# THE DISTRIBUTION OF TOTIENTS

KEVIN FORD

*Dedicated to the memory of Paul Erdős*

ABSTRACT. This paper is a comprehensive study of the set of totients, i.e. the set of values taken by Euler's $\phi$-function. The main functions studied are $V(x)$, the number of totients $\leqslant x$, $A(m)$, the number of solutions of $\phi(x) = m$ (the "multiplicity" of $m$), and $V_k(x)$, the number of $m \leqslant x$ with $A(m) = k$. The first of the main results of the paper is a determination of the true order of $V(x)$. It is also shown that for each $k \geqslant 1$, if there is a totient with multiplicity $k$ then $V_k(x) \gg V(x)$. Sierpiński conjectured that every multiplicity $k \geqslant 2$ is possible, and we deduce this from the Prime $k$-tuples Conjecture. An older conjecture of Carmichael states that no totient has multiplicity 1. This remains an open problem, but some progress can be reported. In particular, the results stated above imply that if there is one counterexample, then a positive proportion of all totients are counterexamples. The lower bound for a possible counterexample is extended to $10^{10^{10}}$ and the bound $\liminf_{x\to\infty} V_1(x)/V(x) \leqslant 10^{-5,000,000,000}$ is shown. Determining the order of $V(x)$ and $V_k(x)$ also provides a description of the "normal" multiplicative structure of totients. This takes the form of bounds on the sizes of the prime factors of a pre-image of a typical totient. One corollary is that the normal number of prime factors of a totient $\leqslant x$ is $c \log\log x$, where $c \approx 2.186$. Lastly, similar results are proved for the set of values taken by a general multiplicative arithmetic function, such as the sum of divisors function, whose behavior is similar to that of Euler's function.

## 1. INTRODUCTION

Let $\mathscr{V}$ denote the set of values taken by Euler's $\phi$-function (totients), i.e.

$$\mathscr{V} = \{1, 2, 4, 6, 8, 10, 12, 16, 18, 20, 22, 24, 28, 30, \cdots\}.$$

Let

$$\begin{aligned} \mathscr{V}(x) &= \mathscr{V} \cap [1, x], \\ V(x) &= |\mathscr{V}(x)|, \\ \phi^{-1}(m) &= \{n : \phi(n) = m\}, \\ A(m) &= |\phi^{-1}(m)|, \\ V_k(x) &= |\{m \leqslant x : A(m) = k\}|. \end{aligned} \tag{1.1}$$

We will refer to $A(m)$ as the multiplicity of $m$. This paper is concerned with the following problems.

1. What is the order of $V(x)$?
2. What is the order of $V_k(x)$ when the multiplicity $k$ is possible?
3. What multiplicities are possible?
4. What is the normal multiplicative structure of totients?

1991 *Mathematics Subject Classification.* Primary 11A25, 11N64.

Keywords: Euler's function, totients, Carmichael's Conjecture, Sierpiński's Conjecture.

Much of the early work for this paper was completed while the author was enjoying the hospitality of the Institute for Advanced Study, supported by National Science Foundation grant DMS 9304580.

1. The fact that $\phi(p) = p-1$ for primes $p$ implies $V(x) \gg x/\log x$ by the Prime Number Theorem. Pillai [27] gave the first non-trivial upper bound on $V(x)$, namely

$$V(x) \ll \frac{x}{(\log x)^{(\log 2)/e}}.$$

Using sieve methods, Erdős [8] improved this to

$$V(x) \ll_{\varepsilon} \frac{x}{(\log x)^{1-\varepsilon}}$$

for every $\varepsilon > 0$, and later in [9] showed

$$V(x) \gg \frac{x \log\log x}{\log x}.$$

Erdős and Hall [11, 12] sharpened the bounds further, showing

$$\frac{x}{\log x} \exp\{c_1 (\log_3 x)^2\} \ll V(x) \ll \frac{x}{\log x} \exp\{c_2 (\log_2 x)^{1/2}\}$$

for certain positive constants $c_1$ and $c_2$. Here and throughout this paper $\log_k x$ denotes the $k$th iterate of the logarithm. The upper bound was improved again by Pomerance [28], who showed that for some positive constant $c_3$,

$$V(x) \ll \frac{x}{\log x} \exp\{c_3 (\log_3 x)^2\}.$$

The gap between $c_1$ and $c_3$ was removed by Maier and Pomerance [25], who showed that

$$V(x) = \frac{x}{\log x} \exp\{(C + o(1))(\log_3 x)^2\}, \tag{1.2}$$

where $C$ is a constant defined as follows. Let

$$F(x) = \sum_{n=1}^{\infty} a_n x^n, \qquad a_n = (n+1)\log(n+1) - n \log n - 1. \tag{1.3}$$

Since $a_n \sim \log n$ and $a_n > 0$, it follows that $F(x)$ is defined and strictly increasing on $[0,1)$, $F(0) = 0$ and $F(x) \to \infty$ as $x \to 1^-$. Thus, there is a unique number $\varrho$ such that

$$F(\varrho) = 1 \qquad (\varrho = 0.542598586098471021959\ldots). \tag{1.4}$$

In addition, $F'(x)$ is strictly increasing, and

$$F'(\varrho) = 5.69775893423019267575\ldots$$

Let

$$C = \frac{1}{2|\log \varrho|} = 0.81781464640083632231\ldots \tag{1.5}$$

and

$$\begin{aligned} D &= 2C(1 + \log F'(\varrho) - \log(2C)) - 3/2 \\ &= 2.17696874355941032173\ldots \end{aligned} \tag{1.6}$$

Our main result is a determination of the true order of $V(x)$.

**Theorem 1.** *We have*

$$V(x) = \frac{x}{\log x} \exp\{C(\log_3 x - \log_4 x)^2 + D \log_3 x - (D + 1/2 - 2C)\log_4 x + O(1)\}.$$

2. Erdős [10] showed by sieve methods that if $A(m) = k$, then for most primes $p$, $A(m(p-1)) = k$. If the multiplicity $k$ is possible, it follows immediately that $V_k(x) \gg x/\log x$. Applying the machinery used to prove Theorem 1, we can show that for each $k$, either $V_k(x) = 0$ for all $x$, or $V_k(x) \gg V(x)$.

**Theorem 2.** *If there is a number $d$ with $A(d) = k$, then*

$$V_k(x) \gg_\varepsilon d^{-1-\varepsilon} V(x) \qquad (x \geqslant x_0(k)).$$

In other words, a positive fraction of totients have multiplicity $k$ if the multiplicity $k$ is possible. This suggests that the multiplicity of "most" totients is bounded.

**Theorem 3.** *We have*

$$\frac{|\{m \in \mathscr{V}(x) : A(m) \geqslant N\}|}{V(x)} = \sum_{k \geqslant N} \frac{V_k(x)}{V(x)} \ll N^{-1} \exp\{O(\sqrt{\log N})\}.$$

A simple modification of the proof of Theorems 1 and 2 also gives bounds for totients in short intervals. First, define $\pi(x)$ to be the number of primes $\leqslant x$. A real number $\theta$ is said to be admissible if $\pi(x + x^\theta) - \pi(x) \gg x^\theta / \log x$ with $x$ sufficiently large. The current record is due to Baker, Harman and Pintz [2], who showed that $\theta = 0.525$ is admissible.

**Theorem 4.** *If $\theta$ is admissible, $y \geqslant x^\theta$ and the multiplicity $k$ is possible, then*

$$V_k(x + y) - V_k(x) \asymp \frac{y}{x + y} V(x + y).$$

*Consequently, for every fixed $c > 1$, $V(cx) - V(x) \asymp V(x)$.*

Erdős has asked if $V(cx) \sim cV(x)$ for each fixed $c > 1$, which would follow from an asymptotic formula for $V(x)$. The method of proof of Theorem 1, however, falls short of answering Erdős' question.

It is natural to ask what the maximum totient gaps are, in other words what is the behavior of the function $M(x) = \max_{v_i \leqslant x}(v_i - v_{i-1})$ if $v_1, v_2, \cdots$ denotes the sequence of totients? Can it be shown, for example, that for $x$ sufficiently large, that there is a totient between $x$ and $x + x^{1/2}$?

3. In 1907, Carmichael [4] announced that for every $m$, the equation $\phi(x) = m$ has either no solutions $x$ or at least two solutions. In other words, no totient can have multiplicity 1. His proof of this assertion was flawed, however, and the existence of such numbers remains an open problem. In [5], Carmichael did show that no number $m < 10^{37}$ has multiplicity 1, and conjectured that no such $m$ exists (this is now known as Carmichael's Conjecture). Klee [23] improved the lower bound for a counterexample to $10^{400}$, Masai and Valette [26] improved it to $10^{10,000}$ and recently Schlafly and Wagon [33] showed that a counterexample must exceed $10^{10,000,000}$. An immediate corollary of Theorems 1 and 2 (take $d = 1, k = 2$) is

**Theorem 5.** *We have*

$$\limsup_{x\to\infty} \frac{V_1(x)}{V(x)} < 1.$$

*Furthermore, Carmichael's Conjecture is equivalent to the bound*

$$\liminf_{x\to\infty} \frac{V_1(x)}{V(x)} = 0.$$

Although this is a long way from proving Carmichael's Conjecture, Theorem 5 show that the set of counterexamples cannot be a "thin" subset of $\mathscr{V}$. Either there are no counterexamples or a positive fraction of totients are counterexamples.

The basis for the computations of lower bounds for a possible counterexample is a lemma of Carmichael, generalized by Klee, which allows one to show that if $A(m) = 1$ then $x$ must be divisible by the squares of many primes. Using the method outlined in [33] and modern computer hardware, we push the lower bound for a counterexample to Carmichael's Conjecture to the aesthetically pleasing bound $10^{10^{10}}$.

**Theorem 6.** *If $A(m) = 1$, then $m$ exceeds $10^{10^{10}}$.*

As a corollary, a variation of an argument of Pomerance [29] gives the following.

**Theorem 7.** *We have*

$$\liminf_{x\to\infty} \frac{V_1(x)}{V(x)} \leqslant 10^{-5,000,000,000}.$$

The proof of these two theorems motivates another classification of totients. Let $V(x;k)$ be the number of totients up to $x$, all of whose pre-images are divisible by $k$. A trivial corollary to the proof of Theorem 2 is

**Theorem 8.** *If $d$ is a totient, all of whose pre-images are divisible by $k$, then*

$$V(x;k) \gg_\varepsilon d^{-1-\varepsilon} V(x).$$

*Thus, for each $k$, either $V(x;k)=0$ for all $x$ or $V(x;k) \gg_k V(x)$.*

In the 1950's, Sierpiński conjectured that all multiplicities $k \geqslant 2$ are possible (see [30] and [10]), and in 1961, Schinzel [31] deduced this conjecture from his well-known Hypothesis H. Schinzel's Hypothesis H [32], a generalization of Dickson's Prime $k$-tuples Conjecture [7], states that any set of polynomials $F_1(n), \dots, F_k(n)$, subject to certain restrictions, are simultaneously prime for infinitely many $n$. Using a much simpler, iterative argument, we show that Sierpiński's Conjecture follows from the Prime $k$-tuples Conjecture.

**Theorem 9.** *The Prime $k$-tuples Conjecture implies that for each $k \geqslant 2$, there is a number $d$ with $A(d)=k$.*

While certainly true, a proof of Hypothesis H remains elusive even in the simple case where $k=2$ and $F_1, F_2$ are linear polynomials (generalized twin primes). However, combining the iterative process used to prove Theorem 9 with the theory of "almost primes", it is possible to prove Sierpińsk's Conjecture unconditionally. Details will appear in a forthcoming paper [15].

4. Establishing Theorems 1 and 2 requires a determination of what a "normal" totient looks like. This will initially take the form of a series of linear inequalities in the prime factors of a pre-image of a totient. An analysis of these inequalities reveals the normal sizes of the prime factors of a pre-image of a typical totient. To state our results, we first define

$$L_0 = L_0(x) = [2C(\log_3 x - \log_4 x)]. \tag{1.7}$$

In a simplified form, we show that for all but $o(V(x))$ totients $m \leqslant x$, every pre-image $n$ satisfies

$$\log_2 q_i(n) \sim \varrho^i (1 - i/L_0) \log_2 x \qquad (0 \leqslant i \leqslant L_0), \tag{1.8}$$

where $q_i(n)$ denotes the $(i+1)$st largest prime factor of $n$. We also expect each prime to be "normal", that is, $\omega(q_i(n)-1) \approx \log_2 q_i(n)$, where $\omega(m)$ is the number of distinct prime factors of $m$. This suggests that for a typical totient $v \leqslant x$,

$$\omega(v) \approx (1+\varrho+\varrho^2+\cdots)\log_2 x = \frac{\log_2 x}{1-\varrho}.$$

The same also should hold with $\omega(v)$ replaced by $\Omega(v)$, the number of prime factors of $v$ counted with multiplicity.

**Theorem 10.** *Suppose $\varepsilon = \varepsilon(x)$ satisfies $0 \leqslant \varepsilon \leqslant 1/3$. Then*

$$\#\left\{ m \in \mathscr{V}(x) : \left| \frac{\Omega(m)}{\log_2 x} - \frac{1}{1-\varrho} \right| \geqslant \varepsilon \right\} \ll \frac{V(x)}{(\log_2 x)^\varepsilon}.$$

*Consequently, if $g(x) \to \infty$ arbitrarily slowly, then almost all totients $m \leqslant x$ satisfy*

$$\left| \frac{\Omega(m)}{\log_2 x} - \frac{1}{1-\varrho} \right| \leqslant \frac{g(x)}{\log_3 x}.$$

*Moreover, the theorem holds with $\Omega(m)$ replaced by $\omega(m)$.*

**Corollary 11.** *If either $g(m)=\omega(m)$ or $g(m)=\Omega(m)$, then*

$$\sum_{m\in\mathscr{V}(x)} g(m)=\frac{V(x)\log_2 x}{1-\varrho}\left(1+O\left(\frac{1}{\log_3 x}\right)\right).$$

By contrast, Erdős and Pomerance [13] showed that the average of $\Omega(\phi(n))$, where the average is taken over all $n\leqslant x$, is $\frac{1}{2}(\log_2 x)^2+O((\log_2 x)^{3/2})$.

We next analyze the distribution of individual prime factors of a preimage of a typical totient.

**Theorem 12.** *Suppose $x$ is large, and $L_0=L_0(x)$. For $i\leqslant L_0-5$, define $\beta_i=\varrho^i(1-i/L_0)$ and let $V(x;\mathscr{C})$ denote the number of totients $m\leqslant x$ with a pre-image $n$ satisfying condition $\mathscr{C}$. If $\varepsilon\leqslant\frac{i}{5L_0}$, then*

$$V\left(x;\left|\frac{q_i(n)}{\beta_i\log_2 x}-1\right|\geqslant\varepsilon\right)\ll V(x)$$
$$\times\exp\left\{-\frac{L_0(L_0-i)}{2i}\varepsilon^2\left(1+O\left(\frac{\varepsilon L_0}{i}+\frac{1}{\varepsilon}\sqrt{\frac{i}{L_0(L_0-i)}}\log(L_0-i)\right)\right)\right\}.$$

*If $\frac{i}{5L_0}\leqslant\varepsilon\leqslant\frac{1}{2}$ and $i\leqslant L_0/2$ then*

$$V\left(x;\left|\frac{q_i(n)}{\beta_i\log_2 x}-1\right|\geqslant\varepsilon\right)\ll V(x)\exp\left\{-\frac{\varepsilon L_0}{4}+O\big(L_0\varepsilon^2+\sqrt{L_0\varepsilon}\log(L_0\varepsilon)\big)\right\}.$$

There are many ways of using Theorem 12 to construct a result concerning the simultaneous approximation of many of the prime factors of a normal totient. We prove a result, where the goal is to obtain near best approximations of the maximum number of prime factors.

**Theorem 13.** *Suppose $g(x)$ is an increasing function of $x$ satisfying $2\leqslant g(x)=o(\log_3 x)$. For a given $x$, set $L_0=L_0(x)$ and $\beta_i=\varrho^i(1-i/L_0)$ for $0\leqslant i\leqslant L_0$. Then the number of totients $m\leqslant x$ with a pre-image $n$ not satisfying both*

$$\left|\frac{\log_2 q_i(n)}{\beta_i\log_2 x}-1\right|\leqslant\sqrt{\frac{i}{L_0(L_0-i)}}\log(L_0-i)\log g(x)\qquad(1\leqslant i\leqslant L_0-g(x))\tag{1.9}$$

*and $L_0-g(x)\leqslant\omega(n)\leqslant\Omega(n)\leqslant L_0+g(x)$ is*

$$\ll V(x)e^{-\frac{1}{4}\log^2 g(x)}.$$

In essence, Theorem 13 says that the set of $n\leqslant x$ having about $L_0(x)$ prime factors distributed according to (1.9) generates almost all totients. It also says that for most totients, all of its pre-images are "virtually" square-free. The function $g(x)$ need not tend to infinity. Notice that the intervals in (1.9) are not only disjoint, but the gaps between them are rather large. In particular, this "discreteness phenomenon" means that for most totients $m\leqslant x$, no pre-image $n$ has any prime factors $p$ in the intervals

$$0.999\geqslant\frac{\log_2 p}{\log_2 x}\geqslant 0.543,\quad 0.542\geqslant\frac{\log_2 p}{\log_2 x}\geqslant 0.295,\ \text{etc.}$$

This should be compared to the distribution of the prime factors of a normal integer $n\leqslant x$ (e.g. Theorem 12 of [19]).

5. As the details of the proofs of these results are extremely complex and require very delicate estimating, we summarize the central ideas here. First, for most integers $m$, the prime divisors of $m$ are "nicely distributed", meaning the number of prime factors of $m$ lying between $a$ and $b$ is about $\log_2 b-\log_2 a$. This is a more precise version of the classical result of Hardy and Ramanujan [21] that most numbers $m$ have about $\log_2 m$ prime factors. Take an integer $n$ with prime factorization $p_0p_1\cdots$, where for simplicity we assume $n$ is square-free, and $p_0>p_1>\cdots$. By sieve methods it can be shown that for most primes $p$, the prime divisors of $p-1$ have the same "nice" distribution. If $p_0,p_1,\ldots$ are such "normal" primes, it follows that

$\phi(n) = (p_0 - 1)(p_1 - 1)\cdots$ has about $\log_2 n - \log_2 p_1$ prime factors in $[p_1, n]$, about $2(\log_2 p_1 - \log_2 p_2)$ prime factors in $[p_2, p_1]$, and in general, $\phi(n)$ will have $k(\log_2 p_{k-1} - \log_2 p_k)$ prime factors in $[p_k, p_{k-1}]$. That is, $n$ has $k$ times as many prime factors in the interval $[p_k, p_{k-1}]$ as does a "normal" integer of its size. If $n$ has many "large" prime divisors, then the prime factors of $m = \phi(n)$ will be much denser than normal, and the number, $N_1$, of such integers $m$ will be "small". On the other hand, the number, $N_2$ of integers $n$ with relatively few "large" prime factors is also "small". Our objective then is to precisely define these concepts of "large" and "small" so as to minimize $N_1 + N_2$.

The argument in [25] is based on the heuristic that a normal totient is generated from a number $n$ satisfying

$$\log_2 q_i(n) \approx \varrho^i \log_2 x \tag{1.10}$$

for each $i$ (compare with (1.8)). As an alternative to this heuristic, assuming all prime factors of a pre-image $n$ of a totient are normal leads to consideration of a series of inequalities among the prime factors of $n$. We show that such $n$ generate "most" totients. By mapping the $L$ largest prime factors of $n$ (excluding the largest) to a point in $\mathbb{R}^L$, the problem of determining the number of such $n$ up to $x$ reduces to the problem of finding the volume of a certain region of $\mathbb{R}^L$, which we call the fundamental simplex. Our result is roughly

$$V(x) \approx \frac{x}{\log x} \max_L T_L (\log_2 x)^L,$$

where $T_L$ denotes the volume of the simplex. It turns out that the maximum occurs at $L = L_0(x) + O(1)$. Careful analysis of these inequalities reveals that "most" of the integers $n$ for which they are satisfied satisfy (1.8). Thus, the heuristic (1.10) gives numbers $n$ for which the smaller prime factors are too large. The crucial observation that the $L$th largest prime factor ($L = L_0 - 1$) satisfies $\log_2 p_L \approx \frac{1}{L}\varrho^L \log_2 x$ is a key to determining the true order of $V(x)$.

In section 2 we define "normal" primes and show that most primes are "normal". The set of linear inequalities used in the aforementioned heuristic are defined and analyzed in section 3. The principal result is a determination of the volume of the simplex defined by the inequalities, which requires excursions into linear algebra and complex analysis. Section 4 is devoted to proving the upper bound for $V(x)$, and in section 5, the lower bound for $V_k(x)$ is deduced. Together these bounds establish Theorems 1 and 2, as well as Theorems 4, 5 and 8 as corollaries. The distribution of the prime factors of a pre-image of a typical totient are detailed in section 6, culminating in the proof of Theorems 12, 13, 10 and Corollary 11.

In section 7, we summarize the computations giving Theorem 6 and present very elementary proofs of Theorems 7 and 9. Section 8 is concerned with the behavior of the ratios $V_k(x)/V(x)$. We prove Theorem 3 and discuss some consequences. We are unable to determine the behavior for any specific $k$ (other than what Theorem 2 gives), but can say a few things about the behavior as $k \to \infty$. Lastly, section 9 outlines an extension of all of these results to more general multiplicative arithmetic functions such as $\sigma(n)$, the sum of divisors function. Specifically, we prove

**Theorem 14.** *Suppose $f : \mathbb{N} \to \mathbb{N}$ is a multiplicative function satisfying*

$$\{f(p) - p : p \textit{ prime}\} \textit{ is a finite set not containing 0,} \tag{1.11}$$

$$\sum_{\substack{h \geqslant 16 \\ h \textit{ square-full}}} \frac{\varepsilon(h)}{f(h)} \ll 1, \qquad \varepsilon(h) = \exp\{\log_2 h (\log_3 h)^{20}\}. \tag{1.12}$$

*Then the analogs of Theorems 1–4, 8, 12–11 and 15 hold with $f(n)$ replacing $\phi(n)$, with the exception of the dependence on* d *in Theorems 2 and 8, which may be different.*

Theorem 15 depends on the definition of the fundamental simplex, and is not stated until section 6.

## 2. Preliminary lemmata

Let $P^+(n)$ denote the largest prime factor of $n$ and let $\Omega(n, U, T)$ denote the total number of prime factors $p$ of $n$ such that $U \leqslant p \leqslant T$, counted according to multiplicity. Constants implied by the Landau $O-$ and

Vinogradov $\ll$ – and $\gg$ – symbols are absolute unless otherwise specified, and $c_1, c_2, \dots$ will denote absolute constants, not depending on any parameter. Symbols in boldface type indicate vector quantities.

A small set of additional symbols will have constant meaning throughout this paper. These include the constants $\mathscr{V}$, $\varrho$, $C$, $D$, $a_i$, defined respectively in (1.1), (1.4), (1.5), (1.6), and (1.3), as well as the constants $\mathscr{S}_L$, $T_L$, $g_i$ and $g_i^*$, defined in section 3. Also included are the following functions: the functions defined in (1.1), $L_0(x)$ (1.7), $F(x)$ (1.3); the functions $Q(\alpha)$ and $W(x)$ defined respectively in Lemma 2.1 and (2.4) below; and $\mathscr{S}_L(\boldsymbol{\xi})$, $T_L(\boldsymbol{\xi})$, $\mathscr{R}_L(\boldsymbol{\xi}; x)$, $R_L(\boldsymbol{\xi}; x)$ and $x_i(n; x)$ defined in section 3. Other variables are considered "local" and may change meaning from section to section, or from lemma to lemma.

A crucial tool in the proofs of Theorems 1 and 2 is a more precise version of the result from [25] that for most primes $p$, the larger prime factors of $p-1$ are nicely distributed (see Lemma 2.6 below). We begin with three basic lemmas.

**Lemma 2.1.** *If $z > 0$ and $0 < \alpha < 1 < \beta$ then*

$$\sum_{k \leqslant \alpha z} \frac{z^k}{k!} < e^{(1-Q(\alpha))z}, \qquad \sum_{k \geqslant \beta z} \frac{z^k}{k!} < e^{(1-Q(\beta))z},$$

*where $Q(\lambda) = \int_1^\lambda \log t \, dt = \lambda \log(\lambda) - \lambda + 1$.*

*Proof.* We have

$$\sum_{k \leqslant \alpha z} \frac{z^k}{k!} = \sum_{k \leqslant \alpha z} \frac{(\alpha z)^k}{k!} \left(\frac{1}{\alpha}\right)^k \leqslant \left(\frac{1}{\alpha}\right)^{\alpha z} \sum_{k \leqslant \alpha z} \frac{(\alpha z)^k}{k!} < \left(\frac{e}{\alpha}\right)^{\alpha z} = e^{(1-Q(\alpha))z}.$$

The second inequality follows in the same way. □

**Lemma 2.2.** *The number of integers $n \leqslant x$ for which $\Omega(n) \geqslant \alpha \log_2 x$ is*

$$\ll_\alpha \begin{cases} x(\log x)^{-Q(\alpha)} & 1 < \alpha < 2 \\ x(\log x)^{1-\alpha \log 2} \log_2 x & \alpha \geqslant 2. \end{cases}$$

*Proof.* This can be deduced from the Theorems in Chapter 0 of [19]. □

**Lemma 2.3.** *The number of $n \leqslant x$ divisible by a number $m \geqslant \exp\{(\log_2 x)^2\}$ with $P^+(m) \leqslant m^{1/\log_2 x}$ is $\ll x/\log^2 x$.*

*Proof.* Let $\Psi(x, y)$ denote the number of integers $\leqslant x$ which have no prime factors $> y$. For $x$ large, standard estimates of $\Psi(x, y)$ ([22], Theorem 1.1 and Corollary 2.3) give

$$\Psi(z, z^{1/\log_2 x}) \ll z \exp\{-(\log_2 x \log_3 x)/2\}$$

uniformly for $z \geqslant \exp\{(\log_2 x)^2\}$. The lemma follows by partial summation. □

We also need basic sieve estimates ([18], Theorems 4.1, 4.2).

**Lemma 2.4.** *Uniformly for $1.9 \leqslant y \leqslant z \leqslant x$, we have*

$$|\{n \leqslant x : p | n \implies p \notin (y, z]\}| \ll x \frac{\log y}{\log z}.$$

**Lemma 2.5.** *Suppose $a_1, \dots, a_h$ are positive integers and $b_1, \dots, b_h$ are integers such that*

$$E = \prod_{i=1}^h a_i \prod_{1 \leqslant i < j \leqslant h} (a_i b_j - a_j b_i) \neq 0.$$

*Then*

$$\#\{n \leqslant x : a_i n + b_i \text{ prime } (1 \leqslant i \leqslant h)\} \ll \frac{x}{(\log z)^h} \prod_{p | E} \frac{1 - \frac{\nu(p)}{p}}{(1 - \frac{1}{p})^h} \ll \frac{x (\log_2(E+2))^h}{(\log z)^h},$$

*where $\nu(p)$ is the number of solutions of the congruence $\prod(an_i+b_i) \equiv 0 \pmod p$, and the implied constant may depend on $h$.*

Next, we examine the normal multiplicative structure of shifted primes $p-1$. When $S \geqslant 2$, a prime $p$ is said to be $S$-normal if

$$\Omega(p-1,1,S) \leqslant 2\log_2 S \tag{2.1}$$

and for every pair of numbers $(U,T)$ with $S \leqslant U < T \leqslant p-1$ we have

$$|\Omega(p-1,U,T) - (\log_2 T - \log_2 U)| < \sqrt{\log_2 S \log_2 T}. \tag{2.2}$$

**Lemma 2.6.** *Uniformly in $S \geqslant 2$, $x \geqslant e^e$, the number of primes $p \leqslant x$ which are not $S$-normal is*

$$\ll \frac{x(\log_2 x)^5}{\log x}(\log S)^{-1/6}.$$

*Proof.* Assume $x$ is sufficiently large and $S \geqslant \log^{1000} x$, otherwise the lemma is trivial. By Lemmas 2.2 and 2.3, the number of primes $p \leqslant x$ with either $P^+(p-1) \leqslant x^{1/\log_2 x}$, $\Omega(p-1) \geqslant 10\log_2 x$ or $p-1$ divisible by the square of a prime $q \geqslant S$ is $O(x/\log^2 x)$. Let $p$ be a prime not in these categories, which is also not $S$-normal. Write $p-1 = qb$, where $q = P^+(p-1) > x^{1/\log_2 x}$. By (2.1) and (2.2), either (i) $\Omega(b,1,S) \geqslant 2\log_2 S - 1$ or (ii) for some $S \leqslant U < T \leqslant x$, $|\Omega(b,U,T) - (\log_2 T - \log_2 U)| \geqslant \sqrt{\log_2 S \log_2 T} - 1$. By Lemma 2.5, for each $b$, the number of $q$ is

$$\ll \frac{x}{\phi(b)\log^2(x/b)} \ll \frac{x(\log_2 x)^3}{b}.$$

The sum of $1/b$ over $b$ satisfying (i) is

$$\begin{aligned}&\leqslant \sum_{\substack{P^+(b')\leqslant S\\ \Omega(b')\geqslant 2\log_2 S-1}} \frac{1}{b'} \prod_{S<p\leqslant x}\left(1+\frac{1}{p}\right) \ll \frac{\log x}{\log S}\left(\frac{3}{2}\right)^{1-2\log_2 S} \sum_{P^+(b')\leqslant S} \frac{(3/2)^{\Omega(b')}}{b'}\\ &\ll (\log x)(\log S)^{1/2-2\log(3/2)} \ll (\log x)(\log S)^{-0.3}.\end{aligned}$$

Consider $b$ satisfying (ii). For positive integers $k$, let $t_k = e^{e^k}$. We claim for some integers $j,k$ satisfying $\log_2 S - 1 \leqslant j < k \leqslant \log_2 x + 1$, that

$$|\Omega(b,t_j,t_k) - (k-j+1)| \geqslant \sqrt{(k-1)\log_2 S} - 4. \tag{2.3}$$

To see this, suppose (2.3) fails for all $j,k$, and suppose $S \leqslant U < T \leqslant x$. Define $j,k$ by $t_j \leqslant U \leqslant t_{j+1}$ and $t_k \leqslant S < t_k$. Then

$$\Omega(b,t_{j+1},t_k) \leqslant \Omega(b,U,T) \leqslant \Omega(b,t_j,t_{k+1}),$$

from which (2.2) follows. Now fix $j,k$ and let $h = \sqrt{(k-1)\log_2 S} - 4$. For any integer $l$,

$$\begin{aligned}\sum_{\Omega(b,t_j,t_k)=l} \frac{1}{b} &\leqslant \prod_{p\leqslant t_j}\left(1+\frac{1}{p}\right) \prod_{t_k<p\leqslant x}\left(1+\frac{1}{p}\right) \frac{1}{l!}\Bigg(\sum_{t_j<p\leqslant t_k} \frac{1}{p}\Bigg)^l\\ &\ll e^{j-k}\log x\,\frac{(k-j+1)^l}{l!}.\end{aligned}$$

The sum of the left side, over $|l-(k-j+1)| \geqslant h$, is then $\ll (\log x)\exp\{(-(k-j)Q(\beta)\}$, where $\beta = 1 + \frac{h}{k-j+1}$. Here we used the fact that $Q(1-\lambda) > Q(1+\lambda)$ for $0 < \lambda \leqslant 1$. By the integral representation of $Q(x)$, we have $Q(1+\lambda) \geqslant \frac{\lambda}{2}\log(1+\lambda)$. Also,

$$h \geqslant 0.99\sqrt{k\log_2 S} \geqslant 0.99\log_2 S \geqslant 990.$$

If $h \geqslant k-j+1$, then

$$(k-j)Q(\beta) \geqslant \frac{h(k-j)\log 2}{2(k-j+1)} \geqslant \frac{h\log 2}{4} \geqslant \frac{\log_2 S}{6},$$

and if $h < k-j+1$, then

$$(k-j)Q(\beta) \geqslant \frac{(k-j)\log 2}{2}\left(\frac{h}{k-j+1}\right)^2 \geqslant \frac{h^2}{3(k-j+1)} \geqslant \frac{\log_2 S}{4}.$$

Thus, the sum of reciprocals of $b$ failing (2.3) is $\ll (\log x)(\log S)^{-1/6}$, uniformly in $j, k$. As there are $\leqslant (\log_2 x)^2$ choices for $j, k$, the proof is complete. □

**Lemma 2.7.** *The number of $m \in \mathscr{V}(x)$ for which either $d^2|m$ or $d^2|n$ for some $n \in \phi^{-1}(m)$ and $d > Y$ is $O(x\log_2 x/Y)$.*

*Proof.* If $\phi(n) = m \leqslant x$, then from the standard result $n/\phi(n) \ll \log_2 n$, we have $n \ll x\log_2 x$. Summing over all possible $d$ gives the result. □

Now define

$$W(x) = \max_{2\leqslant y\leqslant x} \frac{V(y)\log y}{y}. \tag{2.4}$$

**Lemma 2.8.** *The number of $m \in \mathscr{V}(x)$ for which some $n \in \phi^{-1}(m)$ is divisible by a prime which is not $S$-normal is*

$$O\left(\frac{xW(x)(\log_2 x)^6}{\log x}(\log S)^{-1/6}\right).$$

*Proof.* Since $W(x) \gg 1$, we may suppose $S \geqslant \log^3 x$, for otherwise the lemma is trivial. Suppose $p$ is a prime divisor of $n$ for some $n \in \phi^{-1}(m)$. If $n = n'p$ then either $\phi(n) = (p-1)\phi(n')$ or $\phi(n) = p\phi(n')$, so in either case $\phi(n') \leqslant x/(p-1)$. Let $G(t)$ denote the number of primes $p \leqslant t$ which are not $S$-normal. By Lemma 2.6, the number of $m$ in question is at most

$$\begin{aligned}\sum_p V\left(\frac{x}{p-1}\right) &\ll \sum_p \frac{xW(x/(p-1))}{(p-1)\log(x/p)} \\ &\ll xW(x)\int_2^{x/2} \frac{G(t)dt}{t^2\log(x/t)} \ll \frac{xW(x)(\log_2 x)^6}{\log x}(\log S)^{-1/6}.\end{aligned}$$

□

## 3. The fundamental simplex

For a natural number $n$, let $q_0(n) \geqslant q_1(n) \geqslant \cdots$ denote the prime factors of $n$. When $i \geqslant \Omega(n)$, set $q_i(n) = 1$. For a fixed $y$, let

$$x_i(n;y) = \frac{\max(2, \log_2 q_i(n))}{\log_2 y}. \tag{3.1}$$

For $\mathscr{S} \subseteq [0,1]^L$, let $\mathscr{R}_L(\mathscr{S}; y)$ denote the set of integers $n$ with $\Omega(n) \leqslant L$ and

$$(x_0(n;y), x_1(n;y), \ldots, x_{L-1}(n;y)) \in \mathscr{S},$$

and set

$$R_L(\mathscr{S}; y) = \sum_{n\in\mathscr{R}_L(\mathscr{S};y)} \frac{1}{\phi(n)}. \tag{3.2}$$

Heuristically, $R_L(\mathscr{S};x) \approx (\log_2 y)^L \operatorname{Vol}(\mathscr{S})$. Our result in this direction relates $R_L(\mathscr{S};y)$ to the volume of perturbations of $\mathscr{S}$. Specifically, letting $\mathscr{S}+\mathbf{v}$ denote the translation of $\mathscr{S}$ by the vector $\mathbf{v}$, for $\varepsilon > 0$ let

$$\mathscr{S}^{+\varepsilon} = \bigcup_{\mathbf{v}\in[-\varepsilon,\varepsilon]^L} (\mathscr{S}+\mathbf{v}), \qquad \mathscr{S}^{-\varepsilon} = \bigcap_{\mathbf{v}\in[-\varepsilon,\varepsilon]^L} (\mathscr{S}+\mathbf{v}).$$

**Lemma 3.1.** *Let $y \geqslant 2000$, $\varepsilon = 1/\log_2 y$ and suppose $\mathscr{S} \subseteq \{\mathbf{x} \in \mathbb{R}^L : 0 \leqslant x_L \leqslant \cdots \leqslant x_1 \leqslant 1\}$. Then*

$$R_L(\mathscr{S};y) \gg (\log_2 y)^L \operatorname{Vol}\left(\mathscr{S}^{-\varepsilon}\right).$$

*If, for some $\lambda > 0$, $\mathbf{x} \in \mathscr{S}$ implies $x_{L-j} \geqslant \frac{j-\lambda}{\log_2 y}$ for all $j$, then*

$$R_L(\mathscr{S};y) \ll_\lambda (\log_2 y)^L \operatorname{Vol}\left(\mathscr{S}^{+\varepsilon}\right).$$

*Proof.* For positive integers $m_1,\ldots,m_L$, let $B(\mathbf{m}) = \prod_{i=1}^L [(m_i-1)\varepsilon, m_i\varepsilon)$. If $\mathscr{B}$ is the set of boxes $B(\mathbf{m})$ entirely contained in $\mathscr{S}$, then the union of these boxes contains $\mathscr{S}^{-\varepsilon}$. Moreover, for each box, $m_1 > m_2 > \ldots > m_L \geqslant 2$. For $m \geqslant 1$, there is at least one prime in $[\exp(e^{m-1}), \exp(e^m))$, thus

$$\begin{aligned} R_L(\mathscr{S};y) &\geqslant \sum_{B(\mathbf{m})\in\mathscr{B}} \prod_{i=1}^L \sum_{m_i-1\leqslant \log_2 p \leqslant m_i} \frac{1}{p-1} \\ &= \sum_{B(\mathbf{m})\in\mathscr{B}} \prod_{i=1}^L \left(1+O(e^{-m_i})\right) \gg |\mathscr{B}| \geqslant (\log_2 y) \operatorname{Vol}(\mathscr{S}^{-\varepsilon}). \end{aligned}$$

For the second part, let $\mathscr{B}$ be the set of boxes $B(\mathbf{m})$ which intersect $\mathscr{S}$, so that their union is contained in $\mathscr{S}^{+\varepsilon}$. By assumption, $m_{L-j} \geqslant 2(j-\lambda)-1$ for each $\mathbf{m}$. Also, $\phi(p^b) \geqslant (p-1)^b$ for each prime. Therefore,

$$\begin{aligned} R_L(\mathscr{S};y) &\leqslant \sum_{B(\mathbf{m})\in\mathscr{B}} \prod_{i=1}^L \sum_{m_i-1\leqslant \max(\log_2 p, 2) \leqslant m_i} \frac{1}{p-1} \\ &= \sum_{B(\mathbf{m})\in\mathscr{B}} \prod_{i=1}^L \left(1+O(e^{-m_i})\right) \ll |\mathscr{B}| \leqslant (\log_2 y) \operatorname{Vol}(\mathscr{S}^{+\varepsilon}). \end{aligned}$$

Here, if $m_i = 3$, the sum on $p$ includes an additional summand 1 to account for $n$ with $\Omega(n) < i$. □

Suppose $\xi_i \geqslant 0$ for $0 \leqslant i \leqslant L-1$. Recall (1.3) and let $\mathscr{S}_L^*(\boldsymbol{\xi})$ be the set of $(x_1,\ldots,x_L) \in \mathbb{R}^L$ satisfying

$$\begin{array}{ll} (I_0) & a_1x_1 + a_2x_2 + \cdots + a_Lx_L \leqslant \xi_0, \\ (I_1) & a_1x_2 + a_2x_3 + \cdots + a_{L-1}x_L \leqslant \xi_1 x_1, \\ \vdots & \vdots \\ (I_{L-2}) & a_1x_{L-1} + a_2x_L \leqslant \xi_{L-2}x_{L-2}, \\ (I_{L-1}) & 0 \leqslant x_L \leqslant \xi_{L-1}x_{L-1}. \end{array}$$

and let $\mathscr{S}_L(\boldsymbol{\xi})$ be the subset of $\mathscr{S}_L^*(\boldsymbol{\xi})$ satisfying $0 \leqslant x_L \leqslant \cdots \leqslant x_1 \leqslant 1$. Define

$$T_L^*(\boldsymbol{\xi}) = \operatorname{Vol}(\mathscr{S}_L^*(\boldsymbol{\xi})), \qquad T_L(\boldsymbol{\xi}) = \operatorname{Vol}(\mathscr{S}_L(\boldsymbol{\xi})).$$

For convenience, let $\mathbf{1} = (1,1,\ldots,1)$, $\mathscr{S}_L = \mathscr{S}_L(\mathbf{1})$ (the "fundamental simplex"), $T_L = \operatorname{Vol}(\mathscr{S}_L)$, $\mathscr{S}_L^* = \mathscr{S}_L^*(\mathbf{1})$, and $T_L^* = \operatorname{Vol}(\mathscr{S}_L^*)$.

To simplify our later work, we first relate $\mathscr{S}_L(\boldsymbol{\xi})$ to $\mathscr{S}_L$. The next lemma is trivial.

**Lemma 3.2.** *Let $\xi_{L-1} = 1$. If $\xi_i \geqslant 1$ for all $i$, and $\mathbf{x} \in \mathscr{S}_L(\boldsymbol{\xi})$, then $\mathbf{y} \in \mathscr{S}_L$, where $y_i = (\xi_0 \cdots \xi_{i-1})^{-1}x_i$. If $0 < \xi_i \leqslant 1$ for all $i$ and $\mathbf{y} \in \mathscr{S}_L$, then $\mathbf{x} \in \mathscr{S}_L(\boldsymbol{\xi})$, where $x_i = (\xi_0 \cdots \xi_{i-1})y_i$.*

**Corollary 3.3.** *Define* $H(\boldsymbol{\xi}) = \xi_0^L \xi_1^{L-1} \cdots \xi_{L-2}^2$. *We have* $T_L \leqslant T_L(\boldsymbol{\xi}) \leqslant H(\boldsymbol{\xi}) T_L$ *when* $\xi_i \geqslant 1$ *for all* $i$, *and* $H(\boldsymbol{\xi}) T_L \leqslant T_L(\boldsymbol{\xi}) \leqslant T_L$ *when* $0 < \xi_i \leqslant 1$ *for all* $i$.

In applications, $H(\boldsymbol{\xi})$ will be close to 1, so we concentrate on bounding $T_L$.

**Lemma 3.4.** *We have*

$$T_L \asymp \frac{\varrho^{L(L+3)/2}}{L!} (F'(\varrho))^L.$$

Before proceeding to the proof, we note the following immediate corollary.

**Corollary 3.5.** *If* $H(\boldsymbol{\xi}) \asymp 1$, *then*

$$T_L(\boldsymbol{\xi}) \asymp \frac{\varrho^{L(L+3)/2}}{L!} (F'(\varrho))^L.$$

*Furthermore, if* $L = 2C(\log_3 x - \log_4 x) - \Psi$, *where* $0 \leqslant \Psi \ll \sqrt{\log_3 x}$, *then*

$$\begin{aligned}(\log_2 x)^L T_L(\boldsymbol{\xi}) = \exp\{C(\log_3 x - \log_4 x)^2 + D \log_3 x - (D + 1/2 - 2C) \log_4 x \\ - \Psi^2/4C - (D/2C - 1)\Psi + O(1)\}.\end{aligned}$$

*If* $L = [2C(\log_3 x - \log_4 x)] - \Psi > 0$, *then*

$$\begin{aligned}(\log_2 x)^L T_L(\boldsymbol{\xi}) \ll \exp\{C(\log_3 x - \log_4 x)^2 + D \log_3 x - (D + 1/2 - 2C) \log_4 x \\ - \Psi^2/4C - (D/2C - 1)\Psi\}.\end{aligned}$$

*Proof.* The second and third parts follow from (1.5), (1.6) and Stirling's formula. □

Below is the primary tool needed for Lemma 3.4.

**Lemma 3.6.** *Suppose* $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_L$ *are vectors in* $\mathbb{R}^L$, *any* $L$ *of which are linearly independent. Suppose further that*

$$\mathbf{v}_0 + \sum_{i=1}^{L} b_i \mathbf{v}_i = \mathbf{0}, \tag{3.3}$$

*where* $b_i > 0$ *for every* $i$. *Also suppose* $\alpha > 0$. *The volume,* $V$, *of the region defined by*

$$\{\mathbf{v}_i \cdot \mathbf{x} \leqslant 0 \, (1 \leqslant i \leqslant L), \mathbf{v}_0 \cdot \mathbf{x} \leqslant \alpha\}$$

*is*

$$V = \frac{\alpha^L}{L!(b_1 b_2 \cdots b_L) |\det(\mathbf{v}_1, \ldots, \mathbf{v}_L)|}.$$

*Proof.* We may assume without loss of generality that $\alpha = b_1 = b_2 = \cdots = b_L = 1$, for the general case follows by suitably scaling the vectors $\mathbf{v}_i$. Aside from the point $\mathbf{0}$, the other vertices of the simplex are $\mathbf{p}_1, \cdots, \mathbf{p}_L$, where $\mathbf{p}_i$ is the unique vector satisfying

$$\begin{aligned}&\mathbf{p}_i \cdot \mathbf{v}_j = 0 \quad (1 \leqslant j \leqslant L, j \neq i); \\ &\mathbf{p}_i \cdot \mathbf{v}_0 = 1.\end{aligned}$$

Taking the dot product of $\mathbf{p}_i$ with each side of (3.3) yields $\mathbf{v}_i \cdot \mathbf{p}_i = -1$, so $\mathbf{p}_i$ lies in the region $\{\mathbf{v}_i \cdot \mathbf{x} \leqslant 0\}$. The given region is thus an $L$-dimensional "hyper-tetrahedron" with volume $|\det(\mathbf{p}_1, \cdots, \mathbf{p}_L)|/L!$, and

$$(\mathbf{p}_1, \cdots, \mathbf{p}_L)(\mathbf{v}_1, \cdots, \mathbf{v}_L)^T = -I,$$

where $I$ is the identity matrix. Taking determinants gives the lemma. □ □

Having $2L-2$ inequalities defining $\mathscr{S}_L$ creates complications estimating $T_L$, so we devise a scheme where only $L+1$ inequalities are considered at a time, thus allowing the use of Lemma 3.6. The numbers $b_i$ occurring in that lemma will come from the sequence $\{g_i\}$, defined by

$$g_0 = 1, \qquad g_i = \sum_{j=1}^{i} a_j g_{i-j} \quad (i \geqslant 1). \tag{3.4}$$

**Lemma 3.7.** *For every $i \geqslant 1$, $|g_i - \lambda\varrho^{-i}| \leqslant 5$, where $\lambda = \frac{1}{\varrho F'(\varrho)}$.*

*Proof.* Let $G(z) = \sum_{n=0}^{\infty} g_n z^n = (1-F(z))^{-1}$ $(|z| < \varrho)$. Write $1 - F(z) = (1 - z/\varrho)l(z)$ and $l(z) = \sum_{n=0}^{\infty} l_n z^n$, so that $l_n = \sum_{k=1}^{\infty} \varrho^k a_{n+k} > 0$. Next consider $k(z) = (1-z)^2 l(z) = \sum_{n=0}^{\infty} k_n z^n$. We have $k_0 = 1$, $k_1 = h_1 - 2 = \varrho^{-1} - a_1 - 2 < 0$ and for $n \geqslant 2$,

$$k_n = h_n - 2h_{n-1} + h_{n-2} = \sum_{k=1}^{\infty} \varrho^k \left(a_{n+k} - 2a_{n+k-1} + a_{n+k-2}\right) < 0.$$

Also, $k_n = O(1/n^2)$, and $\sum_{n\geqslant 1} k_n = -1$. Thus, $k(z)$ is analytic for $|z| < 1$, continuous on $|z| \leqslant 1$, and nonzero for $|z| \leqslant 1, z \neq 1$. Further,

$$\Re k(z) \geqslant 1 + k_1 \Re z - (1 + k_1) = |k_1|\Re(1-z),$$

so that for $|z| < 1$,

$$\left|\frac{1}{l(z)}\right| \leqslant \frac{|1-z|^2}{|k_1|\Re(1-z)} \leqslant \frac{1}{|k_1|} \max_{|z|=1} \frac{|1-z|^2}{\Re(1-z)} = \frac{2}{|k_1|} < 3.7.$$

Now let

$$e(z) = \sum_{n=0}^{\infty} \left(g_n - \lambda\varrho^{-i}\right) = \frac{1}{1-F(z)} - \frac{\lambda}{1-z/\varrho} = \frac{1/h(z) - 1/h(\varrho)}{1 - z/\varrho}.$$

From the preceding arguments, we see that $e(z)$ is analytic for $|z| < 1$, continuous on $|z| \leqslant 1$, and bounded in absolute value by $(3.7 + \lambda)/|1/\varrho - 1| \leqslant 5$. By Cauchy's integral formula, the Taylor coefficients of $e(z)$ are all bounded by 5 in absolute value. □

*Remark* 1. The above proof is based on [16], and is much simpler than the proof given in the published version of this paper. With more work, one can show that for $i \geqslant 1$, the numbers $g_i - \frac{\varrho^{-i}}{\varrho F'(\varrho)}$ are negative, increasing and have sum $-1 + \lambda/(1-\varrho) = -0.2938\ldots$

Let $\mathbf{e}_1, \cdots, \mathbf{e}_L$ denote the standard basis for $\mathbb{R}^L$, so that $\mathbf{e}_i \cdot \mathbf{x} = x_i$. For $1 \leqslant i \leqslant L-2$, set

$$\mathbf{v}_i = -\mathbf{e}_i + \sum_{j=1}^{L-i} a_j \mathbf{e}_{i+j} \tag{3.5}$$

and also set

$$\mathbf{v}_0 = \sum_{j=1}^{L} a_j \mathbf{e}_j, \qquad \mathbf{v}_{L-1} = -\mathbf{e}_{L-1} + \mathbf{e}_L, \qquad \mathbf{v}_L = -\mathbf{e}_L.$$

For convenience, define

$$g_0^* = 1, \quad g_i^* = g_i + (1 - a_1)g_{i-1}. \tag{3.6}$$

Thus, for $1 \leqslant j \leqslant L-2$, inequality $(I_j)$ may be abbreviated as $\mathbf{v}_j \cdot \mathbf{x} \leqslant 0$. Also, inequality $(I_0)$ is equivalent to $\mathbf{v}_0 \cdot \mathbf{x} \leqslant 1$ and the inequality $x_{L-1} \geqslant x_L \geqslant 0$ is represented by $\mathbf{v}_{L-1} \cdot \mathbf{x} \leqslant 0$ and $\mathbf{v}_L \cdot \mathbf{x} \leqslant 0$. A straightforward calculation using (3.4), (3.5) and (3.6) gives

$$\mathbf{e}_i = -\sum_{j=i}^{L-1} g_{j-i}\mathbf{v}_j - g_{L-i}^* \mathbf{v}_L. \tag{3.7}$$

It follows that

$$\mathbf{v}_0 + \sum_{j=1}^{L-1} g_j \mathbf{v}_j + g_L^* \mathbf{v}_L = \mathbf{0}. \tag{3.8}$$

We now have the ingredients for the proof of Lemma 3.4. The basic idea is that inequalities $(I_0)$–$(I_{L-2})$ by themselves determine a region which is only slightly larger than $\mathscr{S}_L$. In other words, the inequalities $1 \geqslant x_1 \geqslant \cdots \geqslant x_{L-1} \geqslant x_L$ are relatively insignificant. Set

$$\mathscr{U}_0 = \mathscr{S}_L^* \cap \{x_1 \geqslant 1\}, \qquad \mathscr{U}_i = \mathscr{S}_L^* \cap \{x_i \leqslant x_{i+1}\} \quad (1 \leqslant i \leqslant L-2)$$

and $V_i = \mathrm{Vol}(\mathscr{U}_i)$. Evidently

$$T_L^* - \sum_{i=0}^{L-2} V_i \leqslant T_L \leqslant T_L^*. \tag{3.9}$$

Since $|\det(\mathbf{v}_1, \cdots, \mathbf{v}_L)| = 1$, Lemma 3.6 and (3.8) give

$$T_L^* = \frac{1}{L!(g_1 \cdots g_{L-1}) g_L^*}. \tag{3.10}$$

For the remaining argument, assume $L$ is sufficiently large. We shall show that

$$\sum_{i=0}^{L-2} V_i < 0.61 T_L^*, \tag{3.11}$$

which, combined with (3.9), (3.10) and Lemma 3.7, proves Lemma 3.4.

**Lemma 3.8.** *We have*

$$V_0 \ll (7/9)^L T_L^*.$$

*Proof.* The condition $x_1 \geqslant 1$ combined with $\mathbf{v}_0 \cdot \mathbf{x} \leqslant 1$ implies $\mathbf{u} \cdot \mathbf{x} \leqslant 0$, where $\mathbf{u} = \mathbf{v}_0 - \mathbf{e}_1$. By (3.7) and (3.8),

$$\mathbf{u} = \sum_{j=1}^{L-1} (g_{j-1} - g_j) \mathbf{v}_j + (g_{L-1}^* - g_L^*) \mathbf{v}_L.$$

Thus

$$\mathbf{v}_0 + \frac{a_1}{1-a_1} \mathbf{u} + \sum_{j=2}^{L} b_j \mathbf{v}_j = \mathbf{0},$$

where

$$\begin{aligned} b_j &= g_j + \frac{a_1}{1-a_1}(g_j - g_{j-1}) \qquad (2 \leqslant j \leqslant L-1), \\ b_L &= g_L^* + \frac{a_1}{1-a_1}(g_L^* - g_{L-1}^*). \end{aligned}$$

In addition, $|\det(\mathbf{u}, \mathbf{v}_2, \ldots, \mathbf{v}_L)| = (1 - a_1)$. Therefore, by Lemma 3.6,

$$V_0 \ll \frac{1}{L!(b_2 b_3 \cdots b_L)}.$$

Lemma 3.7 implies $b_j > (7/9) g_j$ for large $j$, and the lemma follows from (3.10). □

**Lemma 3.9.** *For $i \geqslant 1$, we have*

$$V_i = \frac{1}{(1-a_1) L! (g_1 \cdots g_{i-1}) A_i B_i} \prod_{j=i+2}^{L-1} \left( \frac{1}{g_j + B_i h_{j-i}} \right) \frac{1}{g_L^* + B_i h_{L-i}^*},$$

*where*

$$A_i = g_i + \frac{g_{i+1}}{1-a_1}, \quad B_i = \frac{g_{i+1}}{1-a_1}, \quad h_l = g_l - g_{l-1}, \quad h_l^* = h_l + (1-a_1)h_{l-1}.$$

*Proof.* In $\mathscr{U}_i$ we have

$$a_1 x_{i+1} + a_2 x_{i+2} + \cdots + a_{L-i} x_L \leqslant x_i \leqslant x_{i+1},$$

which implies

$$\begin{aligned} x_{i+1} &\geqslant \frac{1}{1-a_1}(a_2 x_{i+2} + \cdots + a_{L-i} x_L) \\ &\geqslant x_{i+2} + a_2 x_{i+3} + \cdots + a_{L-i-1} x_L. \end{aligned}$$

The condition $\mathbf{v}_{i+1} \cdot \mathbf{x} \leqslant 0$ is therefore implied by the other inequalities defining $\mathscr{U}_i$, which means

$$V_i = \mathrm{Vol}\{\mathbf{v}_0 \cdot \mathbf{x} \leqslant 1; \mathbf{v}_j \cdot \mathbf{x} \leqslant 0 \; (1 \leqslant j \leqslant L, j \neq i+1); \; (\mathbf{e}_i - \mathbf{e}_{i+1}) \cdot \mathbf{x} \leqslant 0\}.$$

We note $|\det(\mathbf{v}_1, \cdots, \mathbf{v}_i, \mathbf{e}_i - \mathbf{e}_{i+1}, \mathbf{v}_{i+2}, \cdots, \mathbf{v}_L)| = (1-a_1)$. It is also easy to show from (3.8) that

$$\mathbf{0} = \mathbf{v}_0 + \sum_{j=1}^{i-1} g_j \mathbf{v}_j + A_i \mathbf{v}_i + B_i(\mathbf{e}_i - \mathbf{e}_{i+1}) + \sum_{j=i+2}^{L} b_j \mathbf{v}_j,$$

where

$$b_j = g_j + B_i h_{j-i} \quad (i+2 \leqslant j \leqslant L-1), \qquad b_L = g_L^* + B_i h_{L-i}^*.$$

An application of Lemma 3.6 now completes the determination of $V_i$. □

We now deduce numerical estimates for $V_i/T_L^*$. Recall the definitions (3.6). Adopting the notation of Lemma 3.9 and using Lemma 3.7 (plus explicit computation of $g_i$ for small $i$) gives

$$\begin{aligned} A_i &> 4 \qquad (i \geqslant 1), \\ g_j + B_i h_{j-i} &> 1.44 g_j \qquad (i \text{ large, say } i \geqslant L-100), \\ g_j + B_i h_{j-i} &> 1.16 g_j \qquad (i \geqslant 1, j \geqslant i+2), \\ g_L^* + B_i h_{L-i}^* &> 1.44 g_L^* \qquad (i < L-2), \\ g_L^* + B_{L-2} h_2^* &> 1.19 g_L^*. \end{aligned}$$

From these bounds, plus (3.10) and Lemma 3.9, it follows that

$$\begin{aligned} V_{L-2}/T_L^* &< (4 \cdot 1.19)^{-1}, \\ V_i/T_L^* &< (4 \cdot 1.44^{L-i-1})^{-1} \qquad (L-99 \leqslant i \leqslant L-3), \\ V_i/T_L^* &< (4 \cdot 1.44^{99} \cdot 1.16^{L-i-100})^{-1} \qquad (1 \leqslant i \leqslant L-100). \end{aligned}$$

Therefore, by (3.9) and Lemma 3.8,

$$\sum_{i=0}^{L-2} V_i/T_L^* < O((4/5)^L) + \frac{1}{4}\left(\frac{1}{1.19} + \frac{1.44^{-2}}{1-1.44^{-1}} + \frac{1.44^{-99}}{(1-1/1.16)}\right) < 0.61,$$

which implies (3.11). This completes the proof of Lemma 3.4. □

Important in the study of $\mathscr{S}_L$ are both global bounds on the numbers $x_i$ as well as a determination of where "most" of the volume lies.

**Lemma 3.10.** *Let $x_0 = 1$. If $\mathbf{x} \in \mathscr{S}_L$, then $x_i \geqslant g_{j-i}^* x_j$ for $0 \leqslant i \leqslant j \leqslant L$. If $\mathbf{x} \in \mathscr{S}_L(\boldsymbol{\xi})$ and $\xi_i \geqslant 1$ for all i, then $x_j \leqslant 2.63 \xi_i \cdots \xi_{j-1} \varrho^{j-i} x_i$ for $0 \leqslant i < j \leqslant L$.*

*Proof.* Fix $i$ and note that the first inequality is trivial for $j = i$ and $j = i - 1$. Assume it holds now for $j \geqslant k + 1$. Then by $(I_k)$ and the induction hypothesis,

$$x_k \geqslant \sum_{h=1}^{L-k} a_h x_{k+h} \geqslant \sum_{h=1}^{i-k} a_h g^*_{i-k-h} x_i = g^*_{i-k} x_i.$$

By Lemma 3.7, the maximum of $\varrho^{-i}/g_i^*$ is $2.6211\ldots$, occurring at $i = 2$. The second inequality follows by Lemmas 3.2 and 3.10. □

Careful analysis of $\mathscr{S}_L$ reveals that most of the volume occurs with $x_i \approx \frac{L-i}{L}\varrho^i$ for each $i$, with the "standard deviation" from the mean increasing with $i$. This observation plays an important role in subsequent arguments. For now, we restrict our attention to the variable $x_L$, since results concerning the other variables will not be needed until section 6. The next lemma shows that $x_L \approx \varrho^L/L$ for most of $\mathscr{S}_L$, a bound which is significantly smaller than the global upper bound given by Lemma 3.10.

**Lemma 3.11.** *If $\alpha \geqslant 0$, then $\operatorname{Vol}(\mathscr{S}_L \cap \{x_L \leqslant \alpha\}) \ll T_L \alpha L \varrho^{-L}$. If $\alpha \geqslant 0$ and $\xi_i \geqslant 1$ for all $i$, then*

$$\operatorname{Vol}(\mathscr{S}_L^*(\boldsymbol{\xi}) \cap \{x_L \geqslant \alpha\}) \ll H(\boldsymbol{\xi}) e^{-\alpha L g_L^*/(\xi_0 \cdots \xi_{L-1})} T_L.$$

*Proof.* Consider first $\mathbf{x} \in \mathscr{S}_L \cap \{x_L \leqslant \alpha\}$. Since $(x_1, \ldots, x_{L-1}) \in \mathscr{S}_{L-1}$, the volume is $\leqslant \alpha T_{L-1}$. Applying Lemma 3.4 gives the first part.

Next, suppose $\mathbf{x} \in \mathscr{S}_L^*(\boldsymbol{\xi}) \cap \{x_L \geqslant \alpha\}$, set $x_i' = (\xi_0 \cdots \xi_{i-1})^{-1} x_i$ for each $i$, and $\alpha' = \alpha(\xi_0 \cdots \xi_{L-1})^{-1}$. By Lemma 3.2, $\mathbf{x}' \in \mathscr{S}_L \cap \{x_L \geqslant \alpha'\}$. If $\alpha' \geqslant 1/g_L^*$, the volume is zero by Lemma 3.10. Otherwise, set $y_i = x_i' - \alpha g^*_{L-i}$ for each $i$. We have $y_{L-1} \geqslant y_L \geqslant 0$, $\mathbf{v}_j \cdot \mathbf{y} \leqslant 0$ for $1 \leqslant j \leqslant L-2$, and $\mathbf{v}_0 \cdot \mathbf{y} \leqslant 1 - \alpha' g_L^*$. By Lemmas 3.4 and 3.6, the volume of such $\mathbf{y}$ is $\leqslant (1 - \alpha' g_L^*)^L T_L^* \ll (1 - \alpha' g_L^*)^L T_L$. The second part now follows. □

Recall the definition of $L_0(x)$ (1.7).

**Lemma 3.12.** *Suppose $\xi_i \geqslant 1$ for all $i$, $H(\boldsymbol{\xi}) \leqslant 2$, $L \leqslant L_0(y)$ and $\alpha \geqslant 0$. Then*

$$R_L(\mathscr{S}; y) \ll (\log_2 y)^L T_L e^{-\alpha L g_L^*/(\xi_0 \cdots \xi_{L-1})}, \qquad \mathscr{S} = \mathscr{S}_L(\boldsymbol{\xi}) \cap \{x_L \geqslant \alpha\}. \tag{3.12}$$

*If $\xi_i = 1 - \omega_i$, $\omega_i = \frac{1}{10(L_0 - i)^3}$ for each $i \leqslant L - 1$, then there is an absolute constant $M_1$ so that whenever $1 \leqslant A \leqslant (\log y)^{1/2}$, $M = [M_1 + 2C \log A]$ and $L \leqslant L_0(y) - M$, we have*

$$R_L(\mathscr{S}; y) \gg (\log_2 y)^L T_L, \tag{3.13}$$

*where $\mathscr{S}$ is the subset of $\mathscr{S}_L(\boldsymbol{\xi})$ with the additional restrictions*

$$x_{i+1} \leqslant (1 - \omega_i) x_i \quad (i \geqslant 1), \qquad x_L \geqslant \frac{A}{\log_2 y}. \tag{3.14}$$

*Proof.* Let $\varepsilon = 1/\log_2 y$. For (3.12), $R_L(\mathscr{S}; y) = R_L(\mathscr{S} \cap \{x_L \geqslant 2\varepsilon\}; y)$. Thus, without loss of generality we may assume $\alpha \geqslant 2\varepsilon$. For $\mathbf{x} \in \mathscr{S}$, we have by Lemma 3.10,

$$x_{L-j} \geqslant \frac{1}{2.63H} \varrho^{-j} x_L \geqslant \frac{\varrho^{-j}}{2.63 \log_2 x}.$$

By Lemma 3.1, $R_L(\mathscr{S}; y) \ll (\log_2 y)^L \operatorname{Vol}(\mathscr{S}^{+\varepsilon})$. If $\mathbf{x}' \in \mathscr{S}^{+\varepsilon}$, then there is $\mathbf{x} \in \mathscr{S}$ with $|x_j' - x_j| \leqslant \varepsilon$ for all $j$. In particular,

$$x'_{L-j} \geqslant \frac{x_{L-j}}{2} \geqslant \frac{\varrho^{-j}}{5.26H} \alpha.$$

Hence, by $(I_i)$, for $0 \leqslant i \leqslant L-2$,

$$\begin{aligned}\sum_{j=1}^{L-i} a_j x'_{i+j} &\leqslant \sum_{j=1}^{L-i} a_j(x_{i+j}+\varepsilon) \leqslant x_i\xi_i + \varepsilon \sum_{j=1}^{L-i} a_j \\ &\leqslant (x'_i+\varepsilon)\xi_i + \frac{\varepsilon}{2}(L-i)^2 \\ &\leqslant x'_i\xi_i\left(1+\frac{2.64H\varrho^{L-i}}{\alpha/2\varepsilon}\right) =: x'_i\xi'_i.\end{aligned}$$

Also, setting $\xi'_{L-1} = \xi_{L-1}(1+3\varepsilon/\alpha)$, we have

$$0 \leqslant x'_L \leqslant x_L + \varepsilon \leqslant \xi_{L-1}x_{L-1}(1+\varepsilon/\alpha) \leqslant \xi'_{L-1}x'_{L-1}.$$

Thus, $\mathscr{S}^{+\varepsilon} \subseteq \mathscr{S}_L(\boldsymbol{\xi}') \cap \{x'_L \geqslant \alpha - \varepsilon\}$. Finally, (3.12) follows from $H(\boldsymbol{\xi}') \ll 1$ and Lemma 3.11.

Now suppose $\xi_i = 1 - \frac{1}{10(L_0-i)^3}$ for each $i$. By Lemma 3.1, $R_L(\mathscr{S};y) \gg (\log_2 y)^L \operatorname{Vol}(\mathscr{S}^{-\varepsilon})$. For $1 \leqslant i \leqslant L-1$, put

$$\omega'_i = \frac{6(2+(L-i)\log(L-i))\varrho^{L-i}}{100+A}, \qquad \xi'_i = 1-\omega_i-\omega'_i.$$

Let $\mathscr{T}$ be the subset of $\mathscr{S}_L(\boldsymbol{\xi}')$ with the additional restrictions $x_{i+1} \leqslant \xi'_i x_i$ for each $i$ and $x_L \geqslant (200+\log_2 A)/\log_2 y$. Suppose $\mathbf{x} \in \mathscr{T}$ and $|x'_i - x_i| \leqslant \varepsilon$ for each $i$. By Lemma 3.10,

$$x'_i \geqslant \frac{x_i}{2} \geqslant \frac{\varrho^{L-i}}{6}x_L \geqslant \frac{\varrho^{L-i}(A+200)}{6\log_2 y}.$$

Thus, for $0 \leqslant i \leqslant L-1$,

$$x'_{i+1} \leqslant x_i + \varepsilon \leqslant \xi'_i(x'_i+\varepsilon)+\varepsilon \leqslant \left(\xi'_i - \frac{2\varepsilon}{x'_i}\right)x'_i \leqslant \xi_i x'_i$$

and

$$\begin{aligned}a_1x'_{i+1}+\cdots+a_{L-i}x'_L &\leqslant \xi'_i x_i + \varepsilon(a_1+\cdots a_{L-i}) \\ &\leqslant \xi'_i(x'_i+\varepsilon)+\varepsilon(1+(L-i)\log(L-i)) \leqslant \xi_i x'_i.\end{aligned}$$

Therefore, $\mathbf{x}' \in \mathscr{S}$ and hence $\mathscr{T} \subseteq \mathscr{S}^{-\varepsilon}$. Make the substitution $x_i = (\xi'_0 \cdots \xi'_{i-1})y_i$ for $1 \leqslant i \leqslant L$. Then $\mathbf{y} \in \mathscr{S}_L \cap \{y_L \geqslant (A+200)/\log_2 y\}$. By Lemma 3.11,

$$\operatorname{Vol}(\mathscr{S}^{-\varepsilon}) \geqslant \operatorname{Vol}(\mathscr{T}) \geqslant H(\boldsymbol{\xi}')\left[T_L - O(A\varrho^M T_L)\right] \gg T_L$$

if $M_1$ is large enough. □

## 4. The upper bound for $V(x)$

In this section, we prove that

$$V(x) \ll \frac{x}{\log x}\exp\{C(\log_3 x - \log_4 x)^2 + D(\log_3 x) - (D+1/2-2C)\log_4 x\}. \tag{4.1}$$

We begin with the basic tools needed for the proof, which show immediately the significance of the set $\mathscr{S}_L(\boldsymbol{\xi})$. First, recall the definition of an $S$-normal prime (2.1)–(2.2).

**Lemma 4.1.** *Suppose $y$ is sufficiently large, $k \geqslant 2$ and*

$$1 \geqslant \theta_1 \geqslant \cdots \geqslant \theta_k \geqslant \log_2 S/\log_2 y,$$

*where $S \geqslant \exp\{(\log_2 y)^{36}\}$. Let $\log_2 E_j = \theta_j \log_2 y$ for each $j$. The number of totients $v \leqslant y$ with a pre-image $n$ divisible only by $S$-normal primes and satisfying*

$$\log_2 q_j(n) \geqslant E_j \qquad (1 \leqslant j \leqslant k)$$

*is*

$$\ll y(\log y)^{A+B}(\log_2 y)(\log S)^{k\log k} + \frac{y}{(\log y)^2},$$

*where*

$$A = -\sum_{j=1}^{k} a_j\theta_j, \qquad B = 4\sqrt{\frac{\log_2 S}{\log_2 y}}\,\sum_{j=2}^{k}\theta_{j-1}^{1/2} j\log j.$$

*Proof.* Let $F = \min(E_1, y^{1/(20\log_2 y)})$, $E_{k+1} = S$, and $\theta_{k+1} = \log_2 S/\log_2 y$. We may assume that $\Omega(v) \leqslant 10\log_2 y$ and that neither $n$ nor $v$ is divisible by the square of a prime $\geqslant S$. By Lemmas 2.2 and 2.7, the number of exceptional $v$ is $O(y/\log^2 y)$. Let $m$ be the part of $v$ composed of primes in $(S, F]$. Then $m \leqslant F^{\Omega(v)} \leqslant y^{1/2}$. By Lemma 2.4, the number of totients with a given $m$ is

$$\ll \frac{y}{m}\frac{\log S}{\log F} = \frac{y}{m}(\log y)^{\theta_{k+1}-\theta_1}(\log_2 y).$$

Let $\delta_j = \sqrt{\log_2 S\log_2 E_{j-1}}$ for each $j$. Since the primes $q_i(n)$ are $S$-normal, by (2.2)

$$\Omega(m, E_j, E_{j-1}) \geqslant j(\theta_{j-1} - \theta_j - \delta_j)\log_2 y =: R_j \qquad (2 \leqslant j \leqslant k+1).$$

Therefore, the total number, $N$, of totients counted satisfies

$$N \ll y(\log y)^{\theta_{k+1}-\theta_1}(\log_2 y)\prod_{j=2}^{k+1}\sum_{r\geqslant R_j}\frac{t_j^r}{r!},$$

where

$$t_j = \sum_{E_j<p\leqslant E_{j-1}}\frac{1}{p} \leqslant (\theta_{j-1}-\theta_j)\log_2 y + 1 := s_j.$$

If $\delta_j \leqslant \frac{1}{2}(\theta_{j-1}-\theta_j)$, then

$$\frac{s_j}{R_j} \leqslant \frac{1}{j}\left(1 + \frac{3\delta_j}{\theta_{j-1}-\theta_j}\right)$$

and Lemma 2.1 implies

$$\sum_{r\geqslant R_j}\frac{s_j^r}{r!} \leqslant \left(\frac{es_j}{R_j}\right)^{R_j} \leqslant (\log y)^{j(\theta_{j-1}-\theta_j-\delta_j)(1-\log j+3\delta_j/(\theta_{j-1}-\theta_j))}$$

$$\leqslant (\log y)^{(j-j\log j)(\theta_{j-1}-\theta_j)+(j\log j+2j)\delta_j}.$$

If $\delta_j > \frac{1}{2}(\theta_{j-1}-\theta_j)$, then the sum on $r$ is

$$\leqslant e^{s_j} \leqslant (\log y)^{(j-j\log j)(\theta_{j-1}-\theta_j)+(2j\log j)\delta_j}.$$

Therefore,

$$N \ll y(\log y)^{A+B}(\log_2 y)(\log S)^{(k+1)\log(k+1)-(k+1)}.$$

□

**Lemma 4.2.** *Recall definitions* (1.3) *and* (3.1). *Suppose $k \geqslant 2$, $0 < \omega < 1/10$ and $y$ is sufficiently large (say $y \geqslant y_0$). Then the number of totients $v \leqslant y$ with a pre-image $n$ satisfying*

$$a_1x_1(n;y) + \cdots + a_kx_k(n;y) \geqslant 1 + \omega$$

*is*

$$\ll y(\log_2 y)^6 W(y)(\log y)^{-1-\omega^2/(600k^3\log k)}.$$

*Proof.* Assume that

$$\omega^2 > 3600 \frac{\log_3 y}{\log_2 y} k^3 \log k, \tag{4.2}$$

for otherwise the lemma is trivial. Define $S$ by

$$\log_2 S = \frac{\omega^2}{100k^3 \log k} \log_2 y, \tag{4.3}$$

so that $S \geqslant \exp\{(\log_2 y)^{36}\}$. The number of $v$ in question is $\leqslant U_1(y) + U_2(y) + U_3(y) + U_4(y)$, where

$$\begin{aligned} U_1(y) &= |\{\phi(n) \leqslant y : p|n \text{ for some prime } p \text{ which is not } S\text{-normal}\}|, \\ U_2(y) &= |\{v : v < y/\log^2 y \text{ or } \Omega(v) \geqslant 10 \log_2 y\}|, \\ U_3(y) &= |\{v : \text{ for some } d > S, d^2|v \text{ or } d^2|n \text{ for some } n \in \phi^{-1}(v)\}|, \\ U_4(y) &= |\{v : v \text{ not counted in } U_1(y), U_2(y), \text{ or } U_3(y)\}|. \end{aligned}$$

By (4.3) and Lemmas 2.2, 2.7 and 2.8 we have

$$\begin{aligned} U_1(y) + U_2(y) + U_3(y) &\ll \frac{y(\log_2 y)^6 W(y)}{\log y} (\log S)^{-1/6} + \frac{y \log_2 y}{S} \\ &\ll y(\log_2 y)^6 W(y)(\log y)^{-1-\omega^2/(600k^3 \log k)}. \end{aligned} \tag{4.4}$$

Let $\varepsilon = \omega/10$, $\alpha = a_1 + \cdots + a_k < k \log k$, and suppose $n$ is a pre-image of a totient counted in $U_4(y)$. Let $x_i = x_i(n; y)$ for $1 \leqslant i \leqslant k$. Then there are numbers $\theta_1, \ldots, \theta_k$ so that $\theta_i \leqslant x_i$ for each $i$, each $\theta_i$ is an integral multiple of $\varepsilon/\alpha$, $\theta_1 \geqslant \cdots \geqslant \theta_k$, and

$$1 + \omega - \varepsilon \leqslant a_1\theta_1 + \cdots + a_k\theta_k \leqslant 1 + \omega. \tag{4.5}$$

For each admissible $k$-tuple $\boldsymbol{\theta}$, let $T(\boldsymbol{\theta}; y)$ denote the number of totients counted in $U_4(y)$ which have some pre-image $n$ satisfying $x_i(n; y) \geqslant \theta_i$ for $1 \leqslant i \leqslant k$. Let $j$ be the largest index with $\theta_j \geqslant \log_2 S/\log_2 y$. By Lemma 4.1,

$$T(\boldsymbol{\theta}; y) \ll y(\log y)^{A+B}(\log_2 y)(\log S)^{k \log k} + y(\log y)^{-2},$$

where, by (4.5),

$$A = -\sum_{i=1}^{j} a_i\theta_i \leqslant -(1 + 0.9\omega) + \alpha \frac{\log_2 S}{\log_2 y}$$

and, by (4.3), (4.5) and the Cauchy-Schwarz inequality,

$$B \leqslant 4 \left( \frac{\log_2 S}{\log_2 y} (1 + \omega) \sum_{j=2}^{k} \frac{j^2 \log^2 j}{a_{j-1}} \right)^{1/2} \leqslant 5 \left( \frac{k^3 \log k \log_2 S}{\log_2 y} \right)^{1/2} \leqslant \frac{\omega}{2}.$$

Also

$$(\log S)^{2k \log k} = (\log y)^{\omega^2/(50k^2)} \leqslant (\log y)^{\omega/2000}.$$

Using (4.2), the number of vectors $\boldsymbol{\theta}$ is trivially at most

$$\left(\frac{\alpha}{\varepsilon}\right)^k \leqslant \left(\frac{10k \log k}{\omega}\right)^k \leqslant (\log_2 y)^{k/2} \leqslant (\log y)^{\omega^2/3000} \leqslant (\log y)^{\omega/30000}.$$

Therefore,

$$U_4(y) \leqslant \sum_{\boldsymbol{\theta}} T(\boldsymbol{\theta}; y) \ll y(\log y)^{-1-\omega/3},$$

which, together with (4.4), finishes the proof. □

Before proceeding with the main argument, we require a crude upper bound for $V(x)$ to get things started. We use the method of Pomerance [28] to show

$$W(x) \ll \exp\{(\log_3 x)^2\}. \tag{4.6}$$

(We could simply cite the result (1.2), but include a short argument below for the sake of completeness). Suppose $x$ is large, and let $v \leqslant x$ be a totient with pre-image $n$. By Lemmas 2.3 and 2.7, the number of $v$ with an $n$ satisfying either $p^2|n$ for some prime $p > \log^3 x$, or $m|n$ for some $m$ with $P^+(m) < m^{1/\log_2 x}$, is $O(x/\log^2 x)$. Let $\beta = 0.6$ and $\beta' = 0.599$. By Lemma 4.2, the number of $v$ with $\sum_{i=1}^4 a_i x_i(n;x) > 1.02$ is $O(xW(x)(\log x)^{-1.000000007})$. Since $\sum_{i=1}^4 a_i(\beta')^i > 1.02$, $x_i(n;x) \leqslant (\beta')^i$ for some $i \leqslant 4$ for each remaining totient. For $1 \leqslant i \leqslant 4$, let $A_i(x)$ denote the number of remaining totients with a pre-image satisfying $x_i(n;x) \leqslant (\beta')^i$ and $x_j(n;x) > (\beta')^j$ for $1 \leqslant j < i$. Write $v = \phi(q_0 \cdots q_{i-1})m$, so that $m \leqslant \exp\{(\log x)^{\beta^i}\}$. Each prime $q_j$, $0 \leqslant j \leqslant i-1$ occurs to the first power and the number of possible totients for a fixed $m$ is

$$\ll \frac{x}{\log x}\frac{1}{\phi(q_1 \cdots q_{i-1})m}.$$

By partial summation,

$$A_i(x) \ll \frac{x}{\log x}(\log_2 x)^i W(\exp\{(\log x)^{\beta^i}\}),$$

from which it follows that for some $i \leqslant 4$ we have $W(x) \ll (\log_2 x)^i W(\exp\{(\log x)^{\beta^i}\})$. Iterating the above expression as in [28] gives (4.6).

**Lemma 4.3.** *We have*

$$\sum_{\substack{v \in \mathscr{V} \\ P^+(v) \leqslant Z}} \frac{1}{v} \ll W(Z^{\log_2 Z}) \log_2 Z \ll \exp\{(\log_3 Z)^2 + \log_3 Z\}.$$

*Proof.* Let $f(y)$ denote the number of totients $v \leqslant y$ with $P^+(v) \leqslant Z$, and set $Z' = Z^{\log_2 Z}$. First suppose $y \geqslant Z'$. If $v > y^{1/2}$, then $P^+(v) < v^{2/\log_2 Z}$, so Lemma 2.3 gives $f(y) \ll y/\log^2 y$. For $y < Z'$, use the trivial bound $f(y) \leqslant V(y)$. The lemma follows from (4.6), $\log_2 Z' = \log_2 Z + \log_3 Z$ and partial summation. □

*Proof of* (4.1). Let $L = L_0(x)$ and for $0 \leqslant i \leqslant L-2$, let

$$\omega_i = \frac{1}{1000}\exp\left\{-\frac{L-i}{5}\right\}, \qquad \xi_i = 1 + \omega_i. \tag{4.7}$$

Then $H(\boldsymbol{\xi}) \leqslant 1.1$. Let $v$ be a generic totient $\leqslant x$ with a pre-image $n$, and set $x_i = x_i(n;x)$ for $i \geqslant 0$. Evidently

$$V(x) \leqslant \sum_{j=0}^{L-2} M_j(x) + N(x),$$

where $M_j(x)$ denotes the number of totients $\leqslant x$ with a pre-image satisfying inequality $(I_i)$ for $i < j$ but not satisfying inequality $(I_j)$, and $N(x)$ denotes the number of totients with every pre-image satisfying $\mathbf{x} \in \mathscr{S}_L(\boldsymbol{\xi})$. By Lemma 4.2 and (4.6),

$$M_0(x) \ll x(\log_2 x)^6 W(x)(\log x)^{-1-(\log_2 x)^{-2/3}} \ll x/\log x.$$

Now suppose $1 \leqslant j \leqslant L-2$, and set $k = L - j$. Let $n$ be a pre-image of a totient counted in $M_j(x)$, and set

$$w = q_j(n)q_{j+1}(n)\cdots, \qquad m = \phi(w).$$

Since $(I_0)$ holds, $x_2 \leqslant \xi_0/(a_1 + a_2) < 0.9$. It follows that $q_0(n) > x^{1/3}$, whence $m < x^{2/3}$. By the definition of $M_j(x)$ and (4.7),

$$x_j \leqslant \xi_j^{-1}(a_1 x_{j+1} + a_2 x_{j+2} + \cdots) < \xi_{j-1}^{-1}(a_1 x_j + a_2 x_{j+1} + \cdots) \leqslant x_{j-1},$$

whence $q_{j-1} > q_j$ and $\phi(n) = \phi(q_0 \cdots q_{j-1})m$. For each $m$, the number of choices for $q_0, \dots, q_{j-1}$ is

$$\ll \begin{cases} \frac{x}{m \log x} R_{j-1}(\mathscr{S}_{j-1}(\xi_0, \dots, \xi_{j-3}); x) & (j \geqslant 4) \\ \frac{x}{m \log x} (\log_2 x)^{j-1} & (j \leqslant 3). \end{cases}$$

Let $f(y)$ be the number of $m \leqslant y$. Define $Y_j$ by $\log_3 Y_j = 2k/9$. Since $m$ is a totient, we have $f(y) \leqslant V(y)$, but when $y > Y_j$ we can do much better. First note that $w \ll y \log_2 y$. By Lemma 2.3, the number of such $w$ with $P^+(w) < y^{1/\log_2 y}$ is $O(y/(\log y)^3)$. Otherwise, we have $q_j(n) = P^+(w) \geqslant y^{1/\log_2 y}$ and

$$x_j \geqslant \frac{\log_2 y - \log_3 y}{\log_2 x}.$$

For $0 \leqslant i \leqslant k$, let

$$z_i = x_i(w; y) = \frac{\log_2 x}{\log_2 y} x_{i+j}.$$

Since $(I_j)$ fails and $y > Y_j$, it follows that

$$a_1 z_1 + \cdots + a_k z_k \geqslant \frac{\log_2 x}{\log_2 y}(1 + \omega_j) x_j \geqslant (1 + \omega_j/2).$$

By Lemma 4.2 and (4.6), when $y \geqslant \max(y_0, Y_j)$ we have

$$f(y) \ll \frac{y W(y) (\log_2 y)^6}{\log y} \exp\left\{ -\frac{\omega_j^2}{600 k^3 \log k} \log_2 y \right\} \ll \frac{y}{\log y (\log_2 y)^2}.$$

Therefore, by partial summation and Lemma 4.3,

$$\sum_m \frac{1}{m} \ll 1 + \sum_{m \leqslant Y_j} \frac{1}{m} \ll W(Y_j) \log_2 Y_j \ll \exp\{4k^2/81 + 2k/9\}.$$

Therefore, by Corollary 3.5 (with $\Psi = k+1$) and Lemma 3.12,

$$\begin{aligned} M_j(x) &\ll \frac{x}{\log x} R_{j-1}(\mathscr{S}_{j-1}(\xi_0, \dots, \xi_{j-3}); x) \exp\{k^2/20\} \\ &\ll \frac{x}{\log x} T_{j-1} (\log_2 x)^{j-1} \exp\{k^2/20\} \qquad (4.8) \\ &\ll \frac{x}{\log x} \exp\{-k^2/4 - ((D+1)/2C - 1)k\} Z(x), \end{aligned}$$

where

$$Z(x) = \exp\{C(\log_3 x - \log_4 x)^2 + D \log_3 x - (D + 1/2 - 2C) \log_4 x\}.$$

Thus

$$\sum_{j=0}^{L-2} M_j(x) \ll \frac{x}{\log x} Z(x). \qquad (4.9)$$

Next, suppose $n$ is a pre-image of a totient counted in $N(x)$. By Lemma 3.10, $x_L \leqslant 3\varrho^L \leqslant \frac{10 \log_3 x}{\log_2 x}$. If $b$ is a nonnegative integer, let $N_b(x)$ be the number of totients counted in $N(x)$ with a pre-image $n > x/\log^2 x$ satisfying $b/\log_2 x \leqslant x_L \leqslant (b+1)/\log_2 x$. Let $w = q_L(n) q_{L+1}(n) \cdots$ and $q = q_1(n) \cdots q_{L-1}(n) w$. Since $x_2 < 0.9$ we have $q < x^{2/3}$ and for a fixed $q$, the number of possibilities for $q_0(n)$ is

$$\ll \frac{x}{\log x} \frac{1}{\phi(q)}.$$

By Corollary 3.3, Lemma 3.11 and the fact that $x_{L-1} \geqslant x_L \geqslant b/(L g_{L-1}^*)$, we have

$$\sum \frac{1}{\phi(q_1 \cdots q_{L-1})} \ll Z(x) e^{-b/4}.$$

By Lemma 4.3 and (4.6), $\sum \frac{1}{w} \ll \exp\{2\log^2 b\}$. Combining these estimates gives

$$N_b(x) \ll \frac{x}{\log x} Z(x) \exp\{-b/4 + 2\log^2 b\}.$$

Summing on $b$ gives $N(x) \ll \frac{x}{\log x} Z(x)$, which together with (4.9) gives (4.1). □

## 5. The lower bound for $V_k(x)$

Our lower bound for $V_k(x)$ is obtained by constructing a set of numbers with multiplicative structure similar to the numbers counted by $N(x)$ in the upper bound argument. Suppose $A(d) = \kappa$ and $\phi(d_i) = d \quad (1 \leqslant i \leqslant \kappa)$. Assume throughout that $x \geqslant x_0(d)$. The variable $k$ is reserved as an index for certain variables below. Define

$$M = M_2 + [(\log d)^{1/9}], M_2 \text{ is a sufficiently large abs. constant} \tag{5.1}$$

$$L = L_0(x) - M, \tag{5.2}$$

$$\xi_i = 1 - \omega_i, \quad \omega_i = \frac{1}{10(L_0 - i)^3} \qquad (0 \leqslant i \leqslant L). \tag{5.3}$$

Let $\mathscr{B}$ denote the set of integers $n = p_0 p_1 \cdots p_L$ with each $p_i$ prime and

$$\phi(n) \leqslant x/d, \tag{5.4}$$

$$(x_1(n;x), \cdots, x_L(n;x)) \in \mathscr{S}_L(\boldsymbol{\xi}), \tag{5.5}$$

$$\log_2 p_i \geqslant (1 + \omega_i) \log_2 p_{i+1} \qquad (0 \leqslant i \leqslant L-1), \tag{5.6}$$

$$p_L \geqslant \max(d+2, 16). \tag{5.7}$$

We show that for most numbers $n \in \mathscr{B}$, $A(d\phi(n)) = A(d)$, and thus $V_\kappa(x) \gg |\mathscr{B}|$. By Corollary 3.5 and Lemma 3.12,

$$|\mathscr{B}| \gg \frac{x}{d\log(x/d)} (\log_2(x/d))^L T_L \gg \frac{x}{d\log x} (\log_2 x)^L T_L. \tag{5.8}$$

Consider the equation

$$d\phi(n) = \phi(n_1), \tag{5.9}$$

where $n \in \mathscr{B}$. Let $q_0 \geqslant q_1 \geqslant \cdots$ be the prime factors of $n_1$. For $j \geqslant \Omega(n_1)$, put $q_j = 1$. If $n|n_1$, then none of the primes $q_i$ $(0 \leqslant i \leqslant L)$ occur to a power greater than 1, for otherwise (5.7) gives $\phi(n_1) \geqslant \phi(n)p_L > \phi(n)d$. Therefore $\phi(n_1) = \phi(n_1/n)\phi(n) = \phi(n)d$, which implies $n_1 = nd_i$ for some $i$. These we will call the trivial solutions to (5.9). We then have $A(d\phi(n)) = \kappa$ for each $n \in \mathscr{B}$ for which (5.9) has no non-trivial solutions. The numbers $n$ which give rise to non-trivial solutions may be grouped as follows. For $0 \leqslant j \leqslant L$, let $\mathscr{B}_j$ be the set of $n \in \mathscr{B}$ such that (5.9) holds for some $n_1$ with $q_i = p_i$ $(0 \leqslant i \leqslant j-1)$ and $p_j \neq q_j$. We then have

$$V_\kappa(x) \geqslant |\mathscr{B}| - \sum_{j=0}^{L} |\mathscr{B}_j|. \tag{5.10}$$

For $n$ counted in $\mathscr{B}_j$ with $j \geqslant 1$, write $n = p_0 n_2 n_3$, where $n_2 = p_1 \cdots p_{j-1}$ (when $j = 0$ or $j = 1$, put $n_2 = 1$) and $n_3 = p_j \cdots p_L$. When $j = 0$, set $n_3 = n$. Also put

$$h = L - j + 1 = \omega(n_3). \tag{5.11}$$

For $x$ large, (5.5) and (5.6) imply $p_0 > x^{1/2}$. By the Brun-Titchmarsh inequality, for each fixed $n_2 n_3$, the number of choices for $p_0$ is $O(x/(d\phi(n_2 n_3)\log x))$. Hence

$$|\mathscr{B}_j| \ll \frac{x}{d\log x} \sum_{n_2} \frac{1}{\phi(n_2)} \sum_{n_3} \frac{1}{\phi(n_3)} \quad (1 \leqslant j \leqslant L).$$

Since $n_2 \in W_{j-1}((\omega_1, \dots, \omega_{j-3}); x)$ when $j \geqslant 4$, Lemma 3.12 gives

$$\sum_{n_2} \frac{1}{\phi(n_2)} \ll (\log_2 x)^{j-1} T_{j-1} \qquad (1 \leqslant j \leqslant L).$$

To attack the sum on $n_3$, let $B_j(y)$ denote the number of possible $n_3$ with $\phi(n_3) \leqslant y$. In particular, $|\mathscr{B}_0| = B_0(x/d)$. When $j \geqslant 1$, by partial summation,

$$|\mathscr{B}_j| \ll \frac{x(\log_2 x)^{j-1} T_{j-1}}{d \log x} \sum_y \frac{B_j(y)}{y^2}. \tag{5.12}$$

Each $n_3$ counted in $B_j(y)$ satisfies

$$d\phi(n_3) = \phi(n_4) \tag{5.13}$$

for some $n_4$ with $P^+(n_4) \neq p_j$.

For small $y$, we have trivially

$$\sum_{\log_3 n_3 \leqslant M/10} \frac{1}{\phi(n_3)} \leqslant \left( \sum_{\log_2 p \leqslant e^{M/10}} \frac{1}{p-1} \right)^h \leqslant e^{hM/9}. \tag{5.14}$$

**Lemma 5.1.** *When* $\log_3 y \geqslant M/10$ *and* $0 \leqslant j \leqslant L$*,*

$$B_j(y) \ll \frac{y}{\log y (\log_2 y)^2}.$$

In particular, $|\mathscr{B}_0| \leqslant B_0(x/d) \ll x/(d \log x)$. Combining Lemma 5.1 with (5.1), (5.12), Lemma 3.4 and (5.14), we obtain for $j \geqslant 1$,

$$\begin{aligned} |\mathscr{B}_j| &\ll \frac{x}{d \log x} (\log_2 x)^{j-1} T_{j-1} \exp\{hM/9\} \\ &\ll \frac{x}{d \log x} (\log_2 x)^L T_L \exp\{h(M/9 - M/2C - h/4C)\}. \end{aligned}$$

If $M_2$ is sufficiently large, summing over $j$ and using (4.1), (5.8) and (5.10) gives

$$V_\kappa(x) \geqslant \frac{|\mathscr{B}|}{2} \gg_\varepsilon d^{-1-\varepsilon} V(x).$$

This completes the proof of Theorems 1 and 2.

*Proof of Lemma 5.1.* By (5.5), $p_j \leqslant y$ means (for $j \leqslant L-2$) that

$$\left( \frac{\log_2 p_{j+1}}{\log_2 y}, \cdots, \frac{\log_2 p_L}{\log_2 y} \right) \in \mathscr{S}_{L-j}((\xi_j, \dots, \xi_{L-2})). \tag{5.15}$$

Thus, by Lemma 3.10 and (5.7) (and trivially when $j \geqslant L-1$),

$$1 \leqslant \log_2 p_L \leqslant 3\varrho^{L-j} \log_2 y,$$

which implies

$$h \leqslant 2C \log_3 y + 3. \tag{5.16}$$

We first remove from consideration those $n_3$ with $n_3 \leqslant y/\log^2 y$, or $\Omega(\phi(n_4)) \geqslant 10 \log_2 y$, those with $b^2 | \phi(n_3)$ or $b^2 | n_4$ for some $b > \log^3 y$, and those with a divisor $m > \exp\{(\log_2 y)^2\}$ with $P^+(m) < m^{1/\log_2 y}$. By Lemmas 2.2, 2.3 and 2.7, the number of such $n_3$ is $O(y/\log^2 y)$. Next define

$$S = \exp\exp\{(\log_3 y)^3\}. \tag{5.17}$$

We next remove from consideration those $n_3$ with either $n_3$ or $n_4$ divisible by a prime which is not $S$-normal. By Lemma 2.8 and (4.1), the number of $n_3$ removed is

$$\ll \frac{y(\log_2 y)^5 W(y)}{\log y}(\log S)^{-1/6} \ll \frac{y}{\log y(\log_2 y)^2}.$$

Let $B_j^*(y)$ denote the number of remaining $n_3$, so that

$$B_j(y) \ll \frac{y}{\log y(\log_2 y)^2} + B_j^*(y). \tag{5.18}$$

For $j \leqslant L-1$, we have $p_{j+1}\cdots p_L \leqslant p_{j+1}^h$, so by (5.1), (5.6), (5.11), $M \leqslant 10\log_3 y$ and (5.16),

$$\log_2(n_3/p_j) \leqslant \frac{\log_2 p_j}{1+(h+M-1)^{-3}} + \log h \leqslant \log_2 y - 2\log_3 y \leqslant \log_2 y - 10.$$

In particular, since $n_3 > y/\log^2 y$, this shows that $p_j > y^{9/10}$ and $p_{j-1} < y^{1/\log_2 y}$. When $j = L$, $p_j = n_3 > y/\log^2 y$.

We now group the $n_3$ counted in $B_j^*(y)$ according to the sizes of $P^+(p_i-1)$. Set $\varepsilon = 1/\log_2 y$. For each $n_3$, there are numbers $\zeta_{j+1},\ldots,\zeta_L$, each an integral multiple of $\varepsilon$, and with $\zeta_i - \varepsilon \leqslant \frac{\log_2 p}{\log_2 y} \leqslant \zeta_i$ for each $i$. Also set $\zeta_j = 1$. By (5.5),

$$\sum_{i=1}^{L-j} a_i\zeta_{j+i} \leqslant 1 - \omega_j + (L-j)^2\varepsilon \leqslant 1 - \omega_j/2. \tag{5.19}$$

Let $J$ be the minimum index with $\zeta_J \leqslant (\log_2 y)^{-1/3}$. We make a further subdivision of the numbers $n_3$, counting separately those with $(p_j\cdots p_{J-1}, q_j\cdots q_{J-1}) = m_1$ and $p_J\cdots p_L = m_2$ (if $J = L+1$ set $m_2 = 1$). For a given $\boldsymbol{\zeta}, m_1, m_2$ giving rise to at least one $n_3$, let $B_j(\boldsymbol{\zeta}; m_1, m_2; y)$ be the number of $n_3$ counted by $B_j^*(y)$ satisfying

$$y^{9/10} \leqslant p_j \leqslant y, \quad \zeta_i - \varepsilon \leqslant \frac{\log_2 P^+(p_i-1)}{\log_2 y} < \zeta_i \qquad (j+1 \leqslant i \leqslant L).$$

We then have

$$B_j^*(y) = \sum_{\boldsymbol{\zeta}, m_1, m_2} B_j(\boldsymbol{\zeta}; m_1, m_2; y), \tag{5.20}$$

where the sum is over $\boldsymbol{\zeta}$ corresponding to at least one $n_3$. Note that when $j = L$ there is only one possible $\boldsymbol{\zeta}$ in (5.20).

Fix $m_1, m_2, \boldsymbol{\zeta}$ and suppose $n_3$ is counted in $B_j(\boldsymbol{\zeta}; y)$. By (5.6), (5.11), $\log_3 y \geqslant M/10$ and (5.16), for $J \leqslant i \leqslant j+1$,

$$\zeta_i \geqslant (1+\omega_i)\zeta_{i-1} \geqslant \zeta_{i-1} + \frac{(\log_2 y)^{-1/3}}{10(M+h)^3} \geqslant \zeta_{i-1} + (\log_2 y)^{-0.35}. \tag{5.21}$$

In particular, $\zeta_1 \leqslant 1 - (\log_2 y)^{-0.35}$. Now let $\delta = \sqrt{\log_2 S \log_2 y}$. We next claim that

$$\left| \frac{\log_2 P^+(p_i-1) - \log_2 P^+(q_i-1)}{\log_2 y} \right| \leqslant (2(i-j)+1)\delta \qquad (j+1 \leqslant i \leqslant J-1). \tag{5.22}$$

To see this, fix $i$, let $\alpha = \log_2 P^+(p_i-1)/\log_2 y$, $\beta = \log_2 P^+(q_i-1)/\log_2 y$, $E(z) = \exp\{(\log y)^z\}$ and $k = i-j$. By (2.2), if $\beta > a + (2k+1)\delta$, then

$$(k+1)(\beta - \alpha - \delta) \leqslant \Omega(\phi(n_3), E(\alpha), E(\beta)) \leqslant k(\beta - \alpha + \delta),$$

a contradiction. Assuming $\beta < \alpha - (2k+1)\delta$ likewise leads to a contradiction. This establishes (5.22).

Let

$$j = j_0 < j_1 < \cdots < j_{K-1} \leqslant J-1,$$

with $K \geqslant 1$, be the indices with $p_{j_i} \neq q_{j_i}$. Also, define $j_K = J$. Let $\nu_0 = 1$, for $i \geqslant 1$ let $\nu_i = \zeta_{j_i} + (2L+1)\delta$, and for $i \geqslant 0$ let $\mu_i = \nu_i - (4L+2)\delta - \varepsilon$. For brevity, for $0 \leqslant k \leqslant K$ set $u_k = E(\mu_k)$ and $v_k = E(\nu_k)$. Then $B_j(\boldsymbol{\zeta}; m_1, m_2; y)$ is at most the number of solutions of

$$(p_{j_0} - 1) \cdots (p_{j_{K-1}} - 1)\phi(m_2) d = (q_{j_0} - 1) \cdots (q_{j_{K-1}} - 1) e \leqslant y/(d\phi(m_1)), \tag{5.23}$$

where $e$ is a totient satisfying $P^+(e) \leqslant v_K$ and $p_{j_i}$ and $q_{j_i}$ are $S$-normal primes satisfying

$$w_k \leqslant P^+(p_{j_k} - 1), P^+(q_{j_k} - 1) \leqslant v_k \qquad (0 \leqslant k \leqslant K-1). \tag{5.24}$$

For $0 \leqslant k \leqslant K$, let

$$s_k(n) = \prod_{\substack{p^a \| n \\ p \leqslant v_k}} p^a, \qquad t_k(n) = \prod_{\substack{p^a \| n \\ v_k < p \leqslant v_{k-1}}} p^a. \tag{5.25}$$

For each $2K$-tuple

$$\mathscr{A} = (p_{j_0}, \ldots, p_{j_{K-1}}; q_{j_0}, \ldots, q_{j_{K-1}})$$

giving rise to a solution of (5.23), let $b = \phi(p_{j_0} \cdots p_{j_{K-1}} m_2)$ and define

$$\begin{aligned} \sigma_k(\mathscr{A}) &= \{s_k(b); s_k(p_{j_0} - 1), \ldots, s_k(p_{j_{K-1}} - 1); s_k(q_{j_0} - 1), \ldots, s_k(q_{j_{K-1}} - 1)\}, \\ \tau_k(\mathscr{A}) &= \{t_k(b); t_k(p_{j_0} - 1), \ldots, t_k(p_{j_{K-1}} - 1); t_k(q_{j_0} - 1), \ldots, t_k(q_{j_{K-1}} - 1)\}. \end{aligned}$$

Defining multiplication of $(2K+1)$-tuples by component-wise multiplication, we have from (5.25),

$$\sigma_{k-1}(\mathscr{A}) = \sigma_k(\mathscr{A})\tau_k(\mathscr{A}). \tag{5.26}$$

Let $\mathfrak{S}_k$ denote the set of $\sigma_k(\mathscr{A})$ arising from solutions $\mathscr{A}$ of (5.23) and $\mathfrak{T}_k$ the corresponding set of $\tau_k(\mathscr{A})$. By (5.26),

$$B_j(\boldsymbol{\zeta}; m_1, m_2; y) \leqslant |\mathfrak{S}_0| = \sum_{\sigma \in \mathfrak{S}_1} \sum_{\substack{\tau \in \mathfrak{T}_1 \\ \sigma\tau \in \mathfrak{S}_0}} 1. \tag{5.27}$$

Let $\sigma^{(j)}$ denote the $j$th coordinate of a $(2K+1)$-tuple $\sigma$. We will apply an iterative procedure based on the identity

$$\sum_{\sigma_{k-1} \in \mathfrak{S}_{k-1}} \frac{1}{\sigma_{k-1}^{(1)}} = \sum_{\sigma_k \in \mathfrak{S}_k} \frac{1}{\sigma_k^{(1)}} \sum_{\substack{\tau_k \in \mathfrak{T}_k \\ \sigma_k\tau_k \in \mathfrak{S}_{k-1}}} \frac{1}{\tau_k^{(1)}}. \tag{5.28}$$

The following two lemmas provide the necessary estimates for the application of (5.28).

**Lemma 5.2.** *For each $\sigma \in \mathfrak{S}_1$, we have*

$$\sum_{\substack{\tau \in \mathfrak{T}_1 \\ \sigma\tau \in \mathfrak{S}_0}} 1 \ll \frac{y(\log_2 y)^6}{\phi(m_1)\sigma^{(1)}(\log y)^{2+\nu_1}}.$$

*Proof.* Let $r = \sigma^{(2)}, s = \sigma^{(K+2)}$, $t = \tau^{(1)} = \tau^{(2)} = \tau^{(K+2)}$. We have $t \leqslant y/(\sigma^{(1)}\phi(m_1))$ and, by (5.1) and (5.6), $\sigma^{(1)} m_1 \leqslant y^{1/10}$. Note that $t$ is composed only of prime factors $\geqslant v_1$, and $rt+1$ and $st+1$ are unequal primes. If $t > 1$, write $t = t'Q$, where $Q = P^+(t)$. Given $t'$, we use Lemma 2.5 to bound the number of $Q$, then Lemma 2.4 to bound the resulting sum of $1/t'$. This gives the lemma. □

**Lemma 5.3.** *If $2 \leqslant k \leqslant K$ and $\sigma \in \mathfrak{S}_k$, then*

$$\sum_{\substack{\tau \in \mathfrak{T}_k \\ \sigma\tau \in \mathfrak{S}_{k-1}}} \frac{1}{\tau^{(1)}} \ll (\log_2 y)^5 (\log y)^{-2\mu_{k-1} + (k\log k + k)(\nu_{k-1} - \nu_k + \delta)}.$$

*Proof.* By (5.24), $\tau$ must have the form

$$\tau = \{w; f_0, \dots, f_{k-1}, 1, \dots, 1; b_0, \dots, b_{k-1}, 1, \dots, 1\},$$

where $w = f_0 \cdots f_{k-1} = b_0 \cdots b_{k-1}$. Write $f = f_{k-1}$ and $b = b_{k-1}$ for short, and let

$$r = \sigma^{(k+1)} = s_k(p_{j_{k-1}} - 1), \quad s = \sigma^{(K+k+1)} = s_k(q_{j_{k-1}} - 1).$$

Since $\sigma\tau \in \mathfrak{S}_{k-1}$, $rf+1$ and $sb+1$ are unequal primes. Let $Q_1 = P^+(f)$, $Q_2 = P^+(b)$, $f' = f/Q_1$ and $b' = b/Q_2$. We sum separately over $\mathfrak{T}_{k,1}$, the set of $\tau$ with $Q_1 = Q_2$ and $\mathfrak{T}_{k,2}$, the set of $\tau$ with $Q_1 \neq Q_2$.

For the sum over $\mathfrak{T}_{k,1}$, let $w = tQ_1$. We have

$$\Sigma_1 := \sum_{\substack{\tau \in \mathfrak{T}_{k,1} \\ \sigma\tau \in \mathfrak{S}_{k-1}}} \frac{1}{\tau^{(1)}} = \sum_t \frac{h(t)}{t} \sum_{Q_1} \frac{1}{Q_1}.$$

Here $h(t)$ denotes the number of solutions of

$$f_0 \cdots f_{k-2} f' = t = b_0 \cdots b_{k-2} b',$$

and $f'Q_1 + 1$, $b'Q_1 + 1$ are unequal primes. By sieve methods (Lemma 2.5), the number of $Q_1 \leqslant z$ is $\ll z(\log z)^{-3}(\log_2 y)^3$ uniformly in $f', b'$. By partial summation, we have

$$\sum_{Q_1} \frac{1}{Q_1} \ll (\log_2 y)^3 (\log y)^{-2\mu_{k-1}}.$$

Also, $h(t)$ is at most the number of dual factorizations of $t$ into $k$ factors each, i.e. $h(t) \leqslant k^{2\Omega(t)}$. By (2.2),

$$\Omega(t) \leqslant k(\nu_{k-1} - \nu_k + \delta) \log_2 y =: I.$$

Thus

$$\sum_t \frac{h(t)}{t} \leqslant \sum_{i \leqslant I} \frac{k^{2i} H^i}{i!},$$

where

$$\sum_{v_k < p \leqslant v_{k-1}} \frac{1}{p} \leqslant (\nu_{k-1} - \nu_k) \log_2 y + 1 =: H.$$

By (5.21), $\nu_{k-1} - \nu_k > 2\delta$, hence $I < \frac{3}{2}kH \leqslant \frac{3}{4}k^2 H$. Applying Lemma 2.1 (with $\alpha \leqslant \frac{3}{4}$) yields

$$\sum_t \frac{h(t)}{t} \leqslant \left(\frac{eHk^2}{I}\right)^I \leqslant (ek)^I \leqslant (\log y)^{(k + k\log k)(\nu_{k-1} - \nu_k + \delta)}. \tag{5.29}$$

This gives

$$\Sigma_1 \ll (\log_2 y)^3 (\log y)^{-2\mu_{k-1} + (k + k\log k)(\nu_{k-1} - \nu_k + \delta)}.$$

For the sum over $\mathfrak{T}_{k,2}$, set $w = tQ_1Q_2$. Note that

$$tQ_2 = f_0 \cdots f_{k-2} f', \qquad tQ_1 = b_0 \cdots b_{k-2} b',$$

so $Q_1$ (respectively $Q_2$) divides one of the factors on the right side. If we fix the factors divisible by $Q_1$ and by $Q_2$, then the number of possible ways to form $t$ is $\leqslant k^{2\Omega(t)}$ as before. Then

$$\Sigma_2 := \sum_{\substack{\tau \in \mathfrak{T}_{k,2} \\ \sigma\tau \in \mathfrak{S}_{k-1}}} \frac{1}{\tau^{(1)}} \leqslant \sum_t \frac{k^{2\Omega(t)+2}}{t} \sum_{Q_1, Q_2} \frac{1}{Q_1 Q_2},$$

where $h(t)$ is the number of $\mathbf{v}$ corresponding to a given $t$, and $f'Q_1+1$, $b'Q_2+1$ are unequal primes. By sieve methods (Lemma 2.5), the number of $Q_1 \leqslant z$ (respectively $Q_2 \leqslant z$) is $\ll z(\log z)^{-2}(\log_2 y)^2$. By partial summation, we have

$$\sum_{Q_1,Q_2} \frac{1}{Q_1Q_2} = \sum_{Q_1}\frac{1}{Q_1}\sum_{Q_2}\frac{1}{Q_2} \ll (\log_2 y)^4(\log y)^{-2\mu_{k-1}}.$$

Combined with (5.29) this gives

$$\Sigma_2 \ll k^2(\log_2 y)^4(\log y)^{-2\mu_{k-1}+(k+k\log k)(\nu_{k-1}-\nu_k+\delta)}.$$

By (5.16), $k \leqslant K < 2C\log_3 y+3$, so $k^2 \ll \log_2 y$. Adding $\Sigma_1$ and $\Sigma_2$ establishes the lemma. □

Combining Lemmas 5.2 and 5.3 with (5.27) and (5.28) gives

$$B_j(\boldsymbol{\zeta};m_1,m_2;y) \ll \frac{y}{\phi(m_1)}(c_6\log_2 y)^{5K}(\log y)^{-2-\nu_1+\sum_{k=2}^K b_k}\sum_{\sigma\in\mathfrak{S}_K}\frac{1}{\sigma^{(1)}}, \tag{5.30}$$

where

$$\begin{aligned} b_k &= (\nu_{k-1}-\nu_k+\delta)(k\log k+k)-2\mu_{k-1} \\ &= (\zeta_{j_{k-1}}-\zeta_{j_k})(k\log k+k)-2\zeta_{j_{k-1}}+(k\log k+k+2L+1)\delta+\varepsilon. \end{aligned}$$

From the definition of $a_k$, (5.1), and (5.16), the exponent of $\log y$ in (5.30) is at most

$$-2+\sum_{k=1}^{K-1}a_k\zeta_{j_k}+10L^3\delta \leqslant -1-\omega_j/2+(\log_2 y)^{-0.4} \leqslant -1-\omega_j/3. \tag{5.31}$$

It remains to treat the sum on $\sigma\in\mathfrak{S}_k$. Suppose $f=\sigma^{(1)}$ and consider (5.23). For a given $f$, the number of possible pairs $(\sigma,m_2)$ which give $\sigma^{(1)}=f$ is at most the number of factorizations of $f$ into $K+L-J+1$ factors times the number of factorizations of $df$ into $K+1$ factors, which is at most

$$(K+1)^{\Omega(df)}(K+L-J+1)^{\Omega(f)} \leqslant h^{2\Omega(f)+\Omega(d)}.$$

The definition (5.1) and $\log_2 y \geqslant M^{18}$ imply $d \leqslant \exp\{(\log_2 y)^{1/2}\}$. Hence, by (5.1) and (5.16),

$$h^{\Omega(d)} \leqslant h^{2\log d} \leqslant \exp\{3\log_4 y(\log_2 y)^{1/2}\}.$$

By (2.2), (5.11), (5.16) and (5.17),

$$\Omega(f) \leqslant 2h\log_2 E(\nu_J) \leqslant 2(2C\log_3 y+3)(\log_2 y)^{2/3}.$$

and thus

$$h^{2\Omega(f)+\Omega(d)} \ll \exp\{7\log_3 y(\log_2 y)^{2/3}\}.$$

Also,

$$\sum_{P^+(f)\leqslant E(\nu_J)}\frac{1}{f} \ll \log E(\nu_J) \leqslant \exp\{(\log_2 y)^{2/3}\}.$$

Hence,

$$\sum_{m_2}\sum_{\sigma\in\mathfrak{S}_K}\frac{1}{\sigma^{(1)}} \ll \exp\{8\log_3 y(\log_2 y)^{2/3}\}. \tag{5.32}$$

Therefore, by (5.30), (5.31) and (5.32),

$$\sum_{m_2}B_j(\boldsymbol{\zeta};m_1,m_2;y) \ll \frac{y}{\phi(m_1)}(\log y)^{-1-\omega_j/3+8\log_3 y(\log_2 y)^{-1/3}}.$$

Now $m_1$ has at most one prime factor in each interval $[w_i, z_i]$ $(j+1 \leqslant i \leqslant J-1)$ and no other prime factors. Therefore,

$$\sum_{m_1} \frac{1}{\phi(m_1)} \ll \prod_{i=j+1}^{J-1} (\varepsilon \log_2 y) \ll \exp\{(\log_3 y)^2\}.$$

Also, the number of possibilities for $\boldsymbol{\zeta}$ is at most $\varepsilon^{-L} \leqslant \exp\{2(\log_3 y)^2\}$. Summing over all possible $m_1$ and $\boldsymbol{\zeta}$, and applying $\log_2 y \geqslant M^{18}$ and (5.16), we have

$$\begin{aligned} B_j^*(y) &\ll \frac{y}{\log y} (\log y)^{-\omega_j/3 + (\log_2 y)^{-1/4}} \\ &\ll \frac{y}{\log y} \exp\left\{ \frac{-\log_2 y}{20(2C\log_3 y + M + 3)^3} + (\log_2 y)^{3/4} \right\} \\ &\ll \frac{y}{\log y} \exp\{-(\log_2 y)^{9/10}\}. \end{aligned}$$

Combining this with (5.18) completes the proof of Lemma 5.1. □

## 6. THE NORMAL STRUCTURE OF TOTIENTS

The proofs of Theorems 1 and 2 suggest that for most totients $m \leqslant x$, all the pre-images $n$ of $m$ satisfy $(x_1, x_2, \ldots, x_L) \in S_L(\boldsymbol{\xi})$ with $L$ near $L_0$ and $\boldsymbol{\xi}$ defined as in section 4. We prove such a result below in Theorem 15, which is an easy consequence of Theorem 1 and the machinery created for its proof. From this, we deduce the normal size of $\Omega(m)$ (Theorem 10, Corollary 11) from the normal size of $x_1 + \cdots + x_L$ for $\mathbf{x} \in \mathscr{S}_L$. Likewise, the normal sizes of the primes $q_i(n)$ are deduced from results on the normal sizes of $x_i$ for $\mathbf{x} \in \mathscr{S}_L$ (Theorems 12 and 13).

**Theorem 15.** *Suppose* $0 \leqslant \Psi < L_0(x)$, $L = L_0 - \Psi$ *and let*

$$\xi_i = \xi_i(x) = 1 + \frac{1}{1000} e^{-(L_0 - i)/5} \qquad (0 \leqslant i \leqslant L-2). \tag{6.1}$$

*The number of totients* $m \leqslant x$ *with a pre-image* $n$ *with*

$$(x_1(n;x), \ldots, x_L(n;x)) \notin \mathscr{S}_L(\boldsymbol{\xi}) \tag{6.2}$$

*is* $\ll V(x)\exp\{-\Psi^2/4\}$.

*Proof.* As in section 4, define $M_j(x)$ to be the number of totients $m \leqslant x$ with a pre-image satisfying $(I_i)$ for $i < j$, but not satisfying $(I_j)$, where $\mathbf{x} = (x_1(n;x), \ldots, x_{L_0}(n;x))$. By Theorem 1, Corollary 3.5, and (4.8), the number of totients $m \leqslant x$ with a pre-image $n$ satisfying (6.2) is at most

$$\sum_{j \leqslant L} M_j(x) \ll \frac{x}{\log x} Z(x) e^{-\Psi^2/4} \ll V(x) e^{-\Psi^2/4}.$$

□

We show below that most of the contribution to $\mathscr{S}_L$ comes from $\mathbf{x}$ with

$$x_j \approx \varrho^j (1 - j/L) \qquad (1 \leqslant j \leqslant L-2).$$

Thus, for most totients $m \leqslant x$, we should have $\Omega(m) \approx \frac{1}{1-\varrho} \log_2 x$. Proving this requires accurate upper bounds on $T_L(\mathscr{R})$ for various regions $\mathscr{R}$, where for brevity we write

$$T_L(\mathscr{R}) = \mathrm{Vol}\,(\mathscr{S}_L \cap \mathscr{R})\,.$$

For short, we write $\Sigma$ for $\sum_{i=1}^L x_i$ and set $\beta_0 = \varrho/(1-\varrho)$. As with the bounding of $N(x)$ in section 4, we need to group totients by the size of $x_L$.

**Lemma 6.1.** *Suppose* $\alpha \leqslant 1/(3g_L^*)$ *and* $L \geqslant 100$. *For* $\beta_0 \leqslant \beta \leqslant 2$ , *we have*

$$T_L(\Sigma \geqslant \beta, x_L \geqslant \alpha) \ll \left( \frac{1-\alpha g_L^*}{1+a_1(\beta-\beta_0)/(1-a_1\beta_0)} \right)^L T_L. \tag{6.3}$$

*When* $0 \leqslant \beta \leqslant \beta_0$, *we have*

$$T_L(\Sigma \leqslant \beta, x_L \geqslant \alpha) \ll \left( \frac{\beta(1-\alpha g_L^*)}{\beta_0} \right)^L T_L. \tag{6.4}$$

*Proof.* For $j \geqslant 0$, let

$$f_j = \sum_{i=0}^j g_i, \qquad f_j^* = \sum_{i=0}^j g_i^*.$$

Suppose $\mathbf{x} \in \mathscr{S}_L$ and let $\mathbf{1} = \{1, 1, \cdots, 1\}$. By (3.7),

$$-\mathbf{1} = \sum_{i=1}^{L-1} f_{i-1}\mathbf{v}_i + f_{L-1}^*\mathbf{v}_L. \tag{6.5}$$

First suppose $x_1 + \cdots + x_L \geqslant \beta$ and $x_L \geqslant \alpha$. Set $y_i = x_i - \alpha g_{L-i}^*$ for each $i$. By (3.4), (3.5), (3.6) and (6.5),

$$\begin{aligned} \mathbf{v}_i \cdot \mathbf{y} &\leqslant 0 \qquad (1 \leqslant i \leqslant L), \\ \mathbf{v}_0 \cdot \mathbf{y} &\leqslant 1 - \alpha g_L^* =: \alpha', \\ \mathbf{1} \cdot \mathbf{y} &\geqslant \beta - \alpha f_{L-1}^* =: \beta'. \end{aligned}$$

Thus, $\mathbf{u} \cdot \mathbf{y} \leqslant 0$, where $\mathbf{u} = \beta'\mathbf{v}_0 - \alpha'\mathbf{1}$. This relation will replace $\mathbf{v}_1 \cdot \mathbf{y} \leqslant 0$. By (3.8) and (6.5),

$$\mathbf{v}_0 + A\mathbf{u} + \sum_{j=2}^{L-1} (g_j + A(\beta' g_j - \alpha' f_{j-1}))\mathbf{v}_j + (g_L^* + A(\beta' g_L^* - \alpha' f_{L-1}^*))\mathbf{v}_L = \mathbf{0}, \tag{6.6}$$

where $A = a_1/(\alpha' - \beta' a_1)$. The restrictions on $\alpha$ and $\beta$ force $a_1 \leqslant A \leqslant 22a_1$. Since $f_{j-1}/g_j \leqslant f_1/g_2 \leqslant 1.31$ by Lemma 3.7, the coefficient of $\mathbf{v}_j$ is

$$\geqslant g_j(1 + A(\beta' - 1.31\alpha')) \geqslant g_j \left( 1 + \frac{a_1(\beta - 1.36)}{0.819 - a_1\beta} \right) > 0.$$

Lemma 3.7 also gives $f_{j-1} = \beta_0 g_j + O(j)$ and $f_{j-1}^* = \beta_0 g_j^* + O(j)$. Applying Lemmas 3.4, 3.6 and 3.7, we deduce

$$T_L(\Sigma \geqslant \beta, x_L \geqslant \alpha) \ll \left( \frac{\alpha'}{1 + A(\beta' - \beta_0\alpha')} \right)^L T_L.$$

The relation $\beta' = \beta - \alpha\beta_0 + O(\alpha L) = \beta\alpha' + O(\alpha L)$ now gives (6.3).

Now suppose $x_1 + \cdots + x_L \leqslant \beta$ and $x_L \geqslant \alpha$. Adopting the same definitions for $y_i, \alpha'$ and $\beta'$, we have $\mathbf{v}_i \cdot \mathbf{y} \leqslant 0$ $(1 \leqslant i \leqslant L)$ and $\mathbf{1} \cdot \mathbf{y} \leqslant \beta'$. Assume $\beta' > 0$, for otherwise the desired volume is zero by Lemma 3.10. By (6.5) and Lemma 3.6, $T_L(\Sigma \leqslant \beta, x_L \geqslant \alpha) \ll (\beta'/\beta_0)^L T_L$, which gives (6.4). □

Note that Lemma 6.1 gives non-trivial bounds only when $|\beta - \beta_0| \gg L^{-1}$.

*Proof of Theorem 10.* Assume $\varepsilon \geqslant (\log_3 x)^{-1}$, for otherwise the theorem is trivial. Denote by $V^*(x)$ the number of totients $m \leqslant x$ satisfying

$$\left| \Omega(m) - \frac{\log_2 x}{1-\varrho} \right| \geqslant \varepsilon \log_2 x.$$

Let $\Psi = \Psi(x) = [3\sqrt{\varepsilon \log_3 x}]$, $L = L_0(x) - \Psi$, define $\xi_i$ by (6.1) and set $S = \exp\{(\log_2 x)^{100}\}$. Let $n$ be a generic pre-image of a totient $m \leqslant x$, and set $q_i = q_i(n)$ and $x_i = x_i(n; x)$ for $0 \leqslant i \leqslant L$. Also set $v = \phi(n/(q_0 \cdots q_L))$. Let $U(x)$ denote the number of totients $m \leqslant x$ satisfying one of four conditions:

$$(x_1, x_2, \dots, x_L) \notin \mathscr{S}_L(\boldsymbol{\xi}), \tag{6.7}$$

$$p^2 | m \text{ for some prime } p \geqslant \log^2 x, \tag{6.8}$$

$$\text{some prime factor of } n \text{ is not } S\text{-normal}, \tag{6.9}$$

$$\Omega(v) \geqslant (\log_2 x)^{1/2}. \tag{6.10}$$

By Theorem 15, Lemma 2.7 and Lemma 2.8, the number of totients $m \leqslant x$ with a pre-image satisfying either (6.7), (6.8) or (6.9) is

$$\ll V(x) \left\{ \exp\{-\Psi^2/4\} + \frac{1}{\log x} + \frac{(\log_2 x)^5}{(\log S)^{1/6}} \right\} \ll V(x)(\log_2 x)^{-2\varepsilon}.$$

Now suppose (6.10) holds, but (6.7) and (6.9) do not. By Lemma 3.10,

$$x_L \leqslant 3\varrho^L < \frac{10 \log_3 x}{\log_2 x} \varrho^{-\Psi} =: \frac{Y}{\log_2 x}.$$

Thus $\log_2 P^+(v) \leqslant Y$ and $\Omega(v) \geqslant 10Y$. Given $q_1, \cdots, q_L$ and $v$, the number of possible $q_0$ is

$$\ll \frac{x}{\log x} \frac{1}{v\phi(q_1 \cdots q_L)}.$$

By Lemma 3.12,

$$\sum \frac{1}{\phi(q_1 \cdots q_L)} = R_L(\boldsymbol{\xi}) \ll (\log_2 x)^L T_L \ll W(x)(\log_2 x)^{-2\varepsilon}.$$

By Lemmas 2.2, 2.3 and partial summation, $\sum 1/v \ll 1$, and hence

$$U(x) \ll V(x)(\log_2 x)^{-2\varepsilon}. \tag{6.11}$$

Denote by $U_1(x)$ and $U_2(x)$ the number of remaining totients $m \leqslant x$ with $\Omega(m) \geqslant (1 + \beta_0 + \varepsilon) \log_2 x$ and $\Omega(m) \leqslant (1 + \beta_0 - \varepsilon) \log_2 x$, respectively. By (2.1) and (2.2), we have

$$\Omega(q_i - 1) = \log_2 q_i + O(\sqrt{\log_2 x \log_3 x}) \qquad (1 \leqslant i \leqslant L).$$

Therefore, by (6.10),

$$\Omega(m) = (1 + x_1 + x_2 + \cdots + x_L) \log_2 x + O((\log_2 x)^{1/2} (\log_3 x)^{3/2}). \tag{6.12}$$

Write $m = \phi(q_0 q_1 \cdots q_L)v$, and divide the interval $[0, 1/g_L^*]$ into sub-intervals, considering separately the totients having a pre-image with $x_L$ in a particular sub-interval. We note that

$$\frac{1}{Lg_L^*} \ll \frac{\varrho^{-\Psi}}{\log_2 x}.$$

By Lemmas 3.1, 4.3 and 6.1, together with Theorem 1, the number of totients counted in $U_1(x)$ with $y_L \in [u/Lg_L^*, v/Lg_L^*]$ (take $Z = E(v/Lg_L^*) = \exp\exp[O(v\varrho^{-\Psi})]$) is

$$\begin{aligned} &\ll \frac{x}{\log x} (\log_2 x)^L \left( \frac{1 - u/L}{1 + 0.71\varepsilon} \right)^L T_L W(Z^{\log_2 Z}) \log_2 Z \\ &\ll V(x) \exp\{-1.02\varepsilon \log_3 x - \Psi \log \Psi + O(\Psi)\} g(u, v), \end{aligned}$$

where $g(u, v) = \exp\{-u + \Psi \log v + O(\log^2 v)\}$. The sub-intervals we use are $([L/3]/Lg_L^*, 1/g_L^*]$ and $(k/Lg_L^*, (k+1)/Lg_L^*]$ for $0 \leqslant k \leqslant [L/3] - 1$. For the first interval, we have

$$g([L/3], L) = \exp\{-[L/3] + O(\Psi \log_4 x + (\log_4 x)^2)\} \leqslant e^{-L/4}.$$

For the intervals with small $k$ we obtain

$$\sum_{k\leqslant \Psi^2} g(k,k+1) \leqslant \exp\{O(\log^2\Psi)\}\sum_{k=1}^{\infty}\frac{k^\Psi}{e^k}$$
$$= \exp\{\Psi\log\Psi - \Psi + O(\log^2\Psi)\}$$

and for the intervals with large $k$ we have

$$\sum_{k>\Psi^2} g(k,k+1) \leqslant \exp\{-\Psi^2 + O(\Psi\log\Psi)\}.$$

Therefore, $U_1(x) \ll V(x)\exp\{-1.02\varepsilon\log_3 x + O(\Psi)\}$.

By (6.4) and an argument similar to that used to bound $U_1(x)$, we have

$$U_2(x) \ll V(x)\exp\{-1.3\varepsilon\log_3 x + O(\Psi)\}.$$

The first part of the theorem now follows, since $V^*(x) \leqslant U(x) + U_1(x) + U_2(x)$.

For the second part, consider again a totient $m$ not counted in $U(x)$. Then

$$\Omega(m) - \omega(m) \leqslant \sum_{i=0}^{L}\Omega(q_i - 1, 1, S) + \Omega(\phi(r)) \ll (\log_2 x)^{1/2},$$

and the theorem holds with $\Omega(m)$ replaced by $\omega(m)$. □

There is a curious asymmetry between the bounds for $U_1(x)$ and $U_2(x)$, stemming from the asymmetry in the bounds (6.3) and (6.4). This is a real phenomenon, rather than a product of imprecise estimating and is a consequence of an asymmetry in the distribution of the numbers $x_i$ when $\mathbf{x} \in \mathscr{S}_L$ when $i \ll \log L$. The details are found in Lemma 6.2 below.

*Proof of Corollary 11.* It suffices to prove the theorem with $g(m) = \Omega(m)$. Divide the totients $m \leqslant x$ into three sets, $S_1$, those with $\Omega(m) \geqslant 10\log_2 x$, $S_2$, those not in $S_1$ but with $|\Omega(m) - \log_2 x/(1-\varrho)| \geqslant \frac{1}{3}\log_2 x$, and $S_3$, those not counted in $S_1$ or $S_2$. By Lemma 2.2, $|S_1| \ll \frac{x}{\log^2 x}$ and by Theorem 10, $|S_2| \ll V(x)(\log_2 x)^{-1/3}$. Therefore

$$|S_3| = V(x)(1 - O((\log_2 x)^{-1/3})) \tag{6.13}$$

and also

$$\sum_{m\in S_1\cup S_2}\Omega(m) \ll |S_1|\log x + |S_2|\log_2 x \ll V(x)(\log_2 x)^{2/3}. \tag{6.14}$$

For each $m \in S_3$, let

$$\varepsilon_m = \frac{\Omega(m)}{\log_2 x} - \frac{1}{1-\varrho}$$

and for each natural number $N$, let $S_{3,N}$ denote the set of $m \in S_3$ with $N \leqslant |\varepsilon_m|\log_3 x < N+1$. By Theorem 10, (6.13) and (6.14),

$$\sum_{m\in\mathscr{V}(x)}\Omega(m) = O(V(x)\sqrt{\log_2 x}) + \sum_{0\leqslant N\leqslant \frac{1}{2}\log_3 x}\ \sum_{m\in S_{3,N}}\Omega(m)$$
$$= \frac{\log_2 x}{1-\varrho}|S_3| + O\left(V(x)\frac{\log_2 x}{\log_3 x}\sum_N (N+1)e^{-N}\right)$$
$$= \frac{V(x)\log_2 x}{1-\varrho}\left(1 + O\left(\frac{1}{\log_3 x}\right)\right).$$

□

Due to the asymmetry in the estimates (6.3) and (6.4), there is probably a secondary term of order $V(x)\log_2 x/\log_3 x$. We now turn our attention to examining the size of individual prime factors of a pre-image of a normal totient. As before, the key is estimating the volume of the corresponding subset of $\mathscr{S}_L$. Recall definition (3.4) and Lemma 3.7. Define

$$\lambda_i = \varrho^i g_i \quad (i \geqslant 0), \qquad \lambda = \lim_{i\to\infty} \lambda_i = \frac{1}{\varrho F'(\varrho)} < \frac{1}{3}. \tag{6.15}$$

**Lemma 6.2.** *Suppose $i \leqslant L-2$, $0 \leqslant \alpha < 1/g_L^*$ and $\alpha g_{L-i}^* < \beta < 1/g_i$. Define $\theta$ by*

$$\beta = \frac{\varrho^i(1-i/L)}{1+\theta}. \tag{6.16}$$

*If $\theta > 0$, we have*

$$T_L(x_i \leqslant \beta, x_L \geqslant \alpha) \ll T_L \frac{i}{\theta L}\frac{(1+\theta L/i)^i}{(1+\theta)^L} e^{-L\alpha g_L^*}. \tag{6.17}$$

*If $\theta < 0$, we have*

$$T_L(x_i \geqslant \beta, x_L \geqslant \alpha) \ll T_L \frac{(1-\lambda)^i}{(1+\theta)^L}\left(1+\frac{\theta + i\lambda_i/L}{1-\lambda_i}\right)^L e^{-\frac{2}{3}L\alpha g_L^*}. \tag{6.18}$$

*If $-i\lambda_i/L < \theta < 0$ and $\alpha \leqslant -\theta(L-i)/(2ig_L^*)$, then*

$$T_L(x_i \geqslant \beta, x_L \geqslant \alpha) \ll T_L \frac{i}{|\theta'|L}\frac{(1+\theta' L/i)^i}{(1+\theta')^L} e^{-L\alpha g_L^*}, \tag{6.19}$$

*where $\theta' = \theta + i\alpha g_L^*/(L-i)$.*

*Proof.* For each inequality, we show that the region in question lies inside a simplex for which we may apply Lemma 3.6. The volume is then related to $T_L$ via Lemma 3.4.

The basic strategy is similar to the proof of Lemma 6.1. Consider $\mathbf{x} \in \mathscr{S}_L$ with $x_L \geqslant \alpha$ and let $y_j = x_j - \alpha g_{L-j}^*$ for each $j$. Then $\mathbf{v}_j \cdot \mathbf{y} \leqslant 0$ $(1 \leqslant j \leqslant L)$ and $\mathbf{v}_0 \cdot \mathbf{y} \leqslant 1 - \alpha g_L^*$. Let $\alpha' = 1 - \alpha g_L^*$ and $\beta' = \beta - \alpha g_{L-i}^*$. We may assume that $\beta' > 0$, for $x_L \geqslant \alpha$ implies $x_i \geqslant \alpha g_{L-i}^*$ (Lemma 3.10). We now apply a second linear transformation, setting $z_j = y_j - \beta' g_{i-j}$ for $j \leqslant i$ and $z_j = y_j$ for $j > i$. We then have

$$\begin{aligned} &\mathbf{v}_j \cdot \mathbf{z} \leqslant 0 \qquad (1 \leqslant j \leqslant L, j \neq i), \\ &\mathbf{v}_i \cdot \mathbf{z} \leqslant \beta', \\ &\mathbf{v}_0 \cdot \mathbf{z} \leqslant \alpha' - \beta' g_i. \end{aligned} \tag{6.20}$$

With these definitions, $x_i \geqslant \beta$ is equivalent to $z_i \geqslant 0$ and $x_i \leqslant \beta$ is the same as $z_i \leqslant 0$. For any $A$ satisfying

$$A > 0 \quad (\text{when } z_i < 0), \qquad -g_i \leqslant A < 0 \quad (\text{when } z_i \geqslant 0), \tag{6.21}$$

the desired volume is at most the volume of the simplex defined by

$$\begin{gathered} (\mathbf{v}_0 + (g_i + A)\mathbf{v}_i) \cdot \mathbf{z} \leqslant \alpha' + A\beta', \\ \mathbf{v}_j \cdot \mathbf{z} \leqslant 0, \\ \pm \mathbf{e}_i \cdot \mathbf{z} \leqslant 0. \end{gathered} \tag{6.22}$$

By (6.20),

$$(\mathbf{v}_0 + (g_i + A)\mathbf{v}_i) + \sum_{j<i} g_j \mathbf{v}_j + A\mathbf{e}_i + \sum_{j=i+1}^{L-1} (g_j + Ag_{j-i})\mathbf{v}_j + (g_L^* + Ag_{L-i}^*)\mathbf{v}_L = \mathbf{0}. \tag{6.23}$$

By (6.21), each vector on the left of (6.22) has a positive coefficient in the identity (6.23). Therefore, by (3.10) and Lemma 3.6,

$$\begin{aligned} T_L(x_i \lesseqgtr \beta, x_L \geqslant \alpha) &\leqslant T_L^* \frac{g_i}{|A|}(\alpha' + A\beta')^L \prod_{j=i+1}^{L-1} \left(1 + A\frac{g_{j-i}}{g_j}\right)^{-1} \left(1 + A\frac{g^*_{L-i}}{g^*_L}\right)^{-1} \\ &\ll T_L \frac{g_i}{|A|} \frac{(\alpha' + A\beta')^L}{(1 + A\varrho^i)^{L-i}}. \end{aligned} \tag{6.24}$$

The last inequality follows from Lemma 3.7 and (6.21), which give $Ag_{j-i}/g_j \geqslant -\frac{1}{3}$ and

$$1 + A\frac{g_{j-i}}{g_j} = (1 + A\varrho^i)(1 + O(\varrho^{j-i})).$$

Define $\eta$ by

$$\frac{\beta'}{\alpha'} = \frac{\varrho^i(1 - i/L)}{1 + \eta}. \tag{6.25}$$

We work on (6.17) first. Assume that $x_i \leqslant \beta$ and $\theta > 0$. A convenient choice for $A$ (which is optimal if the term $1/|A|$ in (6.24) is ignored) is

$$A = -\frac{L}{i\varrho^i} + \frac{(L-i)\alpha'}{i\beta'} = \frac{\eta L}{i\varrho^i}. \tag{6.26}$$

From (6.16) and (6.25),

$$1 + \eta = \frac{\beta\alpha'(1+\theta)}{\beta'} = (1+\theta)\frac{1 - \alpha g^*_L}{1 - \alpha g^*_{L-i}/\beta}. \tag{6.27}$$

By Lemma 3.7, for $L$ sufficiently large and $1 \leqslant i < L$ we have

$$g^*_{L-i}/\beta \geqslant \varrho^{-i}\frac{L}{L-i} g^*_{L-i} > g^*_L,$$

from which it follows that $\eta > \theta > 0$. Therefore, $A > 0$ and (6.24) gives

$$T_L(x_i \leqslant \beta, x_L \geqslant \alpha) \ll T_L \frac{i}{\eta L}\left(\frac{\alpha'}{1+\eta}\right)^L (1 + \eta L/i)^i, \tag{6.28}$$

which implies (6.17), since the right side of (6.28) is a decreasing function of $\eta$. We do give up some accuracy with this replacement, but this turns is negligible in applications, since $x_L \approx 1/(Lg^*_L)$ for most of $\mathscr{S}_L$.

Now assume $x_i \geqslant \beta$ and $\theta < 0$. When $\beta \geqslant \varrho^i$, the inequality $\mathbf{v}_i \cdot \mathbf{x} \leqslant 0$ is superfluous by Lemma 3.10, so nothing is lost by ignoring this inequality. In any case, taking $A = -g_i$ in (6.24) gives

$$T_L(x_i \geqslant \beta, x_L \geqslant \alpha) \ll T_L \frac{(1 - \alpha g^*_L - g_i(\beta - \alpha g^*_{L-i}))^L}{(1 - \varrho^i g_i)^{L-i}}.$$

Now from Lemma 3.7,

$$\frac{g_i g^*_{L-i}}{g^*_L} < 1/3,$$

so

$$1 - \alpha g^*_L - g_i(\beta - \alpha g^*_{L-i}) < (1 - \beta g_i)(1 - (2/3)\alpha g^*_L)$$

and (6.18) follows.

In many instances we can do better with another choice of $A$. To prove (6.19), assume $-i\lambda_i/L < \theta < 0$ and $\alpha \leqslant -\theta(L-i)/(2ig^*_L)$. Also define $A$ as in (6.26). From (6.27) and Lemma 3.7,

$$\begin{aligned} \eta &< (1+\theta)(1 - \alpha g^*_L) - 1 + \alpha g^*_L(1+\theta)L/(L-i) \\ &< \theta + i\alpha g^*_L/(L-i) = \theta'. \end{aligned}$$

By the restrictions on $\alpha$ and $\theta$, we have

$$-i\lambda_i/L < \theta < \eta < \theta' \leqslant \theta/2 < 0,$$

and therefore (6.21) holds. Inequality (6.24) now gives

$$T_L(x_i \geqslant \beta, x_L \geqslant \alpha) \ll T_L \frac{i}{|\eta| L} \left( \frac{\alpha'}{1+\eta} \right)^L (1+\eta L/i)^i.$$

Since the right side is increasing with $\eta$, we may replace $\eta$ with $\theta'$, which gives (6.19). □

We now apply Lemma 6.2 to the problem of determining the size of $q_i(n)$ when $n$ is a pre-image of a "normal" totient. For convenience, we define $V(x; \mathscr{C})$ to be the number of totients $m \leqslant x$ with a pre-image $n$ satisfying condition $\mathscr{C}$.

**Lemma 6.3.** *Suppose $x$ is sufficiently large, $L_0 = L_0(x)$ and $1 \leqslant i \leqslant L_0 - 2$. Write*

$$\beta = \frac{\varrho^i (1 - i/L_0)}{1+\theta}. \tag{6.29}$$

*If $0 \leqslant \Psi \leqslant (\log_3 x)^{1/2}$, $i \leqslant L_0 - \Psi - 2$ and $\frac{(3/2)i\Psi}{(L_0-\Psi)(L_0-i)} < \theta \leqslant \frac{1}{2}$, then*

$$V(x; q_i(n) \leqslant \beta \log_2 x) \ll V(x) \left( e^{-\Psi^2/4C} + \frac{i}{\eta L} \frac{(1+\eta L/i)^i}{(1+\eta)^L} e^{O(\Psi)} \right), \tag{6.30}$$

*where $\eta = \theta - \frac{(3/2)i\Psi}{(L_0-\Psi)(L_0-i)}$ and $L = L_0 - \Psi$. If $-\frac{1}{2} \leqslant \theta + \frac{1}{5(L_0-i)^2} < 0$, then*

$$V(x; q_i(n) \geqslant \beta \log_2 x) \ll V(x) \exp\left\{ \frac{\lambda_i L_0 \theta}{1-\lambda_i} + K_1 i + O\left( L_0\theta^2 + \sqrt{-L_0\theta} \log(-L_0\theta) \right) \right\}. \tag{6.31}$$

*where $K_1 = \frac{\lambda}{1-\lambda} + \log(1-\lambda) = 0.08734\ldots$. If $\frac{-i\lambda_i}{L_0} \leqslant \theta < 0$ then*

$$V(x; q_i(n) \geqslant \beta \log_2 x) \ll V(x) \exp\left\{ -\frac{1}{2} \frac{L_0(L_0-i)}{i} \theta^2 + O\left( |\theta| \sqrt{\frac{L_0(L_0-i)}{i}} \log(L_0 - i) \right) \right\}. \tag{6.32}$$

*Proof.* Set $L = L_0 - \Psi$ and define $\xi_i$ as in Theorem 15. By Theorem 15, the number of totients $m \leqslant x$ with a pre-image $n$ satisfying $(x_1(n;x), \ldots, x_L(n;x)) \notin \mathscr{S}_L(\boldsymbol{\xi})$ is

$$\ll V(x) \exp\{-\Psi^2/4C\}. \tag{6.33}$$

For remaining totients, let $x_i = x_i(n;x)$ and set $y_i = (\xi_0 \cdots \xi_{i-1})^{-1} x_i$ for each $i$, so that $\mathbf{y} \in \mathscr{S}_L$.

First, assume $\Psi, i$ and $\theta$ satisfy the hypotheses of (6.30). Then $y_i \leqslant \beta$. To set up an application of Lemma 6.2 (6.17), define $\theta'$ by

$$\frac{\varrho^i (1 - i/L)}{1+\theta'} = \beta.$$

Some algebraic manipulation gives

$$\theta' \geqslant \eta, \qquad \eta = \theta - \frac{(3/2)i\Psi}{L(L_0 - i)}.$$

By the hypothesis on $\theta$, $\eta > 0$. For each natural number $k$, we consider separately totients with $y_L$ satisfying

$$\frac{k}{Lg_L^*} \leqslant y_L \leqslant \frac{(k+1)}{Lg_L^*}. \tag{6.34}$$

Arguing as in the proof of Theorem 10, by Lemmas 3.1, 4.3 and 6.2, we have

$$V(x; q_i(n) \leqslant \beta \log_2 x) \ll \frac{x}{\log x} T_L (\log_2 x)^L \frac{i}{\eta L} \frac{(1+\eta L/i)^i}{(1+\eta)^L} \sum_{k \geqslant 0} e^{-k} W(Z_k) \log_2 Z_k,$$

where $\log_2 Z_k = (k+1)(\log_2 x)/(Lg_L^*)$. Estimating the sum on $k$ as in the proof of Theorem 10 and applying Corollary 3.5 gives (6.30).

Now assume that $-\frac{1}{2} \leqslant \theta + \frac{1}{5(L_0-i)^2} \leqslant -\frac{i\lambda_i}{L_0}$. Then $y_i \geqslant (\xi_0 \cdots \xi_{i-1})^{-1}\beta$. From (6.1) and an argument similar to the case $\theta > 0$, we have

$$y_i \geqslant \frac{\varrho^i(1-i/L)}{1+\eta}, \qquad \eta = \theta + \frac{1}{5(L_0-i)^2}.$$

Take $\Psi = [(2CL_0|\eta|)^{1/2}] + 1$. Using Lemma 6.2 (6.18) and the same subdivision of $y_L$, (6.31) follows in the same manner as the proof of (6.30). A couple of inequalities used are

$$\frac{i\lambda_i}{1-\lambda_i} = \frac{i\lambda}{1-\lambda} + O(1)$$

and

$$\left(1 + \frac{\eta + i\lambda_i/L}{1-\lambda_i}\right)^L \leqslant \exp\left\{\frac{L\eta + i\lambda_i}{1-\lambda_i}\right\},$$

the former being a consequence of Lemma 3.7. The factor $2/3$ occurring in (6.18) accounts for the factor $e^{O(\Psi \log \Psi)}$ instead of the factor $e^{O(\Psi)}$.

For the last inequality, assume $-\frac{i\lambda_i}{L_0} \leqslant \theta < 0$. We may also assume that

$$\theta \leqslant -3\left(\frac{i}{L_0(L_0-i)}\right)^{1/2}, \tag{6.35}$$

for otherwise (6.32) is trivial. Define

$$\eta = \theta + \frac{1}{5(L_0-i)^2} \tag{6.36}$$

and

$$\Psi = \left[|\eta|\sqrt{\frac{2CL_0(L_0-i)}{i}}\right]. \tag{6.37}$$

Set $L = L_0 - \Psi$. By (6.35) and (6.36),

$$-\frac{i\lambda_i}{L} < -\frac{i\lambda_i}{L_0} \leqslant \eta < -2\left(\frac{i}{L_0(L_0-i)}\right)^{1/2}. \tag{6.38}$$

Therefore, for $x$ large,

$$2 \leqslant \Psi \leqslant \frac{1}{2}\sqrt{\frac{i(L_0-i)}{L_0}} \leqslant \frac{1}{2}\sqrt{L_0-i}. \tag{6.39}$$

In particular, (6.39) implies $L_0 - i \geqslant 16$ and thus $\Psi \leqslant \frac{1}{8}(L_0-i)$. Consider now a totient $m \leqslant x$, all of whose pre-images $n$ satisfy $(x_1, \dots, x_L) \in \mathscr{S}_L(\boldsymbol{\xi})$. As in the proof of (6.31), we have

$$y_i \geqslant \frac{\varrho^i(1-i/L)}{1+\eta}. \tag{6.40}$$

As before, we group such $n$ as to the size of $y_L$ (see (6.34)). However, the hypotheses of Lemma 6.2 require that

$$k \leqslant \frac{|\eta|}{2}\frac{L(L-i)}{i} =: k_0$$

in order to use (6.19). By (6.37),

$$k_0 < \frac{2C|\eta|L_0(L_0-i)}{4Ci} \leqslant \frac{\Psi^2}{4C}.$$

For $k < \Psi^2/4C$, we shall use the bound from Lemma 3.11 in place of (6.19). By previous estimates for the sum over $k$, the number of $n$ counted with $k > \Psi^2/4C$ is

$$\ll V(x)\exp\{-\Psi^2/4C + O(\Psi\log\Psi)\}. \tag{6.41}$$

For $k \leqslant \Psi^2/4C$, define $\eta_k = \eta + \frac{ki}{L(L-i)}$. By (6.37) and (6.39),

$$\eta_k \leqslant \eta + \frac{|\eta|}{2}\frac{L_0(L_0-i)}{i}\frac{i}{L(L-i)} < \eta/3 < 0.$$

Hence, by Lemmas 3.1, 4.3 and 6.2 (6.19), the number of $n$ with $k \leqslant \Psi^2/4C$ is

$$\ll \frac{x}{\log x}T_L(\log_2 x)^L \frac{i}{\eta L}\sum_{k\leqslant\Psi^2}\frac{(1+\eta_k L/i)^i}{(1+\eta_k)^L}e^{-k}W(Z_k)\log_2 Z_k, \tag{6.42}$$

where again $\log_2 Z_k = (k+1)(\log_2 x)/(Lg_L^*)$. To estimate the sum on $k$, we use

$$\frac{(1+\eta_k L/i)^i}{(1+\eta_k)^L}e^{-k} \leqslant \exp\left\{-\frac{1}{2}\frac{L(L-i)}{i}\eta^2 - (1+\eta)k\right\}$$

and

$$\sum_{k=0}^{\infty}\frac{k^\Psi}{e^{(1+\eta)k}} \ll \Gamma\left(\frac{\Psi}{1+\eta}\right) \ll \exp\{(1-2\eta)\Psi\log\Psi\}.$$

Combining (6.33), (6.41) and (6.42) together with the bounds (6.35)–(6.39) readily gives (6.32). □

*Proof of Theorem 12.* For the first part, we may assume $\varepsilon > 2(\frac{L_0(L_0-i)}{i})^{-1/2}$, for otherwise the result is trivial. Let $\beta = \frac{\beta_i}{1+\varepsilon} > \beta_i(1-\varepsilon)$ and take $\Psi = [\varepsilon\sqrt{L_0(L_0-i)/i}]$. Note that the bounds on $\varepsilon$ give

$$2 \leqslant \Psi < \frac{1}{6}(L_0-i) \leqslant L_0 - i - 2$$

and

$$\eta := \varepsilon - \frac{(3/2)i\Psi}{(L_0-\Psi)(L_0-i)} \geqslant \varepsilon\left(1 - 2\sqrt{\frac{i}{L_0(L_0-i)}}\right) > 0.$$

An application of Lemma 6.3 (6.30) now gives the first part of the theorem in the case $q_i(n)/\log_2 x \leqslant \beta_i(1-\varepsilon)$. Now set $\beta = \beta_i(1+\varepsilon) = \frac{\beta_i}{1+\theta}$, so that

$$-\frac{i\lambda_i}{L_0} \leqslant -\varepsilon \leqslant \theta \leqslant -\varepsilon + \varepsilon^2.$$

An application of Lemma 6.3 (6.32) completes the first part of the theorem.

The second part follows in a similar manner, except that now the bound for $V(x; \frac{q_i(n)}{\log_2 x} \geqslant \beta_i(1+\varepsilon))$ will be weaker than the bound for $V(x; \frac{q_i(n)}{\log_2 x} \leqslant \beta_i(1-\varepsilon))$. For the former, Lemma 6.3 (6.31) gives the bound

$$V(x)\exp\left\{-\frac{\lambda_i}{1-\lambda_i}L_0\varepsilon + K_1 i + O\left(L_0\varepsilon^2 + (L_0\varepsilon)^{1/2}\log(L_0\varepsilon)\right)\right\},$$

while for the latter, using $\Psi = [\sqrt{4CL_0\varepsilon}]$ in (6.30) we obtain the upper bound

$$V(x)\exp\{-L_0\varepsilon + i\log(1+\varepsilon L_0/i) + O(L_0\varepsilon^2)\}.$$

□

*Proof of Theorem 13.* Assume first that $g(x) \geqslant G_0$, some large absolute constant, for otherwise the conclusion is trivial. for each $i$ set

$$\varepsilon_i = \sqrt{\frac{i}{L_0(L_0-i)}}\log(L_0-i)\log g(x).$$

When $1 \leqslant i \leqslant \log^3 L_0$, the second part of Theorem 12 gives

$$V\left(x; \frac{q_i(n)}{\beta_i \log_2 x} \leqslant 1-\varepsilon_i\right) \ll V(x)\exp\left\{-K_2\sqrt{i}\log_4 x \log g(x) + O(i^{1/4}\sqrt{\log_4 x \log g(x)})\right\},$$

where $K_2 = \lambda_1/(1-\lambda_1) \geqslant 0.265$. Now suppose $\log^3 L_0 \leqslant i \leqslant L_0 - g(x)$ and set $H = L_0 - i$. The first part of Theorem 12 gives

$$\begin{aligned} V\left(x; \frac{q_i(n)}{\beta_i \log_2 x} \geqslant 1+\varepsilon_i\right) &\ll V(x)\exp\left\{-\frac{1}{2}\log^2 H \log^2 g(x)\left(1+O\left(\frac{1}{\log g(x)}\right)\right)\right\} \\ &\ll V(x)\exp\{-(1/3)\log^2 H \log^2 g(x)\} \end{aligned}$$

for $x$ sufficiently large. Since $H \geqslant g(x)$, summing on $i$ gives the desired upper bound on totients with an $n$ not satisfying (1.9). Note that with $n$ satisfying (1.9), we have $\Omega(n) \geqslant \omega(n) \geqslant L_0(x) - g(x)$. By Theorem 15, the number of totients $m \leqslant x$ with $\mathbf{x} \notin \mathscr{S}_L(\boldsymbol{\xi})$ is $O(V(x)e^{-\Psi^2/4C})$, where $L = L_0 - \Psi$, $\Psi = \log g(x)$. Suppose $\mathbf{x}(n) \in \mathscr{S}_L(\boldsymbol{\xi})$, $n = q_0 \cdots q_L r$ and $\Omega(r) > g(x)$. By Lemma 3.10,

$$\log_2 P^+(r) \leqslant \log_2 q_L \leqslant 3\Psi\varrho^{-\Psi} \leqslant 3(\log g(x))g(x)^{1/2C} \leqslant \frac{g(x)}{100}$$

if $x$ is large. By Lemmas 2.2 and 2.3, $\sum 1/\phi(r) \ll 1$. By Lemma 3.12 and Corollary 3.5, the number of totients $\leqslant x$ with such a pre-image $n$ is

$$\ll \frac{x}{\log x}\sum \frac{1}{\phi(q_0\cdots q_L r)} \ll V(x)e^{-\Psi^2/4C} \ll V(x)e^{-\frac{1}{4C}\log^2 g(x)}.$$

□

As a final remark, it is trivial that Theorems 12–10 and Corollary 11 hold with $V(x)$ and $V(x;\mathscr{C})$ replaced by the corresponding functions $V_k(x)$ and $V_k(x;\mathscr{C})$ (here the implied constants depend on $k$).

## 7. Conjectures of Sierpiński and Carmichael

### 7.1. **Sierpiński's Conjecture.**

**Conjecture 1** (Prime $k$-tuples Conjecture (Dickson [7])). *Suppose $a_1, a_2, \ldots a_k$ are positive integers and $b_1, b_2, \ldots, b_k$ are integers so that no prime $p$ divides $(a_1 n + b_1)\cdots(a_k n + b_k)$ for every integer $n$. Then the numbers $a_1 n + b_1, \ldots, a_k n + b_k$ are simultaneously prime for infinitely many $n$.*

Schinzel's argument deducing Sierpiński's Conjecture from Hypothesis H requires $\gg k$ polynomials of degrees up to $k$ to be simultaneously prime to show the existence of a number with multiplicity $k$. Below we follow a completely different approach, which is considerably simpler and requires only the simultaneous primality of three linear polynomials (the Prime 3-tuples Conjecture). The idea is to take a number $m$ with multiplicity $k$ and construct a multiple of it with multiplicity $k+2$. This is motivated by the technique used in section 5 where many numbers with multiplicity $\kappa$ are constructed from a single example.

**Lemma 7.1.** *Suppose $A(m) = k$ and $p$ is a prime satisfying*

(1) *(i) $p > 2m+1$,*
(2) *(ii) $2p+1$ and $2mp+1$ are prime,*
(3) *(iii) $dp+1$ is composite for all $d|2m$ except $d=2$ and $d=2m$.*

*Then $A(2mp) = k+2$.*

*Proof.* Suppose $\phi^{-1}(m) = \{x_1, \ldots, x_k\}$ and $\phi(x) = 2mp$. Condition (i) implies $p \nmid x$, hence $p|(q-1)$ for some prime $q$ dividing $x$. Since $(q-1)|2mp$, we have $q = dp+1$ for some divisor $d$ of $2m$. We have $q > 2p$, so $q^2 \nmid x$ and $\phi(x) = (q-1)\phi(x/q)$. By conditions (ii) and (iii), either $q = 2p+1$ or $q = 2mp+1$. In the former case, $\phi(x/q) = m$, which has solutions $x = (2p+1)x_i$ $(1 \leqslant i \leqslant k)$. In the latter case, $\phi(x/q) = 1$, which has solutions $x = q$ and $x = 2q$. □

Now suppose $A(m) = k$, $m \equiv 1 \pmod 3$, and let $d_1, \dots, d_j$ be the divisors of $2m$ with $3 \leqslant d_i < 2m$. Let $p_1, \dots, p_j$ be distinct primes satisfying $p_i > d_i$ for each $i$. Using the Chinese Remainder Theorem, let $a \mod b$ denote the intersection of the residue classes $-d_i^{-1} \mod p_i$ $(1 \leqslant i \leqslant j)$. Then for every $h$ and $i$, $(a + bh)d_i + 1$ is divisible by $p_i$, hence composite for large enough $h$. The Prime $k$-tuples Conjecture implies that there are infinitely many numbers $h$ so that $p = a + hb$, $2p + 1$ and $2mp + 1$ are simultaneously prime. By Lemma 7.1, $A(2mp) = k + 2$ for each prime $p$. Observe that $p \equiv 2 \pmod 3$ implies $2mp \equiv 1 \pmod 3$. Starting with $A(1) = 2$, $A(2) = 3$, and $A(220) = 5$, Sierpiński's Conjecture follows by induction on $k$.

Table 2 of [33] lists the smallest $m$ for which $A(m) = k$ for $2 \leqslant k \leqslant 100$. In all cases, $m$ is less than 100,000. A modest computer search revealed that for each $k$, $2 \leqslant k \leqslant 1000$, there is an $m < 23,000,000$ with $A(m) = k$. The smallest of these values (denoted $m_k$) are listed in Tables 7.1 and 7.2.

### 7.2. **Carmichael's Conjecture.**

The basis for computations of lower bounds for a counterexample to Carmichael's Conjecture is the following Lemma of Carmichael [5], as refined by Klee [23]. For short, let $s(n) = \prod_{p|n} p$ denote the square-free kernel of $n$.

**Lemma 7.2.** *Suppose $\phi(x) = m$ and $A(m) = 1$. If $d|x$, $e|\frac{x/d}{s(x/d)}$ and $P = 1 + e\phi(d)$ is prime, then $P^2|x$.*

From Lemma 7.2 it is easy to deduce $2^2 3^2 7^2 43^2 | x$. Here, following Carmichael, we break into two cases: (I) $3^2 \parallel x$ and (II) $3^3 | x$. In case (I) it is easy to show that $13^2|x$. From this point onward Lemma 7.2 is used to generate a virtually unlimited set of primes $P$ for which $P^2|x$. In case (I) we search for $P$ using $d = 1, e = 6k$ or $d = 9, e = 2k$, where $k$ is a product of distinct primes (other than 2 or 3) whose squares we already know divide $x$. That is, if $6k + 1$ or $12k + 1$ is prime, its square divides $x$. In case (II) we try $d = 9, e = 2k$ and $d = 27, e = k$, i.e. we test whether or not $6k + 1$ and $18k + 1$ are primes.

As in [33], certifying that a number $P$ is prime is accomplished with the following lemma of Lucas, Lehmer, Brillhart and Selfridge.

**Lemma 7.3.** *Suppose, for each prime $q$ dividing $n - 1$, there is a number $a_q$ satisfying $a_q^{n-1} \equiv 1$ and $a_q^{(n-1)/q} \not\equiv 1 \pmod n$. Then $n$ is prime.*

The advantage of using Lemma 7.3 in our situation is that for a given $P$ we are testing, we already know the prime factors of $P - 1$ (i.e. 2,3 and the prime factors of $k$).

Our overall search strategy differs from [33]. In each case, we first find a set of 32 "small" primes $P$ (from here on, $P$ will represent a prime generated from Lemma 7.2 for which $P^2|x$, other than 2 or 3). Applying Lemma 7.2, taking $k$ to be all possible products of 1,2,3 or 4 of these 32 primes yields a set $S$ of 1000 primes $P$, which we order $p_1 < \cdots < p_{1000}$. This set will be our base set. In particular, $p_{1000} = 796486033533776413$ in case (I) and $p_{1000} = 7839942895076974350 7519$ in case (II). The calculations are then divided into "runs". For run #0, we take for $k$ all possible combinations of 1,2 or 3 of the primes in $S$. For $j \geqslant 1$, run #$j$ tests every $k$ which is the product of $p_j$ and three larger primes in $S$. Each candidate $P$ is first tested for divisibility by small primes and must pass the strong pseudoprime test with bases 2,3,5,7,11 and 13 before attempting to certify that it is prime.

There are two advantages to this approach. First, the candidates $P$ are relatively small (the numbers tested in case (I) had an average of 40 digits and the numbers tested in case (II) had an average of 52 digits). Second, $P - 1$ has at most 6 prime factors, simplifying the certification process.

To achieve $\prod P^2 > 10^{10^{10}}$, 13 runs (run #0 to run #12) were required in case (I) and 14 runs were required in case (II). Together these runs give Theorem 6. A total of 126,520,174 primes were found in case (I), and 104,942,148 primes were found in case (II). The computer program was written in GNU C, utilizing Arjen Lenstra's Large Integer Package. Hardware consisted of a network of 200MHz Pentium PCs running LINUX O/S. Each processor was given one "run" (with up to 14 runs executing concurrently) and the total CPU time used for all 27 runs was 4,765 hours.

| $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 77 | 9072 | 152 | 10080 | 227 | 26880 | 302 | 218880 | 377 | 165888 | 452 | 990720 | 527 | 2677248 |
| 3 | 2 | 78 | 38640 | 153 | 13824 | 228 | 323136 | 303 | 509184 | 378 | 436800 | 453 | 237600 | 528 | 5634720 |
| 4 | 4 | 79 | 9360 | 154 | 23760 | 229 | 56160 | 304 | 860544 | 379 | 982080 | 454 | 69120 | 529 | 411840 |
| 5 | 8 | 80 | 81216 | 155 | 13440 | 230 | 137088 | 305 | 46080 | 380 | 324000 | 455 | 384000 | 530 | 2948400 |
| 6 | 12 | 81 | 4032 | 156 | 54720 | 231 | 73920 | 306 | 67200 | 381 | 307200 | 456 | 338688 | 531 | 972000 |
| 7 | 32 | 82 | 5280 | 157 | 47040 | 232 | 165600 | 307 | 133056 | 382 | 496800 | 457 | 741888 | 532 | 2813184 |
| 8 | 36 | 83 | 4800 | 158 | 16128 | 233 | 184800 | 308 | 82944 | 383 | 528768 | 458 | 86400 | 533 | 3975552 |
| 9 | 40 | 84 | 4608 | 159 | 48960 | 234 | 267840 | 309 | 114048 | 384 | 1114560 | 459 | 1575936 | 534 | 368640 |
| 10 | 24 | 85 | 16896 | 160 | 139392 | 235 | 99840 | 310 | 48384 | 385 | 1609920 | 460 | 248832 | 535 | 529920 |
| 11 | 48 | 86 | 3456 | 161 | 44352 | 236 | 174240 | 311 | 43200 | 386 | 485760 | 461 | 151200 | 536 | 2036736 |
| 12 | 160 | 87 | 3840 | 162 | 25344 | 237 | 104832 | 312 | 1111968 | 387 | 1420800 | 462 | 1176000 | 537 | 751680 |
| 13 | 396 | 88 | 10800 | 163 | 68544 | 238 | 23040 | 313 | 1282176 | 388 | 864864 | 463 | 100800 | 538 | 233280 |
| 14 | 2268 | 89 | 9504 | 164 | 55440 | 239 | 292320 | 314 | 239616 | 389 | 959616 | 464 | 601344 | 539 | 463680 |
| 15 | 704 | 90 | 18000 | 165 | 21120 | 240 | 93600 | 315 | 1135680 | 390 | 1085760 | 465 | 216000 | 540 | 2042880 |
| 16 | 312 | 91 | 23520 | 166 | 46656 | 241 | 93312 | 316 | 274560 | 391 | 264960 | 466 | 331776 | 541 | 3018240 |
| 17 | 72 | 92 | 39936 | 167 | 15840 | 242 | 900000 | 317 | 417600 | 392 | 470016 | 467 | 337920 | 542 | 2311680 |
| 18 | 336 | 93 | 5040 | 168 | 266400 | 243 | 31680 | 318 | 441600 | 393 | 400896 | 468 | 95040 | 543 | 1368000 |
| 19 | 216 | 94 | 26208 | 169 | 92736 | 244 | 20160 | 319 | 131040 | 394 | 211200 | 469 | 373248 | 544 | 3120768 |
| 20 | 936 | 95 | 27360 | 170 | 130560 | 245 | 62208 | 320 | 168480 | 395 | 404352 | 470 | 559872 | 545 | 1723680 |
| 21 | 144 | 96 | 6480 | 171 | 88128 | 246 | 37440 | 321 | 153600 | 396 | 77760 | 471 | 228096 | 546 | 1624320 |
| 22 | 624 | 97 | 9216 | 172 | 123552 | 247 | 17280 | 322 | 168000 | 397 | 112320 | 472 | 419328 | 547 | 262080 |
| 23 | 1056 | 98 | 2880 | 173 | 20736 | 248 | 119808 | 323 | 574080 | 398 | 1148160 | 473 | 762048 | 548 | 696960 |
| 24 | 1760 | 99 | 26496 | 174 | 14400 | 249 | 364800 | 324 | 430560 | 399 | 51840 | 474 | 342720 | 549 | 1889280 |
| 25 | 360 | 100 | 34272 | 175 | 12960 | 250 | 79200 | 325 | 202752 | 400 | 152064 | 475 | 918720 | 550 | 734400 |
| 26 | 2560 | 101 | 23328 | 176 | 8640 | 251 | 676800 | 326 | 707616 | 401 | 538560 | 476 | 917280 | 551 | 842400 |
| 27 | 384 | 102 | 28080 | 177 | 270336 | 252 | 378000 | 327 | 611520 | 402 | 252000 | 477 | 336000 | 552 | 874368 |
| 28 | 288 | 103 | 7680 | 178 | 11520 | 253 | 898128 | 328 | 317952 | 403 | 269568 | 478 | 547200 | 553 | 971520 |
| 29 | 1320 | 104 | 29568 | 179 | 61440 | 254 | 105600 | 329 | 624960 | 404 | 763776 | 479 | 548352 | 554 | 675840 |
| 30 | 3696 | 105 | 91872 | 180 | 83520 | 255 | 257040 | 330 | 116640 | 405 | 405504 | 480 | 129600 | 555 | 4306176 |
| 31 | 240 | 106 | 59040 | 181 | 114240 | 256 | 97920 | 331 | 34560 | 406 | 96768 | 481 | 701568 | 556 | 1203840 |
| 32 | 768 | 107 | 53280 | 182 | 54432 | 257 | 176256 | 332 | 912000 | 407 | 1504800 | 482 | 115200 | 557 | 668160 |
| 33 | 9000 | 108 | 82560 | 183 | 85536 | 258 | 264384 | 333 | 72576 | 408 | 476928 | 483 | 1980000 | 558 | 103680 |
| 34 | 432 | 109 | 12480 | 184 | 172224 | 259 | 244800 | 334 | 480000 | 409 | 944640 | 484 | 1291680 | 559 | 2611200 |
| 35 | 7128 | 110 | 26400 | 185 | 136800 | 260 | 235872 | 335 | 110880 | 410 | 743040 | 485 | 1199520 | 560 | 820800 |
| 36 | 4200 | 111 | 83160 | 186 | 44928 | 261 | 577920 | 336 | 1259712 | 411 | 144000 | 486 | 556416 | 561 | 663552 |
| 37 | 480 | 112 | 10560 | 187 | 27648 | 262 | 99360 | 337 | 1350720 | 412 | 528000 | 487 | 359424 | 562 | 282240 |
| 38 | 576 | 113 | 29376 | 188 | 182400 | 263 | 64800 | 338 | 250560 | 413 | 1155840 | 488 | 1378080 | 563 | 3538944 |
| 39 | 1296 | 114 | 6720 | 189 | 139104 | 264 | 136080 | 339 | 124416 | 414 | 4093440 | 489 | 2088000 | 564 | 861120 |
| 40 | 1200 | 115 | 31200 | 190 | 48000 | 265 | 213120 | 340 | 828000 | 415 | 134400 | 490 | 399168 | 565 | 221760 |
| 41 | 15936 | 116 | 7200 | 191 | 102816 | 266 | 459360 | 341 | 408240 | 416 | 258048 | 491 | 145152 | 566 | 768000 |
| 42 | 3312 | 117 | 8064 | 192 | 33600 | 267 | 381024 | 342 | 74880 | 417 | 925344 | 492 | 2841600 | 567 | 2790720 |
| 43 | 3072 | 118 | 54000 | 193 | 288288 | 268 | 89856 | 343 | 1205280 | 418 | 211680 | 493 | 1622880 | 568 | 953856 |
| 44 | 3240 | 119 | 6912 | 194 | 286848 | 269 | 101376 | 344 | 192000 | 419 | 489600 | 494 | 1249920 | 569 | 7138368 |
| 45 | 864 | 120 | 43680 | 195 | 59904 | 270 | 347760 | 345 | 370944 | 420 | 1879200 | 495 | 2152800 | 570 | 655200 |
| 46 | 3120 | 121 | 32400 | 196 | 118800 | 271 | 124800 | 346 | 57600 | 421 | 1756800 | 496 | 2455488 | 571 | 3395520 |
| 47 | 7344 | 122 | 153120 | 197 | 100224 | 272 | 110592 | 347 | 1181952 | 422 | 90720 | 497 | 499200 | 572 | 3215520 |
| 48 | 3888 | 123 | 225280 | 198 | 176400 | 273 | 171360 | 348 | 1932000 | 423 | 376320 | 498 | 834624 | 573 | 2605824 |
| 49 | 720 | 124 | 9600 | 199 | 73440 | 274 | 510720 | 349 | 1782000 | 424 | 1461600 | 499 | 1254528 | 574 | 1057536 |
| 50 | 1680 | 125 | 15552 | 200 | 174960 | 275 | 235200 | 350 | 734976 | 425 | 349920 | 500 | 2363904 | 575 | 1884960 |
| 51 | 4992 | 126 | 4320 | 201 | 494592 | 276 | 25920 | 351 | 473088 | 426 | 158400 | 501 | 583200 | 576 | 3210240 |
| 52 | 17640 | 127 | 91200 | 202 | 38400 | 277 | 96000 | 352 | 467712 | 427 | 513216 | 502 | 1029600 | 577 | 1159200 |
| 53 | 2016 | 128 | 68640 | 203 | 133632 | 278 | 464640 | 353 | 556800 | 428 | 715392 | 503 | 2519424 | 578 | 4449600 |
| 54 | 1152 | 129 | 5760 | 204 | 38016 | 279 | 200448 | 354 | 2153088 | 429 | 876960 | 504 | 852480 | 579 | 272160 |
| 55 | 6000 | 130 | 49680 | 205 | 50688 | 280 | 50400 | 355 | 195840 | 430 | 618240 | 505 | 1071360 | 580 | 913920 |
| 56 | 12288 | 131 | 159744 | 206 | 71280 | 281 | 30240 | 356 | 249600 | 431 | 772800 | 506 | 3961440 | 581 | 393120 |
| 57 | 4752 | 132 | 16800 | 207 | 36288 | 282 | 157248 | 357 | 274176 | 432 | 198720 | 507 | 293760 | 582 | 698880 |
| 58 | 2688 | 133 | 19008 | 208 | 540672 | 283 | 277200 | 358 | 767232 | 433 | 369600 | 508 | 1065600 | 583 | 2442240 |
| 59 | 3024 | 134 | 24000 | 209 | 112896 | 284 | 228480 | 359 | 40320 | 434 | 584640 | 509 | 516096 | 584 | 6914880 |
| 60 | 13680 | 135 | 24960 | 210 | 261120 | 285 | 357696 | 360 | 733824 | 435 | 708480 | 510 | 616896 | 585 | 695520 |
| 61 | 9984 | 136 | 122400 | 211 | 24192 | 286 | 199584 | 361 | 576576 | 436 | 522720 | 511 | 639360 | 586 | 497664 |
| 62 | 1728 | 137 | 22464 | 212 | 57024 | 287 | 350784 | 362 | 280800 | 437 | 884736 | 512 | 4014720 | 587 | 808704 |
| 63 | 1920 | 138 | 87120 | 213 | 32256 | 288 | 134784 | 363 | 63360 | 438 | 1421280 | 513 | 266112 | 588 | 2146176 |
| 64 | 2400 | 139 | 228960 | 214 | 75600 | 289 | 47520 | 364 | 1351296 | 439 | 505440 | 514 | 2386944 | 589 | 2634240 |
| 65 | 7560 | 140 | 78336 | 215 | 42240 | 290 | 238464 | 365 | 141120 | 440 | 836352 | 515 | 126720 | 590 | 4250400 |
| 66 | 2304 | 141 | 25200 | 216 | 619920 | 291 | 375840 | 366 | 399360 | 441 | 60480 | 516 | 2469600 | 591 | 2336256 |
| 67 | 22848 | 142 | 84240 | 217 | 236160 | 292 | 236544 | 367 | 168960 | 442 | 1836000 | 517 | 2819520 | 592 | 1516320 |
| 68 | 8400 | 143 | 120000 | 218 | 70560 | 293 | 317520 | 368 | 194400 | 443 | 866880 | 518 | 354816 | 593 | 268800 |
| 69 | 29160 | 144 | 183456 | 219 | 291600 | 294 | 166320 | 369 | 1067040 | 444 | 1537920 | 519 | 1599360 | 594 | 656640 |
| 70 | 5376 | 145 | 410112 | 220 | 278400 | 295 | 312000 | 370 | 348480 | 445 | 1219680 | 520 | 295680 | 595 | 1032192 |
| 71 | 3360 | 146 | 88320 | 221 | 261360 | 296 | 108864 | 371 | 147840 | 446 | 349440 | 521 | 1271808 | 596 | 4743360 |
| 72 | 1440 | 147 | 12096 | 222 | 164736 | 297 | 511488 | 372 | 641520 | 447 | 184320 | 522 | 304128 | 597 | 4101120 |
| 73 | 13248 | 148 | 18720 | 223 | 66240 | 298 | 132480 | 373 | 929280 | 448 | 492480 | 523 | 3941280 | 598 | 2410560 |
| 74 | 11040 | 149 | 29952 | 224 | 447120 | 299 | 354240 | 374 | 1632000 | 449 | 954720 | 524 | 422400 | 599 | 9922560 |
| 75 | 27720 | 150 | 15120 | 225 | 55296 | 300 | 84480 | 375 | 107520 | 450 | 1435200 | 525 | 80640 | 600 | 427680 |
| 76 | 21840 | 151 | 179200 | 226 | 420000 | 301 | 532800 | 376 | 352512 | 451 | 215040 | 526 | 508032 |  |  |

TABLE 1. Smallest solution to $A(m) = k$

| $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ | $k$ | $m_k$ |
|---|---:|---|---:|---|---:|---|---:|---|---:|---|---:|---|---:|
| 601 | 662400 | 659 | 1749600 | 717 | 6905088 | 775 | 3346560 | 833 | 1440000 | 891 | 4504320 | 949 | 2963520 |
| 602 | 1486080 | 660 | 1096704 | 718 | 161280 | 776 | 5948640 | 834 | 3564288 | 892 | 2239488 | 950 | 3091200 |
| 603 | 2227680 | 661 | 2724480 | 719 | 201600 | 777 | 8305920 | 835 | 3840000 | 893 | 11100672 | 951 | 1419264 |
| 604 | 1149120 | 662 | 6255360 | 720 | 2877120 | 778 | 3307392 | 836 | 599040 | 894 | 881280 | 952 | 9914400 |
| 605 | 138240 | 663 | 456192 | 721 | 435456 | 779 | 12403200 | 837 | 3928320 | 895 | 2462400 | 953 | 1788480 |
| 606 | 752640 | 664 | 1915200 | 722 | 3252480 | 780 | 5087232 | 838 | 2439360 | 896 | 9085440 | 954 | 5107200 |
| 607 | 397440 | 665 | 172800 | 723 | 1403136 | 781 | 2703360 | 839 | 9767520 | 897 | 2820096 | 955 | 2203200 |
| 608 | 1880064 | 666 | 739200 | 724 | 3556224 | 782 | 6815232 | 840 | 3345408 | 898 | 3897600 | 956 | 4133376 |
| 609 | 155520 | 667 | 1491840 | 725 | 1943040 | 783 | 19325952 | 841 | 564480 | 899 | 6791040 | 957 | 2188800 |
| 610 | 1404000 | 668 | 608256 | 726 | 453600 | 784 | 1016064 | 842 | 1753920 | 900 | 7689600 | 958 | 5581440 |
| 611 | 554400 | 669 | 1094400 | 727 | 4012800 | 785 | 2842560 | 843 | 927360 | 901 | 10133760 | 959 | 9938880 |
| 612 | 3120000 | 670 | 4078080 | 728 | 3182400 | 786 | 2678400 | 844 | 3407040 | 902 | 774144 | 960 | 5448960 |
| 613 | 1559040 | 671 | 374400 | 729 | 4548960 | 787 | 6138720 | 845 | 4579200 | 903 | 2405376 | 961 | 322560 |
| 614 | 1672704 | 672 | 596160 | 730 | 709632 | 788 | 7666560 | 846 | 2736000 | 904 | 1503360 | 962 | 3196800 |
| 615 | 336960 | 673 | 3304800 | 731 | 4736160 | 789 | 5400000 | 847 | 5744640 | 905 | 912384 | 963 | 6105600 |
| 616 | 2908224 | 674 | 3052800 | 732 | 2794176 | 790 | 3729600 | 848 | 1321920 | 906 | 1990656 | 964 | 1958400 |
| 617 | 332640 | 675 | 756000 | 733 | 4018560 | 791 | 1566720 | 849 | 3072000 | 907 | 8458560 | 965 | 3886080 |
| 618 | 1098240 | 676 | 1568160 | 734 | 181440 | 792 | 2423520 | 850 | 524160 | 908 | 3400320 | 966 | 17860608 |
| 619 | 998400 | 677 | 1327104 | 735 | 699840 | 793 | 1853280 | 851 | 3245760 | 909 | 4432320 | 967 | 10195200 |
| 620 | 230400 | 678 | 2204928 | 736 | 2069760 | 794 | 4039200 | 852 | 4176000 | 910 | 1689600 | 968 | 4953600 |
| 621 | 1545600 | 679 | 936000 | 737 | 285120 | 795 | 1670400 | 853 | 1324800 | 911 | 1641600 | 969 | 1976832 |
| 622 | 1056000 | 680 | 190080 | 738 | 993600 | 796 | 1720320 | 854 | 7992000 | 912 | 6098400 | 970 | 7996800 |
| 623 | 4610304 | 681 | 1026432 | 739 | 4134240 | 797 | 1437696 | 855 | 8726400 | 913 | 8114400 | 971 | 4804800 |
| 624 | 1077120 | 682 | 1870848 | 740 | 5298048 | 798 | 2407680 | 856 | 1010880 | 914 | 15593472 | 972 | 9561600 |
| 625 | 3394560 | 683 | 979200 | 741 | 6868800 | 799 | 2196480 | 857 | 466560 | 915 | 4868640 | 973 | 4821120 |
| 626 | 1188000 | 684 | 5520960 | 742 | 561600 | 800 | 5248800 | 858 | 7231680 | 916 | 3440640 | 974 | 3010560 |
| 627 | 4331520 | 685 | 2066688 | 743 | 943488 | 801 | 460800 | 859 | 10444800 | 917 | 4919040 | 975 | 14276736 |
| 628 | 299520 | 686 | 712800 | 744 | 1850688 | 802 | 2237760 | 860 | 2861568 | 918 | 443520 | 976 | 633600 |
| 629 | 3219840 | 687 | 1059840 | 745 | 4490640 | 803 | 16298496 | 861 | 7171200 | 919 | 3758400 | 977 | 748800 |
| 630 | 3590400 | 688 | 1733760 | 746 | 3516480 | 804 | 1078272 | 862 | 302400 | 920 | 2545920 | 978 | 2672640 |
| 631 | 940800 | 689 | 728640 | 747 | 504000 | 805 | 2527200 | 863 | 823680 | 921 | 2112000 | 979 | 13298688 |
| 632 | 677376 | 690 | 2673216 | 748 | 4416000 | 806 | 2550240 | 864 | 3409920 | 922 | 345600 | 980 | 2875392 |
| 633 | 1742400 | 691 | 1410048 | 749 | 6438528 | 807 | 1996800 | 865 | 2119680 | 923 | 3880800 | 981 | 3133440 |
| 634 | 4589568 | 692 | 120960 | 750 | 1278720 | 808 | 5110560 | 866 | 3055104 | 924 | 1468800 | 982 | 3815424 |
| 635 | 3292800 | 693 | 6292800 | 751 | 532224 | 809 | 475200 | 867 | 576000 | 925 | 22492800 | 983 | 16524000 |
| 636 | 7104000 | 694 | 3875040 | 752 | 3015936 | 810 | 2115072 | 868 | 2972160 | 926 | 2756160 | 984 | 8631360 |
| 637 | 737280 | 695 | 684288 | 753 | 2201472 | 811 | 2635776 | 869 | 13167360 | 927 | 8125920 | 985 | 6716160 |
| 638 | 1080000 | 696 | 423360 | 754 | 2845440 | 812 | 3060288 | 870 | 2949120 | 928 | 3598560 | 986 | 4945920 |
| 639 | 224640 | 697 | 3504384 | 755 | 3198720 | 813 | 1393920 | 871 | 4292352 | 929 | 4680000 | 987 | 10386432 |
| 640 | 2587200 | 698 | 4915200 | 756 | 6306048 | 814 | 1002240 | 872 | 311040 | 930 | 4104000 | 988 | 1166400 |
| 641 | 2370816 | 699 | 5456640 | 757 | 5428800 | 815 | 2479680 | 873 | 9803520 | 931 | 2751840 | 989 | 4872960 |
| 642 | 2706048 | 700 | 6829056 | 758 | 2503872 | 816 | 2923200 | 874 | 6547968 | 932 | 10178784 | 990 | 1028160 |
| 643 | 653184 | 701 | 2946240 | 759 | 4623360 | 817 | 4593600 | 875 | 13258080 | 933 | 2086560 | 991 | 1965600 |
| 644 | 1477440 | 702 | 2864160 | 760 | 1336320 | 818 | 1283040 | 876 | 1542240 | 934 | 11625120 | 992 | 22958208 |
| 645 | 1848000 | 703 | 2352000 | 761 | 1975680 | 819 | 2426112 | 877 | 5280000 | 935 | 552960 | 993 | 1344000 |
| 646 | 3446784 | 704 | 2064384 | 762 | 2483712 | 820 | 1909440 | 878 | 19885824 | 936 | 7683840 | 994 | 2059200 |
| 647 | 792000 | 705 | 387072 | 763 | 6741504 | 821 | 5510400 | 879 | 1722240 | 937 | 7064064 | 995 | 2449440 |
| 648 | 1938816 | 706 | 1169280 | 764 | 207360 | 822 | 1918080 | 880 | 10632960 | 938 | 1905120 | 996 | 7925760 |
| 649 | 316800 | 707 | 984960 | 765 | 4884480 | 823 | 798336 | 881 | 1959552 | 939 | 5308416 | 997 | 12109824 |
| 650 | 3075840 | 708 | 5193216 | 766 | 591360 | 824 | 14424480 | 882 | 6884352 | 940 | 2099520 | 998 | 997920 |
| 651 | 860160 | 709 | 2125440 | 767 | 1907712 | 825 | 3172608 | 883 | 6240000 | 941 | 580608 | 999 | 12633600 |
| 652 | 746496 | 710 | 6948864 | 768 | 432000 | 826 | 1397760 | 884 | 2306304 | 942 | 5195520 | 1000 | 1360800 |
| 653 | 2708640 | 711 | 253440 | 769 | 276480 | 827 | 4245696 | 885 | 5879808 | 943 | 12531456 | | |
| 654 | 5466240 | 712 | 2970240 | 770 | 1827840 | 828 | 6199200 | 886 | 10964160 | 944 | 506880 | | |
| 655 | 635040 | 713 | 822528 | 771 | 3302208 | 829 | 1179360 | 887 | 2496000 | 945 | 6439680 | | |
| 656 | 4327680 | 714 | 4727808 | 772 | 259200 | 830 | 1615680 | 888 | 5144832 | 946 | 3306240 | | |
| 657 | 1603584 | 715 | 3689280 | 773 | 685440 | 831 | 5679360 | 889 | 449280 | 947 | 4723200 | | |
| 658 | 988416 | 716 | 844800 | 774 | 3009600 | 832 | 794880 | 890 | 16057440 | 948 | 4663296 | | |

TABLE 2. Smallest solution to $A(m) = k$

Aside from Theorem 5, the only other known result concerning the behavior of $V_1(x)$ as $x \to \infty$ is the bound

$$\liminf_{x\to\infty} \frac{V_1(x)}{V(x)} \leqslant \frac{1}{2}, \tag{7.1}$$

established by very elementary means in an unpublished note of Pomerance (see [29] and [24]). A modification of his argument, combined with the results of the above computations, yields the much stronger bound in Theorem 7. The following lemma is the key. Recall the definition of $V(x;k)$ given in the introduction.

**Lemma 7.4.** *We have $V(x;a^2) \leqslant V(x/a)$.*

*Proof.* The lemma is trivial when $a = 1$ so assume $a \geqslant 2$. Let $n$ be a totient with $x/a < n \leqslant x$. First we show that for some integer $s \geqslant 0$, $a^{-s}n$ is a totient with an pre-image not divisible by $a^2$. Suppose $\phi(m) = n$. If $a^2 \nmid m$, take $s = 0$. Otherwise we can write $m = a^t r$, where $t \geqslant 2$ and $a \nmid r$. Clearly $\phi(ar) = a^{1-t}n$, so we take $s = t - 1$. Next, if $n_1$ and $n_2$ are two distinct totients in $(x/a, x]$, then $a^{-s_1}n_1 \neq a^{-s_2}n_2$ (since $n_1/n_2$ cannot be a power of $a$), so the mapping from totients in $(x/a, x]$ to totients $\leqslant x$ with a pre-image not divisible by $a^2$ is one-to-one. Thus $V(x) - V(x;a^2) \geqslant V(x) - V(x/a)$. □

The above computations show that if $\phi(x) = n$ and $A(n) = 1$, then $x$ is divisible by either $a^2$ or $b^2$, where $a$ and $b$ are numbers greater than $10^{5,001,850,000}$.

Without loss of generality, suppose $a \leqslant b$. By Lemma 7.4, we have

$$V_1(x) \leqslant V(x/a) + V(x/b) \leqslant 2V(x/a). \tag{7.2}$$

**Lemma 7.5.** *Suppose $V_1(x) \leqslant bV(x/a)$, where $a > 1$ and $b > 0$. Then*

$$\liminf_{x\to\infty} \frac{V_1(x)}{V(x)} \leqslant \frac{b}{a}.$$

*Proof.* Suppose $c = \liminf_{x\to\infty} \frac{V_1(x)}{V(x)} > 0$. For every $\varepsilon > 0$ there is a number $x_0$ such that $x \geqslant x_0$ implies $V_1(x)/V(x) \geqslant c - \varepsilon$. For large $x$, set $n = [\log(x/x_0)/\log a]$. Then

$$\begin{aligned} V(x) &= \frac{V(x)}{V(x/a)} \frac{V(x/a)}{V(x/a^2)} \cdots \frac{V(x/a^{n-1})}{V(x/a^n)} V(x/a^n) \\ &\leqslant b^n \frac{V(x)}{V_1(x)} \frac{V(x/a)}{V_1(x/a)} \cdots \frac{V(x/a^{n-1})}{V_1(x/a^{n-1})} V(ax_0) \\ &\leqslant b^n (c-\varepsilon)^{-n}(ax_0) = O(x^{-\log((c-\varepsilon)/b)/\log a}). \end{aligned}$$

This contradicts the trivial bound $V(x) \gg x/\log x$ if $c > \frac{b}{a} + \varepsilon$. Since $\varepsilon$ is arbitrary, the lemma follows. □

Theorem 7 follows immediately. Further improvements in the lower bound for a counterexample to Carmichael's Conjecture will produce corresponding upper bounds on $\liminf_{x\to\infty} V_1(x)/V(x)$. Explicit bounds for the $O(1)$ term appearing in Theorem 1 (which would involve considerable work to obtain) combined with (7.2) should give $\limsup_{x\to\infty} V_1(x)/V(x) \leqslant 10^{-5,000,000,000}$ as well.

Lemma 7.4 raises another interesting question. First, suppose $d$ is a totient, all of whose pre-images $m_i$ are divisible by $k$. The lower bound argument given in Section 5 shows that for at least half of the numbers $b \in \mathscr{B}$, the totient $\phi(b)d$ has only the pre-images $bm_i$. In particular, all of the pre-images of such totients are divisible by $k$ and Theorem 8 follows.

It is also natural to ask for which $k$ do there exist totients, all of whose pre-images are divisible by $k$. A short search reveals examples for each $k \leqslant 11$ except $k = 6$ and $k = 10$. For $k = 2, 4$ and 8, take $d = 2^{18} \cdot 257$, for $k = 3$ or 9 take $d = 54 = 2 \cdot 3^3$, for $k = 5$ take $d = 12500 = 4 \cdot 5^5$, for $k = 7$, take $d = 294 = 6 \cdot 7^2$ and for $k = 11$, take $d = 110$. It appears that there might not be any totient, all of whose pre-images are divisible by 6, but I cannot prove this. Any totient with a unique pre-image must have that pre-image divisible by 6, so the non-existence of such numbers implies Carmichael's Conjecture.

| $x$ | $V(x)$ | $V_2/V$ | $V_3/V$ | $V_4/V$ | $V_5/V$ | $V_6/V$ | $V_7/V$ |
|---|---|---|---|---|---|---|---|
| 1M | 180,184 | 0.380727 | 0.140673 | 0.098988 | 0.042545 | 0.062730 | 0.020790 |
| 5M | 840,178 | 0.379462 | 0.140350 | 0.102487 | 0.042687 | 0.063193 | 0.020373 |
| 10M | 1,634,372 | 0.378719 | 0.140399 | 0.103927 | 0.042703 | 0.063216 | 0.020061 |
| 25M | 3,946,809 | 0.378198 | 0.140233 | 0.105466 | 0.042602 | 0.063414 | 0.019819 |
| 125M | 18,657,531 | 0.377218 | 0.140176 | 0.107873 | 0.042560 | 0.063742 | 0.019454 |
| 300M | 43,525,579 | 0.376828 | 0.140170 | 0.108933 | 0.042517 | 0.063818 | 0.019284 |
| 500M | 71,399,658 | 0.376690 | 0.140125 | 0.109509 | 0.042493 | 0.063851 | 0.019194 |

TABLE 3. $V_k(x)/V(x)$ for $2 \leqslant k \leqslant 7$

I believe that obtaining the asymptotic formula for $V(x)$ will require simultaneously determining the asymptotics of $V_k(x)/V(x)$ (more will be said in section 8) and $V(x;k)/V(x)$ for each $k$. It may even be necessary to classify totients more finely. For instance, taking $d = 4, k = 4$ in the proof of Theorem 2 (section 5), the totients $m$ constructed have $\phi^{-1}(m) = \{5n, 8n, 10n, 12n\}$ for some $n$. On the other hand, taking $d = 6, k = 4$ produces a different set of totients $m$, namely those with $\phi^{-1}(m) = \{7n, 9n, 14n, 18n\}$ for some $n$. Likewise, for any given $d$ with $A(d) = k$, the construction of totients in section 5 may miss whole classes of totients with multiplicity $k$. There is much further work to be done in this area.

## 8. THE AVERAGE OF $A(m)$ AND THE RATIO $V_k(x)/V(x)$

*Proof of Theorem 3.* Suppose $x$ is sufficiently large and set $M = [\sqrt{4C\log N}]$, $L = L_0(x) - M$. Define $\xi_i$ as in Theorem 15. By Theorem 15, the number of totients $m \leqslant x$ with a pre-image $n$ satisfying $\mathbf{x}(n) \notin \mathscr{S}_L(\boldsymbol{\xi})$ is $O(V(x)/N)$ (here $\mathbf{x}(n) = (x_1(n;x), \dots, x_L(n;x))$). For $n \in \mathscr{R}_L(\boldsymbol{\xi})$, set $q_i = q_i(n)$ and $x_i = x_i(n;x)$ for each $i$. For $0 \leqslant b < L$, let $N_b(x)$ denote the number of such $n$ with

$$\frac{b}{Lg_L^*} \leqslant x_L < \frac{b+1}{Lg_L^*} =: \frac{Z_b}{\log_2 x}.$$

Write $n = q_0 \cdots q_{L-1} r$, so that $\log_2 P^+(r) \leqslant Z_b$. Also, $\log_2 Z_b \ll b\varrho^M$. By Corollary 3.5 and Lemmas 3.11 and 4.3,

$$\begin{aligned} N_b(x) &\ll \frac{x}{\log x} R_{L-1}(\boldsymbol{\xi};x) e^{-b} (\log_2 Z_b) W(E(Z_b)) \\ &\ll V(x) \exp\{-b + M\log b - M\log M + O(M + \log^2 b)\}. \end{aligned}$$

Summing on $b$ (as in the proof of Theorem 10) gives $\sum_b N_b(x) \ll V(x)\exp\{O(M)\}$. Therefore, the number of totients $m \leqslant x$ with $A(m) \geqslant N$ is $O(V(x)N^{-1}\exp\{O(M)\})$, and the theorem follows. □

In contrast, the average value of $A(m)$ over totients $m \leqslant x$ is clearly $\geqslant x/V(x) = (\log x)^{1+o(1)}$. The vast differences between the "average" behavior and the "normal" behavior is a result of some totients having an enormous number of pre-images. In fact, Erdős [8] showed that there are infinitely many totients for which $A(m) \geqslant m^{c_4}$ for some positive constant $c_4$. The current record is $c_4 = 0.7039$ [1]. It is probable that $V_k(x)/V(x)$ approaches a limit for each $k$. Based on Theorem 9 and computations, we make the

**Conjecture 2.** *For $k \geqslant 2$,*

$$\lim_{x\to\infty} \frac{V_k(x)}{V(x)} = C_k.$$

Theorem 3 implies that $V_k(x)/V(x) \ll k^{-2+\varepsilon}$ on average. Table 3 lists values of $V(x)$ and the ratios $V_k(x)/V(x)$ for $2 \leqslant k \leqslant 7$.

Numerical investigations seem to indicate that $C_k \asymp 1/k^2$ for large $k$. In fact, at $x = 500,000,000$ we have $1.75 \leqslant R_k \leqslant 2.05$ for $20 \leqslant k \leqslant 200$. This data is very misleading, however. It is an immediate corollary of Theorem 2 that for infinitely many $k$,

$$\frac{V_k(x)}{V(x)} \gg k^{-1/c_4+\varepsilon} \gg k^{-1.42} \qquad (x > x_0(k)),$$

so $V_k(x)/V(x)$ is much larger than average for many $k$. Erdős has conjectured that every $c_4 < 1$ is admissible, which implies that for every $\varepsilon > 0$,

$$\frac{V_k(x)}{V(x)} \gg k^{-1-\varepsilon} \qquad (x > x_0(k))$$

for infinitely many $k$. The behavior of the constants $C_k$ is probably quite complicated.

## 9. Generalization to other multiplicative functions

The proofs of the theorems depend entirely on the definition of $\phi(n)$ as a multiplicative function, and not on the arithmetical interpretation of $\phi(n)$ as the number of positive integers less than $n$ which are coprime to $n$. This suggests that corresponding results should hold for a wide class of similar multiplicative arithmetic functions, such as $\sigma(n)$, the sum of divisors function.

Suppose $f : \mathbb{N} \to \mathbb{N}$ is a multiplicative arithmetic function. We first define

$$\begin{aligned} \mathscr{V}_f &= \{f(n) : n \in \mathbb{N}\}, \\ V_f(x) &= |\mathscr{V}_f \cap [1, x]|, \\ f^{-1}(m) &= \{n : f(n) = m\}, \\ A_f(m) &= |f^{-1}(m)|, \\ V_{f,k}(x) &= |\{m \leqslant x : A_f(m) = k\}|, \end{aligned} \tag{9.1}$$

which are the analogous functions to the definitions (1.1). With only minor modifications to the arguments given in previous sections, we prove Theorem 14. By itself, condition (1.11) is enough to prove the lower bound for $V_f(x)$. Condition (1.12) is used only for the upper bound argument and the lower bound for $V_{f,k}(x)$.

The function $f(n) = n$, which takes all positive integer values, is an example of why zero must be excluded from the set in (1.11). Condition (1.12) insures that the values of $f(p^k)$ for $k \geqslant 2$ are not too small too often, and thus have little influence on the size of $V_f(x)$. It essentially forces $f(h)$ to be a bit larger than $h^{1/2}$ on average. It's probable that (1.12) can be relaxed, but not too much. For example, the multiplicative function defined by $f(p) = p - 1$ for prime $p$, and $f(p^k) = p^{k-1}$ for $k \geqslant 2$ clearly takes all integer values, while

$$\sum_{h \text{ square-full}} \frac{1}{f(h)(\log_2 h)^2} \ll 1.$$

Condition (1.12) also insures that $A(m)$ is finite for each $f$-value $m$. For example, a function satisfying $f(p^k) = 1$ for infinitely many prime powers $p^k$ has the property that $A(m) = \infty$ for every $f$-value $m$.

Some functions appearing in the literature which satisfy the conditions of Theorem 14 are $\sigma(n)$, the sum of divisors function, $\phi^*(n)$, $\sigma^*(n)$ and $\psi(n)$. Here $\phi^*(n)$ and $\sigma^*(n)$ are the unitary analogs of $\phi(n)$ and $\sigma(n)$, defined by $\phi^*(p^k) = p^k - 1$ and $\sigma^*(p^k) = p^k + 1$ [6], and $\psi(n)$ is Dedekind's function, defined by $\psi(p^k) = p^k + p^{k-1}$.

In general, implied constants will depend on the function $f(n)$. One change that must be made throughout is to replace every occurrence of "$p - 1$" (when referring to $\phi(p)$) with "$f(p)$", for instance in the definition of $S$-normal primes in section 2. Since the possible values of $f(p) - p$ is a finite set, Lemma 2.6 follows easily with the new definitions. The most substantial change to be made in section 2, however, is to Lemma 2.7, since we no longer have the bound $n/f(n) \ll \log_2 n$ at our disposal.

**Lemma 2.7$^*$.** *The number of $m \in \mathscr{V}_f(x)$ for which either $d^2|m$ or $d^2|n$ for some $n \in f^{-1}(m)$ and $d > Y$ is $O(x(\log_2 x)^K/\varepsilon(Y^2))$, where $K = \max_p(p - f(p))$.*

*Proof.* The number of $m$ with $d^2|m$ for some $d > Y$ is $O(x/Y)$. Now suppose $d^2|n$ for some $d > Y$, and let $h = h(n)$ denote the square-full part of $n$, that is

$$h(n) = \prod_{\substack{p^a \| n \\ a \geqslant 2}} p^a.$$

In particular, $h(n) > Y^2$. From the fact that $f(p) \geqslant p - K$ for all primes $p$, we have

$$f(n) = f(h)f(n/h) \gg \frac{f(h)n}{h}(\log_2(n/h))^{-K}.$$

Thus, if $f(n) \leqslant x$, then

$$\frac{n}{h}\left(\log_2 \frac{n}{h}\right) \ll \frac{x}{f(h)}.$$

Therefore, the number of possible $n$ with a given $h$ is crudely $\ll x(\log_2 x)^K / f(h)$. By (1.12), the total number of $n$ is at most

$$\ll x(\log_2 x)^K \sum_{h \geqslant Y^2} \frac{1}{f(h)} \ll \frac{x(\log_2 x)^K}{\varepsilon(Y^2)} \sum_{h \geqslant 16} \frac{\varepsilon(h)}{f(h)} \ll \frac{x(\log_2 x)^K}{\varepsilon(Y^2)}.$$

□

The final modification to section 2 is to Lemma 2.8, which needs an additional term $O(x(\log_2 x)^K/\varepsilon(S^2))$ in the conclusion. This comes from the use of Lemma 2.7* to count the number of totients with a pre-image divisible by a square of a prime $p > S$.

In section 3, only the proof of Lemma 3.1 requires modification. For the upper bound (3.12), we must take into account contributions of numbers which are not square-free using only (1.12), whereas before we used $\phi(p^b) \geqslant (p-1)^b$. First change the definition of $R_L(\mathscr{S}; x)$ to

$$R_L(\boldsymbol{\xi}; x) = \sum_{n \in \mathscr{R}_L(\boldsymbol{\xi}; x)} \frac{1}{f(n)}.$$

The proof of Lemma 3.1 handles the sum over squarefree $n$, the estimates being identical by (1.11). Only squarefree $n$ are considered for the lower bound, so we need only handle non-squarefree $n$ in the upper bound. Note first that $\mathbf{x} \in \mathscr{S}_L(\boldsymbol{\xi})$ implies $x_i > x_{i+2}$ for each $i$, hence any $n \in \mathscr{R}_L(\mathscr{S}; x)$ is cube-free. For non-squarefree $n$, let $I(n)$ denote the set of indices $i$ $(1 \leqslant i \leqslant j-1)$ with $p_i \parallel n$, and let $\mathscr{I}$ denote the set of possible $I(n)$. Note that $I(n)$ contains at most $j - 3$ indices. Also, by (1.12),

$$\sum_{t=1}^{\infty} \frac{1}{f(t^2)} \ll 1.$$

Also define $r = \prod q_i$, the product being over all $i \notin I(n)$. In particular, $r$ is a square. With $\mathbf{m}$ fixed, as in the proof of Lemma 3.1,

$$\begin{aligned} \sum_n \frac{1}{f(n)} &= \sum_{I \in \mathscr{I}} \sum_{\substack{\mathbf{x}(n) \in B(\mathbf{m}) \\ I(n) = I}} \frac{1}{g(r)} \prod_{i \in I} \frac{1}{f(p_i)} \leqslant \sum_{I \in \mathscr{I}} \sum_{i \in I} \left( \sum_{p_i} \frac{1}{f(p_i)} \right) \left( \sum_{t=1}^{\infty} \frac{1}{f(t^2)} \right) \\ &\ll \sum_{I \in \mathscr{I}} \prod_{i \in I} \left(1/2 + O(e^{-m_i})\right) \qquad (9.2) \\ &\ll 2^{1-L} \sum_{I \in \mathscr{I}} \prod_{i \notin I} (4/5)^{L_0 - i} \leqslant 2^{1-L} \sum_{k \geqslant 2} \frac{5^k}{k!} \ll 2^{-L}. \end{aligned}$$

where the range of $p_i$ is $(m_i - 1)/2 \leqslant \log_2 p_i \leqslant m_i/2$. If, as in (3.13), we have a lower bound on $p_L$, the above estimate is replaced with $o(2^{-L})$ as $p_L \to \infty$.

For the lower bound (3.13), we first modify the definitions of $L$ and $\mathscr{R}_L^*(\boldsymbol{\xi};x)$. One complication is that the mapping $p \to f(p)$ may not be one-to-one. We say a prime $p$ is "bad" if $f(p) = f(p')$ for some prime $p' \neq p$ and say $p$ is "good" otherwise. By (1.11) and Lemma 2.5, the number of bad primes $\leqslant y$ is $O(y/\log^2 y)$, so $\sum_{p\, bad} 1/p$ converges. The other complication has to do with "small" values of $f(p^k)$ for some prime powers $p^k$ with $k \geqslant 2$. For each prime $p$, define

$$Q(p) := \min_{k\geqslant 2} \frac{f(p^k)}{f(p)}. \tag{9.3}$$

Introduce another parameter $d$ (which will be the same $d$ as in Theorem 2) and suppose $L \leqslant L_0 - M$ where $M$ is a sufficiently large constant depending on $P_0$ and $d$. If follows from (1.12) and (9.3) that

$$\sum_{Q(p)\leqslant d} \frac{1}{p} = O(d).$$

Specify in the definition of $\mathscr{S}$ the additional restrictions that every $p_i$ is "good" and satisfies $Q(p_i) > d$. Therefore, by the proof of Lemma 3.12 (3.13), if $M_1$ is large enough, we have $R_L(\mathscr{S};x) \gg V_f(x)$.

In section 4, we need to add a term $O(y(\log_2 y)^K/\varepsilon(S^2))$ to the conclusion of Lemma 4.1 due to the use of Lemma 2.7$^*$. A modified Lemma 4.2 is stated below.

**Lemma 4.2$^*$.** *Suppose $k \geqslant 300$, $\omega = 1/(10k^3)$, $y \geqslant y_0$ and $\log_3 y \geqslant k/3$. Then the number of totients $\leqslant y$ with a preimage $n$ satisfying*

$$a_1x_1(n;y) + \cdots + a_kx_k(n;y) \geqslant 1 + \omega$$

*is*

$$\ll y(\log_2 y)^5 W(y)(\log y)^{-1-1/(60000k^9 \log k)}.$$

The only difference in the proof is that $\log_2 S = \frac{\log_2 y}{10000k^9 \log k}$, so in $U_1(y)$, $U_3(y)$ and $T(\boldsymbol{\theta};x)$ we have

$$\frac{y(\log_2 y)^K}{\varepsilon(S^2)} \ll y \exp\left\{ -\frac{\log_2 y}{2500k^9 \log k}(\log_3 y - 10\log k - 8)^{20} + K\log_3 y \right\} \ll \frac{y}{\log^2 y}.$$

In the main upper bound argument, we first eliminate the possibility of large squares dividing $n$ using Lemma 2.7$^*$. Defining $Y_j$ by $\log_3 Y_j = k/3$ insures that (4.8) implies (4.9).

The additional restrictions $p_i$ "good" and $Q(p_i) > d$ introduced for the proof of Lemma 3.12 are needed for the lower bound argument in section 5. First, we modify slightly the definition of the set $B$. In place of (5.4) use $f(n) \leqslant x/d$ and in place of (5.7) put $Q(p_i) > d\,(0 \leqslant i \leqslant L)$. Also add the condition that none of the primes $p_i$ are bad. Fortunately, the numbers in $B$ are square-free by definition. The equation (5.9) becomes

$$df(n) = f(n_1). \tag{9.4}$$

Since $Q(p_i) > d$ for each $p_i$, if $n|n_1$ and one of the primes $q_i$ $(0 \leqslant i \leqslant L)$ occurs to a power grater than 1, then $\phi(n_1) > d\phi(n)$. Therefore, the $L+1$ largest prime factors of $n_1$ occur to the first power only, which forces $n_1 = nm_i$ for some $i$ (the trivial solutions). For nontrivial solutions, we have at least one index $i$ for which $p_i \neq q_i$, and hence $f(p_i) \neq f(q_i)$ (since each $p_i$ is "good"). Obvious changes are made to (5.23) and the definitions of $\sigma_k(\mathscr{A})$ and $\tau_k(\mathscr{A})$. In Lemma 5.2, the phrase "$rt+1$ and $st+1$ are unequal primes" is replace by "$rt+a$ and $st+a'$ are unequal primes for some pair of numbers $(a,a')$ with $a,a' \in \mathscr{P}$." Here $\mathscr{P}$ denotes the set of possible values of $f(p)-p$. As the number of pairs $(a,a')$ is finite, this poses no problem in the argument. Similar changes are made in several places in Lemma 5.3. It is not possible to prove analogs of Theorems 5–9 for general $f$ satisfying the hypotheses of Theorem 14. One reason is that there might not be any "Carmichael Conjecture" for $f$, e.g. $A_\sigma(3) = 1$, where $\sigma$ is the sum of divisors function. Furthermore, the proof of Theorem 9 depends on the identity $\phi(p^2) = p\phi(p)$ for primes $p$. If, for some $a \neq 0$, $f(p) = p + a$ for all primes $p$, then the argument of [14] shows that if the multiplicity $k$ is possible and $r$ is a positive integer, then the multiplicity $rk$ is possible. For functions such as $\sigma(n)$, for which the multiplicity 1 is possible, this completely solves the problem of the possible multiplicities. For other functions, it shows at least that a positive proportion of multiplicities are possible. If multiplicity 1 is not possible, and $f(p^2) = pf(p)$, the argument in [15] shows that all multiplicities beyond some point are possible.

We can, however, obtain information about the possible multiplicities for more general $f$ by an induction argument utilizing the next lemma. Denote by $a_1, \dots, a_K$ the possible values of $f(p) - p$ for prime $p$.

**Lemma 7.1**$^*$. *Suppose $A_f(m) = k$. Let $p, q, s$ be primes and $r \geqslant 2$ an integer so that*

(1) *(i) $s$ and $q$ are "good" primes,*
(2) *(ii) $mf(s) = f(q)$,*
(3) *(iii) $f(s) = rp$,*
(4) *(iv) $p \nmid f(\pi^b)$ for every prime $\pi$, integer $b \geqslant 2$ with $f(\pi^b) \leqslant mf(s)$,*
(5) *(v) $dp - a_i$ is composite for $1 \leqslant i \leqslant K$ and $d|rm$ except $d = r$ and $d = rm$.*

*Then $A_f(mrp) = k + A_f(1)$.*

*Proof.* Let $f^{-1}(m) = \{x_1, \dots, x_k\}$ and suppose $f(x) = mrp$. By condition (iv), $p|f(\pi)$ for some prime $\pi$ which divides $x$ to the first power. Therefore, $f(\pi) = dp$ for some divisor $d$ of $mr$. Condition (v) implies that the only possibilities for $d$ are $d = r$ or $d = rm$. If $d = r$, then $f(\pi) = rp = f(p)$ which forces $\pi = s$ by condition (i). By conditions (ii) and (iii), we have $f(x/s) = m$, which gives solutions $x = sx_i$ $(1 \leqslant i \leqslant k)$. Similarly, if $d = rm$, then $\pi = q$ and $f(x/q) = 1$, which has $A_f(1)$ solutions. □

By the Chinese Remainder Theorem, there is an arithmetic progression $\mathscr{A}$ so that condition (v) is satisfied for each number $p \in \mathscr{A}$, while still allowing each $rp + a_i$ and $rmp + a_i$ to be prime. To eliminate primes failing condition (iv), we need the asymptotic form of the Prime $k$-tuples Conjecture due to Hardy and Littlewood [20] (actually only the case where $a_i = 1$ for each $i$ is considered in [20]; the conjectured asymptotic for $k$ arbitrary polynomials can be found in [3]).

**Conjecture 3** (Prime $k$-tuples Conjecture (asymptotic version)**.** *Suppose $a_1, \dots, a_k$ are positive integers and $b_1, \dots, b_k$ are integers so that no prime divides $(a_1n + b_1) \cdots (a_kn + b_k)$ for every integer $n$. Then for some constant $C(\mathbf{a}, \mathbf{b})$, the number of $n \leqslant x$ for which $a_1n + b_1, \dots, a_kn + b_k$ are simultaneously prime is*

$$\sim C(\mathbf{a}, \mathbf{b}) \frac{x}{\log^k x} \qquad (x \geqslant x_0(\mathbf{a}, \mathbf{b})).$$

A straightforward calculation using (1.12) gives $|\{\pi^b : f(\pi^b) \leqslant y\}| \ll y/\varepsilon(y)$. If $s$ is taken large enough, the number of possible $p \leqslant x$ satisfying condition (iv) (assuming $r$ and $m$ are fixed and noting condition (iii)) is $o(x/\log^3 x)$. The procedure for determining the set of possible multiplicities with this lemma will depend on the behavior of the particular function. Complications can arise, for instance, if $m$ is even and all of the $a_i$ are even (which makes condition (ii) impossible) or if the number of "bad" primes is $\gg x/\log^3 x$.

## References

[1] R. C. Baker and G. Harman, *Shifted primes without large prime factors*, Acta Arith. **83** (1998), 331–361.

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

[3] P. T. Bateman and R. A. Horn , *A heuristic asymptotic formula concerning the distribution of prime numbers* , Math. Comp. **16** (1962), 363–367.

[4] R. D. Carmichael , *On Euler's $\phi$-function* , Bull. Amer. Math. Soc. **13** (1907) , 241–243.

[5] ———, *Note on Euler's $\phi$-function* , Bull. Amer. Math. Soc. **28** (1922) , 109–110.

[6] E. Cohen , *Arithmetical functions associated with the unitary divisors of an integer* , Math. Z. **74** (1960) , 66–80.

[7] L. E. Dickson , *A new extension of Dirichlet's theorem on prime numbers* , Messenger of Math. **33** (1904) , 155–161.

[8] P. Erdős , *On the normal number of prime factors of $p - 1$ and some related problems concerning Euler's $\phi$-function* , Quart. J. Math. Oxford **6** (1935), 205-213.

[9] ———, *Some remarks on Euler's $\phi$-function and some related problems* , Bull. Amer. Math. Soc. **51** (1945), 540–544.

[10] ———, *Some remarks on Euler's $\phi$-function* , Acta Arith. **4** (1958), 10–19.

[11] P. Erdős and R.R.Hall , *On the values of Euler's $\phi$-function* , Acta Arith. **22** (1973), 201-206.

[12] ———, *Distinct values of Euler's $\phi$-function* , Mathematika **23** (1976), 1–3.

[13] P. Erdős and C. Pomerance, *On the normal number of prime factors of $\phi(n)$* , Rocky Mountain J. of Math. **15** (1985), 343–352.

[14] K. Ford and S. Konyagin, *On two conjectures of Sierpiński concerning the arithmetic functions $\sigma$ and $\phi$*, Number Theory in Progress (Zakopane, Poland, 1997), vol. II, de Gruyter (1999), 795-803.

[15] K. Ford , *The number of solutions of $\phi(x) = m$* , Annals of Math. **150** (1999), 283–311.

[16] K. Ford and K.-W. Lau, *Asymptotics of a recurrence sequence: Solution of Problem 10682*, Amer. Math. Monthly **107** (2009), 374–375.

[17] J. Friedlander, *Shifted primes without large prime factors*, in Number theory and applications (Banff, AB, 1988) , Kluwer Acad. Publ., Dorbrecht (1989), 393–401.

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

[19] R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge University Press, 1988.

[20] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio Numerorum': III. On the representation of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

[21] G. H. Hardy and S. Ramanujan , *The normal number of prime factors of a number $n$* , Quart. J. Math. **48** (1917), 76–92.

[22] A. Hildebrand and G. Tenenbaum , *Integers without large prime factors* , J. Théor. Nombres Bordeaux **5** (1993), 411–484.

[23] V. Klee , *On a conjecture of Carmichael* , Bull. Amer. Math. Soc. **53** (1947), 1183–1186.

[24] Mattics, L. E. , *A half step towards Carmichael's Conjecture, Solution to Problem 6671* , Amer. Math. Monthly **100** (1993), 694–695.

[25] H. Maier and C. Pomerance , *On the number of distinct values of Euler's $\phi$-function* , Acta Arith. **49** (1988), 263–275.

[26] P. Masai and A. Valette , *A lower bound for a counterexample to Carmichael's Conjecture*, Bollettino U.M.I. (6) **1** (1982), 313–316.

[27] S. Pillai , *On some functions connected with $\phi(n)$* , Bull. Amer. Math. Soc. **35** (1929), 832–836.

[28] C. Pomerance , *On the distribution of the values of Euler's function* , Acta Arith. **47** (1986), 63–70.

[29] ———, *Problem 6671*, Amer. Math. Monthly **98** (1991), 862.

[30] A. Schinzel , *Sur l'equation $\phi(x) = m$* , Elem. Math. **11** 1956, 75–78.

[31] ———, *Remarks of the paper "Sur certaines hypothèses concernant les nombres premiers"* , Acta Arith. **7** (1961/62), 1–8.

[32] A. Schinzel and W. Sierpiński , *Sur certaines hypothèses concernant les nombres premiers* , Acta Arith. **4** (1958), 185–208.

[33] A. Schlafly and S. Wagon , *Carmichael's conjecture on the Euler function is valid below $10^{10,000,000}$* , Math. Comp. **63** (1994), 415–419.