and a
square-free number \( \eta > 1 \) with at most two prime factors.

It is easy to draw a comparison between this result and Goldbach’s weak conjecture, also known as
the ternary Goldbach problem.

Theorem 2 (Vinogradov–Helfgott). For any odd number \( N \geq 7 \) there exist three primes \( p_1, p_2 \) and \( p_3 \),
such that

\[
N = p_1 + p_2 + p_3.
\]

This result was first proved by Vinogradov [19] with a non explicit constant and, after different
papers in this direction, proved for the conjectured range of natural numbers by Helfgott [10]. We are
mainly interested in two notable papers. Kaniecki showed in [12] that, under the Generalized Riemann
Hypothesis (GRH), every odd integer is a sum of at most five primes. Deshouillers, Effinger, te Riele
and Zinoviev demonstrated in [7] that, under GRH, Goldbach’s weak conjecture holds for all \( N > 7 \).
Drawing inspiration from these papers we will improve the range of Theorem 1 assuming GRH and
reworking its proof. With this aim we will use the new explicit and complete version of the prime
number theorem under GRH proven by Ernvall-Hytönen and Palojärvi [8]. We will now state our main
result.

Theorem 3. Let \( \pi_2(N) \) denote the number of representations of a given even integer \( N \) as the sum of
a prime number and a product of at most two prime factors. If \( N > \exp(\exp(15.85)) \), then assuming
GRH we have

\[
\pi_2(N) > 4 \cdot 10^{-4} \frac{U_N N}{\log^2 N},
\]

∗The research was supported by OP RDE project No. CZ.02.2.69/0.0/0.0/18_053/0016976 International mobility of
research, technical and administrative staff at Charles University.
where, with $\gamma$ the Euler–Mascheroni constant,

$$U_N = 2e^{\gamma} \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{p>2, p|N} \frac{p-1}{p-2}.$$ 

There are two main reasons why the assumption of GRH allows to improve upon the unconditional result from Theorem 1: the potential Siegel zeros contribute significantly to the error terms in [1], thus, the GRH simplifies and reduces the error terms we are dealing with in this work. Secondly, the error terms $r(d)$ appearing in the linear sieve (Theorem 4) can be expressed via the error terms from the prime number theorem in arithmetic progressions. The assumption of GRH lowers such error terms as shown in [8].

We will derive an easy corollary from Theorem 3.

**Corollary 1.** Every even integer $N > \exp(\exp(15.85))$ can be represented as the sum of a prime and a square-free number $\eta > 1$ with at most two prime factors.

We should note that the main obstruction at proving Corollary 1 for smaller $N$ can be found in the explicit version of the linear sieve, this suggests to the researchers interested in improving the explicit version of Chen’s theorem to focus on improving this result. Another obstruction, and thus a point of interest to improve the explicit version of Chen’s theorem, can be found in the size of $\epsilon(u)$ such that

$$\prod_{u \leq p < z} \left(1 - \frac{1}{p-1}\right)^{-1} < (1 + \epsilon(u)) \frac{\log z}{\log u},$$

see Lemma 8 for the bound we prove for such function.

An outline of the paper is as follows. In Section 2 we state several lemmas from the existing literature that will be used in the later sections of the paper. In Section 3 we introduce an explicit version of the linear sieve and some other preliminary results and definitions. In Section 4 we prove an explicit bound for a bilinear form assuming GRH. In Section 5 we first set up all the required preliminaries for sieve methods and introduce some related lemmas. In Subsections 5.1, 5.2 and 5.3 we obtain upper and lower bounds for the sifted integer sets. We conclude in Subsections 5.4 and 5.5 by proving Theorem 3 and Corollary 1.

## 2 General lemmas

We require some lemmas in the arguments that follow. Lemmas 2 and 3 are conditional on GRH, Lemmas 4 and 5 could have been slightly improved under GRH, though the improvement would not affect the final result much. Here and below $p$ will denote a prime number.

**Lemma 1.** [18, Corollary 1] If the Riemann hypothesis (RH) holds, then

$$|\pi(x) - \text{li}(x)| \leq \frac{\sqrt{x} \log x}{8\pi}, \quad \text{for} \quad 3 \leq x,$$

where

$$\text{li}(x) = \int_0^x \frac{dt}{\log t}.$$ 

**Lemma 2.** [18, Theorem 18] Let $\theta(x) = \sum_{p \leq x} \log p$. Then

$$\theta(x) \leq x \quad \text{for} \quad 0 < x \leq 10^{18}.$$
We now introduce the conditional result under GRH that will be fundamental in our argument. Here and below we assume \( N > 4 \cdot 10^{18} \), since [14] asserts the Chen’s Theorem is true unconditionally for \( N \) up to \( 4 \cdot 10^{18} \).

**Lemma 3.** Let \( d \) be a positive integer and \( X_2 \) be a real number such that \( 3 \leq d \leq \sqrt{X_2} \). Assume that GRH holds, then for all \( N \geq X_2 \) and integer \( a \) with \( \bar{a}_d(X_2) = |a| \neq 1 \)

\[
\left| \pi(N;d,a) - \frac{\text{li}(N)}{\varphi(d)} \right| \leq q_G(X_2) \sqrt{N \log N},
\]

where

\[
q_G(X_2) = 0.165 + \frac{12.683}{\log X_2} + \frac{254.980}{\log^2 X_2} + \frac{2607.854}{\log^3 X_2} + \frac{11605.056}{\log^4 X_2} + \frac{1.314 \log X_2}{X_2^{1/4}} + \frac{0.092 \log \log X_2}{X_2^{1/4}}
\]

\[+ \frac{60.883}{X_2^{1/4}} + \frac{8.250 \log \log X_2}{X_2^{1/4} \log X_2} + \frac{939.260}{X_2^{1/4} \log^2 X_2} - \frac{237.934}{X_2^{1/2} \log X_2} \leq 0.640 \quad \text{for} \quad X_2 \geq 4 \cdot 10^{18}.
\]

**Proof.** Apply \( q \leq \sqrt{N} \) to [8, Theorem 1].

We will use the two following unconditional results as they suffice for our aim.

**Lemma 4.** [15, Theorem 11] Let \( \omega(p) \) count the number of prime divisors of an integer \( N \) without multiplicities, then

\[
\omega(N) < \frac{1.3841 \log N}{\log \log N} \quad \text{for} \quad N \geq 3.
\]

**Lemma 5** (8 (4.6) and Lemma 4.5]. For \( x \geq 10^9 \),

\[
\sum_{n \leq x} \mu^2(n) \leq 0.608 x,
\]

\[
\sum_{n \leq x} \frac{\mu^2(n)}{\varphi(n)} \leq \log x + 1.333 + \frac{58}{\sqrt{x}} \leq 1.1 \log x.
\]

## 3 An explicit formula for the linear sieve

We now introduce the version of the linear sieve proven by Johnston and the authors in [1, Theorem 6 & Table 1].

**Theorem 4.** Let \( A = \{a(n)\}_{n=1}^\infty \) be an arithmetic function such that

\[
a(n) \geq 0 \quad \text{for all} \quad n, \quad \text{and} \quad |A| = \sum_{n=1}^\infty a(n) < \infty.
\]

Let \( \mathbb{P} \) be a set of prime numbers, and for \( z \geq 2 \) let

\[P(z) = \prod_{\substack{p \in \mathbb{P} \\ p < z}} p.
\]

Let

\[S(A, \mathbb{P}, z) = \sum_{n=1}^\infty a(n).
\]
For every \( n \geq 1 \), let \( g_n(d) \) be a multiplicative function such that
\[
0 \leq g_n(p) < 1 \quad \text{for all } p \in \mathbb{P}.
\]

Define \( r(d) \) by
\[
|A_d| = \sum_{n=1}^{\infty} a(n) = \sum_{n=1}^{\infty} a(n)g_n(d) + r(d).
\]

Let \( Q \subseteq \mathbb{P} \), and \( Q \) be the product of primes from \( Q \). Suppose that, for some \( \epsilon \) such that \( 0 < \epsilon < 1/63 \), the inequality
\[
\prod_{p \in P \setminus Q \atop u \leq p < z} (1 - g_n(p))^{-1} \leq (1 + \epsilon) \frac{\log z}{\log u},
\]
holds for all \( n \) and \( 1 < u < z \). Then, for any \( D \geq z \) we have an upper bound
\[
S(A, P, z) < (F(s) + \epsilon C_1(\epsilon) e^2 h(s))X + R,
\]
and for any \( D \geq z^2 \) we have a lower bound
\[
S(A, P, z) > (f(s) - \epsilon C_2(\epsilon) e^2 h(s))X - R,
\]
where
\[
s = \frac{\log D}{\log z},
\]
\[
h(s) = \begin{cases} 
  e^{-2} & 1 \leq s \leq 2 \\
  e^{-s} & 2 < s \leq 3 \\
  3s^{-1}e^{-s} & s \geq 3,
\end{cases}
\]
\( f(s) \) and \( F(s) \) are two functions defined in [13, (9.27) and (9.28)], \( C_{1,2}(\epsilon) \) come from Table 1,
\[
X = \sum_{n=1}^{\infty} a(n) \prod_{p \in P(z)} (1 - g_n(p)),
\]
and the remainder term is
\[
R = \sum_{d \mid P(z) \atop d < QD} |r(d)|.
\]

If there is a multiplicative function \( g(d) \) such that \( g_n(d) = g(d) \) for all \( n \), then
\[
X = V(z)|A|, \quad \text{where} \quad V(z) = \prod_{p \mid P(z)} (1 - g(p)).
\]

For our applications we will choose \( g_n \) such that \( g_n(p) = \frac{1}{p^n} \). With this choice we will provide the value of \( \epsilon \), for which (2) is satisfied under certain conditions. The smaller \( \epsilon \) we will be able to take, the better lower bound for \( N \) we will achieve. Lemma 8 provides the value of \( \epsilon \) we will use in the paper. Any improvement of the lemma would lead to a strong improvement in the range of \( N \) in Theorem 3 and it is thus something that would be interesting to pursue with a computational approach.

**Lemma 6.** For all \( x \geq 9551 \), there exists a prime in the interval \([0.996x, x]\).
Table 1: Values for $C_1(\epsilon)$ and $C_2(\epsilon)$

| $\epsilon^{-1}$ | $C_1(\epsilon)$ | $C_2(\epsilon)$ | $\epsilon^{-1}$ | $C_1(\epsilon)$ | $C_2(\epsilon)$ | $\epsilon^{-1}$ | $C_1(\epsilon)$ | $C_2(\epsilon)$ |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| 63               | 32881            | 32875            | 75               | 678              | 679              | 143              | 198              | 200              |
| 64               | 7582             | 7580             | 76               | 635              | 636              | 200              | 162              | 164              |
| 65               | 3890             | 3890             | 77               | 598              | 600              | 249              | 149              | 150              |
| 66               | 2542             | 2542             | 78               | 566              | 568              | 300              | 141              | 142              |
| 67               | 1880             | 1881             | 79               | 538              | 540              | 400              | 132              | 134              |
| 68               | 1480             | 1500             | 80               | 514              | 515              | 500              | 127              | 129              |
| 69               | 1254             | 1255             | 81               | 492              | 493              | 600              | 124              | 126              |
| 70               | 1084             | 1086             | 84               | 438              | 439              | 700              | 122              | 124              |
| 71               | 960              | 962              | 87               | 398              | 400              | 800              | 121              | 122              |
| 72               | 865              | 867              | 93               | 341              | 343              | 900              | 120              | 121              |
| 73               | 790              | 791              | 99               | 303              | 305              | 1000             | 119              | 120              |
| 74               | 729              | 730              | 114              | 247              | 249              | 1100             | 118              | 120              |

Proof. We use [14, Table 8] for $9551 \leq x \leq 4 \cdot 10^{18}$ and [11, Theorem 1.1] for $x > 4 \cdot 10^{18}$. \hfill \Box

By increasing the value of $x$ one can obtain much stronger results than Lemma 6. See [11] and [5] for the recent improvements on the length of the intervals containing primes.

Lemma 7. We have

\[
\sum_{p \leq x} \frac{1}{p} \geq \log \log x + M, \quad \text{for} \quad 3 \leq x \leq 10^8,
\]

\[
\sum_{p \leq x} \frac{1}{p} \geq \log \log x + M - \frac{1.4998 \cdot 10^{-4}}{\log x}, \quad \text{for} \quad x > 10^8,
\]

\[
\sum_{p \leq x} \frac{1}{p} \leq \log \log x + M + \frac{4.47 \cdot 10^{-9}}{\log x}, \quad \text{for} \quad x > \exp(1000).
\]

(3)

Proof. By [16] (4.20) we have

\[
\sum_{p \leq x} \frac{1}{p} = \log \log x + M + \frac{\theta(x) - x}{x \log x} + \int_x^\infty \frac{(y - \theta(y))(1 + \log y)}{y^2 \log^2 y} dy,
\]

(4)

where $\theta$ is Chebyshev’s theta function. To obtain (3), we note that

\[
\sum_{p \leq x} \frac{1}{p} \leq \log \log x + M + \frac{M_1}{\log x} \left(1 + \frac{3}{2 \log x}\right),
\]

where $M_1$ is from [2, Table 15] and depends on the lower bound for $x$. In particular, $x > \exp(1000)$ gives

\[
\sum_{p \leq x} \frac{1}{p} \leq \log \log x + M + \frac{4.47 \cdot 10^{-9}}{\log x}.
\]

The lower bound for $3 \leq x \leq 10^8$ follows from [16, Theorem 20]. To prove the lower bound for $x > 10^8$ we use again [2, Table 15] for $x = 10^8$ and get

\[
\left|\frac{\theta(x) - x}{x \log x}\right| \leq \frac{2.7457 \cdot 10^{-3}}{\log^2 x} \leq \frac{1.491 \cdot 10^{-4}}{\log x}.
\]

(5)
for $x > 10^8$. We now split into the cases $x \leq 10^{10}$ and $x > 10^{10}$. In the first case, we have

$$\int_{x}^{\infty} \frac{(y - \theta(y))(1 + \log y)}{y^2 \log^2 y} \, dy = \int_{x}^{10^{10}} \frac{(y - \theta(y))(1 + \log y)}{y^2 \log^2 y} \, dy + \int_{10^{10}}^{\infty} \frac{(y - \theta(y))(1 + \log y)}{y^2 \log^2 y} \, dy.$$ 

The first integral on the right-hand side is non-negative by [4, Theorem 2]. For the second integral we again use [2, Table 15] to obtain

$$\left| \int_{10^{10}}^{\infty} \frac{(y - \theta(y))(1 + \log y)}{y^2 \log^2 y} \, dy \right| \leq 8.6315 \cdot 10^{-7} \left( \frac{1}{2 \log^2(10^{10})} + \frac{1}{\log(10^{10})} \right) \leq 8.6315 \cdot 10^{-7} \left( \frac{1}{2 \log(10^{10})} + 1 \right) \frac{1}{\log x} \leq \frac{8.73016 \cdot 10^{-7}}{\log x}.$$ 

for $x \leq 10^{10}$. We combine the last bound with (5) to get the lower bound for $x > 10^8$.

To get the lower bound for $x > 10^{10}$ we just note that by [2, Table 15],

$$\left| \int_{x}^{\infty} \frac{(y - \theta(y))(1 + \log y)}{y^2 \log^2 y} \, dy \right| \leq 8.6315 \cdot 10^{-7} \left( \frac{1}{2 \log^2 x} + \frac{1}{\log x} \right) \leq 8.6315 \cdot 10^{-7} \left( \frac{1}{2 \log(10^{10})} + 1 \right) \frac{1}{\log x} \leq \frac{8.73016 \cdot 10^{-7}}{\log x}.$$ 

It is important to mention that the unconditional bounds above, being based on the computational results from [2], are better for smaller values of $x$ than the conditional bounds [13, Corollary 2] as this last result aims to give an asymptotic explicit improvement which isn’t necessarily optimal in the initial range. While it should be possible to blend these two results together we believe this would not give a strong improvement in the range we are interested in, and we will thus not further pursue this idea.

**Lemma 8.** Let $z > \exp(1000)$. Then for all $9551 \leq u < z$ we have

$$\prod_{u \leq p < z} \left( 1 - \frac{1}{p-1} \right)^{-1} < (1 + \epsilon(u)) \frac{\log z}{\log u},$$

with

$$\epsilon(u) = \left( 1 + \frac{1}{0.996u - 1} \right) \left( 1 + \frac{1.5 \cdot 10^{-4}}{\log u} + \frac{2.25 \cdot 10^{-8}}{\log^2 u} \right) \left( 1 + \frac{1}{u - 1} + \frac{1}{(u-1)^2} \right) - 1. \quad (6)$$

**Proof.** We first note that

$$\prod_{u \leq p < z} \left( 1 - \frac{1}{p-1} \right)^{-1} = \prod_{u \leq p < z} \frac{(p-1)^2}{p(p-2)} \prod_{u \leq p < z} \left( 1 - \frac{1}{p} \right)^{-1}.$$
By Lemma 6 we then have
\[
\prod_{u \leq p < z} \left( \frac{(p-1)^2}{p(p-2)} \right) = \prod_{u \leq p < z} \left( 1 + \frac{1}{p(p-2)} \right) \\
\leq \prod_{0.996u \leq p < z} \left( 1 + \frac{1}{p^2} \right) \\
\leq \prod_{0.996u < p < z} \left( 1 + \frac{1}{p^2} \right) \\
\leq 1 + \sum_{n > 0.996u} \frac{1}{n^2} \leq 1 + \frac{1}{0.996u - 1}.
\]

Thus,
\[
\prod_{u \leq p < z} \left( 1 - \frac{1}{p} \right)^{-1} < \left( 1 + \frac{1}{0.996u - 1} \right) \prod_{u \leq p < z} \left( 1 - \frac{1}{p} \right)^{-1}.
\]

Next, we note that
\[
\prod_{u \leq p < z} \left( 1 - \frac{1}{p} \right) = \exp \left( - \sum_{u \leq p < z} \log \left( 1 - \frac{1}{p} \right) \right).
\]

Now, by Lemma 7
\[
\sum_{u < p < z} \frac{1}{p} \leq \sum_{p < u} \frac{1}{p} - \sum_{p < u} \frac{1}{p} \leq \log \log z - \log \log u + \frac{1.5 \cdot 10^{-4}}{\log u}.
\]

Using (9), (10) and that for \(0 < x \leq 1/2\),
\[
\log(1 - x) \geq -x - x^2, \quad e^x \leq 1 + x + x^2,
\]
we have,
\[
\prod_{u \leq p < z} \left( 1 - \frac{1}{p} \right)^{-1} \leq \frac{\log z}{\log u} \exp \left( \frac{1.5 \cdot 10^{-4}}{\log u} \right) \exp \left( \sum_{p > u} \frac{1}{p^2} \right) \\
\leq \frac{\log z}{\log u} \left( 1 + \frac{1.5 \cdot 10^{-4}}{\log u} + \frac{2.25 \cdot 10^{-8}}{\log^2 u} \right) \exp \left( \sum_{n > u} \frac{1}{n^2} \right) \\
\leq \frac{\log z}{\log u} \left( 1 + \frac{1.5 \cdot 10^{-4}}{\log u} + \frac{2.25 \cdot 10^{-8}}{\log^2 u} \right) \left( 1 + \frac{1}{u - 1} + \frac{1}{(u - 1)^2} \right).
\]

Using (7), (10) and \(u > 10^8\) then gives the desired result. \(\square\)

4 An explicit bilinear form assuming GRH

We modify the explicit upper bound for the bilinear form from [1] under GRH.

**Lemma 9** ([1, Lemma 31]). Let \(a(n)\) be an arithmetic function with \(|a(n)| \leq 1\) for all \(n\). Let \(X, Y, Z \geq 25\) be real numbers with \(Z < Y\), and \(q_G\) be as in Lemma 3. Let \(10^9 \leq D^s\) be such that
\[
\left( \frac{\sqrt{Y} \log D^s}{q_G(Y) \log Y \log 2} \right)^{2/3} \leq Z,
\]
then for all even \( a \in \mathbb{Z} \)

\[
\sum_{d < D^*} \mu^2(d) \max \left| \sum_{n < X} \sum_{Z \in p < Y \atop np \equiv a \pmod{d}} a(n) - \frac{1}{\varphi(d)} \sum_{n < X} \sum_{Z \in p < Y \atop (up,d)=1} a(n) \right| \leq m(X,Y,D^*) \frac{XY \log^{5/3} D^* \log^{1/3} Y}{Y^{1/6}},
\]

where

\[
m(X,Y,D^*) = 2.81 \frac{q^{1/3}_G(Y)}{q^{1/3}_G(Y) Y^{1/3}} + \frac{2.809}{q^{1/3}_G(Y) Y^{1/3}} + \frac{1.721}{q^{1/3}_G(Y) Y^{2/3}} Y \log \log D^*
\]

\[+ 5.498 \left( \frac{1}{\sqrt{X}} + \frac{1}{\sqrt{Y}} \right) \frac{Y^{1/6} \log^{1/3} D^*}{\log^{1/3} Y} + 19.044 \frac{D^* \log^{1/3} D^*}{XY^{5/6} \log^{1/3} Y}.
\]

We note that \( m(X,Y,D^*) \) is decreasing in \( X,Y \) and increasing in \( D^* \).

**Proof.** By the orthogonality property of the Dirichlet characters

\[
\sum_{n < X} \sum_{Z \in p < Y \atop np \equiv a \pmod{d}} a(n) = \sum_{n < X} \sum_{Z \in p < Y} \frac{a(n)}{\varphi(d)} \sum_{\chi \pmod{d}} \chi(a) \chi(np)
\]

\[= \frac{1}{\varphi(d)} \sum_{\chi \pmod{d}} \sum_{n < X} a(n) \chi(n) \sum_{Z \in p < Y} \chi(p)
\]

\[= \frac{1}{\varphi(d)} \sum_{\chi \neq \chi_{0,s} \pmod{d}} \sum_{n < X} a(n) \chi(n) \sum_{Z \in p < Y} \chi(p) + \frac{1}{\varphi(d)} \sum_{n < X} \sum_{Z \in p < Y \atop (up,d)=1} a(n),
\]

so the bilinear form is bounded by

\[
\frac{\mu^2(d)}{\varphi(d)} \sum_{\chi \neq \chi_{0,s} \pmod{d}} \left| \sum_{n < X} a(n) \chi(n) \right| \left| \sum_{Z \in p < Y} \chi(p) \right|.
\]

Every non-trivial character \( \chi \pmod{d} \) can be uniquely factorized into \( \chi = \chi_{0,s} \chi_1 \), with \( d = sr, r \neq 1, \chi_{0,s} \) the principal character mod \( s \) and \( \chi_1 \) a primitive character mod \( r \). Thus the upper bound above can be rewritten as follows

\[
\sum_{rs \in D^* \atop r \neq 1} \frac{\mu^2(rs)}{\varphi(rs)} \sum_{\chi \equiv 0 \pmod{r}}^{*} \sum_{n < X} a(n) \chi(n) \left| \sum_{Z \in p < Y} \chi(p) \right|,
\]

where \(* \) means that the sum is restricted to primitive characters. Let us note that conditions \( r \neq 1, (r,a) = 1 \) imply \( r \geq 3 \) for even \( a \). We begin by estimating the sum restricted to \( 3 \leq r \leq D_0 \), with \( D_0 = \sqrt{X} \) to be defined later. In this case

\[
\sum_{Z \leq p < Y} \chi(p) = \sum_{a \pmod{r}} \chi(a) \sum_{Z \leq p < Y \atop p \equiv a \pmod{r}} 1 = \sum_{a \pmod{r}} \chi(a)(\pi(Y;r,a) - \pi(Z;r,a)) + E(r),
\]
with $|E(r)| \leq 2\varphi(r)$ covering the potential cases when $Y$ or $Z$ is prime. Thus by Lemma 5

$$\left| \sum_{Z \leq p < Y} \chi(p) \right| \leq \left| \sum_{a \mod r} \chi(a) \left( \frac{\text{li}(Y) - \text{li}(Z)}{\varphi(r)} \right) \right| + \varphi(r) \max \left\{ q_G(Z)\sqrt{Z \log Z}, q_G(Y)\sqrt{Y \log Y} \right\} + 2\varphi(r)$$

$$= \varphi(r) \left( 2 + \max \left\{ q_G(Z)\sqrt{Z \log Z}, q_G(Y)\sqrt{Y \log Y} \right\} \right)$$

$$\leq \varphi(r) \left( 2 + q_G(Y)\sqrt{Y \log Y} \right) \quad \text{for} \quad 25 \leq Z < Y,$$

hence

$$\sum_{\chi \mod r} \left| \sum_{Z \leq p < Y} \chi(p) \right| \leq \varphi^2(r) \left( 2 + q_G(Y)\sqrt{Y \log Y} \right) + \sum_{\chi \mod r} \omega(s) \leq D_0 \left( \frac{1.3841 \log D^*}{\log \log D^*} + 2D_0 + D_0 q_G(Y)\sqrt{Y \log Y} \right).$$

By Lemma 5, we get

$$\sum_{r \leq D^*} \frac{\mu^2(rs)}{\varphi(rs)} \left| \sum_{n < X} \frac{a(n)\chi(n)}{\varphi(n)} \right| \leq D_0 \chi \left( \frac{1.3841 \log D^*}{\log \log D^*} + 2D_0 + D_0 q_G(Y)\sqrt{Y \log Y} \right) \left( \sum_{l \leq D^*} \frac{\mu^2(l)}{\varphi(l)} \right)$$

$$\leq 1.1D_0 \chi \frac{\log D^*}{\log \log D^*} + 2D_0 + D_0 q_G(Y)\sqrt{Y \log Y}. \quad (14)$$

We are now left with estimating the sum in (12) restricted to $r \geq D_0$, which can be bounded by

$$\sum_{s < D^*} \frac{\mu^2(s)}{\varphi(s)} \left| \sum_{D_0 < r < D^*} \frac{1}{\varphi(r)} \right| \sum_{n < X} \frac{a(n)\chi(n)}{\varphi(n)} \left| \sum_{Z \leq p < Y} \chi(p) \right| \leq D^* \chi \left( \frac{1.3841 \log D^*}{\log \log D^*} + 2D_0 + D_0 q_G(Y)\sqrt{Y \log Y} \right). \quad (15)$$

To do so, we divide the interval $D_0 \leq r \leq D^*$ into dyadic subintervals

$$D_k \leq r \leq 2D_k, \quad \text{where} \quad D_k = 2^k D_0, \quad 0 \leq k \leq \frac{\log(D^*/D_0)}{\log 2}.$$

Using Cauchy’s inequality and the large sieve inequality [4, Theorem 4, p. 160] we obtain

$$\sum_{D_k \leq r < 2D_k} \frac{1}{\varphi(r)} \left| \sum_{n < X} \frac{a(n)\chi(n)}{\varphi(n)} \right| \leq \frac{1}{D_k} \sum_{D_k \leq r < 2D_k} \chi \left( \frac{r}{\varphi(r)} \right)^{1/2} \left| \sum_{n < X} \frac{a(n)\chi(n)}{\varphi(n)} \right|^{1/2} \leq \sum_{Z \leq p < Y} \chi(p).$$
We choose the optimal value for $D$, whereas (17) is bounded by

$$
\frac{1}{D_k} \left( \sum_{D_k \leq r < 2D_k} \frac{r}{\varphi(r)} \sum_{n \leq X \pmod{r}} a(n) \chi(n) \left| \sum_{n \leq X \pmod{r}} \chi(n) \right|^2 \right)^{1/2} \cdot \left( \sum_{D_k \leq r < 2D_k} \frac{r}{\varphi(r)} \sum_{n \leq X \pmod{r}} a(n) \chi(n) \left| \sum_{Z \leq p < Y \pmod{r}} \chi(p) \right|^2 \right)^{1/2}
$$

$$
\leq \frac{1}{D_k} ((X + 2D_k^2)(Y + 2D_k^2)XY)^{1/2}
$$

$$
\leq \left( 144(D^*)^2 + 12(X + Y) + \frac{XY}{D_0^2} \right)^{1/2} (XY)^{1/2}
$$

$$
\leq \left( 12D^* + 2\sqrt{3}(\sqrt{X} + \sqrt{Y})\sqrt{XY} + \frac{XY}{D_0} \right).
$$

(16)

Using the bound above for $0 \leq k \leq \frac{\log(D^*/D_0)}{\log 2}$ and Lemma 5 we can bound (15) as follows

$$
\sum_{s < D^*} \frac{\mu^2(s)}{\varphi(s)} \sum_{D_0 < r < D^*} \frac{1}{\varphi(r)} \sum_{\chi \pmod{r}} a(n) \chi(n) \left| \sum_{n \leq X \pmod{r}} \chi(n) \right| \left| \sum_{Z \leq p < Y \pmod{r}} \chi(p) \right| \leq \frac{1}{\log 2} \log^2 D^* \left( 12D^* + 2\sqrt{3}(\sqrt{X} + \sqrt{Y})\sqrt{XY} + \frac{XY}{D_0} \right).
$$

(17)

We choose the optimal value for $D_0$ by making the asymptotically main terms of (14) and (17) equal, and get

$$
D_0 = \left( \frac{\sqrt{Y} \log D^*}{q_G(Y) \log Y} \right)^{1/3}.
$$

Then (14) is less than

$$
\frac{XY \log^{5/3} D^* \log^{1/3} Y}{Y^{1/6}} \left( 1.405 q_G^{1/3}(Y) + \frac{2.809}{q_G^{2/3}(Y) \sqrt{Y} \log Y} + \frac{1.721 \log^{2/3} D^*}{q_G^{1/3}(Y) \sqrt{Y} \log \log D^*} \right),
$$

whereas (17) is bounded by

$$
\frac{XY \log^{5/3} D^* \log^{1/3} Y}{Y^{1/6}} \left( 1.405 q_G^{1/3}(Y) + 5.498 \left( \frac{1}{\sqrt{X}} + \frac{1}{\sqrt{Y}} \right) \frac{Y^{1/6} \log^{1/3} D^*}{\log^{1/3} Y} + 19.044 \frac{D^* \log^{1/3} D^*}{XY^{5/6} \log^{1/3} Y} \right).
$$

The bounds above with (12) allow us to complete the proof of the lemma.

\[ \square \]

5 Outline of the proof

We will use the ideas from \[p1\] and from \[p3\] and keep notations form these works.

Some of the definitions that will be presented in this section were previously introduced. We decided to include them here to ease readability and make this section self-contained.
Fix \( z = N^\frac{1}{\pi} \), \( y = N^\frac{1}{\pi} \).

We shall consider the sets
\[
\mathcal{P} = \{ p \mid N \}, \quad A = \{ N - p : p \leq N, p \in \mathcal{P} \}, \quad A_p = \{ a \in A : p \mid a \}, \quad A_d = \bigcap_{p \mid d} A_p,
\]
\[
B = \{ N - p_1 p_2 p_3 : z \leq p_1 < y \leq p_2 < p_3, p_1 p_2 p_3 < N, (p_1 p_2 p_3, N) = 1 \},
\]
where \( p, p_1, p_2, p_3 \) denote prime numbers. We clearly have \(|A| = \pi(N) - \omega(N)\) and \(|A_d| = \pi(N; d, N) - \omega(N; d, N)\), where \( \omega(n; q, a) \) denotes the number of prime factors of \( n \) equal \( a \mod q \). We also set
\[
P(z) = \prod_{p < z, p \in \mathcal{P}} p, \quad V(z) = \prod_{p \mid P(z)} \left( 1 - \frac{1}{p - 1} \right), \quad S(A, n) = \left| A - \bigcup_{p \mid n} A_p \right|.
\]

The following result of Chen is a key component in the proof of our main theorem.

**Lemma 10.** [13, Theorem 10.2]

\[
\pi_2(N) > S(A, \mathcal{P}, z) - \frac{1}{2} \sum_{z < q < y} S(A_q, \mathcal{P}, z) - \frac{1}{2} S(B, \mathcal{P}, y) - 2N^\frac{1}{\pi} - N^\frac{1}{\pi^2}.
\]

Thus, to prove Theorem 3 it suffices to give a good lower bound for \( S(A, \mathcal{P}, z) \) and upper bounds for \( S(B, \mathcal{P}, z) \) and \( S(A_q, \mathcal{P}, z) \) for each prime \( q \) with \( z \leq q < y \) using the linear sieve. We will now introduce some useful lemmas in the context of Chen’s theorem.

**Lemma 11.** For \( x \geq 8 \) we have
\[
V(x) = \frac{U_N}{\log x} \left( 1 + \frac{\theta(3 \log x + 5)}{8\pi \sqrt{x}} \right) \left( 1 + \frac{2\theta}{x} \right) \left( 1 + \frac{8\theta \log N}{x} \right) \left( 1 + \frac{\theta}{x - 1} \right),
\]
where \( |\theta| \leq 1 \) and
\[
U_N = 2e^\gamma \prod_{p > 2} \left( 1 - \frac{1}{(p - 1)^2} \right) \prod_{p \mid N} \frac{p - 1}{p - 2}.
\]

In particular, when \( x = N^{1/8} \geq (4 \cdot 10^{18})^{1/8} \), we have
\[
\frac{U_N}{\log x} \left( 1 - \frac{0.62 \log N}{N^{\frac{1}{10}}} \right) < V(x) < \frac{U_N}{\log x} \left( 1 + \frac{0.62 \log N}{N^{\frac{1}{10}}} \right)
\]
and, when \( x = N^{1/3} \geq (4 \cdot 10^{18})^{1/3} \), we have
\[
\frac{U_N}{\log x} \left( 1 - \frac{0.06 \log N}{N^{\frac{1}{5}}} \right) < V(x) < \frac{U_N}{\log x} \left( 1 + \frac{0.06 \log N}{N^{\frac{1}{5}}} \right).
\]

**Proof.** Follows from the proof of [13, Theorem 10.3], making use of [18, Corollary 3].

**Lemma 12.** For all \( N \geq \exp(\exp(15)) \) we have
\[
\frac{N}{\log N} \leq |A| \leq \left( 1 + 4 \cdot 10^{-7} \right) \cdot \frac{N}{\log N}.
\]

\(^1\)We can assume \( N > \exp(\exp(15)) \) since the final bound for \( N \) is higher.
Proof. By definition of the set $A$, for $N \geq 3$,
\[ ||A| - \pi(N)|| \leq \omega(N) \leq \frac{1.3841 \log N}{\log \log N}, \]
hence by (1),
\[ \text{li}(N) - \frac{\sqrt{N \log N}}{8\pi} - \frac{1.3841 \log N}{\log \log N} \leq |A| \leq \text{li}(N) + \frac{\sqrt{N \log N}}{8\pi} + \frac{1.3841 \log N}{\log \log N}. \]
The statement of the lemma now follows from the bound for $\text{li}(N)$ below
\[ 1 + \frac{1}{\log N} \leq \text{li}(N) \leq 1 + \frac{1}{\log N} + \frac{3}{\log^2 N} \quad \text{for} \quad N \geq \exp(11), \]
combined with $N \geq \exp(\exp(15))$. \hfill \qed

5.1 Lower bound for $S(A, \mathbb{P}, z)$

Theorem 5. Assume that GRH holds. Let $\alpha_1, N \geq X_2 \geq 4 \cdot 10^{18}$, $z = N^{1/8}$, and $y = N^{1/3}$ be such that
\[ \frac{\sqrt{X_2}}{\log^{A+1} X_2} \geq 45, \quad \frac{N^{\alpha_1}}{\log^{A+1} N} \geq \exp(u), \quad 0 < \alpha_1 < \frac{1}{8}, \quad (18) \]
with $u \leq 10^{18}$ such that
\[ \epsilon := \epsilon(u) < \frac{1}{63}, \quad (19) \]
where $\epsilon(u)$ is defined by (3). Then
\[ S(A, \mathbb{P}, z) > \frac{8U_N N}{\log^2 N} \left( 1 - \frac{0.62 \log X_2}{X_2^{1/2}} \right) \left( 2e^\gamma \log(3 - 8\alpha_1) - \epsilon C_2(\epsilon) \frac{3e^{8\alpha_1 - 2}}{4 - 8\alpha_1} \right) - \frac{1}{8} \left( 1 - \frac{0.62 \log X_2}{X_2^{1/2}} \right) \left( 2e^\gamma \prod_{p > 2} \left( 1 - \frac{1}{(p-1)^2} \right) \right)^{-1} \frac{c_G(X_2)}{\log^{A+2} N}, \]
with $c_G$ defined in Lemma 14 below, $q_G$ is defined by Lemma 3 and $C_2(\cdot)$ from Table 7.

Proof. We will apply Theorem 3 with $A$, $z$, $\mathbb{P}$ defined above, and thus, with $r(d) = |A_d| - |A|/\varphi(d)$. We set
\[ Q := \{ p \in \mathbb{P}, p < \epsilon \}, \quad D = N^{\frac{1}{2} - \alpha}, \quad g(p) = \frac{1}{p - 1} \quad \text{for} \quad p \in \mathbb{P}. \]

We can check that all the assumptions of Theorem 3 hold: $D \geq z^2$ by (18), and (2) holds with $\epsilon = \epsilon(u)$ by (19) and (3). Hence, we get
\[ S(A, \mathbb{P}, z) > (f(s) - \epsilon C_2(\epsilon)e^2h(s))V(z)|A| - \sum_{d | P(z)} |r(d)| \]
\[ \geq \frac{8U_N N}{\log^2 N} \left( 1 - \frac{0.62 \log N}{N^{1/2}} \right) \left( f(s) - \epsilon C_2(\epsilon)e^2h(s) \right) - \sum_{d | P(z)} |r(d)| \]
\[ \geq \frac{8U_N N}{\log^2 N} \left( 1 - \frac{0.62 \log N}{N^{1/2}} \right) \left( 2e^\gamma \log(3 - 8\alpha_1) - \epsilon C_2(\epsilon) \frac{3e^{8\alpha_1 - 2}}{4 - 8\alpha_1} \right) - \sum_{d | P(z)} |r(d)|. \]
In the last line we used \( s = \frac{\log D}{\log z} = 4 - 8\alpha_1 \in [3, 4] \), so that \( f(s) = \frac{2^{s-1}\log(s-1)}{s} \) and \( h(s) = 3s^{-1}e^{-s} \). It remains to bound the error term. We note that by Lemma 2 and (18),

\[
QD \leq N^{\frac{1}{2} - \alpha_1} \prod_{p \leq N^{\frac{1}{2} - \alpha_1}} \exp(\theta(u)) \leq N^{\frac{1}{2} - \alpha_1} \exp(u) \leq \frac{\sqrt{N}}{\log^{A+1} N},
\]

so Lemmas 13 and 14 below complete the proof of the theorem.

For the next lemma we define

\[
E_{\pi}(x; k, l) = \pi(x; k, l) - \frac{\pi(x)}{\varphi(k)}.
\]

Lemma 13. Suppose \( N \geq X_2 \geq 4 \cdot 10^{18}, A \geq -1, \) and \( H = \frac{\sqrt{N}}{\log^{A+1} N} \). Then

\[
\sum_{\substack{d \leq H \\ (d, N) = 1}} \mu^2(d)|E_{\pi}(N; d, N)| \leq \frac{p_G(X_2)N}{\log^{A} N}
\]

with

\[
p_G(X_2) = 0.65(0.02 + q_G(X_2)) \leq 0.429 \quad \text{for} \quad X_2 \geq 4 \cdot 10^{18},
\]

and \( q_G(X_2) \) defined in Lemma 3.

Proof. If \( d = 1 \), then \( |E_{\pi}(N; d, N)| = 0 \), so the upper bound from the statement holds. Suppose \( d > 1 \), then \( 3 \leq d \) since \( (d, N) = 1 \) and \( N \) is even. By the triangle inequality

\[
|E_{\pi}(N; d, N)| \leq \left| \pi(N; d, N) - \frac{\text{li}(N)}{\varphi(d)} \right| + \frac{1}{\varphi(d)}|\text{li}(N) - \pi(N)|,
\]

where the first term can be bounded by Lemma 3 and the bound for the second one is provided by [18, Corollary 1]

\[
\frac{1}{\varphi(d)}|\text{li}(N) - \pi(N)| \leq \frac{1}{10\pi}\sqrt{N} \log N \quad \text{for} \quad 2.657 \leq N.
\]

Thus,

\[
|E_{\pi}(N; d, N)| \leq (0.02 + q_G(X_2))\sqrt{N} \log N,
\]

which leads to the following bound

\[
\sum_{\substack{d \leq H \\ (d, N) = 1}} \mu^2(d)|E_{\pi}(N; d, N)| \leq 0.65H(0.02 + q_G(X_2))\sqrt{N} \log N
\]

\[
\leq 0.65(0.02 + q_G(X_2))\frac{N}{\log^{A} N},
\]

by Lemma 5.

Lemma 14. Suppose all the conditions from Lemma 13 are satisfied. Then

\[
\sum_{\substack{d \leq H \\ (d, N) = 1}} \mu^2(d)|r(d)| \leq \frac{c_G(X_2)N}{\log^{A} N},
\]

(20)

with

\[
c_G(X_2) = p_G(X_2) + \frac{0.9}{\sqrt{X_2 \log \log X_2}} \leq 0.429 \quad \text{for} \quad X_2 \geq 4 \cdot 10^{18}.
\]
Proof. We note that
\[ r(d) = |A_d| - \frac{|A|}{\varphi(d)}, \]
hence
\[ |r(d)| \leq |E_\pi(N; d, N)| + \omega(N). \]

From Lemmas 4, 5 and 13 we have
\[
\sum_{\substack{d \leq H \atop (d, N) = 1}} \mu^2(d)|r(d)| \leq \frac{p_G(X_2)N}{\log^4 N} + 0.65 \frac{\sqrt{N}}{\log^{A+1} N} \cdot \frac{1.3841 \log N}{\log \log N}
\leq \frac{N}{\log^4 N} \left( p_G(X_2) + \frac{0.9}{\sqrt{X_2 \log \log X_2}} \right).
\]
\[
\square
\]

5.2 An upper bound for \( \sum_{z \leq q \leq y} S(A_q, \mathbb{P}, z) \)

Theorem 6. Assume that GRH holds. Let \( \alpha_2, N \geq X_2 \geq \exp(\exp(15)), z = N^{1/8}, \) and \( y = N^{1/3} \) be such that
\[
\frac{\sqrt{X_2}}{\log^{A+1} X_2} \geq 10^9, \quad \frac{N^{\alpha_2}}{\log^{A+1} N} \geq \exp(u), \quad 0 < \alpha_2 < \frac{1}{24},
\]
with \( u \leq 10^{18} \) such that \( \epsilon := \epsilon(u) < \frac{1}{63}, \)
where \( \epsilon(u) \) is defined by (3). Let
\[
k_x = 8 \left( \frac{1}{6} - x \right),
\]
then
\[
\sum_{\substack{z \leq q \leq y \atop q \mid N}} S(A_q, \mathbb{P}, z) \leq 8U_N \frac{N}{\log^2 N} \left( l_1(X_2) + \frac{l_2(X_2)}{\log^{A-3} N} + \frac{l_3(X_2)}{\log^{A-1} N} \right),
\]
where
\[
l_1(X_2) = (1 + 3 \cdot 10^{-7}) \left( 1 + \frac{0.62 \log X_2}{X_2^2} \right) \frac{X_2^{1/8}}{X_2^{1/8} - 1} \left( \frac{\log(6) + \log \left( \frac{3 - 8\alpha_2}{\alpha_2 - 18\alpha_2} \right)}{\frac{3}{2} - \alpha_2} \right),
\]
\[
l_2(X_2) = \frac{c_G(X_2)}{8} \left( 0.55 + \frac{1}{\log X_2} \right) \left( 2e^{\gamma} \prod_{p > 2} \left( 1 - \frac{1}{p - 1} \right) \right)^{-1},
\]
\[
l_3(X_2) = \left( 1 + \frac{0.62 \log X_2}{X_2^2} \right) \left( 2e^{\gamma} \frac{k_\alpha}{c_G(X_2)} + eC_1(\epsilon)e^2 h(c_G(X_2)) \right)\frac{c_G(X_2)}{k_\alpha},
\]
with \( p_G(X_2) \) defined in Lemma 13 and \( C_1(\cdot) \) from Table 1.
Proof. Let \( q \) be a prime with \( z \leq q < y \). If \( q|N \), then for every \( N - p \in A_q \), \( q \) divides both \( N \) and \( p \). The latter cannot be achieved since \((p, N) = 1\). Thus, \( S(A_q, \mathbb{P}, z) = 0 \) for \((q, N) \neq 1\). Now we assume \((q, N) = 1\).

We will proceed as in the proof of \([1, Theorem 47]\). Namely, we apply the upper bound linear sieve to get

\[
S(A_q, \mathbb{P}, z) < (F(s_q) + \epsilon C_1(\epsilon)\varepsilon N h(s_q)) V(z)|A_q| + R_q,
\]

with

\[
R_q = \sum_{d \in \mathcal{P}(z)} |r_q(d)|, \quad r_q(d) = |A_{qd}| \frac{|A_q|}{\varphi(d)},
\]

\[
Q = Q(u), \quad D_q = \frac{D}{q} = \frac{N^{\frac{1}{2}-\alpha_2}}{q} \gg z, \quad \text{and} \quad s_q = \frac{\log D_q}{\log z}.
\]

Hence, we are left with bounding

\[
\sum_{z \leq q < y \atop q|N} S(A_q, \mathbb{P}, z) \leq 8U_N \left( 1 + \frac{0.62 \log N}{N^{\frac{1}{2}}} \right) \sum_{z \leq q < y \atop q|N} |A_q| \left( \frac{F(s_q) + \epsilon C_1(\epsilon)\varepsilon N h(s_q)}{\log N} \right)
\]

\[
+ \sum_{z \leq q < y \atop q|N} R_q.
\]

We will bound the error term \((25)\). Since \( d|P(z) \), \( d \) is coprime to \( q \gg z \), and thus \( \varphi(qd) = \varphi(q)\varphi(d) \), so we have

\[
r_q(d) = |A_{qd}| \frac{|A_q|}{\varphi(d)} = |A_{qd}| \left( \frac{|A|}{\varphi(qd)} + \frac{|A|}{\varphi(qd)} - \frac{|A_q|}{\varphi(d)} \right) = r(qd) + \frac{r(q)}{\varphi(d)}
\]

and hence

\[
\sum_{z \leq q < y \atop (q, N) = 1} R_q \leq \sum_{z \leq q < y \atop (q, N) = 1} \sum_{d \in \mathcal{P}(z) \atop d \leq QD_q} |r(qd)| + \sum_{z \leq q < y \atop (q, N) = 1} |r(q)| \sum_{d \in \mathcal{P}(z) \atop d \leq QD_q} \frac{1}{\varphi(d)}
\]

\[
\leq \sum_{d' < QD \atop (d', N) = 1} |r(d')| + \sum_{z \leq q < y \atop (q, N) = 1} |r(q)| \sum_{d < QD} \frac{\mu^2(d)}{\varphi(d)}.
\]

To obtain the last line we note that from \( qd = q'd' \) with \( z \leq q, q' < y \) and \( d, d'|P(z) \) we get \( q = q' \) and \( d = d' \). By \((21)\) and Lemma \(2\) \( QD \leq \frac{\sqrt{N}}{\log^{14} N} \), so

\[
\sum_{d' < QD \atop (d', N) = 1} |r(d')| \leq \frac{c_2(X_2)N}{\log A N},
\]

15
with \( c_G \) defined in Lemma \([14]\) By \([21]\), \( y \leq \sqrt[4]{N} \frac{\log^{3/2} N}{\log A^{1/2} N} \), so we can use Lemmas \([5]\) and \([14]\) to get

\[
\sum_{z \leq q < y} \frac{|r(q)|}{(q, N)} \sum_{d \leq QD} \frac{\mu^2(d)}{\varphi(d)} \leq 0.55 \frac{c_G(X_2)N}{\log^{4/3} N}.
\]

Let us bound the main term \([21]\). Since \( 0 < \alpha_2 < \frac{1}{3} \),

\[
s_q = \frac{\log D_q}{\log z} = 8 \left( \frac{1}{2} - \alpha_2 - \frac{\log q}{\log N} \right) \in [1, 3],
\]

so \( F(s_q) = \frac{2\gamma}{s_q} \) by \([13]\) Theorem 9.8, hence by using \( A_q = \frac{|A|}{q-1} + r(q) \) we get

\[
\sum_{z \leq q < y} \frac{|A_q|}{q(N)} \left( \frac{F(s_q) + \epsilon C_1(\epsilon) e^2 h(s_q)}{\log N} \right) \leq \sum_{z \leq q < y} \frac{|A|}{q-1} \left( \frac{\epsilon^2}{4 \log D/q} + \frac{\epsilon C_1(\epsilon) e^2 h(s_q)}{\log N} \right)
\]

\[
+ \left( \frac{\epsilon^2}{4 \log D/y} + \frac{\epsilon C_1(\epsilon) e^2 h(k_\alpha)}{\log N} \right) \frac{c_G(X_2)N}{\log^{4/3} N}.
\]

We repeat the argument from the proof of \([1]\) Theorem 47\] to bound the first line in \((26)\) above. By \([9]\) Lemma 1 (ii) we get

\[
\sum_{z \leq p < y} \frac{1}{q \log D/q} \leq \sum_{z \leq p < y} \frac{1}{q \log D/y} \leq \int_{z}^{y} \frac{1}{t \log t \log D/t} \, dt + \frac{64}{\log^2 N \log D/y}
\]

\[
= \frac{\log(6) + \log \left( \frac{3 - 8\alpha_2}{3 - 18\alpha_2} \right)}{(\frac{1}{2} - \alpha_2) \log N} + \frac{64}{(\frac{1}{4} - \frac{3}{4} - \alpha_2) \log^{3} N},
\]

where we have used a substitution \( t \rightarrow N^\gamma \) to evaluate the integral. Using \([10]\) Theorem 5], we have that

\[
\sum_{z \leq p \leq y} \frac{1}{p} < \log \log y - \log \log z + \frac{1}{\log^2 z},
\]

since \( y \geq 286 \) by the initial conditions on \( N \). We note that \( z = N^{1/8} \) would be a prime number for an even integer \( N \) only if \( N = 2^8 = 256 \), which would contradict \([21]\). Thus, \( z \) is not a prime, and

\[
\sum_{z < p \leq y} \frac{1}{p} < \log \log y - \log \log z + \frac{1}{\log^2 z}.
\]

Hence, we complete the proof by noting that \([26]\) is at most

\[
|A| \frac{N^{1/4}}{N^{1/8} + 1} \left( \frac{\epsilon^2}{4} \left( \frac{\log(6) + \log \left( \frac{3 - 8\alpha_2}{3 - 18\alpha_2} \right)}{(\frac{1}{2} - \alpha_2) \log N} + \frac{512}{k_{\alpha_2} \log^4 N} \right) + \left( \frac{\log \frac{8}{3}}{\log^2 N} + \frac{64}{\log^2 N} \right) \frac{\epsilon C_1(\epsilon) e^2 h(k_{\alpha_2})}{\log N} \right),
\]

where \(|A|\) is bounded in Lemma \([12]\) \]}
5.3 An upper bound for $S(B, \mathbb{P}, y)$

We will now prove an upper bound for $S(B, \mathbb{P}, y)$. This is obtained using Theorem 4 together with a bound on $B$ and introducing a bilinear form that allows to bound the error term using Lemma 9.

**Theorem 7.** Set $0 < \epsilon_1 < 1$. Let $N$ be a positive integer, and $\alpha_3$ be a real number satisfying

$$\frac{\sqrt{N}}{\log^{\alpha_1+1} N} \geq 10^9, \quad \frac{N^{\alpha_3}}{\log^{\alpha_1+1} N} \geq \exp(u), \quad 0 \leq \alpha_3 < 1/16,$$

with $u \leq 10^{18}$ such that

$$\epsilon(u) < \frac{1}{63},$$

where $\epsilon(u)$ is defined by (5). Then we have

$$S(B, \mathbb{P}, y) \leq \frac{U_N N}{\log^2 N} \left( 1 + \frac{0.06 \log N}{N^{\alpha_3}} \right) \cdot \left( \frac{2}{1 - \alpha_3} e^{\gamma} + 3 \epsilon C_1(\epsilon) \right) (1 + \epsilon_1) \left( \log \left( \text{li}(y) - \sqrt{y \log y/(8\pi)} \right) \right) \cdot \left( \frac{64 \log \frac{26}{3}}{\log^2 N} + \log \left( \frac{8}{3} \right) + \frac{64}{\log^2 N} \left( 3 \log (1 + \epsilon_1) + \frac{27}{\log^2 N} \right) \right) \cdot \left( 2^{\gamma} \prod_{p > 2} \left( 1 - \frac{1}{p - 1} \right)^{1} \cdot N^{-1/48} (1 + \epsilon_1) \log^5 N \frac{\sqrt{N}}{\log^{\alpha_1+1} N} (1 + \epsilon_1)^{-1/6} + 0.159 N^{-\frac{1}{16}} \right),$$

with $\gamma$ defined in Lemma 9, $m(\cdot)$ defined in Lemma 4, and $C_1(\cdot)$ from Table 7.

We start by recalling that

$$B = \{N - p_1p_2p_3 : z \leq p_1 < y \leq p_2 \leq p_3, \ p_1p_2p_3 < N, (p_1p_2p_3, N) = 1 \},$$

where $p_1, p_2, p_3$ denote prime numbers. We now need to drop the restriction $(p_1, N) = 1$ and relax the condition $p_1p_2p_3 < N$, so that $p_1$ and $p_2p_3$ will range over independent intervals introducing a bilinear form that we will bound using Lemma 9.

Let us fix $0 < \epsilon_1 < 1$ and let $j$ be a non-negative integer, such that

$$w_j := z(1 + \epsilon_1)^j < y,$$

which implies

$$j < \frac{\log(y/z)}{\log(1 + \epsilon_1)}.$$

We introduce sets

$$B^{(j)} = \{N - p_1p_2p_3 : z \leq p_1 < y \leq p_2 \leq p_3, w_jp_2p_3 < N, (p_2p_3, N) = 1, \ w_j \leq p_1 < w_j(1 + \epsilon_1) \},$$

for

$$0 \leq j \leq j_0 = \left\lfloor \frac{\log y/z}{\log(1 + \epsilon_1)} \right\rfloor - 1.$$
Let $\mathcal{B} = \bigcup_j B^{(j)}$, then $B \subseteq \mathcal{B}$, so we have

$$S(B, \mathcal{P}, y) \leq S(\mathcal{B}, \mathcal{P}, y) = \sum_{j \leq j_0} S(B^{(j)}, \mathcal{P}, y),$$

(30)

since $B^{(j)}$ are disjoint. We will apply Theorem 4 to sets $B^{(j)}$. Let us take $0 < \alpha_3 < 1/16$, $N$ and $u$ satisfying (27) and (28), and denote $D = N^{2 - \alpha_3}$, $Q = Q(u)$, then we have

$$D \geq y, \quad QD \leq \frac{\sqrt{N}}{\log^{1+1} N},$$

since $\alpha_3 < 1/6$ and by Lemma 2 respectively. We set

$$s = \frac{\log D}{\log y} = \frac{3}{2} - 3\alpha_3.

(31)

Then Theorem 4 gives

$$S(B^{(j)}, \mathcal{P}, y) < \left| B^{(j)} \right| V(y)(F(s) + \epsilon C_1(\epsilon)e^2h(s)) + R^{(j)},$$

with

$$R^{(j)} = \sum_{d < D} \left| t^{(j)}(d) \right|, \quad t^{(j)}(d) = \left| B^{(j)} \right| - \frac{B^{(j)}}{\varphi(d)}, \quad \text{and} \quad B^{(j)} = \sum_{p_1p_2p_3 \equiv N \mod d} \frac{1}{p_1p_2p_3 < N, (p_2p_3, N) = 1}.

By Lemma 11 we have $V(y) < 3 \frac{U_N}{\log N} \left( 1 + \frac{0.06 \log N}{N^{1/2}} \right)$. By the definition of $s$, $1 \leq s \leq 2$, and therefore $F(s) = 2^{-2}$ by [13] p. 259. Thus, for every $j$

$$S(B^{(j)}, \mathcal{P}, y) < \left| B^{(j)} \right| \frac{U_N}{\log N} \left( 1 + \frac{0.06 \log N}{N^{1/2}} \right) \left( \frac{2}{2 - \alpha_3} e^2 + 3\epsilon C_1(\epsilon)e^2h \left( \frac{3}{2} - 3\alpha_3 \right) \right) + R^{(j)}.$$

Since $1 \leq \frac{3}{2} - 3\alpha_3 \leq 2$, we note that $h \left( \frac{3}{2} - 3\alpha_3 \right) = e^{-2}$. Therefore, by (32),

$$S(\mathcal{B}, \mathcal{P}, y) < \left| \mathcal{B} \right| \frac{U_N}{\log N} \left( 1 + \frac{0.06 \log N}{N^{1/2}} \right) \left( \frac{2}{2 - \alpha_3} e^2 + 3\epsilon C_1(\epsilon) \right) + R,$

(32)

where $R = \sum_{j \leq j_0} R^{(j)}$. We bound the main term first. We can see that

$$\left| B^{(j)} \right| \leq (\pi(w_j(1 + \epsilon_1)) - \pi(w_j) + 1) \cdot \left| \{(p_2, p_3) : y \leq p_2 \leq p_3, w_jp_2p_3 < N, (p_2p_3, N) = 1 \} \right|,$$

$$B \subseteq \mathcal{B} \subseteq \{ N - p_1p_2p_3 : z \leq p_1 < y \leq p_2 \leq p_3, p_1p_2p_3 < (1 + \epsilon_1)N \}.$$

(33)

We will now need an explicit upper bound for the cardinality of $\mathcal{B}$ that is obtained in a similar way as done by Nathanson in [13] pp. 289–291.
Lemma 15. We have, assuming GRH and that $y \geq 2657$,

$$|\mathcal{B}| \lesssim (1 + \epsilon_1) \left( \frac{\log y \left( \text{li}(y) - \sqrt{y} \log y / (8\pi) \right)}{y} \right) N$$

$$\cdot \left( \frac{\tau}{\log N} + \frac{64 \log 26}{\log^3 N} + \left( \log \left( \frac{8}{3} \right) + \frac{64}{\log^2 N} + \left( 3 \log (1 + \epsilon_1) \right) \left( \frac{27}{\log^2 N} + \frac{27}{\log^3 N} \right) \right) \right),$$

with

$$\tau = \int_{1/8}^{1/3} \frac{\log(2 - 3\beta)}{\beta(1 - \beta)} d\beta < 0.363084.$$

Proof. With $(1 + \epsilon_1)N/(p_1p_2) > p_3 \geq y \geq 2657$, by [18, Corollary 1] we obtain

$$\pi \left( \frac{(1 + \epsilon_1)N}{p_1p_2} \right) \lesssim (1 + \epsilon_1) \left( \frac{\log y \left( \text{li}(y) - \sqrt{y} \log y / (8\pi) \right)}{y} \right) N \frac{p_1p_2 \log(N/p_1p_2)}{p_1p_2 (1 + \epsilon_1)N}.$$

Now since $p_1 < p_2 \leq p_3$ and $p_1p_2p_3 < (1 + \epsilon_1)N$, we have $p_1p_2^2 < (1 + \epsilon_1)N$ and

$$p_3 \leq \frac{(1 + \epsilon_1)N}{p_1p_2}.$$

Thus, from the definition (33) of $\mathcal{B}$,

$$|\mathcal{B}| \lesssim \sum_{z \leq p_1 < y \leq p_2 \leq p_3 < (1 + \epsilon_1)N} 1 \leq \sum_{z \leq p_1 < y \leq p_2 \leq (1 + \epsilon_1)N} \pi \left( \frac{(1 + \epsilon_1)N}{p_1p_2} \right)$$

$$\lesssim (1 + \epsilon_1) \left( \frac{\log y \left( \text{li}(y) - \sqrt{y} \log y / (8\pi) \right)}{y} \right) N \sum_{z \leq p_1 < y} \frac{1}{p_1} \sum_{y < p_2 < w} \frac{1}{p_2 \log(N/p_1p_2)},$$

with $w = \sqrt{\frac{(1 + \epsilon_1)N}{p_1}}$.

We now introduce the functions $h_p(t) = (\log N/pt)^{-1}$ and

$$H(u) = \int_y^{\sqrt{N/u}} h_u(t) d \log t.$$

Using [9, Lemma 1 (ii)] and [17, Theorem 5],

$$\sum_{y \leq p_2 < w} \frac{1}{p_2 \log(N/p_1p_2)} \leq \int_y^{\sqrt{N/w}} h_{p_1}(t) d \log t + \frac{h_{p_1}(w)}{\log^2 y}$$

$$= H(p_1) + \int_{\sqrt{N/p_1}}^{w} h_{p_1}(t) d \log t + \frac{h_{p_1}(w)}{\log^2 y}$$

$$\leq H(p_1) + \frac{10 \log(1 + \epsilon_1)}{\log^2 N} + \frac{27}{\log^3 N}.$$

Where, in the last step, we used

$$\int_{\sqrt{N/p_1}}^{w} h_{p_1}(t) d \log t \leq \frac{10 \log(1 + \epsilon_1)}{\log^2 N},$$

19
obtained by the change of variables \( t = \sqrt{N/p_1}s \), as done in [13] p. 290, and

\[
h_{p_1}(w) = \frac{2}{\log \left( \frac{N}{(1+\epsilon)p_1} \right)} \leq \frac{3}{\log N}.
\]

Therefore, also using [17] Theorem 5,

\[
\sum_{z \leq p_1 < y} \frac{1}{p_1} \sum_{y \leq p_2 < \sqrt{N}} \frac{1}{p_2} \log(N/p_1 p_2) \leq \sum_{z \leq p_1 < y} \frac{H(p_1)}{p_1} + \left( \log \frac{8}{3} + \frac{64}{\log^2 N} \right) \left( \frac{10 \log(1+\epsilon)}{\log^2 N} + \frac{27}{\log^3 N} \right).
\]

We now note that \( H(y) = 0 \) and, upon using the substitution \( t = N^\tau \) in the definition of \( H, H(z) \leq \frac{0.56}{\log N} \).

Using [9] Lemma 1 (ii) and [17] Theorem 5] and \( \int_y^z H(u) d\log u = \frac{\pi}{\log N} \) (see [13] p. 291] we thus have

\[
\sum_{z \leq p_1 < y} \frac{H(p_1)}{p_1} < \int_y^z H(u) d\log u + \frac{H(z)}{\log^2 z} \leq \frac{\pi}{\log N} + \frac{36}{\log^2 N}.
\]

This concludes the proof of the lemma.

Now let us bound the error term of [123]. By the definition of \( r_d^{(j)} \), we have

\[
r_d^{(j)} = \sum_{p_1 p_2 p_3 \equiv N \pmod{d}} \sum_{z \leq p_1 < y} \frac{1 - \frac{1}{\varphi(d)}}{\varphi(d)} \sum_{p_1 | d, p_1 \geq z} 1.
\]

We now add the condition \((p_1 p_2 p_3, d) = 1\) in the second sum. This is equivalent to \((p_1, d) = 1\), since the condition \((p_2 p_3, d) = 1\) already follows from the fact that \(d\) divides \(P(y)\). Thus, adding the condition \((p_1 p_2 p_3, d) = 1\) in the second sum makes it decrease by at most

\[
\frac{1}{\varphi(d)} \sum_{p_1 p_2 p_3 < (1+\epsilon_1)N} 1 \leq \frac{(1 + \epsilon_1)N}{\varphi(d)} \sum_{p_1 | d, p_1 \geq z} \frac{1}{p_1} \leq \frac{(1 + \epsilon_1)\theta Nw(d)}{z\varphi(d)}.
\]

We set \(a(n)\) to be the characteristic function of the set of integers \(n = p_2 p_3\) with \(y \leq p_2 < p_3\) and \((N, p_2 p_3) = 1\). Then, by applying Lemma [9] we get for some \(0 \leq \theta \leq 1\)

\[
r_d^{(j)} = \sum_{a(n)} \sum_{n < X} \frac{a(n)}{\varphi(n)} \sum_{n < X, Z \leq Y} \frac{a(n)}{\varphi(n)} \sum_{(np, d) = 1} (1 + \epsilon_1)\theta Nw(d),
\]

with

\[
X = \frac{N}{w_j}, \quad Y = \min(y, (1 + \epsilon_1)w_j), \quad Z = w_j, \quad a = N.
\]
Let us define $D^* = \sqrt{N}/\log^{A+1}N \geq DQ$. It is easy to see that with the above choices, the conditions (27) and $N \geq 4 \cdot 10^{18}$, Lemma 9 holds. Therefore,

$$R^{(j)} \leq \sum_{d < D^*} \left| r_d^{(j)} \right|$$

$$\leq \sum_{d < D^*} \left( \sum_{n < \frac{N}{\delta^j}} \sum_{y \leq Y \pmod{d}} \sum_{n < \frac{N}{\delta^j}} \frac{a(n)}{\varphi(n)} \sum_{y \leq Y \pmod{d}} \sum_{n < \frac{N}{\delta^j}} a(n) \right) + \sum_{d < D^*} \left( \frac{(1 + \epsilon_1)Nw(d)}{\delta \varphi(d)} \right)

\leq m(X, Y, D^*) \frac{XY \log^{5/3} D^* \log^{1/3} Y}{Y^{1/6}} + 0.762(1 + \epsilon_1)N^{7/8} \log^2 N

\leq m \left( N^{2/3}, (1 + \epsilon_1)N^{1/8}, \frac{\sqrt{N}}{\log^{A+1}N} \right) \frac{N^{47/48}(1 + \epsilon_1)^{5/6} \log^2 N}{25/3 \ 31/3} + 0.762(1 + \epsilon_1)N^{7/8} \log^2 N,$$

for each $j = 0, 1, \ldots, j_0$. Since $j_0 \leq \frac{5 \log N}{24 \log(1 + \epsilon_1)}$ by (29), we get

$$R = \sum_{j=0}^{j_0} R^{(j)}$$

$$\leq 0.046 \ m \left( N^{2/3}, (1 + \epsilon_1)N^{1/8}, \frac{\sqrt{N}}{\log^{A+1}N} \right) \frac{N^{47/48}(1 + \epsilon_1)^{5/6} \log^3 N}{\log(1 + \epsilon_1)} + 0.159 \frac{(1 + \epsilon_1)}{\log(1 + \epsilon_1)} N^{7/8} \log^3 N

\leq \frac{N^{47/48}(1 + \epsilon_1)^{5/6} \log^3 N}{\log(1 + \epsilon_1)} \left[ 0.046 \ m \left( N^{2/3}, (1 + \epsilon_1)N^{1/8}, \frac{\sqrt{N}}{\log^{A+1}N} \right) (1 + \epsilon_1)^{-1/6} + 0.159 N^{-\frac{4 \epsilon}{81 \epsilon}} \right].$$

(34)

The desired result now follows from (32), Lemma 15 and (34).

5.4 Proof of Theorem 8

We combine Lemma 10 with the bounds obtained in Theorems 5, 6, and 7. We choose the parameters $N \geq X_2 = \exp(\exp(15.85)), \ u = 10^{4.27}, \ \alpha_1 = \alpha_2 = \alpha_3 = 10^{-2.61}, \ A = 4, \ \epsilon_1 = 10^{-20}, \ \epsilon = \frac{1}{81 \epsilon}, \ C_1 = 112$, and $C_2 = 114$. All computations were implemented in SageMath 9.3.

5.5 Proof of Corollary 1

Let $\pi_2'(N)$ be a number of representations of $N$ as a sum of a prime number and a square-free semi-prime number. Then

$$\pi_2'(N) \geq \pi_2(N) - \sqrt{N} - 1,$$

where 1 counts the representation $N = p + 1$, if such exists, and $\sqrt{N}$ bounds the number integers up to $N$ with two equal prime factors.

With the choice of parameters from the previous section we obtain:

$$\pi_2'(N) \geq 4 \cdot 10^{-4} \frac{U_NN}{\log^2 N} - \sqrt{N} - 1 > 0,$$

the last inequality being checked using SageMath 9.3.
References

[1] M. Bordignon, D.R. Johnston, and V. Starichkova, An explicit version of Chen’s theorem, arxiv:2207.09452 (2022).

[2] S. Broadbent, H. Kadiri, A. Lumley, N. Ng, and K. Wilk, Sharper bounds for the Chebyshev function $\theta(x)$, Math. Comp. 90 (2021), no. 331, 2281–2315.

[3] J. Büthe, A Brun-Titchmarsh inequality for weighted sums over prime numbers, Acta Arith. 166 (2014), no. 3, 289–299.

[4] J. Büthe, An analytic method for bounding $\psi(x)$, Math. Comp. 87 (2018), no. 312, 1991–2009.

[5] M. Cully-Hugill and E. S. Lee, Explicit interval estimates for prime numbers, Preprint on arXiv:2103.05986 (2021), Submitted.

[6] H. Davenport, Multiplicative Number Theory, Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000, Third Edition.

[7] J.-M. Deshouillers, G. Effinger, H. te Riele, and D. Zinoviev, A complete Vinogradov 3-primes theorem under the Riemann hypothesis, Electron. Res. Announc. Amer. Math. Soc. 3 (1997), 99–104. MR 1469323

[8] A.-M. Ernvall-Hytönen and N. Palojärvi, Explicit bound for the number of primes in arithmetic progressions assuming the generalized Riemann hypothesis, Math. Comp. 91 (2022), no. 335, 1317–1365.

[9] G. Greaves, Sieves in Number Theory, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 43, Springer-Verlag, Berlin, 2001. MR 1836967

[10] H. Helfgott, The ternary Goldbach problem, arXiv:1501.05438 (2015).

[11] H. Kadiri and A. Lumley, Short effective intervals containing primes, Integers 14 (2014), Paper No. A61, 18.

[12] L. Kaniecki, On Šnirelman’s constant under the Riemann hypothesis, Acta Arith. 72 (1995), no. 4, 361–374. MR 1348203

[13] M.B. Nathanson, Additive Number Theory, Graduate Texts in Mathematics, vol. 164, Springer-Verlag, New York, 1996, The classical bases.

[14] T. Oliveira e Silva, S. Herzog, and S. Pardi, Empirical verification of the even Goldbach conjecture and computation of prime gaps up to $4 \cdot 10^{18}$, Math. Comp. 83 (2014), no. 288, 2033–2060.

[15] G. Robin, Estimation de la fonction de Tchebychev $\theta$ sur le k-ième nombre premier et grandes valeurs de la fonction $\omega(n)$ nombre de diviseurs premiers de n, Acta Arith. 42 (1983), no. 4, 367–389.

[16] J.B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, Illinois Journal of Mathematics 6 (1962), no. 1, 64–94.

[17] J.B. Rosser and L. Schoenfeld, Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$, Math. Comp. 29 (1975), no. 129, 243–269.
[18] L. Schoenfeld, *Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. II*, Math. Comp. **30** (1976), no. 134, 337–360.

[19] I.M. Vinogradov, *Representation of an odd number as a sum of three primes*, Dokl. Akad. Nauk. SSR **15** (1937), 291–294.

[20] T. Yamada, *Explicit Chen’s theorem*, https://arxiv.org/abs/1511.03409 (2015).

Charles University, Faculty of Mathematics and Physics, Department of Algebra,
Sokolovská 83, 186 00 Praha 8, Czech Republic
E-mail address: matteobordignon91@gmail.com

UNSW Canberra, Department of Science, Northcott Drive, Canberra,
ACT Australian Capital Territory 2612
E-mail address: v.starichkova@adfa.edu.au