

# Bertrand's Postulate for Carmichael Numbers

Daniel Larsen

## Abstract

Alford, Granville, and Pomerance proved in [1] that there are infinitely many Carmichael numbers. In the same paper, they ask if a statement analogous to Bertrand's postulate could be proven for Carmichael numbers. In this paper, we answer this question, proving the stronger statement that for all  $\delta > 0$  and  $x$  sufficiently large in terms of  $\delta$ , there exist at least  $e^{\frac{\log x}{(\log \log x)^{2+\delta}}}$  Carmichael numbers between  $x$  and  $x + \frac{x}{(\log x)^{\frac{1}{2+\delta}}}$ .

## 1 Introduction

A composite number  $n$  which divides  $a^n - a$  for every integer  $a$  is called a Carmichael number (named after Robert Carmichael, although the concept is decades older [9]). In a celebrated paper, Alford, Granville, and Pomerance proved that there are infinitely many Carmichael numbers [1]. In fact, they were able to prove a much stronger result, namely that there exist at least  $x^{2/7}$  of them up to sufficiently large  $x$ . Despite such a strong lower bound on the number of Carmichael numbers up to  $x$  (which has subsequently been improved to above  $x^{1/3}$  [4]), little is currently known about their distribution. Indeed, in their paper, Alford, Granville, and Pomerance note, "It seems to be difficult to prove a 'Bertrand's postulate for Carmichael numbers,' that is, that there is always a Carmichael number between  $x$  and  $2x$  once  $x$  is sufficiently large." In this paper, we settle this question, in fact proving that the ratio of successive Carmichael numbers tends to 1.

The method of Alford, Granville, and Pomerance is to choose  $L$  such that  $(\mathbb{Z}/L\mathbb{Z})^\times$  has no elements of large order and then find  $k$  such that there are many primes of the form  $dk + 1$  where  $d$  divides  $L$ . For the rest of the argument to work, it is important that the number of divisors which produce primes is much bigger than the order of any element in the multiplicative group mod  $L$ . The maximal order of elements of  $(\mathbb{Z}/L\mathbb{Z})^\times$ , assuming  $L$  is square-free, is the least common multiple of the numbers of the form  $p - 1$

where  $p$  is a prime divisor of  $L$ . Therefore,  $L$  is (roughly speaking) chosen to be the product of primes less than  $y$  such that  $p - 1$  possesses only small prime factors.

The next step is to use character theory to show that there are subsets of the primes of the form  $dk + 1$  such that their product is  $1 \bmod L$ . Call one such product  $\Pi$ . Assuming  $k$  is relatively prime to  $L$ ,  $\Pi$  is in fact  $1 \bmod kL$  (since each prime factor of  $\Pi$  is  $1 \bmod k$ ). By Korselt's criterion [5], a square-free number is a Carmichael number if and only if  $p - 1 \mid n - 1$  whenever  $p \mid n$ . Since  $dk \mid kL$ , it follows that  $\Pi$  is a Carmichael number.

A critical step of their argument is finding a suitable  $k$ . To do this, they look at all the pairs  $(k, d)$  such that  $dk + 1$  is prime, where  $k$  is drawn from some bounded interval and  $d$  is chosen from  $\mathbf{D}$ , some subset of the divisors of  $L$ . Assuming  $\mathbf{D}$  contains enough elements, there exists a particular  $k$  with many primes of the form  $dk + 1$  ( $d \in \mathbf{D}$ ) by the pigeonhole principle. This is a rough sketch of the method used by Alford, Granville, and Pomerance, although of course their actual method is considerably more complicated.

Unfortunately, this approach cannot be used to produce Carmichael numbers in short intervals. This is because the use of the pigeonhole principle relinquishes control over which elements of  $\mathbf{D}$  correspond to primes. In the extreme case, this could result in all the primes being almost exactly the same size, in which case, there might not be any products in a particular short interval.

Here is a flawed approach which nevertheless points in the right direction. Split the divisors of  $L$  into pairs of the form  $(d_i, d'_i)$  with the property that  $d_i > 2d'_i$ . (Of course, some divisors will need to be discarded to make this work, but that is of no concern.) Then use the pigeonhole principle on triples  $(k, d_i, d'_i)$  where both  $d_i k + 1$  and  $d'_i k + 1$  are prime. This guarantees that no more than half of the divisors corresponding to primes will be of the same general size. In fact, as long as  $k$  is chosen properly, the products of the primes will spread out evenly.

The problem with this method is there are no known lower bounds on the number of  $k$  such that  $dk + 1$  and  $d'k + 1$  are both prime for fixed  $d$  and  $d'$ . However, there has been considerable progress in this direction. In a landmark 2014 paper [11], Zhang showed that there exists  $h \leq 70000000$  such that  $x$  and  $x + h$  are simultaneously prime infinitely often. Then in a further breakthrough, Maynard [7] and independently Tao (unpublished) found a different and more flexible approach to the problem, which allowed more general results to be proven.

The connection between prime tuples and Carmichael numbers has long been known. In 1939, Chernick observed that  $(6k + 1)(12k + 1)(18k + 1)$

is a Carmichael number whenever those three factors are all prime [2]. The prime tuples conjecture (or rather a slightly more general statement known as Dickson's conjecture) asserts that every finite set of linear forms can be made simultaneously prime infinitely often unless the set is inadmissible, which means that some prime divides all the values of their product at integers. Clearly, then, the prime tuples conjecture would imply that there are infinitely many Carmichael numbers. In fact, it would imply something stronger (and as yet unproven), that there are infinitely many Carmichael numbers with a bounded number of prime factors. This result can actually be proven under the assumption of hypotheses which are much weaker than the prime tuples conjecture, although still unproven. (See [10].) However, to the author's knowledge, none of the unconditional progress towards the prime tuples conjecture had ever been used to answer questions regarding Carmichael numbers before this paper.

We are now in a position to revise our previous strategy. We split the divisors into  $M$ -tuples for some large number  $M$ , instead of just into pairs. A 2016 paper of Maynard [8], which extends his previous work, will then give us lower bounds on the number of triples  $(k, d, d')$  with  $d$  and  $d'$  in the same  $M$ -tuple and  $dk + 1$  and  $d'k + 1$  both prime, producing enough triples that the pigeonhole principle will be able to find a specific  $k$  with many primes of the form  $dk + 1$ . Crucially, this can be done while making sure the primes do not cluster around powers of any  $\alpha > 1 + \epsilon$ . (Here  $\epsilon$  represents some positive quantity which goes to zero as  $y$  (the starting parameter) goes to infinity.) Then a standard argument from Fourier analysis will guarantee Carmichael numbers of controllable size. By slowly sliding  $y$  to infinity, the following theorem can be proven.

**Theorem 1.** *For all  $\delta > 0$  and  $x$  sufficiently large in terms of  $\delta$ , there are at least  $e^{\frac{\log x}{(\log \log x)^{2+\delta}}}$  Carmichael numbers between  $x$  and  $x + \frac{x}{(\log x)^{\frac{1}{2+\delta}}}$ .*

## 2 Preliminaries

To start with, we give some standard notation in analytic number theory. Let  $\mathbb{P}$  be the set of primes. Let  $(a, b)$  be the greatest common divisor of  $a$  and  $b$ , and let  $\phi(n) = \#\{0 < m < n : (m, n) = 1\}$ . Let

$$\pi(x) = \sum_{p \leq x} 1,$$

$$\pi(x; q, a) = \sum_{\substack{p \leq x \\ p \equiv a \pmod{q}}} 1,$$

and

$$E(x; q, a) = \left| \pi(x; q, a) - \frac{\pi(x)}{\phi(q)} \right|.$$

In the above expressions,  $p$  refers to a prime. Indeed, throughout the paper,  $p$  will always represent a prime. In contrast,  $q$  in the above expressions is not necessarily prime.

The final arithmetic function we will use is Carmichael's  $\lambda$  function, defined on  $n$  as the largest order of an element that appears in  $(\mathbb{Z}/n\mathbb{Z})^\times$ . We start with a standard result about  $\phi$ .

**Lemma 1.**

$$\sum_{k \leq x} \frac{1}{\phi(k)} = O(\log x).$$

*Proof.* This is proved in, for example, Lemma 12.1 of [6], where it is further shown that the constant of proportionality is  $\frac{\zeta(2)\zeta(3)}{\zeta(6)}$  where  $\zeta$  is the Riemann zeta function.  $\square$

Maynard's sieve requires a version of the Bombieri-Vinogradov theorem. The standard error term of  $\frac{x}{(\log x)^A}$  won't suffice for the purposes of this paper. However, a better bound can be attained if the effects of at most one Siegel zero are disregarded.

**Theorem 2.** *There exists a positive constant  $C$  such that for every sufficiently large value of  $x$ , there exists an integer  $s \in [\sqrt{\log x}, e^{\sqrt{\log x}}]$  with*

$$\sum_{\substack{q \leq x^{2/5} \\ s \nmid q}} \max_{2 \leq t \leq x} \max_{(a,q)=1} \left| \pi(t; q, a) - \frac{\pi(t)}{\phi(q)} \right| \leq \frac{x}{e^{C\sqrt{\log x}}}.$$

*Proof.* This follows from Theorem 11.2 from [6] by summation by parts.  $\square$

A result very much like Theorem 2 also appears in [6], but the above version is slightly more general. For the rest of the paper, we fix a particular value of  $s$  for each  $x$ , which we label  $s_x$ . Now, given a set of linear forms, consider its product, which will be some polynomial  $P(x)$ . If there is no prime  $p$  such that  $p \mid P(x)$  for all integer values of  $x$  then we say that the set of linear forms is *admissible*.

**Theorem 3.** *For every  $m \in \mathbb{N}$  and  $\epsilon > 0$ , there exists a positive integer  $N$  such that for every admissible set*

$$\mathbf{L} = (L_1(k), L_2(k), \dots, L_N(k)) = (d_1k + c_1, d_2k + c_2, \dots, d_Nk + c_N)$$

*of  $N$  linear forms where  $d_i > c_i > 0$  for all  $i$ ,*

$$\#\{x \leq k < 2x : \#\{i : L_i(k) \in \mathbb{P}\} \geq m\} \geq \frac{x}{(\log x)^N} \quad (1)$$

*for sufficiently large  $x$  in terms of  $\mathbf{L}$ . In fact, (1) holds for every  $x$  with an associated  $d_+$  where the pair of  $x$  and  $d_+$  satisfy the following properties:*

1.  $x$  is sufficiently large in terms of  $m$  and  $\epsilon$ .
2.  $d_+ \geq d_i$  for all  $i$ .
3.  $\log x \geq (\log d_+)^{2+\epsilon}$
4.  $s_{2xd_+}$  (as taken from the statement of Theorem 2) has a prime factor which does not divide  $d_i$  for any  $i$ .

*Proof.* This theorem is a particular case of Theorem 3.1 from [8], corresponding to  $\mathbf{P} = \mathbb{P}$  and  $\mathbf{A} = \mathbb{N}$ , as well as  $\alpha = 1$ ,  $\theta = \frac{1}{3}$ ,  $\delta = \frac{1}{6}$ , and  $B$  equal to the prime factor of  $s_{2xd_+}$  which does not divide any of the  $d_i$ . Some work is still required, though, as various conditions must be checked. One beautiful aspect of Maynard's result is that it turns regularity conditions regarding the individual linear forms into a statement about how the linear forms interact with each other. The two conditions which need to be established for each  $i$  are, letting

$$\Pi(x; q, a) = \#\{d_ik + c_i \in \mathbb{P} : x \leq k < 2x, k \equiv a \pmod{q}\}$$

and

$$\Pi(x) = \#\{d_ik + c_i \in \mathbb{P} : x \leq k < 2x\},$$

that

$$\sum_{\substack{q \leq x^\theta \\ (q, B)=1}} \max_{(d_ia+c_i, q)=1} \left| \Pi(x; q, a) - \Pi(x) \frac{\phi(d_i)}{\phi(d_iq)} \right| \ll \frac{\Pi(x)}{(\log x)^{100N^2}} \quad (2)$$

and

$$\frac{\phi(B)}{B} \frac{\phi(d_i)}{d_i} \Pi(x) \geq \frac{\delta x}{\log x}. \quad (3)$$

In fact, the latter statement need only be true on average, but we will show that it is true for all  $i$ . The other conditions for Maynard's theorem are easily seen to be true; for example, the other two conditions of Hypothesis 1 require  $\mathbf{A}$  to be well-distributed in arithmetic progressions, which is definitely true for  $\mathbf{A} = \mathbb{N}$ .

Fix  $i$  for which we will prove equations (2) and (3). Let

$$E = \sum_{\substack{q \leq x^{1/3} \\ (q, B)=1}} \max_{(d_i a + c_i, q)=1} \left| \Pi(x; q, a) - \frac{\phi(d_i)}{\phi(d_i q)} \Pi(x) \right|.$$

To clean up the following equations, we will drop the subscripts from  $d_i$  and  $c_i$ . First of all, we claim that

$$\Pi(x; q, a) = \pi(2dx; dq, da + c) - \pi(dx; dq, da + c).$$

Indeed,  $\Pi(x; q, a)$  counts the number of primes of the form  $d(a + nq) + c$  with  $x \leq a + nq < 2x$ , or equivalently,  $dx + c \leq d(a + nq) + c < 2dx + c$ . Since  $0 < c < d$  by hypothesis, this is equivalent to  $dx < d(a + nq) + c < 2dx$ , which establishes our claim. Now

$$\begin{aligned} E &= \sum_{\substack{q \leq x^{1/3} \\ (q, B)=1}} \max_{(da+c, q)=1} \left| \pi(2dx; dq, da + c) - \pi(dx; dq, da + c) - \frac{\phi(d)}{\phi(dq)} (\pi(2dx; d, c) - \pi(dx; d, c)) \right| \\ &\leq \sum_{\substack{q \leq x^{1/3} \\ (q, B)=1}} \max_{(b, dq)=1} \left| \pi(2dx; dq, b) - \pi(dx; dq, b) - \frac{\phi(d)}{\phi(dq)} (\pi(2dx; d, c) - \pi(dx; d, c)) \right| \end{aligned}$$

since  $c$  and  $d$  are relatively prime (which follows from the admissibility of  $\mathbf{L}$ ) so that for  $b = da + c$  with  $(da + c, q) = 1$ ,  $b$  is relatively prime to both  $q$  and  $d$ . Now,

$$\begin{aligned} \left| \pi(dx; dq, b) - \frac{\phi(d)}{\phi(dq)} \pi(dx; d, c) \right| &\leq \left| \pi(dx; dq, b) - \frac{\pi(dx)}{\phi(dq)} \right| + \frac{\phi(d)}{\phi(dq)} \left| \pi(dx; d, c) - \frac{\pi(dx)}{\phi(d)} \right| \\ &= E(dx; dq, b) + \frac{\phi(d)}{\phi(dq)} E(dx; d, c), \end{aligned}$$

and the analogous inequality with  $2dx$  instead of  $dx$  holds as well. Hence

$$E \leq \sum_{\substack{q \leq x^{1/3} \\ (q, B)=1}} \max_{(b, dq)=1} (E(2dx; dq, b) + E(dx; dq, b)) + (E(2dx; d, c) + E(dx; d, c)) \sum_{\substack{q \leq x^{1/3} \\ (q, B)=1}} \frac{\phi(d)}{\phi(dq)}.$$

By definition,  $B$  is relatively prime to  $d$ , so the first part of the above expression is bounded by

$$\sum_{\substack{q \leq dx^{1/3} \\ (q, B)=1}} \max_{(b, q)=1} (E(2dx; q, b) + E(dx; q, b)),$$

while the second part is less than

$$(E(2dx; d, c) + E(dx; d, c)) \sum_{q \leq x^{1/3}} \frac{1}{\phi(q)}$$

since  $\phi(d)\phi(q) \leq \phi(dq)$ . For  $x \leq z \leq 2x$ , Theorem 2 implies that

$$\begin{aligned} \sum_{\substack{q \leq dx^{1/3} \\ (q, B)=1}} \max_{(a, q)=1} E(dz; q, a) &\leq \sum_{\substack{q \leq (2d_+x)^{2/5} \\ (q, B)=1}} \max_{2 \leq t \leq 2d_+x} \max_{(a, q)=1} E(t; q, a) \leq \frac{2d_+x}{e^{C\sqrt{\log(2d_+x)}}} \\ &\leq \frac{2x}{e^{C\sqrt{\log x} - (\log x)^{\frac{1}{2+\epsilon}}}} \ll \frac{x}{(\log x)^{100N^2+1}}. \end{aligned}$$

Additionally,

$$E(dz; d, c) < \sum_{\substack{q \leq dx^{1/3} \\ (q, B)=1}} \max_{(a, q)=1} E(dz; q, a) \ll \frac{x}{(\log x)^{100N^2+1}}$$

for  $x \leq z \leq 2x$ . Putting everything together (and using Lemma 1),

$$\begin{aligned} E &\leq \sum_{\substack{q \leq dx^{1/3} \\ (q, B)=1}} \max_{(b, q)=1} (E(2dx; q, b) + E(dx; q, b)) + (E(2dx; d, c) + E(dx; d, c)) \sum_{q \leq x^{1/3}} \frac{1}{\phi(q)} \\ &\ll \frac{x}{(\log x)^{100N^2+1}} + \frac{x}{(\log x)^{100N^2+1}} + \log x \left( \frac{x}{(\log x)^{100N^2+1}} + \frac{x}{(\log x)^{100N^2+1}} \right) \\ &\ll \frac{x}{(\log x)^{100N^2}}. \end{aligned}$$

This shows that Hypothesis 1 is satisfied. Additionally,

$$\begin{aligned} \Pi(x) &= \pi(2dx; d, c) - \pi(dx; d, c) \geq \frac{\pi(2dx) - \pi(dx)}{\phi(d)} - E(2dx; d, c) - E(dx; d, c) \\ &> \frac{dx}{2\phi(d) \log x} - O\left(\frac{x}{(\log x)^{100N^2+1}}\right) > \frac{dx}{3\phi(d) \log x} \end{aligned}$$

for sufficiently large  $x$ , using the Prime Number Theorem. Since  $B$  is a prime,  $\phi(B) \geq \frac{B}{2}$  so

$$\frac{\phi(B)}{B} \frac{\phi(d)}{d} \Pi(x) > \frac{x}{6 \log x}.$$

We have thus established equations (2) and (3). The theorem now follows from Theorem 3.1 of [8] by taking  $N$  to be large compared to the  $C$  corresponding to  $\alpha = 1$  and  $\theta = \frac{1}{3}$  in Maynard's result.  $\square$

### 3 Choosing Primes

For a given set  $\mathbf{S}$ , let  $\mathbf{D_S} = \{d : d \mid \prod_{s \in \mathbf{S}} s\}$  and  $n\mathbf{S} = \{ns : s \in \mathbf{S}\}$ . Let the *logarithmic spread* of  $S$  be the smallest non-zero value  $|\log s_1 - \log s_2|$  takes as  $s_1$  and  $s_2$  range over  $\mathbf{S}$ , and let the *logarithmic diameter* be the largest value the same expression takes.

Now, suppose  $E > 0$  and  $\gamma > 0$  are such that

$$\left\{ p \leq x : P^+(p-1) \leq x^{1-E} \right\} \geq \frac{\gamma x}{\log x}$$

for sufficiently large  $x$ , where  $P^+(n)$  is the largest prime factor of  $n$ . Indeed, such values exist and can be computed effectively, as in [1]. (This, in fact, goes back to work of Erdős [3].) Unlike in the paper of Alford, Granville, and Pomerance, the exact values of  $E$  and  $\gamma$  will make no difference in the end, so we will not specify them. Note that the smaller  $\delta$  is taken in Theorem 1, the stronger the result becomes. Therefore, for the rest of the paper, we will assume that  $\delta$  is some small but positive quantity. Let  $y$  be a large parameter (large in terms of both  $E$  and  $\delta$ ). Let  $\Delta = \frac{\delta}{24}$ . Let  $M$  be the smallest power of two greater than the value of  $N$  in Theorem 3 corresponding to  $m = 2$  and  $\epsilon = \Delta$ . Let  $\Upsilon = \lceil e^{y^{2+2\Delta}} \rceil$  and

$$L^* = \prod_{\substack{\frac{y}{\log y} \leq p \leq y \\ P^+(p-1) \leq y^{1-E}}} p.$$

Let  $p^*$  be the largest prime factor of  $s_x$  (as defined following Theorem 2) for  $x = 2\Upsilon(L^*)^2$ . Finally, let

$$\mathbf{P} = \left\{ \frac{y}{\log y} \leq p \leq y : P^+(p-1) \leq y^{1-E}, p \neq p^* \right\}.$$

Taking  $\eta = \frac{\gamma}{2}$ , it is evident that  $|\mathbf{P}| \geq \frac{\eta y}{\log y}$ . We now give a combinatorial lemma about  $\mathbf{D_P}$  which explains why we took  $M$  to be a power of two.



**Lemma 2.** *There exists a subset of  $\mathbf{D_P}$  (containing half as many elements) which can be partitioned into  $\frac{2^{|\mathbf{P}|}}{2M}$  sets of  $M$  elements, each with logarithmic spread greater than 1.*

*Proof.* Let  $k$  be the smallest integer such that  $2^{2k} > 2M \binom{2k}{k}$ . Since  $M$  is fixed, we can safely assume that  $y$  is arbitrarily large in terms of  $k$ . Let  $\mathbf{P}_0$  be an arbitrary subset of  $\mathbf{P}$  with  $2k$  elements. Then any subset of  $\mathbf{D_{P_0}}$  with at least  $\frac{2^{|\mathbf{P}_0|}}{2M}$  elements contains two elements such that one of them has strictly more prime factors than the other. When  $y$  is large, such a subset of  $\mathbf{D_{P_0}}$  will have logarithmic diameter greater than 1, considering that the primes in  $\mathbf{P}_0$  are within a factor of  $\log y$  of each other. We will say that a set  $\mathbf{S}$  possesses “the wide subset property” if every subset of  $\mathbf{D_S}$  with at least  $\frac{2^{|\mathbf{S}|}}{2M}$  elements has logarithmic diameter greater than 1. In particular,  $\mathbf{P}_0$  has this property.

Now let  $\mathbf{P}_1 = \mathbf{P}_0 \cup p_1$  for some  $p_1 \in \mathbf{P} \setminus \mathbf{P}_0$ . Suppose  $\mathbf{R}$  is a subset of  $\mathbf{D_{P_1}}$  with at least  $\frac{2^{|\mathbf{P}_1|}}{2M}$  elements. Decompose  $\mathbf{R}$  as  $\mathbf{R}_1 \cup (p\mathbf{R}_2)$  where  $\mathbf{R}_1 = \mathbf{R} \cap \mathbf{D_{P_0}}$  and  $\mathbf{R}_2 = (p^{-1}\mathbf{R}) \cap \mathbf{D_{P_0}}$ . Either  $\mathbf{R}_1$  or  $\mathbf{R}_2$  (perhaps both) has at least  $\frac{2^{|\mathbf{P}_0|}}{2M}$  elements, and they are subsets of  $\mathbf{D_{P_0}}$ . Hence, at least one of  $\mathbf{R}_1$  and  $\mathbf{R}_2$  has logarithmic diameter greater than 1, which means that  $\mathbf{R}$  has logarithmic diameter greater than 1. Hence  $\mathbf{D_{P_1}}$  also has “the wide subset property.” By induction,  $\mathbf{D_P}$  has this property as well.

Recall that  $M$  is a power of two. It therefore makes sense to define  $\mathbf{D}_1$  as the set consisting of the  $\frac{|\mathbf{D_P}|}{2M}$  smallest elements of  $\mathbf{D_P}$ ,  $\mathbf{D}_2$  as the set consisting of the next  $\frac{|\mathbf{D_P}|}{2M}$  smallest elements, and so on, resulting in sets  $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_{2M}$ . Let  $\mathbf{S}_j$  be the set which contains the  $j$ th element of  $\mathbf{D}_2, \mathbf{D}_4, \dots, \mathbf{D}_{2M}$ . Since  $\mathbf{D_P}$  has “the wide subset property,”  $\mathbf{D}_3, \mathbf{D}_5, \dots, \mathbf{D}_{2M-1}$  all have logarithmic diameter greater than 1. Therefore  $\mathbf{S}_j$  has logarithmic spread greater than 1 for all  $j$ . As promised, we now have  $\frac{|\mathbf{D_P}|}{2M}$  sets, each containing  $M$  different elements of  $\mathbf{D_P}$  and each with logarithmic spread greater than 1.  $\square$

Let  $L = \prod_{p \in \mathbf{P}} p$  and  $Y = \Upsilon L$ . Fix  $j$ . Let  $\mathbf{T}_j$  be the set of integers  $k$  between  $Y$  and  $2Y$  with  $(k, L) = 1$  such that there exist  $d$  and  $d'$ , distinct elements of  $\mathbf{S}_j$ , with both  $dk + 1$  and  $d'k + 1$  prime. Label the elements of  $\mathbf{S}_j$  as  $d_1, d_2, \dots, d_M$ . Let

$$\Omega_p = \{a \in (\mathbb{Z}/p\mathbb{Z})^\times : (d_i a + 1, p) = 1 \ \forall i\}$$

for  $p \mid L$  and then let

$$\Omega = \prod_{p \mid L} \Omega_p.$$

By the Chinese remainder theorem, each element of  $\mathbf{\Omega}$  can be thought of as an integer  $a_L$  with  $1 \leq a_L < L$  such that for every  $p \mid L$ ,  $a_L \equiv a_p \pmod{p}$ , where  $a_p$  is the element in  $\mathbf{\Omega}_p$  which corresponds to  $a_L$ .

**Lemma 3.** *For all  $a_L \in \mathbf{\Omega}$ , the set of linear forms  $\{(d_i L)k + (d_i a_L + 1)\}_{i=1}^M$  is admissible.*

*Proof.* If  $p \mid L$  then  $p \nmid d_i a_L + 1$  for all  $i$ , so  $p$  never divides  $(d_i L)k + (d_i a_L + 1)$  regardless of  $i$  and  $k$ . On the other hand, if  $p \nmid L$  then  $L$  has a multiplicative inverse  $L'$  in  $(\mathbb{Z}/p\mathbb{Z})^\times$ , and when  $k \equiv -L' a_L \pmod{p}$ ,

$$(d_i L)k + (d_i a_L + 1) \equiv d_i(-LL' a_L + a_L) + 1 \equiv d_i(-a_L + a_L) + 1 \equiv 1 \pmod{p}$$

for all  $i$ . Hence, no prime always divides one of the elements of the given set of linear forms.  $\square$

The reason for using this set of linear forms as opposed to  $\{d_i k + 1\}_{i=1}^M$  is we want to ensure that  $k$  is relatively prime to  $L$ . However, we also can't afford to only restrict our attention to  $\{(d_i L)k + (d_i a_L + 1)\}_{i=1}^M$  for a particular value of  $a_L$ . Therefore, we will use all values of  $a_L \in \mathbf{\Omega}$ , almost breaking even as a result since

$$\begin{aligned} 1 - \frac{|\mathbf{\Omega}|}{L} &\leq 1 - \prod_{p \mid L} \left(1 - \frac{M+1}{p}\right) = 1 - e^{-O(1) \sum_{p \mid L} \frac{1}{p}} \\ &= 1 - e^{-O(1)(\log \log y - \log \log \frac{y}{\log y})} = 1 - e^{-o(1)} = o(1). \end{aligned}$$

Fix a particular  $a_L \in \mathbf{\Omega}$ . We apply Theorem 3, taking  $m = 2$  and  $\epsilon = \Delta$ . Recall that  $M$  is at least as large as the value of  $N$  in the theorem corresponding to these values of  $m$  and  $\epsilon$ . Therefore, we can use the linear forms  $\{(d_i L)k + (d_i a_L + 1)\}$  for  $1 \leq i \leq M$ . We will also take  $x = \Upsilon$  and  $d_+ = (L^*)^2$ . Indeed, with these choices of variables,

$$e^{(\log d_+)^{2+\epsilon}} \leq e^{(\log(y^2 L^2))^{2+\Delta}} < e^{(\log(y^{\frac{4y}{\log y}}))^{2+\Delta}} = e^{(4y)^{2+\Delta}} < e^{y^{2+2\Delta}} \leq x$$

so the conditions of the theorem do indeed hold. Now, this implies that  $\mathbf{T}_j$  has at least  $\frac{\Upsilon}{(\log \Upsilon)^M}$  elements of the form  $Lk + a_L$ . We can then repeat this process for every element in  $\mathbf{\Omega}$ , getting different elements of  $\mathbf{T}_j$  each time. Therefore,

$$|\mathbf{T}_j| \geq \frac{\Upsilon |\mathbf{\Omega}|}{(\log \Upsilon)^M} = (1 - o(1)) \frac{\Upsilon L}{(\log \Upsilon)^M} \geq (1 - o(1)) \frac{Y}{(\log Y)^M}.$$

We now have a large set of primes, divided into pairs  $dk + 1$  and  $d'k + 1$ . Before applying the Pigeonhole Principle to find a particular  $k$  with many such pairs of primes, we need to filter out those values of  $d$ ,  $d'$ , and  $k$  for which  $dk + 1$  and  $d'k + 1$  “cluster.” To do this, we introduce two parameters,  $V$  and  $W$ . Recall that the goal is to ensure that there is no  $\alpha$  slightly greater than 1 such that the primes are very close to powers of  $\alpha$ . Roughly speaking,  $V$  defines what “very close” means, while  $W$  defines what “slightly greater than 1” means.

To make things more precise, consider, for some real numbers  $V$  and  $W$  which will be chosen later, the set  $\mathbf{U}_j$  of  $k$  between  $Y$  and  $2Y$  such that there exist  $d$  and  $d'$ , two elements of  $\mathbf{S}_j$ , with  $d > d'$ , as well as  $\omega$  with  $|\omega| < W$ , such that

$$|\omega \log(dk + 1) - m| \leq \frac{1}{V}$$

and

$$|\omega \log(d'k + 1) - m'| \leq \frac{1}{V}$$

for some non-zero integers  $m$  and  $m'$ . Heuristically, for any given choice of  $\omega$ , the chance of a random integer  $k$  satisfying these two equations is  $\frac{1}{V^2}$ . On the other hand, this chance essentially resets every time  $\omega$  goes up by  $\frac{1}{V \log Y}$ . Hence the chance that for a random integer  $k$ , there exists  $\omega$  satisfying these two equations is roughly  $\frac{W \log Y}{V}$ . If  $W$  is significantly smaller than  $V$ , one would expect that  $\mathbf{U}_j$  would have relatively few elements. We now work to make this heuristic rigorous, starting by formalizing the two relevant properties that the sequences  $\log(dk + 1)$  and  $\log(d'k + 1)$  possess.

**Condition (Parallel).** *Two sequences  $\{a_i\}_{i=1}^n$  and  $\{a'_i\}_{i=1}^n$  are said to satisfy the Parallel Condition if there exists  $\alpha > 1$  such that  $|a_i - a'_i - \alpha| \leq \frac{1}{n}$  for all  $i$ .*

**Condition (Smoothness).** *A sequence  $\{a_i\}_{i=1}^n$  is said to satisfy the Smoothness Condition if the number of  $i$  such that  $|a_i - \beta| \leq \frac{1}{n}$  is at most 12 for all  $\beta$ .*

**Lemma 4.** *Suppose  $A > 1$  is some real number and  $\{a_i\}_{i=1}^n$  and  $\{a'_i\}_{i=1}^n$  are sequences of real numbers between 1 and  $A$  which together satisfy the Parallel Condition. Assume further that  $\{a_i\}$  satisfies the Smoothness Condition. For any positive real numbers  $W$  and  $V$  with  $VW \ll n$ , the number of  $i$  for which there exist  $\omega$  with  $|\omega| < W$  and non-zero integers  $m$  and  $m'$  such that*

$$|\omega a_i - m| \leq \frac{1}{V}$$

and

$$|\omega a'_i - m'| \leq \frac{1}{V}$$

is  $O(\frac{A^3 W \log(AW)n}{V})$ .

*Proof.* For integers  $\ell$  and  $m$ , let  $\mathbf{K}_{\ell,m}$  be the number of indices  $i$  for which there exists  $\omega$  with  $|\omega| < W$  such that

$$|\omega a_i - m| \leq \frac{1}{V}$$

and

$$|\omega a'_i - m + \ell| \leq \frac{1}{V},$$

where both  $m$  and  $m - \ell$  are both non-zero. Fix integers  $\ell$  and  $m$  with  $m$  and  $m - \ell$  both non-zero. Suppose  $i_1$  and  $i_2$  are two indices with corresponding  $\omega_1$  and  $\omega_2$  less than  $W$  in absolute value such that

$$|\omega_k a_{i_k} - m| \leq \frac{1}{V}$$

and

$$|\omega_k a'_{i_k} - m + \ell| \leq \frac{1}{V}$$

for both choices of  $k \in \{1, 2\}$ . Then

$$|\omega_k(a_{i_k} - a'_{i_k}) - \ell| \leq \frac{2}{V}$$

for both  $k$ . By assumption, there exists  $\alpha > 1$  such that  $|a_i - a'_i - \alpha| \leq \frac{1}{n}$  for all  $i$ . It follows that

$$|\omega_k \alpha - \ell| \leq |\omega_k(a_{i_k} - a'_{i_k}) - \ell| + |\omega_k(a_{i_k} - a'_{i_k} - \alpha)| \leq \frac{2}{V} + \frac{W}{n} = O(\frac{1}{V})$$

for both  $k$ . In particular,

$$|\omega_1 - \omega_2| \leq \frac{1}{\alpha} |(\omega_1 \alpha - \ell) - (\omega_2 \alpha - \ell)| < |\omega_1 \alpha - \ell| + |\omega_2 \alpha - \ell| = O(\frac{1}{V}).$$

Now,

$$\begin{aligned} |a_{i_1} - a_{i_2}| &= \frac{|(\omega_1 a_{i_1} - m) - (\omega_2 a_{i_2} - m) + (\omega_2 - \omega_1) a_{i_2}|}{|\omega_1|} = O(\frac{a_{i_2}}{V|\omega_1|}) \\ &= O(\frac{a_{i_1} a_{i_2}}{V|m|}) = O(\frac{A^2}{V|m|}) \end{aligned}$$

since  $|\omega_1| \gg \frac{|m|}{a_{i_1}}$ , using the fact that  $m$  is non-zero. By assumption, for all  $\epsilon \geq \frac{1}{n}$ , the number of  $i$  such that  $|a_i - \beta| \leq \epsilon$  is  $O(\epsilon n)$  regardless of  $\beta$ . Let  $\epsilon = \frac{A^2}{V|m|}$ . Since  $|m| = O(a_{i_1}\omega_1) = O(AW)$ , it follows that

$$\epsilon \gg \frac{A}{VW} \gg \frac{A}{n} > \frac{1}{n}.$$

Therefore,  $\mathbf{K}_{\ell,m}$  has  $O(\frac{A^2 n}{V|m|})$  elements. Hence, setting  $\mathbf{K} = \cup_{\ell,m} \mathbf{K}_{\ell,m}$ , we have that

$$\begin{aligned} |\mathbf{K}| &= \sum_{\ell=O(AW)} \sum_{\substack{m=O(AW) \\ m \neq 0}} |\mathbf{K}_{\ell,m}| = O(AW \log(AW) \frac{A^2 n}{V}) \\ &= O(\frac{A^3 W \log(AW) n}{V}) \end{aligned}$$

as promised.  $\square$

Let  $n = Y$  and  $A = 2 \log Y$ . Let  $V = e^{\frac{\eta y}{(4+2\Delta) \log y}}$  and  $W = e^{\frac{\eta y}{(4+4\Delta) \log y}}$ . Let  $a_k = \log(d(Y+k-1)+1)$  and  $a'_k = \log(d'(Y+k-1)+1)$  for  $1 \leq k \leq n$ . The Parallel Condition holds with  $\alpha = \log d - \log d'$  since

$$\begin{aligned} \log(d(Y+k-1)+1) - \log(d'(Y+k-1)+1) &= \log\left(\frac{d(Y+k-1)+1}{d'(Y+k-1)+1}\right) \\ &= \log\left(\frac{d}{d'} + O\left(\frac{1}{d'Y}\right)\right) \\ &= \alpha + O\left(\frac{1}{nd'}\right). \end{aligned}$$

Furthermore, if we define  $b_i = a_{i+1} - a_i$  for  $1 \leq i < n$  then we see that  $b_i$  is a decreasing sequence with  $b_1 < 3b_{n-1}$ , while  $a_i$  is an increasing sequence with  $a_n > a_1 + \frac{1}{2}$ . From these two facts, it follows that the average value of  $b_i$  is greater than  $\frac{1}{2n}$  and the smallest value of  $b_i$  is greater than  $\frac{1}{6n}$ . This shows that  $\{a_i\}$  satisfies the Smoothness Condition. Since our values of  $V$  and  $W$  satisfy the hypotheses of Lemma 4, we obtain the bound

$$|\mathbf{U}_j| = \binom{M}{2} O\left(\frac{W \log(W \log Y) \log^3(Y) Y}{V}\right) = o\left(\frac{Y}{(\log Y)^M}\right),$$

remembering that  $M$  is fixed. Therefore, removing the elements of  $\mathbf{U}_j$  from  $\mathbf{T}_j$  still leaves  $(1 - o(1)) \frac{Y}{(\log Y)^M}$  elements.

Now, so far all this work has been done for a particular choice  $j$ , but there are  $\frac{|\mathbf{D}_{\mathbf{P}}|}{2M} - 1$  other values of  $j$  which could have been chosen. This gives in total  $(1 - o(1)) \frac{|\mathbf{D}_{\mathbf{P}}| Y}{2M (\log Y)^M}$  pairs  $(k, j)$  with the following properties.

1.  $k \notin \mathbf{U}_j$
2.  $(k, L) = 1$
3.  $Y \leq k < 2Y$
4. There exist distinct elements of  $\mathbf{S}_j$ , which we will call  $d$  and  $d'$ , such that  $dk + 1$  and  $d'k + 1$  are both prime.

Hence, the pigeonhole principle tells us that there exists  $k_0$  between  $Y$  and  $2Y$  relatively prime to  $L$  with  $(1 - o(1)) \frac{|\mathbf{D_P}|}{2M(\log Y)^M}$  pairs of “non-clustering” primes, each of the form  $dk + 1$  for  $d$  dividing  $L$ . Let  $\mathbf{Q}$  be the set of these pairs. Then,

$$|\mathbf{Q}| \geq (1 - o(1)) \frac{|\mathbf{D_P}|}{2M(\log Y)^M} \geq (1 - o(1)) \frac{2^{\frac{\eta y}{\log y}}}{2M(\log Y)^M} > e^{\frac{\eta y}{2 \log y}}.$$

## 4 Building Carmichael Numbers

In this section, we will construct Carmichael numbers of controlled size. We start with an application of Fourier analysis.

**Theorem 4.** *Let  $q_1, q_2, \dots, q_N$  be a set of primes with product  $Q$ . Label the largest of the primes  $q_0$ . Assume that  $q_i^2 \geq q_0$  for all  $i$ . Let  $\ell$  be some positive integer which is not divisible by any of the  $q_i$ . Assume that*

$$N > \max \left( \log q_0, \lambda(\ell) \log (\phi(\ell)) \right)^5.$$

*Further assume that for all  $\omega$  with  $\frac{2}{\lambda(\ell) \log q_0} < |\omega| < 4A \log \Phi$  and  $\alpha$  with  $|\alpha| \geq \frac{2\pi}{\lambda(\ell)}$ , at most half of the primes  $q_i$  satisfy*

$$|\omega \log q_i - k\alpha| < 8\sqrt{\frac{\log \Phi}{N}} \quad (4)$$

*for some integer  $k$ , where  $A$  is some real number greater than 1 and*

$$\Phi = (AN\phi(\ell) \log q_0)^2.$$

*For all sufficiently large  $N$ , the number of values of  $d$  which divide  $Q$ , while satisfying both  $d \equiv 1 \pmod{\ell}$  and  $|\log d - \frac{\log Q}{2} - B| \leq \frac{1}{2A}$ , is  $2^{N - O(\sqrt{N} + \log \Phi)}$  for all  $B \leq \frac{\sqrt{N} \log q_0}{36}$ .*

*Proof.* Let  $H = (\mathbb{Z}/\ell\mathbb{Z})^\times$ . Let  $h_i$  be the element of  $H$  corresponding to  $q_i$ . We will construct a sequence of nested groups starting with  $H$ , which we relabel  $G_0$ . Starting with  $j = 1$  and then increasing  $j$  by increments of one, let  $G_j$  be a proper subgroup of  $G_{j-1}$  such that

$$\#\{i : h_i \in G_{j-1} \setminus G_j\} < \lambda(G_{j-1})^2 \log(\#\{i : h_i \in G_{j-1}\} |G_{j-1}|),$$

with the process stopping when no such proper subgroup exists. The result is a chain of proper subgroups

$$G_0 \supset G_1 \supset \cdots \supset G_m$$

where for every proper subgroup of  $G_m$ , at least  $\lambda(G_m)^2 \log(n|G_m|)$  of the  $h_i$  are in  $G_m$  but not in the proper subgroup (letting  $n$  be the number of  $i$  such that  $h_i$  is in  $G_m$ ). At each step in this process, at most  $\lambda(G_0)^2 \log(N|G_0|)$  of the  $h_i$  are discarded. Additionally,  $m$  is bounded by  $2 \log |G_0|$ . As a result

$$\begin{aligned} N - n &< 2\lambda(G_0)^2 \log(N|G_0|) \log |G_0| < 2\lambda(\ell)^2 \log^2(\phi(\ell)) \log N \\ &< 2N^{\frac{2}{5}} \log N < \frac{\sqrt{N}}{6}. \end{aligned}$$

In particular,  $n > \frac{5N}{6}$ . Let  $q_{a_1}, q_{a_2}, \dots, q_{a_{N-n}}$  be the primes which were discarded, and let  $p_1, p_2, \dots, p_n$  be the remaining primes, the largest of which we will call  $p_0$ . Let  $b$  be the recentered version of  $B$ , that is,

$$b = B + \frac{\sum_{i=1}^{N-n} \log q_{a_i}}{2}.$$

Then

$$\begin{aligned} |b| &\leq |B| + \frac{\sum_{i=1}^{N-n} \log q_{a_i}}{2} \leq |B| + \frac{\sqrt{N} \log q_0}{12} \leq \frac{\sqrt{N} \log q_0}{9} \\ &\leq \frac{2\sqrt{N} \log p_0}{9} < \frac{\sqrt{n} \log p_0}{4}. \end{aligned}$$

In what follows, we will refer to  $G_m$  as  $G$ , and we will write it additively, despite its origin as a subgroup of a multiplicative group. We now introduce some final notation before the main part of the proof begins. Let  $K$  be the interval  $[b - \frac{1}{2A}, b + \frac{1}{2A}]$ . Let  $P = \prod_i p_i$ . Let  $r_i = \frac{\log p_i}{2}$  and  $r_0 = \frac{\log p_0}{2}$ . Let  $g_i$  be the representative of  $p_i$  in  $G$ . Let  $\chi_0$  be the trivial character of  $G$ . Let  $X_i$  be the random variable on the space  $G \times \mathbb{R}$  which is  $(0, -r_i)$  or  $(g_i, r_i)$  with equal probability.

Now, we wish to show that

$$\Pr_{d|P}\left(\log d - \frac{\log P}{2} \in K, \sum_{i: p_i|d} g_i = 0\right)$$

is positive, where  $\Pr_{d|P}$  denotes the probability of a random divisor of  $P$  satisfying some condition. Let  $\mathbb{1}_e$  be the characteristic function of the identity in  $G$  and  $\mathbb{1}_K$  be the characteristic function of the interval  $K$  in  $\mathbb{R}$ . Note that for every  $d$  that divides  $P$ ,

$$\sum_{i: p_i|d} r_i + \sum_{i: p_i \nmid d} (-r_i) = \sum_{i: p_i|d} \log p_i - \frac{1}{2} \sum_i \log p_i = \log d - \frac{\log P}{2}.$$

Therefore, we can write

$$\Pr_{d|P}\left(\log d - \frac{\log P}{2} \in K, \sum_{i: p_i|d} g_i = 0\right) = \mathbb{E}\left(\mathbb{1}_{e \times K}\left(\sum_j X_j\right)\right)$$

where  $\mathbb{1}_{e \times K} = (\mathbb{1}_e, \mathbb{1}_K)$ , in other words,  $\mathbb{1}_{e \times K}$  acts on  $G$  as  $\mathbb{1}_e$  and acts on  $\mathbb{R}$  as  $\mathbb{1}_K$ . We are about to take a Fourier transform, so we will replace  $\mathbb{1}_K$  with a Gaussian, making the higher frequencies easier to deal with. Let  $\Psi(x) = e^{-a^2(x-b)^2}$  where  $a = 2A\sqrt{\log \Phi}$  and let  $\Psi_e = (\mathbb{1}_e, \Psi)$ . Now,

$$\mathbb{1}_e(g) = \frac{1}{|G|} \sum_{\chi \in \hat{G}} \chi(g)$$

while

$$\hat{\Psi}(\omega) = \int_{-\infty}^{\infty} \Psi(x) e^{-i\omega x} dx = \frac{\sqrt{\pi}}{a} e^{-\frac{\omega^2}{4a^2} - ib\omega}$$

so

$$\Psi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\Psi}(\omega) e^{i\omega x} d\omega = \frac{1}{2a\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{4a^2} - ib\omega} e^{i\omega x} d\omega.$$

Thus

$$\Psi_e(x, g) = \frac{1}{2a\sqrt{\pi}|G|} \sum_{\chi} \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{4a^2} - ib\omega} \chi_{\omega}(x, g) d\omega$$

where  $\chi_{\omega}$  is the character on  $G \times \mathbb{R}$  corresponding to  $\chi$  on  $G$  and  $e^{i\omega x}$  on  $\mathbb{R}$ , that is,  $\chi_{\omega} = \chi \boxtimes e^{i\omega x}$ . Let

$$C = 2a\sqrt{\pi}|G|$$



and

$$E = \mathbb{E}\left(\Psi_e\left(\sum_j X_j\right)\right).$$

Note that

$$\mathbb{E}(\chi_\omega(X_j)) = \frac{e^{i\omega r_j} \chi(g_j) + e^{-i\omega r_j}}{2}.$$

We now claim that

$$\begin{aligned} CE &= C \mathbb{E}\left(\frac{1}{2a\sqrt{\pi}|G|} \sum_\chi \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{4a^2} - i b \omega} \chi_\omega\left(\sum_j X_j\right) d\omega\right) \\ &= \mathbb{E}\left(\sum_\chi \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{4a^2} - i b \omega} \prod_j (\chi_\omega(X_j)) d\omega\right) \\ &= \sum_\chi \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{4a^2} - i b \omega} \prod_j \left(\frac{e^{i\omega r_j} \chi(g_j) + e^{-i\omega r_j}}{2}\right) d\omega \\ &= \Re\left(\sum_\chi \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{4a^2} - i b \omega} \prod_j \left(\frac{e^{i\omega r_j} \chi(g_j) + e^{-i\omega r_j}}{2}\right) d\omega\right). \end{aligned}$$

The second equality follows since the random variables  $X_j$  are independent, while in the next step, we interchange the expectation operator with the integral and the product. (Note that the integrand has absolute value bounded by a Gaussian and is therefore absolutely convergent.) Finally, the last equality follows since  $C$  and  $E$  are both real.

We will show that the main contribution to the above expression comes from when  $\chi = \chi_0$  and  $\omega$  is very small. The challenge is to show that every other contribution is small. The main way we do this is by showing that  $\prod_j \left(\frac{e^{i\omega r_j} \chi(g_j) + e^{-i\omega r_j}}{2}\right)$  is small. When  $\chi$  is non-trivial and  $\omega$  is very small, this product will be small if we can show that  $\chi(g_j)$  is sometimes not equal to 1. Fortunately,  $G$  was constructed entirely for that purpose. Then as  $\omega$  becomes larger (in absolute value), the set  $\{e^{i\omega r_j}\}_{j=1}^n$  should spread out, providing cancellation. Once  $\omega$  reaches a certain size, the contribution becomes negligible for a different reason, namely because of the  $e^{-\frac{\omega^2}{4a^2}}$  term.

Now, for values of  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$  which will be chosen shortly, we can write that

$$CE > 2\left(I_1 + I_2 - I_3 - \sum_{\chi \neq \chi_0} J_1^\chi - \sum_\chi (J_2^\chi + J_3^\chi)\right)$$

where

$$I_1 = \int_0^{c_1} e^{-\frac{\omega^2}{4a^2}} \cos(b\omega) \prod_j \cos(\omega r_j) d\omega,$$

$$I_2 = \int_{c_1}^{c_2} e^{-\frac{\omega^2}{4a^2}} \cos(b\omega) \prod_j \cos(\omega r_j) d\omega,$$

and

$$I_3 = \int_{c_2}^{c_3} \prod_j |\cos(\omega r_j)| d\omega,$$

while

$$J_1^X = \max_{s \in \{\pm 1\}} \int_0^{c_3} \prod_j \left| \frac{e^{is\omega r_j} \chi(g_j) + e^{-is\omega r_j}}{2} \right| d\omega,$$

$$J_2^X = \max_{s \in \{\pm 1\}} \int_{c_3}^{c_4} \prod_j \left| \frac{e^{is\omega r_j} \chi(g_j) + e^{-is\omega r_j}}{2} \right| d\omega,$$

and

$$J_3^X = \int_{c_4}^{\infty} e^{-\frac{\omega^2}{4a^2}} d\omega.$$

(Note that the maximum over  $s$  in  $J_1^X$  and  $J_2^X$  comes from having to account for the contribution from negative values of  $\omega$ .)

Again, the claim is that the main contribution comes from  $I_1$ . We take  $c_1 = \frac{1}{r_0 \sqrt{n}}$ , which will be small enough to ensure that the minimum value of the integrand over that range will still be large. Then  $c_2$  will be chosen to be  $\frac{3}{r_0 \sqrt{n}}$ , ensuring that  $I_2$  is at least positive. Next,  $c_3 = \frac{1}{\lambda(G)r_0}$ . (Since  $n > 9\lambda(G)^2$ ,  $c_2$  is indeed less than  $c_3$ .) Elementary methods will show that  $I_3$  is small compared to  $I_1$ . Then  $J_1^X$  can be shown to be quite small because of the work done previously in making sure that there is no proper subgroup of  $G$  which contains almost all of the  $g_i$ . Bounding  $J_3^X$  will be quite easy as long as  $c_4$  is large enough for the Gaussian term to become vanishingly small. To that end, we take  $c_4 = 4A \log \Phi$ . That leaves the task of bounding  $J_2^X$ . A priori, this part of the integral is very difficult to control because it depends on there being no grand conspiracy in the choice of our primes. This is where our non-clustering hypothesis (4) comes in.

We will start with a lower bound on  $I_1$ . Let  $0 \leq \omega \leq c_1$ . By assumption,  $A > 1$  so  $a > 1$ . Therefore,  $e^{-\frac{\omega^2}{4a^2}} > \frac{99}{100}$  for large  $n$ . Also,

$$\cos(b\omega) \geq \cos\left(\frac{1}{2}\right) > \frac{5}{6}$$

since

$$|b\omega| \leq \frac{\sqrt{n} \log p_0}{4r_0 \sqrt{n}} = \frac{1}{2}.$$

Now

$$\cos(\omega r_j) \geq 1 - \frac{\omega^2 r_j^2}{2} \geq 1 - \frac{1}{2n}$$

so

$$\prod_{j=1}^n (\cos(\omega r_j)) \geq \left(1 - \frac{1}{2n}\right)^n > \frac{20}{33}$$

for large  $n$ . Hence

$$I_1 = \int_0^{c_1} e^{-\frac{\omega^2}{4a^2}} \cos(b\omega) \prod_j (\cos(\omega r_j)) d\omega > \int_0^{c_1} \frac{1}{2} d\omega = \frac{1}{2r_0\sqrt{n}}.$$

Now, the integrand

$$e^{-\frac{\omega^2}{4a^2}} \cos(b\omega) \prod_j (\cos(\omega r_j))$$

will be positive if  $|b\omega| < \frac{\pi}{2}$  (which will be true if  $|\omega| < \frac{\pi}{r_0\sqrt{n}}$ ) and  $|\omega r_j| < \frac{\pi}{2}$  for all  $j$  (which will be true if  $|\omega| < \frac{\pi}{2r_0}$ ). Since  $c_2 = \frac{3}{r_0\sqrt{n}}$ , it follows that  $I_2 > 0$ .

Now suppose  $c_2 \leq |\omega| \leq c_3$ . Then  $|\omega r_j| \leq 1$  for all  $j$ . Hence

$$I_3 = \int_{c_2}^{c_3} \prod_j |\cos(\omega r_j)| d\omega < \int_{c_2}^{c_3} \left(1 - \frac{\omega^2 r_0^2}{4}\right)^n d\omega < \int_{c_2}^{c_3} e^{-\frac{n\omega^2 r_0^2}{4}} d\omega.$$

Letting  $u = \frac{\sqrt{n}\omega r_0}{2}$ ,

$$I_3 < \frac{2}{r_0\sqrt{n}} \int_{\frac{3}{2}}^{\frac{\sqrt{n}}{2\lambda(G)}} e^{-u^2} du < \frac{2}{r_0\sqrt{n}} \int_{\frac{3}{2}}^{\infty} e^{-u^2} du = \frac{\sqrt{\pi}}{r_0\sqrt{n}} \operatorname{erfc}\left(\frac{3}{2}\right) < \frac{1}{5r_0\sqrt{n}},$$

where

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$$

with the elementary bound  $\operatorname{erfc}(x) \leq e^{-x^2}$  for  $x \geq 0$ .

Let  $\chi$  be any non-trivial character. Take  $s$  to be either  $-1$  or  $1$ . Let  $0 \leq |\omega| \leq c_3$ . Note that

$$\left| \frac{\chi(g_j) + e^{-2is\omega r_j}}{2} \right| = \sqrt{\frac{1 + \Re(\chi(g_j)e^{2is\omega r_j})}{2}} = \sqrt{\frac{1 + \cos(\frac{2\pi k_j}{\lambda(G)} + 2s\omega r_j)}{2}}$$

where  $-\frac{\lambda(G)}{2} < k_j \leq \frac{\lambda(G)}{2}$  is the integer such that  $\chi(g_j) = e^{\frac{2\pi i k_j}{\lambda(G)}}$ . Indeed, such an integer exists since the order of  $g_j$  divides  $\lambda(G)$ , so  $\chi(g_j)^{\lambda(G)} = 1$ . If  $|k_j| \geq 1$  then

$$\left| \frac{2\pi k_j}{\lambda(G)} + 2s\omega r_j \right| \geq \left| \frac{2\pi k_j}{\lambda(G)} \right| - |2s\omega r_j| \geq \frac{2\pi}{\lambda(G)} - \frac{2\pi}{3\lambda(G)} = \frac{4\pi}{3\lambda(G)} > \frac{4}{\lambda(G)}$$

because  $|\omega r_j| \leq \frac{1}{\lambda(G)} < \frac{\pi}{3\lambda(G)}$ . On the other hand,

$$\left| \frac{2\pi k_j}{\lambda(G)} + 2s\omega r_j \right| \leq \pi + \frac{2}{\lambda(G)}.$$

Therefore, if  $|k_j| \geq 1$  then the distance between  $\frac{2\pi k_j}{\lambda(G)} + 2s\omega r_j$  and the nearest multiple of  $2\pi$  is greater than  $\frac{4}{\lambda(G)}$  because then  $\lambda(G) \geq 2$  and even in the extreme case that  $\lambda(G) = 2$ ,

$$2\pi - \left( \pi + \frac{2}{\lambda(G)} \right) = \pi - 1 > 2 = \frac{4}{\lambda(G)}.$$

Hence,

$$\begin{aligned} \left| \frac{\chi(g_j) + e^{-2is\omega r_j}}{2} \right| &< \sqrt{\frac{1 + \cos\left(\frac{4}{\lambda(G)}\right)}{2}} < \sqrt{\frac{1 + 1 - \frac{4}{\lambda(G)^2}}{2}} \\ &= \sqrt{1 - \frac{2}{\lambda(G)^2}} < 1 - \frac{1}{\lambda(G)^2} \end{aligned}$$

if  $|k_j| \geq 1$ . Recall that  $G$  was constructed so that for every proper subgroup of  $G$ , at least  $\lambda(G)^2 \log(n|G|)$  of the  $g_j$  are outside it. The kernel of a non-trivial character is a proper subgroup. Hence, for at least  $\lambda(G)^2 \log(n|G|)$  values of  $j$ ,  $\chi(g_j) \neq 1$ , or alternatively,  $|k_j| \geq 1$ . Therefore, since even when  $k_j = 0$ , the corresponding term in the product is still at most 1,

$$\begin{aligned} J_1^X &= \max_{s \in \{-1, 1\}} \int_0^{c_3} \prod_j \left| \frac{e^{is\omega r_j} \chi(g_j) + e^{-is\omega r_j}}{2} \right| d\omega \\ &< \int_0^{c_3} \left( 1 - \frac{1}{\lambda(G)^2} \right)^{\lambda(G)^2 \log(n|G|)} d\omega < \int_0^{c_3} e^{-\log(n|G|)} d\omega \\ &= \frac{e^{-\log(n|G|)}}{\lambda(G)r_0} = \frac{1}{n|G|\lambda(G)r_0} < \frac{1}{100|G|r_0\sqrt{n}}. \end{aligned}$$

Now let  $\chi$  be any character. Seeing as  $\lambda(G) \leq \lambda(\ell)$ , the non-clustering hypothesis (4) implies that for every

$$\frac{2}{\lambda(G) \log p_0} \leq |\omega| \leq 4A \log \Phi,$$

at least a third of the possible values of  $j$  are such that

$$\left| 2\omega r_j - \frac{2\pi k}{\lambda(G)} \right| \geq 8\sqrt{\frac{\log \Phi}{N}} > 7\sqrt{\frac{6}{5N} \log \Phi} > 7\sqrt{\frac{\log \Phi}{n}}$$

for all integers  $k$ . (This would have been true for at least half of the original primes, but we have since discarded up to a sixth of the primes.) For  $\omega$  in the same interval, it follows that letting  $k$  be the integer that minimizes  $2s\omega r_j - \frac{2\pi k}{\lambda(G)}$  for  $s = 1$ ,

$$\cos \left( 2s\omega r_j - \frac{2\pi k}{\lambda(G)} \right) < \cos \left( 7\sqrt{\frac{\log \Phi}{n}} \right) < 1 - \frac{49 \log \Phi}{4n} < 1 - \frac{12 \log \Phi}{n}.$$

Moreover, the same inequality holds for  $s = -1$  and  $k$  which minimizes  $-2\omega r_j - \frac{2\pi k}{\lambda(G)}$  since cosine is even. We have thus established an upper bound on  $\cos \left( 2s\omega r_j - \frac{2\pi k}{\lambda(G)} \right)$  for all integers  $k$  and both choices of  $s$ . Hence, it follows that in this range of  $\omega$  between  $c_3$  and  $c_4$ ,

$$\left| \frac{\chi(g_j) + e^{-2is\omega r_j}}{2} \right| = \sqrt{\frac{1 + \cos \left( \frac{2\pi k_j}{\lambda(G)} + 2s\omega r_j \right)}{2}} < \sqrt{1 - \frac{6 \log \Phi}{n}} < 1 - \frac{3 \log \Phi}{n}$$

where  $k_j$  is chosen to be an integer such that  $\chi(g_j) = e^{\frac{2\pi i k_j}{\lambda(G)}}$ . Therefore

$$\begin{aligned} J_2^\chi &= \max_{s \in \{-1, 1\}} \int_{c_3}^{c_4} \prod_j \left| \frac{e^{is\omega r_j} \chi(g_j) + e^{-is\omega r_j}}{2} \right| d\omega < \int_{c_3}^{c_4} \left( 1 - \frac{3 \log \Phi}{n} \right)^{\frac{n}{3}} d\omega \\ &< \int_{c_3}^{c_4} e^{-\log \Phi} d\omega < c_4 e^{-\log \Phi} = \frac{4A \log \Phi}{\Phi} = \frac{4A}{\sqrt{\Phi}} \frac{\log \Phi}{\sqrt{\Phi}} \\ &< \frac{4}{N\phi(\ell) \log q_0} < \frac{1}{100|G|r_0\sqrt{n}}, \end{aligned}$$

recalling that  $\sqrt{\Phi} = AN\phi(\ell) \log q_0$ .

Finally, letting  $u = \frac{\omega}{2a}$ ,

$$\begin{aligned}
J_3^\chi &= \int_{c_4}^{\infty} e^{-\frac{\omega^2}{4a^2}} d\omega = 2a \int_{\frac{c_4}{2a}}^{\infty} e^{-u^2} du = a\sqrt{\pi} \operatorname{erfc}\left(\frac{c_4}{2a}\right) \leq a\sqrt{\pi} e^{-\frac{c_4^2}{4a^2}} \\
&= a\sqrt{\pi} e^{-\frac{16A^2 \log^2 \Phi}{16A^2 \log \Phi}} = \frac{a\sqrt{\pi}}{\Phi} = \frac{2\sqrt{\pi}A}{\sqrt{\Phi}} \sqrt{\frac{\log \Phi}{\Phi}} \\
&< \frac{2\sqrt{\pi}}{N\phi(\ell) \log q_0} < \frac{1}{100|G|r_0\sqrt{n}}.
\end{aligned}$$

Putting everything together,

$$CE > \frac{2}{r_0\sqrt{n}} \left( \frac{1}{2} - \frac{1}{5} - \frac{|G|}{100|G|} - \frac{|G|}{100|G|} - \frac{|G|}{100|G|} \right) > \frac{1}{2r_0\sqrt{n}}$$

so

$$E > \frac{1}{2r_0\sqrt{n}} \frac{1}{2a\sqrt{\pi}|G|} > \frac{1}{15A|G|r_0\sqrt{n} \log \Phi},$$

using that  $a = 2A\sqrt{\log \Phi}$ . Now,

$$\begin{aligned}
\Pr_{d|P} \left( \log d - \frac{\log P}{2} \in K, \sum_{i: p_i|d} g_i = 0 \right) &= \mathbb{E} \left( \mathbb{1}_{e \times K} \left( \sum_j X_j \right) \right) \\
&= \mathbb{E} \left( \Psi_e \left( \sum_j X_j \right) \right) - \mathbb{E} \left( \Psi_e \left( \sum_j X_j \right) - \mathbb{1}_{e \times K} \left( \sum_j X_j \right) \right) \\
&\geq E - \max(\Psi_e - \mathbb{1}_{e \times K}) = E - e^{-\frac{a^2}{4A^2}} = E - e^{-\log \Phi} \\
&> \frac{1}{15A|G|r_0\sqrt{n} \log \Phi} - \frac{1}{AN\phi(\ell) \log q_0 \sqrt{\Phi}} \\
&> \frac{1}{15A|G|r_0\sqrt{n} \log \Phi} - \frac{1}{100A|G|r_0\sqrt{n} \log \Phi} \\
&> \frac{1}{18A|G|r_0\sqrt{n} \log \Phi}.
\end{aligned}$$

Hence there are at least

$$\frac{2^n}{18A|G|r_0\sqrt{n} \log \Phi} > \frac{2^n}{\Phi} = \frac{2^{N-O(\sqrt{N})}}{\Phi}$$

values of  $d$  dividing  $P$  (and hence  $Q$ ) such that  $d \equiv 1 \pmod{\ell}$  and

$$\left| \log d - \frac{\log Q}{2} - B \right| = \left| \log d - \frac{\log P}{2} - b \right| \leq \frac{1}{2A}. \quad \square$$

The proof of this technical result out of the way, we now return to the main argument. In Section 3, we constructed a set  $\mathbf{Q}$ , consisting of pairs of “non-clustering” primes. These primes are going to be the prime divisors of our Carmichael numbers.

Let  $N_- = \lceil \frac{1}{4} e^{\frac{\eta y}{2 \log y}} \rceil$ , and let  $N_+ = \lfloor e^{\frac{\eta y}{2 \log y}} \rfloor$ . Recall that  $|\mathbf{Q}| \geq N_+$ . We label the first  $N_+$  pairs in  $\mathbf{Q}$  as  $(q_1, q'_1), \dots, (q_{N_+}, q'_{N_+})$ . For  $N \leq N_+$ , let  $\mathbf{Q}_N$  be the set of primes of the form  $q_i$  or  $q'_i$  with  $1 \leq i \leq N$ . Remember that these primes are all of the form  $dk_0 + 1$  for some  $d \mid L$  and hence are distinct from each other. Let  $Q_N = \prod_{q \in \mathbf{Q}_N} q$ . Let

$$Z_-(y) = \left\lceil e^{y^{2+2\Delta}} \right\rceil^{\frac{1}{3} e^{\frac{\eta y}{2 \log y}}}$$

and

$$Z_+(y) = \left\lceil e^{y^{2+2\Delta}} \right\rceil^{\frac{11}{12} e^{\frac{\eta y}{2 \log y}}}.$$

Let  $z$  be an arbitrary integer between  $Z_-(y)$  and  $Z_+(y)$ . First we claim that there exists an integer  $n$  between  $N_-$  and  $N_+$  such that

$$\left| \log z - \frac{\log Q_n}{2} \right| < \log Y.$$

Indeed, recalling that  $Y = \Upsilon L$  where  $\Upsilon = \lceil e^{y^{2+2\Delta}} \rceil$  and

$$L = \prod_{p \in \mathbf{P}} p < y^{\frac{2y}{\log y}} = e^{2y},$$

it follows that

$$Q_{N_+} > Y^{2e^{\frac{\eta y}{2 \log y}}} > (Z_+(y))^2$$

and

$$Q_{N_-} < (2YL)^{\frac{1}{2} e^{\frac{\eta y}{2 \log y}}} < \Upsilon^{\frac{2}{3} e^{\frac{\eta y}{2 \log y}}} = (Z_-(y))^2,$$

while

$$\frac{\log Q_{N+1} - \log Q_N}{2} < \frac{\log((2LY)^2)}{2} = \log(2LY) < 2 \log Y$$

for all  $N < N_+$ . Therefore, such  $n$  does indeed exist. Let

$$b = \log z - \frac{\log Q_n}{2} + \frac{1}{2e^{\frac{\eta y}{(4+6\Delta) \log y}}}.$$

Then  $|b| < \log Y + 1$ .

**Lemma 5.** *The hypotheses of Theorem 4 hold for  $N = 2n$ ,  $B = b$ ,  $\ell = L$ ,  $A = e^{\frac{\eta y}{(4+6\Delta)\log y}}$ , and  $\{q_i\} = \mathbf{Q}_n$ .*

*Proof.* Since  $L$  is small compared to  $Y$ , it is indeed true that for all  $i$ ,  $q_i^2 \geq q_0$ , where  $q_0$  is the largest element of  $\mathbf{Q}_n$ . Also,  $N > \log^5 q_0$ . Letting  $v_q(m)$  be the  $p$ -adic valuation of  $m$ ,

$$\begin{aligned} \lambda(L) &= \prod_{\substack{q \leq y^{1-E} \\ q \in \mathbb{P}}} \max_{p \in \mathbf{P}} q^{v_q(p-1)} \leq \prod_{\substack{q \leq y^{1-E} \\ q \in \mathbb{P}}} y \\ &= y^{\pi(y^{1-E})} < y^{\frac{2y^{1-E}}{(1-E)\log y}} = e^{\frac{2y^{1-E}}{1-E}}. \end{aligned} \quad (5)$$

Combining this with the fact that

$$\log \phi(L) < \log L < \log \left( y^{\frac{2y}{\log y}} \right) = 2y,$$

it follows that

$$N \geq \left( \lambda(L) \log \phi(L) \right)^5.$$

Also,  $|B| < \log Y + 1 < \frac{\sqrt{N} \log q_0}{36}$ .

Now, considering how  $\mathbf{Q}$  was constructed, at most half of the primes  $q \in \mathbf{Q}_n$  are such that

$$\min_{\substack{k \in \mathbb{Z} \\ k \neq 0}} (|\omega \log q - k|) \leq \frac{1}{e^{\frac{\eta y}{(4+2\Delta)\log y}}}$$

for any particular  $\omega$  with  $|\omega| \leq e^{\frac{\eta y}{(4+4\Delta)\log y}}$ . As a consequence of this and (5), for all  $\omega$  with  $|\omega| \leq e^{\frac{\eta y}{(4+5\Delta)\log y}}$  and  $\alpha \geq \frac{2\pi}{\lambda(L)}$ , at most half of the primes  $q \in \mathbf{Q}_n$  are such that

$$|\omega \log q - k\alpha| \leq \frac{1}{e^{\frac{\eta y}{(4+\Delta)\log y}}}$$

for some non-zero integer  $k$ . Additionally, if  $|\omega| \geq \frac{2}{\lambda(L) \log q_0}$  then

$$|\omega \log q| \geq \frac{1}{4\lambda(L)} > \frac{1}{4e^{\frac{2y^{1-E}}{1-E}}} > \frac{1}{e^{\frac{\eta y}{(4+\Delta)\log y}}}.$$

Therefore, for  $\frac{2}{\lambda(L) \log q_0} \leq |\omega| \leq e^{\frac{\eta y}{(4+5\Delta)\log y}}$  and  $\alpha \geq \frac{2\pi}{\lambda(L)}$ , at most half of the primes  $q \in \mathbf{Q}_n$  are such that

$$|\omega \log q - k\alpha| \leq \frac{1}{e^{\frac{\eta y}{(4+\Delta)\log y}}}$$



for any integer  $k$ . Let  $\Phi = (AN\phi(L)\log q_0)^2$ . Then

$$\log \Phi = 2(\log A + \log N + \log \phi(L) + \log \log q_0) < 4\left(\frac{\eta y}{\log y} + y + \log \log Y\right) < 8y.$$

Hence

$$8\sqrt{\frac{\log \Phi}{N}} < \frac{32\sqrt{y}}{e^{\frac{\eta y}{4\log y}}} < \frac{1}{e^{\frac{\eta y}{(4+\Delta)\log y}}}$$

and

$$4A\log \Phi < 32Ay < e^{\frac{\eta y}{(4+5\Delta)\log y}}.$$

Therefore, for all  $\omega$  with  $\frac{2}{\lambda(L)\log q_0} \leq |\omega| \leq 4A\log \Phi$  and  $\alpha \geq \frac{2\pi}{\lambda(L)}$ , at most half of the primes  $q \in \mathbf{Q}_n$  are such that

$$|\omega \log q - k\alpha| < 8\sqrt{\frac{\log \Phi}{N}}$$

for some integer  $k$ . □

Let  $\Pi$  be one of the products guaranteed by the theorem. Each of its prime factors is congruent to 1 mod  $k_0$ . Hence  $\Pi \equiv 1 \pmod{k_0}$ . By construction, it is also true that  $\Pi \equiv 1 \pmod{L}$ . Since  $(k_0, L) = 1$ , these can be combined:  $\Pi \equiv 1 \pmod{k_0 L}$ . Each prime factor of  $\Pi$  looks like  $dk_0 + 1$  for some  $d \mid L$ . It follows that  $p - 1 \mid \Pi - 1$  for every  $p \mid \Pi$ , which means that  $\Pi$  is a Carmichael number. Also,

$$\left| \log \Pi - \log z - \frac{1}{2A} \right| = \left| \log \Pi - \frac{\log Q_n}{2} - B \right| \leq \frac{1}{2A}.$$

Therefore, there exist at least

$$2^{2N_- - O(\sqrt{N_+} + \log \Phi)} > e^{\frac{1}{3}e^{\frac{\eta y}{2\log y}}}$$

Carmichael numbers with a logarithm in the interval  $(\log z, \log z + \frac{1}{A})$ . We now want to remove all mention of  $y$  in the quantities in the preceding statement and write everything in terms of  $z$ . We have that

$$z \leq \left\lceil e^{y^{2+2\Delta}} \right\rceil^{\frac{11}{12}e^{\frac{\eta y}{2\log y}}} \leq e^{e^{\left(2-\frac{\delta}{2}\right)\log y}}$$

so recalling that  $\Delta = \frac{\delta}{24}$ ,

$$\frac{A}{2} = \frac{e^{\frac{\eta y}{(4+6\Delta)\log y}}}{2} > \frac{1}{2}(\log z)^{\frac{2-\frac{\delta}{2}}{4+\frac{\delta}{4}}} > (\log z)^{\frac{1}{2+\delta}}$$

since our assumption that  $\delta$  is small means that

$$(2 - \frac{\delta}{2})(2 + \delta) = 4 + \delta - \frac{\delta^2}{2} > 4 + \frac{\delta}{4}.$$

Therefore

$$e^{\log z + \frac{1}{A}} < z \left(1 + \frac{2}{A}\right) < z + \frac{z}{(\log z)^{\frac{1}{2+\delta}}}.$$

Also,  $\log z < y^{2+2\Delta} e^{\frac{\eta y}{2 \log y}}$  and  $\log \log z > y^{1-\Delta}$  so

$$\frac{\log z}{(\log \log z)^{2+\delta}} < \frac{y^{2+2\Delta} e^{\frac{\eta y}{2 \log y}}}{y^{(1-\Delta)(2+\delta)}} = e^{\frac{\eta y}{2 \log y}} y^{4\Delta + \Delta\delta - \delta} = e^{\frac{\eta y}{2 \log y}} y^{\frac{\delta^2}{24} - \frac{5\delta}{6}} < \frac{1}{3} e^{\frac{\eta y}{2 \log y}},$$

which means that  $e^{\frac{1}{3} e^{\frac{\eta y}{2 \log y}}} > e^{\frac{\log z}{(\log \log z)^{2+\delta}}}$ . Therefore, there are at least  $e^{\frac{\log z}{(\log \log z)^{2+\delta}}}$  Carmichael numbers in the interval  $(z, z + \frac{z}{(\log z)^{\frac{1}{2+\delta}}})$  for all  $z$  between  $Z_-(y)$  and  $Z_+(y)$ . Note that

$$Z_-(y+1) = \left\lceil e^{(y+1)^{2+2\Delta}} \right\rceil^{\frac{1}{3} e^{\frac{\eta(y+1)}{2 \log(y+1)}}} < e^{\frac{101}{100} \frac{\epsilon}{3} y^{2+2\Delta} e^{\frac{\eta y}{2 \log y}}} < e^{\frac{11}{12} y^{2+2\Delta} e^{\frac{\eta y}{2 \log y}}} \leq Z_+(y).$$

As a result, the intervals  $[Z_-(y), Z_+(y)]$  and  $[Z_-(y+1), Z_+(y+1)]$  overlap. Hence, sliding  $y$  to infinity by increments of one, we conclude that for sufficiently large  $z$ , there are at least  $e^{\frac{\log z}{(\log \log z)^{2+\delta}}}$  Carmichael numbers in the interval  $(z, z + \frac{z}{(\log z)^{\frac{1}{2+\delta}}})$ . As a final remark, we note that Theorem 2 is effective and therefore so is our main result.

## 5 Acknowledgements

The author would like to acknowledge a helpful letter from James Maynard and a number of useful conversations with Michael Larsen. The author would also like to thank Andrew Granville, Simon Höra, and the anonymous referee for their many helpful comments and suggestions.

## References

- [1] Alford, W. R., Granville, A., and Pomerance, C. “There are infinitely many Carmichael numbers” *Ann. of Math.* 140 (1994): 703–722.

- [2] Chernick, J. “On Fermat’s simple theorem” *Bull. Amer. Math. Soc.* 45, no. 4 (1939): 269–274.
- [3] Erdős, P. “On the normal number of prime factors of  $p-1$  and some other related problems concerning Euler’s  $\phi$ -function” *Q. J. Math.* 6 (1935): 205–213.
- [4] Harman, G. “Watt’s mean value theorem and Carmichael numbers” *Int. J. Number Theory* 4, no. 2 (2008): 241–248.
- [5] Korselt, A. “Problème chinois” *L’Intermédiaire des Mathématiciens* 6 (1899): 142–143.
- [6] Lenstra Jr., H. W. and Pomerance, C. “Primality testing with Gaussian periods” *J. Eur. Math. Soc.* 21 (2019): 1229–1269.
- [7] Maynard, J. “Small gaps between primes” *Ann. of Math.* 181 (2015): 383–413.
- [8] Maynard, J. “Dense clusters of primes in subsets” *Compos. Math.* 152 (2016): 1517–1554.
- [9] Šimerka, V. “Zbytky z arithmetické posloupnosti (On the remainders of an arithmetic progression)” *Časopis Pro Pěstování Matematiky a Fysiky* 14, no. 5, (1885): 221–225.
- [10] Wright, T. “Factors of Carmichael numbers and an even weaker  $k$ -tuples conjecture” *Bull. Aust. Math. Soc.* 99 (2019): 376–384.
- [11] Zhang, Y. “Bounded gaps between primes” *Ann. of Math.* 179 (2014): 1121–1174.