Journal of Number Theory **84**, 93–118 (2000)
doi:10.1006/jnth.2000.2512, available online at http://www.idealibrary.com on IDEAL®

Artin's Conjecture on Average for Composite Moduli

Shuguang Li

Natural Sciences Division, University of Hawaii at Hilo, 200 W. Kawili Street, Hilo, Hawaii 96720

Communicated by A. Granville

Received November 18, 1998; revised October 8, 1999

Let a be an integer $\neq -1$ and not a square. Let $P_a(x)$ be the number of primes up to x for which a is a primitive root. Goldfeld and Stephens proved that the average value of $P_a(x)$ is about a constant multiple of $x/\ln x$. Carmichael extended

prove that the average value of $N_a(x)$ oscillates. That is, $\lim_{x\to\infty} 1/x^2 \sum_{1\leqslant a\leqslant x} N_a(x) > 0$ and $\underline{\lim}_{x\to\infty} 1/x^2 \sum_{1\leqslant a\leqslant x} N_a(x) = 0$.

1. INTRODUCTION

A primitive root for a prime p is any integer a such that the order of $a \bmod p$ is $\phi(p)$, where ϕ is the Euler function. Problems on the distribution of primitive roots for primes have been reduced to problems on the distribution of values of a few well-studied arithmetic functions:

The proportion of primitive roots modulo p in the group $(\mathbb{Z}/p\mathbb{Z})^*$ is $\phi(p-1)/(p-1)$. Elliott has proved that $\phi(p-1)/(p-1)$ has a limiting distribution function [3], in the sense that $\lim_{x\to\infty} \frac{1}{\pi(x)} \#\{p \leqslant x : \phi(p-1)/(p-1) \leqslant u\}$ exists for all real numbers u.

We can also fix an integer a and count the number $P_a(x)$ of primes p up to x for which a is a primitive root. In 1927, Emil Artin conjectured that for every integer a which is $\neq -1$ and not a square, $\lim_{x\to\infty} P_a(x) = \infty$. Moreover, he formulated his conjecture as an asymptotic formula of the type

$$P_a(x) \sim \frac{A(a)\, x}{\log x} \qquad (x \to \infty),$$

where $A(a) > 0$ is a number depending on a. The value of $A(a)$ in the conjecture was revised by Heilbronn (see [19]) due to the work of D. H. Lehmer.

0022-314X/00 $35.00

The Artin conjecture was proved by C. Hooley [9] under the assumption of the extended Riemann hypothesis for Dedekind zeta functions over certain Kummerian fields. Goldfeld [7] proved it on average unconditionally. P. J. Stephens [18] sharpened Goldfeld's result as the following: for $N > \exp(4 \ln^{1/2} x \ln_2^{1/2} x)$,

$$\frac{1}{N} \sum_{a \leqslant N} P_a(x) = A \operatorname{li}(x) + O\left(\frac{x}{\ln^D x}\right),$$

where $A = \prod_p (1 - \frac{1}{p(p-1)})$ and D is any constant > 1. Note that A is the average, asymptotically, of the numbers $A(a)$ in Hooley's result.

It is well known that primitive roots exist only for integers $n = 2, 4, p^r$, and $2p^r$ where p is an odd prime, as this is when the group $(\mathbb{Z}/n\mathbb{Z})^*$ is cyclic. However, we can extend the concept as follows, due to Carmichael [1].

DEFINITION 1.1. Suppose that a and n are coprime integers. If a has maximal order modulo n, we call a a primitive λ-root for n.

Note that primitive λ-roots are primitive roots if the modulus n is of the form before Definition 1.1. Here λ refers to the function $\lambda(n)$, the maximal exponent modulo n of integers coprime to n. Let $l_a(n)$ be the order of a modulo n when $\gcd(a, n) = 1$. Then $\lambda(n)$ is the maximal value of such $l_a(n)$ when a runs over all of its possible values. The function $\lambda(n)$ is easily evaluated from the prime factorization of n. Let us use the notation $p^e \parallel n$ to mean that the prime power p^e divides n while p^{e+1} does not. Then $\lambda(n) = \operatorname{lcm}_{p^e \parallel n}\{\lambda(p^e)\}$ where $\lambda(p^e) = \phi(p^e)$ for all prime power p^e (where ϕ is Euler's function), except $\lambda(2^e) = \frac{1}{2}\phi(2^e)$ for $e \geqslant 3$. In [6] one can find results concerning the size of the function $\lambda(n)$.

We can ask similar questions of the distribution of primitive λ-roots as we did for primitive roots for primes.

If we fix the modulus n and count the number of primitive λ-roots a in $[1, n]$ for n, we are led to the function $r(n)$, the proportion of primitive λ-roots modulo n in $(\mathbb{Z}/n\mathbb{Z})^*$. The function $r(n)$ has a closed form [11, 13]

$$r(n) = \prod_{q \mid \lambda(n)} \left(1 - \frac{1}{q^{\Delta_q(n)}}\right), \tag{1}$$

where q runs over primes and $\Delta_q(n)$ is defined as follows: Factor the group $(\mathbb{Z}/n\mathbb{Z})^*$ into the direct sum of cyclic subgroups with prime power orders. Then $\Delta_q(n)$ is the number of the direct summands whose order is the maximal

such power of q. In [11] we proved that the function $r(n)$ does not possess a limiting distribution function in the sense that for the function

$$D(x, u) := \frac{1}{x} \sum_{\substack{n \leqslant x \\ r(n) \leqslant u}} 1,$$

for any real numbers $x > 0$ and u, the limit $\lim_{x \to \infty} D(x, u)$ does not exist for all $u \in (0, u_0)$, for some $u_0 > 0$. The function $D(x, u)$ can be extended as

$$D^{(k)}(x, u) := \frac{1}{\ln_k x} \sum_{\substack{n \leqslant x \\ r(n) \leqslant u}} \frac{1}{n \ln n \cdots \ln_{k-1} n},$$

where $\ln_k x$ is a short form for k-fold iteration of natural log of x. It is not difficult to see that if any such function $D^{(k_0)}(x, u)$ has a limit as x goes to infinity, so do all the functions $D^{(k)}(x, u)$ with $k \geqslant k_0$. We showed that $D^{(2)}(x, u)$ does not exist [12], and we do not yet know whether $\lim_{x \to \infty} D^{(3)}(x, u)$ exists for all u. But we do know that the oscillations of $D^{(3)}(x, u)$ must be "much smoother" than that of $D^{(2)}(x, u)$ if there does exist such an oscillation.

Let a be an integer and $N_a(x)$ denote the number of positive integers n up to x such that a is a primitive λ-root modulo n. We then ask whether or not we have $\lim_{x \to \infty} N_a(x) = \infty$? As $N_a(x) \geqslant P_a(x)$, by Hooley's result, if a is not a square and $\neq -1$, then we do have the above limit assuming ERH. Furthermore if a is a primitive root for p^2, then a is a primitive root for all powers of p. Secondly we may ask whether we would have some asymptotic formula for the function $N_a(x)$. In particular, do we have $N_a(x) \sim B(a)\, x$ for some positive constant $B(a)$ depending on a, perhaps with some exceptional values of a, in analogy with the Artin conjecture?

The purpose of this paper is to show the following theorem on the average of $N_a(x)$, which suggests that $N_a(x)/x$ does not typically tend to a limit.

THEOREM 1.1. *We have that*

$$\overline{\lim_{x \to \infty}} \frac{1}{x^2} \sum_{1 \leqslant a \leqslant x} N_a(x) > 0 \qquad \text{and} \qquad \underline{\lim_{x \to \infty}} \frac{1}{x^2} \sum_{1 \leqslant a \leqslant x} N_a(x) = 0.$$

Remark. It would be desirable to let a run over a smaller interval, as in Stephens' result. We only average $N_a(x)$ for positive values of a because of the following observation. If a is a primitive λ-root modulo n, $-a$ is too unless every odd prime factor of n is $\equiv 3 \bmod 4$ and 8 does not divide n. Since the number of such integers up to x is at most $O(x/\log^{1/2} x)$, one can deduce an estimate for $N_{-a}(x)$ if one knows $N_a(x)$.

One may ask what is the order for each individual function $N_a(x)$. We conjecture that $\overline{\lim}_{x\to\infty} \frac{1}{x} N_a(x) > 0$ and $\underline{\lim}_{x\to\infty} \frac{1}{x} N_a(x) = 0$ for all but certain special values of a. Although we cannot prove the conjecture completely, in a coming paper, we will show that one half of it is true. That is, there exists an unbounded set of numbers x such that on this set, for every integer a, we have $N_a(x) = O(x(\ln_5 x)^{-1/(5e^2)})$ where the implied constant is independent of a.

Theorem 1.1 follows from the following theorem and another result concerning the distribution of $r(n)$, [11].

THEOREM 1.2. *We have*

$$\overline{\lim_{x\to\infty}} D(x, \ln_5^{-c_1} x) = 1$$

for some constant $c_1 > 0$.

This is stronger than Corollary 5.4 in [11], which asserts that $D(x, \ln_5^{-\rho} x) > \delta$ for some positive constants ρ and δ. The proof of Theorem 1.2 is much harder than that of Corollary 5.4. From Theorems 1.1 and 1.2 we can directly retrieve a main result, Theorem 5.5 in [11], which asserts that $\lim_{x\to\infty} D(x, u)$ does not exist for each u in certain interval $(0, u_0)$. Indeed if c_2 is the positive constant in Theorem 1.1, then it immediately follows from Theorem 1.2 that $\overline{\lim}_{x\to\infty} D(x, u) = 1$ for all u, $0 < u \leqslant 1$, and $\underline{\lim}_{x\to\infty} D(x, u) < 1$ for all u, $0 < u < c_2/2$. Thus, $\lim_{x\to\infty} D(x, u)$ does not exist for all u, $0 < u < c_2/2$.

Let $\tilde{f}(n)$ be the sum of $1/q$ for primes q: $q \leqslant \ln_4 n$, $q \mid \lambda(n)$ and $\Delta_q(n) = 1$, except the case that n is too small for $\ln_4 n$ to be defined where $\tilde{f}(n)$ is defined be 0. A major step in proving Theorem 1.2 is to show the following relation between the first and the second moments of $\tilde{f}$,

$$\sum_{n\leqslant x} \tilde{f}^2(n) = \frac{1}{x}\left(\sum_{n\leqslant x} \tilde{f}(n)\right)^2 + O(x),$$

for all $x \geqslant 1$. The relation is proved by comparing the major terms in each sum, instead of estimating the first moment and the second moment separately. Sieve methods and some elementary methods are used to achieve the relation.

Note that from the above relation on $\tilde{f}$ one can imply that $\tilde{f}(N) = (1 + o(1)) \frac{1}{x} \sum_{n\leqslant x} \tilde{f}(n)$ for almost all $N \leqslant x$ and an unbounded set of numbers x.

The author is grateful to Carl Pomerance for asking me to study this problem and for his suggestions that significantly simplified the original proofs. The author also thanks Andrew Granville for making valuable

suggestions. In the paper, the constants named with letter c are singled out by additional subscripts in the order as they appear.

2. AVERAGE ORDER FOR $N_a(x)$

We will prove Theorem 1.1 in this section after the necessary preparation. Let $R(n)$ be the number of primitive λ-roots modulo n in $[1, n]$. Thus, $R(n) = r(n)\,\phi(n)$. It follows from the definition of $N_a(x)$ that

$$\sum_{a \leqslant y} N_a(x) = \sum_{a \leqslant y} \sum_{\substack{n \leqslant x \\ l_a(n) = \lambda(n)}} 1 = \sum_{n \leqslant x} \sum_{\substack{a \leqslant y \\ l_a(n) = \lambda(n)}} 1$$

$$= \sum_{n \leqslant x} \left[\frac{y}{n} \right] R(n) + \sum_{n \leqslant x} \sum_{\substack{1 \leqslant a \leqslant \{y/n\}\, n \\ l_a(n) = \lambda(n)}} 1. \tag{2}$$

It is easy to see that for any $y \geqslant x \geqslant 1$

$$\sum_{a \leqslant y} N_a(x) \geqslant \sum_{n \leqslant x} \left[\frac{y}{n} \right] R(n) \geqslant \frac{y}{2} \sum_{n \leqslant x} \frac{R(n)}{n}$$

and

$$\sum_{a \leqslant y} N_a(x) \leqslant \sum_{n \leqslant x} \left\lceil \frac{y}{n} \right\rceil R(n) \leqslant 2y \sum_{n \leqslant x} \frac{R(n)}{n}$$

Thus we have proved the following lemma.

LEMMA 2.1. *For any $y \geqslant x$ we have*

$$\frac{1}{2} \sum_{n \leqslant x} \frac{R(n)}{n} \leqslant \frac{1}{y} \sum_{a \leqslant y} N_a(x) \leqslant 2 \sum_{n \leqslant x} \frac{R(n)}{n}.$$

Note that from (2) we have $\sum_{a \leqslant y} N_a(x) = y \sum_{n \leqslant x} \frac{R(n)}{n} + O(\sum_{n \leqslant x} R(n))$. But $\sum_{n \leqslant x} R(n) \leqslant x \sum_{n \leqslant x} R(n)/n$. Therefore we have

$$\frac{1}{y} \sum_{a \leqslant y} N_a(x) = \left(1 + O\left(\frac{x}{y} \right) \right) \sum_{n \leqslant x} \frac{R(n)}{n}.$$

When $\lim_{x \to \infty} x/y = 0$, we have $1/y \sum_{a \leqslant y} N_a(x) \sim \sum_{n \leqslant x} \frac{R(n)}{n}$. It would be desirable to let a run over a smaller interval, $y \leqslant \sqrt{x}$, say, in analogy with the Stephens' result mentioned in the introduction.

In [11] we have obtained extreme orders for the function $r(n)$. Using the same method we can prove the following extreme orders for the function $R(n)/n$.

THEOREM 2.2.

$$\overline{\lim_{x\to\infty}}\ R(n)/n = 1 \qquad \text{and} \qquad \underline{\lim_{x\to\infty}}\ \frac{R(n)}{n}(\ln_2 n)^2 = e^{-2\gamma},$$

where γ is Euler's constant.

Proof. (i) Since $R(n)/n \leqslant 1$, we only need to find a sequence of integers n_x such that $\lim_{x\to\infty} n_x = \infty$ and $\lim_{x\to\infty} R(n_x)/n_x = 1$. Let $\mathscr{B}$ be the set of primes $p \leqslant x$ such that $p \equiv 3 \bmod 4$ and $\gcd(p-1, P(x^{1/5})) = 1$, where $P(z) : \prod_{2<p\leqslant z} p$. Then apply Theorem 7.4 in [8] to sieve the set $\mathscr{A} = \{p-1 : p \leqslant x \text{ and } p \equiv 3 \bmod 4\}$ with the set of primes $\mathscr{P} = \{p : 2 < p \leqslant x^{1/5}\}$, taking $\kappa = 1$ and $\alpha = 1/2$. We have

$$\#\mathscr{B} \geqslant \delta \frac{x}{\ln^2 x}$$

for some constant $\delta > 0$ and all $x \geqslant 3$.

Let $p \in \mathscr{B}$. If q is an odd prime factor of $p-1$ then $q > x^{1/5}$. Since $p \leqslant x$, $p-1$ has at most 5 odd prime factors, counting multiplicity. Choose $[\ln x]$ such primes $p_i \in \mathscr{B}$, $i = 1, ..., [\ln x]$. Let $n_x = \prod_{i=1}^{[\ln x]} p_i$. Then by definition and (1)

$$r(n_x) = \prod_{q \mid \lambda(n_x)} \left(1 - \frac{1}{q^{A_q(n_x)}}\right) \geqslant \left(1 - \frac{1}{2^{[\ln x]}}\right)\left(1 - \frac{1}{x^{1/5}}\right)^{5[\ln x]},$$

while

$$\frac{\phi(n_x)}{n_x} = \prod_{i=1}^{[\ln x]} \left(1 - \frac{1}{p_i}\right) \geqslant \left(1 - \frac{1}{x^{1/5}}\right)^{[\ln x]}.$$

Clearly the right sides have limit 1 as $x \to \infty$, and the left sides are $\leqslant 1$. As $\frac{R(n)}{n} = r(n) \cdot \frac{\phi(n)}{n}$, we have $\lim_{x\to\infty} R(n_x)/n_x = \lim_{x\to\infty} r(n_x) \cdot \phi(n_x)/n_x = 1$. We obtain the upper limit.

(ii) Note that

$$r(n) \geqslant \prod_{q \mid \lambda(n)} \left(1 - \frac{1}{q}\right) \geqslant \prod_{q \leqslant N(n)} \left(1 - \frac{1}{q}\right),$$

where $N(n)$ is chosen to be the least number such that the product of the primes up to $N(n)$ is greater than or equal to $\lambda(n)$. From prime number theory, $N(n) = (1 + o(1)) \ln \lambda(n) \leqslant (1 + o(1)) \ln n$. From Mertens' theorem, it follows that $r(n) \geqslant (e^{-\gamma} + o(1))/\ln_2 n$. Thus $R(n)/n = r(n)\,\phi(n)/n \geqslant (1 + o(1))/(e^{2\gamma} \ln_2^2 n)$. In the rest we are going to show that this is the best possible.

For each prime $q < \ln x$, let a be the least integer such that $q^a > \ln x$. Let m be the product of all such q^a. Then, for large x, $x^{1/2} \leqslant (\ln x)^{\pi(\ln x)} \leqslant m \leqslant (\ln x)^{2\pi(\ln x)} \leqslant x^3$. By Linnik's Theorem there exists a prime p_0 such that $m < p_0 \leqslant m^{c_3}$ and $p_0 \equiv 1 \bmod m$, where c_3 is an absolute constant. Let $n'_x = p_0 \prod_{p \leqslant \ln x} p$. Then $x < n'_x \leqslant x^{3c_3 + 2}$ since $x^{1/2} < \prod_{p \leqslant \ln x} p < x^2$ and $x^{1/2} < p_0 \leqslant x^{3c_3}$.

Let $q \leqslant \ln x$ be a prime. If it is a prime factor of $p - 1$ with $p \leqslant \ln x$ then its maximal power in $p - 1$ is less than that in $p_0 - 1$. Thus, by (1), $\Delta_q(n'_x) = 1$ for all $q \leqslant \ln x$. Therefore

$$r(n'_x) = \prod_{q \mid \lambda(n'_x)} \left(1 - \frac{1}{q^{\Delta_q(n'_x)}}\right) \leqslant \prod_{q \leqslant \ln x} \left(1 - \frac{1}{q}\right)$$

and

$$\frac{\phi(n'_x)}{n'_x} = \prod_{p \mid n'_x} \left(1 - \frac{1}{p}\right) \leqslant \prod_{p \leqslant \ln x} \left(1 - \frac{1}{p}\right).$$

By the prime number theorem and the fact that $\ln_2 x = \ln_2 n'_x + O(1)$, we have that

$$\prod_{p \leqslant \ln x} \left(1 - \frac{1}{p}\right) = \frac{1 + o(1)}{e^{\gamma} \ln_2 x} = \frac{1 + o(1)}{e^{\gamma} \ln_2 n'_x}.$$

Our result on the lower bound of $R(n)/n$ follow immediately. ∎

Using Theorem 2.2 one can obtain a lower bound for $\frac{1}{x} \sum_{1 \leqslant a \leqslant x} N_a(x)$ as $\frac{1}{x} \sum_{1 \leqslant a \leqslant x} N_a(x) \gg x/\ln_2^2 x$. However this can be improved as in the next theorem.

THEOREM 2.3. *For a positive constant x_0 we have*

$$\frac{1}{x} \sum_{1 \leqslant a \leqslant x} N_a(x) \gg \frac{x}{\ln_3 x}$$

for all $x \geqslant x_0$.

We need the next lemma to prove Theorem 2.3.

LEMMA 2.4. *For any* $t \in (0, 1)$ *the limit*

$$\lim_{x \to \infty} \frac{1}{x} \left| \left\{ n \leqslant x : \frac{\phi(n)}{n} \leqslant t \right\} \right| = w(t)$$

exists. And the function $w(t)$ *is continuous and strictly increasing in the interval* (0,1) *with* $\lim_{t \to 0^+} w(t) = 0$ *and* $\lim_{t \to 1^-} w(t) = 1$.

Proof. See [17]. ∎

Proof of Theorem 2.3. By Lemma 2.1,

$$\frac{1}{x} \sum_{1 \leqslant a \leqslant x} N_a(x) \geqslant \frac{1}{2} \sum_{n \leqslant x} \frac{R(n)}{n} = \frac{1}{2} \sum_{n \leqslant x} r(n) \cdot \frac{\phi(n)}{n}.$$

By formula (1),

$$r(n) \geqslant \prod_{q \mid \lambda(n)} \left(1 - \frac{1}{q}\right) = \prod_{q \mid \phi(n)} \left(1 - \frac{1}{q}\right).$$

Let S be the set of integers $n \geqslant 1$ such that $\phi(n)$ has at most $\ln_2^2 n$ distinct prime factors. By [5, 4], S has density 1. By an elementary estimate, we have

$$r(n) \gg \frac{1}{\ln_3 n}$$

uniformly for all $n \in S$ and $n \geqslant n_0$ for some $n_0 > 0$.

On the other hand, let $S' = \{n : \phi(n)/n \geqslant 1/2\}$. Then S' has density $1 - \omega(1/2) \in (0, 1)$ by Lemma 2.4. Therefore we have

$$\frac{1}{x} \sum_{1 \leqslant a \leqslant x} N_a(x) \gg \sum_{\substack{n_0 \leqslant n \leqslant x \\ n \in S \cap S'}} \frac{1}{\ln_3 n} \gg \frac{x}{\ln_3 x}.$$

This concludes the proof. ∎

LEMMA 2.5. *There exists a positive constant* c_4 *and an unbounded set of numbers* x *such that*

$$D(x, u) \leqslant \frac{c_4}{|\ln u|}$$

for all u *with* $0 < u < 1$.

Proof. See Corollary 5.2 in [11]. ∎

Proof of Theorem 1.1. Let $u, t \in (0, 1)$ which will be fixed later. We have

$$\sum_{n \leqslant x} \frac{R(n)}{n} = \sum_{n \leqslant x} r(n) \frac{\phi(n)}{n} \geqslant u \sum_{\substack{n \leqslant x \\ r(n) \geqslant u}} \frac{\phi(n)}{n} \geqslant ut \sum_{\substack{n \leqslant x \\ r(n) \geqslant u \\ \phi(n)/n \geqslant t}} 1$$

$$\geqslant ut \left([x] - \sum_{\substack{n \leqslant x \\ r(n) < u}} 1 - \sum_{\substack{n \leqslant x \\ \phi(n)/n < t}} 1 \right) \geqslant ut \left(x - \frac{c_4 x}{|\ln u|} - 2w(t)\, x \right),$$

for x sufficiently large and x in a certain unbounded set, where the last inequality follows from Lemmas 2.4 and 2.5.

By Lemma 2.4, we have $\lim_{t \to 0^+} w(t) = 0$. Thus we can choose u and t small enough to ensure that $c_2 = \frac{ut}{2}(1 - c_4/|\ln u| - 2w(t)) > 0$. By Lemma 2.1 we have that

$$\frac{1}{x} \sum_{1 \leqslant a \leqslant x} N_a(x) \geqslant c_2 x$$

for the numbers x mentioned above.

On the other hand, by Lemma 2.1 again, we have

$$\frac{1}{x} \sum_{1 \leqslant a \leqslant x} N_a(x) \leqslant 2 \sum_{n \leqslant x} \frac{R(n)}{n}.$$

By Theorem 1.2, on an unbounded set of numbers x, we have

$$\sum_{n \leqslant x} \frac{R(n)}{n} \leqslant \sum_{\substack{n \leqslant x \\ r(n) \leqslant \ln_5^{-c_1} x}} \frac{1}{\ln_5^{c_1} x} + \sum_{\substack{n \leqslant x \\ r(n) > \ln_5^{-c_1} x}} 1$$

$$\leqslant \frac{x}{\ln_5^{c_1} x} + x(1 - D(x, \ln_5^{-c_1} x)) = o(x).$$

This concludes our proof. ∎

3. PROOF OF THEOREM 1.2

Theorem 1.2 is equivalent to

$$\underline{\lim}_{x \to \infty} \frac{1}{x} \# \{ n \leqslant x : r(n) > \ln_5^{-c_1} x \} = 0.$$

From (1) we have

$$r(n)=\prod_{q\mid\lambda(n)}\left(1-\frac{1}{q^{\Delta_q(n)}}\right)\leqslant\exp\left(-\sum_{\Delta_q(n)=1}\frac{1}{q}\right).$$

Recall that $\tilde{f}(n)=\sum_{q\leqslant\ln_4 n,\,\Delta_q(n)=1}\frac{1}{q}$ if $n>e^e$, and $\tilde{f}(n):=0$ otherwise. Then $r(n)\leqslant\exp(-\tilde{f}(n))$. A nice fact about the function $\tilde{f}$ is that we can get control of the upper order of its first and second moments as one can see below. This allows us to get some information back for $r(n)$. So we play with $\tilde{f}$ before coming back to Theorem 1.2.

LEMMA 3.1. *Let $\varepsilon\in(0,1/4]$ be any number. There exists an x_0 such that*

$$\sum_{n\leqslant x}\tilde{f}(n)$$
$$=x\sum_{q\leqslant\ln_4 x}\frac{1}{q}\sum_{|k-(\ln_3 x/\ln q)|\leqslant(\varepsilon\ln_3 x/\ln q)}\frac{\ln_2 x}{q^k}\exp\left(-\frac{\ln_2 x}{\phi(q^k)}\right)+O_\varepsilon\left(\frac{x}{\ln_4 x}\right)$$

when $x\geqslant x_0$.

We have earlier dealt with the first moments of the functions $\tilde{f}(n)$ and $f(n)=\sum_{q:\,\Delta_q(n)=1}1/q$ in [11]. We investigated two series almost the same as the one in Lemma 3.1. There we were only interested in lower and upper bounds of the first moments. Here we need a more accurate estimate for the first moment of $\tilde{f}(n)$. As compensation we sacrifice simplicity of the involved series.

LEMMA 3.2. *Let $\varepsilon(0,1/5]$ be any number. There exists an x_0 such that*

$$\sum_{n\leqslant x}\tilde{f}^2(n)=x\sum_{q_1,\,q_2\leqslant\ln_4 x}{\sum_{k_i}}'\frac{\ln_2^2 x}{q_1^{k_1+1}q_2^{k_2+1}}\exp\left(-\sum_{j=1,\,2}\frac{\ln_2 x}{\phi(q_j^{k_j})}\right)+O(x),$$

for all $x\geqslant x_0$, where $\sum'_{k_i}$ means that the sum is taken over all the k_i with $|k_i-\ln_3 x/\ln q_i|\leqslant\varepsilon\ln_3 x/\ln q_i$.

We leave the proofs of the two lemmas to the next two sections.

COROLLARY 3.3. *We have for all $x\geqslant1$,*

$$\sum_{n\leqslant x}\tilde{f}^2(n)=\frac{1}{x}\left(\sum_{n\leqslant x}\tilde{f}(n)\right)^2+O(x).$$

Proof. Taking $\varepsilon = 1/5$ in Lemma 3.1 we have

$$\sum_{n \leqslant x} \tilde{f}(n) = x \sum_{q \leqslant \ln_4 x} \frac{1}{q} \sum_{k}' \frac{\ln_2 x}{q^k} \exp\left(-\frac{\ln_2 x}{\phi(q^k)}\right) + O\left(\frac{x}{\ln_4 x}\right),$$

where $\sum_k'$ means that the sum taking over all the k in the interval $[\frac{4 \ln_3 x}{5 \ln q}, \frac{6 \ln_3 x}{5 \ln q}]$. Square both sides. Note that

$$\sum_{q \leqslant \ln_4 x} \frac{1}{q} \sum_{k \geqslant 1} \frac{\ln_2 x}{q^k} \exp\left(-\frac{\ln_2 x}{\phi(q^k)}\right) = O\left(\sum_{q \leqslant \ln_4 x} \frac{1}{q}\right) = O(\ln_6 x).$$

Thus we have $\frac{1}{x}(\sum_{n \leqslant x} \tilde{f}(n))^2$ is equal to

$$x \sum_{q_1, q_2 \leqslant \ln_4 x} \sum_{k_i}' \frac{\ln_2^2 x}{q_1^{k_1+1} q_2^{k_2+1}} \exp\left(-\sum_{j=1,2} \frac{\ln_2 x}{\phi(q_j^{k_j})}\right) + O\left(\frac{x \ln_6 x}{\ln_4 x}\right),$$

where $\sum_{k_i}'$ carries similar meaning as above. The main term of the above is exactly the same as that of $\sum_{n \leqslant x} \tilde{f}^2(n)$ given in Lemma 3.2 with the corresponding $\varepsilon = 1/5$. Therefore we have proved Corollary 3.3. ∎

COROLLARY 3.4. *There is an unbounded set of numbers x such that on the set we have*

$$\sum_{n \leqslant x} (\tilde{f}(n) - c_5 \ln_6 x)^2 = o(x \ln_6^2 x)$$

for some constant $c_5 > 0$, depending on the unbounded set of x.

Proof. By Theorem 5.1 in [11] there exists an unbounded set S of numbers x such that on the set S

$$\sum_{n \leqslant x} \tilde{f}(n) \geqslant bx \ln_6 x$$

for some constant $b > 0$. On the other hand it is clear that $\tilde{f}(n) \leqslant \ln_6 n + O(1) \leqslant 2 \ln_6 x$ for all $n \leqslant x$ and x sufficiently large, and so $\sum_{n \leqslant x} \tilde{f}(n) \leqslant 2x \ln_6 x$. Thus for all large $x \in S$.

$$\sum_{n \leqslant x} \tilde{f}(n) = b_x x \ln_6 x$$

for some $b_x \in [b, 2]$. By compactness of the interval $[b, 2]$ the sequence $\{b_x : x \in S\}$ has a limit point c_5 in the interval. Without loss of generality we assume that

$$\lim_{\substack{x \in S \\ x \to \infty}} b_x = c_5.$$

Then when $x \in S$, by Corollary 3.3, we have

$$\sum_{n \leqslant x} (\tilde{f}(n) - c_5 \ln_6 x)^2 = (b_x - c_5)^2 \, x \ln_6^2 x + O(x) = o(x \ln_6^2 x). \quad \blacksquare$$

Proof of Theorem 1.2. As we noted at the beginning of the section, Theorem 1.2 is equivalent to that $|\{n \leqslant x : r(n) > \ln_5^{-c_5/2} x\}| = o(x)$ on an unbounded set of x. Let S be the unbounded set of numbers x in Corollary 3.4. Then by Corollary 3.4 we have

$$\frac{1}{x} \left| \left\{ n \leqslant x : \tilde{f}(n) \leqslant \frac{c_5}{2} \ln_6 x \right\} \right| \to 0$$

as $x \in S$ goes to infinity. Since $r(n) \leqslant \exp(-\tilde{f}(n))$,

$$\{n \leqslant x : r(n) > \ln_5^{-c_5/2} x\} \subset \left\{ n \leqslant x : \tilde{f}(n) \leqslant \frac{c_5}{2} \ln_6 x \right\}.$$

Let $c_1 = c_5/2$. Our theorem follows from the above argument. $\blacksquare$

4. A FEW RESULTS OF SIEVE METHODS

In this section we give a few results in the form which are best for our applications in the following sections. These results are basic and crucial to our later arguments. One may want to refer to [8] for other or general results of sieve methods. First let us introduce some notations of sieve methods.

Let $\mathscr{A}$ be the set of positive integers up to $x \geqslant 1$. Throughout the paper set $\mathscr{A}$ will always be of this type. Let $\mathscr{P}$ be a set of primes. Define $P(z) := \prod_{p \leqslant z, p \in \mathscr{P}}$ for any $z \leqslant x$, and $W(z) := \prod_{p \leqslant z, p \in \mathscr{P}} (1 - 1/p)$. We are interested in an estimate for the counting function

$$S(\mathscr{A}, \mathscr{P}, z) := \sum_{\substack{n \in \mathscr{A} \\ (n, P(z)) = 1}} 1.$$

LEMMA 4.1 (See [16, 15]). *For any integer $k \geqslant 2$ and any $x \geqslant 2$,*

$$\sum_{\substack{p \leqslant x \\ p \equiv 1 \bmod k}} \frac{1}{p} = \frac{\ln_2 x}{\phi(k)} + O\left(\frac{\ln k}{\phi(k)}\right),$$

where the implied constant is uniform and effectively computable.

Proof (based on the proof in [16]). If $x \leqslant k$, the left side of the above formula is 0 and the formula is trivially true. Thus in the following we always assume $x > k$.

Let $\pi(t, k, 1)$ denote the number of primes p up to t so that $p \equiv 1 \bmod k$. Trivially we have

$$\pi(t, k, 1) < \frac{t}{k} \tag{3}$$

for all $t \geqslant 2$. By partial summation, we have

$$\sum_{\substack{p \leqslant x \\ p \equiv 1 \bmod k}} \frac{1}{p} = \frac{\pi(x, k, 1)}{x} + \int_k^x \frac{\pi(t, k, 1)}{t^2}\, dt. \tag{4}$$

Assume $k < x \leqslant e^{k^2}$. First we have, by (3), that for all $x \geqslant 2$,

$$0 \leqslant \frac{1}{x}\pi(x, k, 1) \leqslant \frac{1}{k}. \tag{5}$$

Next we claim that

$$\int_k^{e^{k^2}} \frac{\pi(t, k, 1)}{t^2}\, dt \leqslant \frac{9 \ln k}{\phi(k)}. \tag{6}$$

To see this let us note that by the Montgomery–Vaughan version of the Brun–Titchmarsh inequality [14], for $t \geqslant k^2$,

$$\pi(t, k, 1) < \frac{4t}{\phi(k) \ln t}.$$

Thus

$$\int_{k^2}^{e^{k^2}} \frac{\pi(t, k, 1)}{t^2}\, dt \leqslant \frac{8 \ln k}{\phi(k)}.$$

Using (3), we have

$$\int_k^{k^2} \frac{\pi(t,k,1)}{t^2}\,dt < \int_k^{k^2} \frac{dt}{kt} = \frac{\ln k}{k}.$$

Our claim (6) follows from the two estimates. So follows the lemma for the case $k < x \leqslant e^{k^2}$ from (4), (5) and (6).

When $x > e^{k^2}$ we have, from (4),

$$\sum_{\substack{p \leqslant x \\ p \equiv 1 \bmod k}} \frac{1}{p} = \frac{\pi(x,k,1)}{x} + \int_k^{e^{k^2}} \frac{\pi(t,k,1)}{t^2}\,dt + \int_{e^{k^2}}^{x} \frac{\pi(t,k,1)}{t^2}\,dt. \tag{7}$$

By (11) on p. 123 of Davenport [2], there exists an absolute and computable constant A_1 such that for all $t > e^{k^2}$

$$\left|\pi(t,k,1) - \frac{t}{\phi(k)\ln t}\right| < \frac{A_1 t}{\phi(k)\ln^2 t}. \tag{8}$$

Since

$$\int_{e^{k^2}}^{x} \frac{dt}{\phi(k)\,t\ln t} = \frac{\ln_2 x}{\phi(k)} - \frac{2\ln k}{\phi(k)},$$

we have, by (8), that

$$\begin{aligned}\left|\int_{e^{k^2}}^{x} \frac{\pi(t,k,1)}{t^2}\,dt - \frac{\ln_2 x}{\phi(k)}\right| &< \frac{2\ln k}{\phi(k)} + \int_{e^{k^2}}^{x} \frac{A_1\,dt}{\phi(k)\,t\ln^2 t} \\ &< \frac{2\ln k}{\phi(k)} + \frac{A_1}{k^2\phi(k)}.\end{aligned} \tag{9}$$

The lemma for the case $x \geqslant e^{k^2}$ follows from (7), (5), (6) and (9). This concludes the proof. ∎

THEOREM 4.2. *Let $\mathscr{P}$ be a set of primes. Assume that $2 \leqslant z \leqslant t$. Let $u = \ln t/\ln z$. Then we have*

$$\sum_{\substack{n \leqslant t \\ (n,\,P(z))=1}} 1 = tW(z)\,(1 + O(\exp(-\tfrac{1}{2}u\ln u)) + O(\exp(-\sqrt{\ln t}))),$$

where $W(z)$ is defined as above, and the implied constants are absolute.

Proof. This lemma follows from a direct application of Theorem 7.2 in [8] to $S(\mathscr{A}, \mathscr{P}, z)$ where $\mathscr{A} = \{n \leqslant t\}$. ∎

LEMMA 4.3. *Let $\mathscr{P}$ be a set of primes and assume $\varepsilon > 0$ is a number depending on $\mathscr{P}$ such that*

$$\sum_{\substack{w < p \leqslant ew \\ p \in \mathscr{P}}} \frac{1}{p} \leqslant \frac{\varepsilon}{\ln w}$$

for all $w \geqslant w_0$, w_0 depending on $\mathscr{P}$. Then if $z = \exp(\ln t / \ln_2^2 t) \geqslant w_0$ and $\varepsilon > \exp(-\ln_2^2 t \ln_3 t)$, we have

$$\sum_{\substack{m \leqslant t \\ (m, P(t)) = 1}} 1 = tW(z) + O(\varepsilon t \ln_3 t \cdot W(z)) + O\left(\frac{\varepsilon t}{\ln_2^2 t}\right),$$

where the implied constants are absolute and $W(z)$ is the same as defined at the beginning of the section.

Proof. First we have

$$\sum_{\substack{m \leqslant t \\ (m, P(t)) = 1}} 1 = \sum_{\substack{m \leqslant t \\ (m, P(z)) = 1}} 1 + O\left(\sum_{\substack{m \leqslant t \\ (m, P(z)) = 1 \\ (m, P(t)) > 1}} 1\right). \tag{10}$$

The sum within the parenthesis is no bigger than $E_1 + E_2$ where

$$E_1 = \sum_{\substack{z < p \leqslant t/z \\ p \in \mathscr{P}}} \sum_{\substack{m \leqslant t/p \\ (m, P(z)) = 1}} 1 \quad \text{and} \quad E_2 = \sum_{\substack{t/z < p \leqslant t \\ p \in \mathscr{P}}} \sum_{\substack{m \leqslant t/p \\ (m, P(t/p)) = 1}} 1.$$

Note that in E_1 we have $t/p \geqslant z$. The inner sum of E_1 is $S(\mathscr{A}, \mathscr{P}, z)$ where $\mathscr{A} = \{n : n \leqslant t/p\}$. Applying Theorem 4.2 to this sum, we can rewrite (10) as

$$\sum_{\substack{m \leqslant t \\ (m, P(t)) = 1}} 1 = \sum_{\substack{m \leqslant t \\ (m, P(z)) = 1}} 1 + O\left(W(z) \sum_{\substack{z < p \leqslant t/z \\ p \in \mathscr{P}}} \frac{t}{p}\right) + O\left(\sum_{\substack{t/z < p \leqslant t \\ p \in \mathscr{P}}} \frac{t}{p}\right),$$

where the implied constants are absolute. We claim that the two error terms are equal to the ones in the statement of the lemma. Because $z \geqslant w_0$, by the conditions of the lemma, we have

$$\begin{aligned} \sum_{\substack{z < p \leqslant t/z \\ p \in \mathscr{P}}} \frac{1}{p} &\leqslant \frac{\varepsilon}{\ln z} + \frac{\varepsilon}{\ln z + 1} + \cdots + \frac{\varepsilon}{\ln z + k} \\ &\leqslant \varepsilon \int_{\ln z - 1}^{\ln z + k} \frac{du}{u} \leqslant \varepsilon \int_{\ln z - 1}^{\ln(t/z)} \frac{du}{u} \ll \varepsilon \ln_3 t, \end{aligned}$$

where $e^k z \leqslant t/z < e^{k+1} z$. The same method applies to the other error term. Thus we have

$$\sum_{\substack{m \leqslant t \\ (m, P(t))=1}} 1 = \sum_{\substack{m \leqslant t \\ (m, P(z))=1}} 1 + O(\varepsilon t \ln_3 t W(z)) + O\left(\frac{\varepsilon t}{\ln_2^2 t}\right). \tag{11}$$

By Theorem 4.2, the main term of (11) is equal to

$$tW(z)(1 + O(\exp(-\ln_2^2 t \ln_3 t)),$$

where the implied constant is absolute. Since we assume that $\varepsilon > \exp(-\ln_2^2 t \ln_3 t)$, the above error can be taken in the first error in (11). Therefore we have proved Lemma 4.3. ∎

The condition in Lemma 4.3 is crucial to its application. It is not our intention to discuss the condition for general sets $\mathscr{P}$ of primes. We are interested in investigating the condition for the set of primes in an arithmetic progression, because this is where our applications of Lemma 4.3 are.

LEMMA 4.4. *Let $m \geqslant 2$ be an integer. For all $w \geqslant m^{12}$ we have*

$$\sum_{\substack{w < p \leqslant ew \\ p \equiv 1 \bmod m}} \frac{1}{p} \leqslant \frac{6}{\phi(m) \ln w}.$$

Proof. By the Montgomery–Vaughan version of the Brun–Titchmarsh inequality [14], we have

$$\sum_{\substack{w < p \leqslant ew \\ p \equiv 1 \bmod m}} \frac{1}{p} \leqslant \frac{1}{w} \sum_{\substack{w < p \leqslant ew \\ p \equiv 1 \bmod m}} 1 \leqslant \frac{1}{w} \frac{2ew}{\phi(m) \ln(ew/m)}$$

$$< \frac{2e}{\phi(m) \ln w} \cdot \frac{12}{11} < \frac{6}{\phi(m) \ln w}. \quad \blacksquare$$

5. PRIME FACTORIZATION OF $\lambda(n)$

Let q be a fixed prime and m be an integer. Let $v_q(m)$ be the index of q in the prime factorization of m. That is $q^{v_q(m)} \| m$.

THEOREM 5.1. *Let ε be a number in the interval $(0,1)$ and let $q \leqslant \ln_2^{\varepsilon/2} x$ be a prime. Then for $x \geqslant 16$*

$$\#\left\{n \leqslant x : \left|v_q(\lambda(n)) - \frac{\ln_3 x}{\ln q}\right| > \varepsilon \frac{\ln_3 x}{\ln q}\right\} = O\left(\frac{x}{\ln_2^{\varepsilon} x}\right),$$

where the O-constant depends only on ε.

Proof. We prove the theorem in two parts. Let $K = \varepsilon \frac{\ln_3 x}{\ln q}$.

1. $\#\{n \leqslant x : v_q(\lambda(n)) < \frac{\ln_3 x}{\ln q} - K\} = O(x \exp(-\ln_2^{\varepsilon/2} x))$.

Let $K_1 = \lceil \frac{\ln_3 x}{\ln q} - K \rceil$. Since $q \leqslant \ln_2^{\varepsilon/2} x$ we have that $K \geqslant 2$ and $K_1 \geqslant 1$ when x is sufficiently large. Then

$$\begin{aligned}
&\#\left\{n \leqslant x : v_q(\lambda(n)) < \frac{\ln_3 x}{\ln q} - K\right\} \\
&\quad = \#\{n \leqslant x : v_q(\lambda(n)) < K_1\} \\
&\quad \leqslant \#\{n \leqslant x : p \not\equiv 1 \bmod q^{K_1} \text{ for all } p \mid n\} \\
&\quad = S(\mathscr{A}, \mathscr{P}_{q^{K_1}}, x),
\end{aligned}$$

where $\mathscr{A} = \{n \leqslant x\}$ and $\mathscr{P}_{q^{K_1}} = \{p : q^{K_1} \mid p-1\}$.

Applying Theorem 4.2 and Lemma 4.1, we have that $S(\mathscr{A}, \mathscr{P}_{q^{K_1}}, x)$ is

$$\begin{aligned}
&\ll x \prod_{\substack{p \leqslant x \\ q^{K_1} \mid p-1}} \left(1 - \frac{1}{p}\right) = x \exp\left(-\sum_{\substack{p \leqslant x \\ q^{K_1} \mid p-1}} \frac{1}{p} + O\left(\frac{1}{q^{2K_1}}\right)\right) \\
&= x \exp\left(-\frac{\ln_2 x}{\phi(q^{K_1})} + O\left(\frac{K_1 \ln q}{q^{K_1}}\right)\right) \ll x \exp\left(-\frac{\ln_2 x}{q^{K_1}(1-1/q)}\right) \\
&\leqslant x \exp\left(-\frac{\ln_2 x}{q^{K_1}}\right) \leqslant x \exp(-q^{K-1}) \leqslant x \exp(-\ln_2^{\varepsilon/2} x).
\end{aligned}$$

We have proved this case.

2. $\#\{n \leqslant x : v_q(\lambda(n)) > \frac{\ln_3 x}{\ln q} + K\} = O(x/\ln_2^{\varepsilon} x)$.

Let $K_2 = [\frac{\ln_3 x}{\ln q} + K]$. Then our counting function takes the form $\#\{n \leqslant x : v_q(q^{K_1}(n)) > K_2\}$, which is

$$\leqslant \#\{n \leqslant x : q^{K_2+2} \mid n\} + \#\{n \leqslant x : q^{K_2+1} \mid p-1 \text{ for some } p \mid n\}.$$

Then first term is bounded by $x/q^{K_2+2} \leqslant x/(q^{K+1} \ln_2 x)$. The second term is bounded by

$$\sum_{\substack{p \leqslant x \\ q^{K_2+1} \mid p-1}} \frac{x}{p} = x\left(\frac{\ln_2 x}{q^{K_2+1}(1-1/q)} + O\left(\frac{K_2 \ln q}{q^{K_2+1}}\right)\right) \ll \frac{x}{\ln_2^{\varepsilon} x}.$$

To see the last inequality, note that $K_2 + 1 \geqslant (1+\varepsilon)\frac{\ln_3 x}{\ln q}$. Thus the first term in the above big parentheses is bounded by $x/\ln_2^{\varepsilon} x$. For the second term we have

$$\frac{K_2 \ln q}{q^{K_2+1}} \leqslant \frac{(1+\varepsilon)\ln_3 x}{\ln_2^{1+\varepsilon} x} \leqslant \frac{1}{\ln_2^{\varepsilon} x}$$

if x is large enough. Thus our claim is true in this case. We are done. ∎

6. PROOF OF LEMMA 3.1

It follows from the definition of $\tilde{f}(n)$ that

$$\sum_{n \leqslant x} \tilde{f}(n) = \sum_{n \leqslant x} \sum_{\substack{q \leqslant \ln_4 n \\ \Delta_q(n)=1}} \frac{1}{q} = \sum_{q \leqslant \ln_4 x} \frac{1}{q} \sum_{\substack{n \leqslant x \\ \Delta_q(n)=1}} 1 + O\left(\frac{x}{\ln_4 x}\right) \tag{12}$$

as the difference between the two sums in the second equality is bounded by

$$\sum_{n \leqslant x} \sum_{\ln_4 n < q \leqslant \ln_4 x} \frac{1}{q} = \sum_{n \leqslant x^{1/2}} \sum_{\ln_4 n < q \leqslant \ln_4 x} \frac{1}{q} + \sum_{x^{1/2} < n \leqslant x} \sum_{\ln_4 n < q \leqslant \ln_4 x} \frac{1}{q}$$

$$\ll x^{1/2} \ln_6 x + \frac{x}{\ln_4 x^{1/2}} \ll \frac{x}{\ln_4 x}.$$

Applying Theorem 5.1 to the inner sum on the right side of (12), we can write $\sum_{n \leqslant x} \tilde{f}(n)$ as

$$\sum_{q \leqslant \ln_4 x} \frac{1}{q} \sum_{\substack{n \leqslant x \\ \Delta_q(n)=1 \\ |v_q(\lambda(n)) - \ln_3 x/\ln q| \leqslant \varepsilon \ln_3 x/\ln q}} 1 + O\left(\frac{x \ln_6 x}{\ln_2^{\varepsilon} x}\right) + O\left(\frac{x}{\ln_4 x}\right), \tag{13}$$

where $\varepsilon \in (0, 1)$ will be determined later.

In the same way as in the proof of Lemma 3.4 in [11], the inner sum of (13) is equal to

$$\sum_{|k-\ln_3 x/\ln q|\leqslant \varepsilon \ln_3 x/\ln q} \sum_{\substack{p\leqslant x^{1/4}\\ q^k \parallel p-1}} \sum_{\substack{m\leqslant x/p\\ (m,\, P_q k(x/p))=1}} 1 + O\left(\frac{x\ln_3 x}{\ln_2^{1-\varepsilon} x}\right).$$

Put this in (13). We have

$$\sum_{n\leqslant x} \tilde{f}(n) = \sum_{q\leqslant \ln_4 x} \frac{1}{q} \sum_k{}' \sum_{\substack{p\leqslant x^{1/4}\\ q^k \parallel p-1}} \sum_{\substack{m\leqslant x/p\\ (m,\, P_q k(x/p))=1}} 1 + E(x), \tag{14}$$

where $\sum_k'$ is the sum over all the k with $|k-\frac{\ln_3 x}{\ln q}|\leqslant \frac{\varepsilon \ln_3 x}{\ln q}$, and

$$E(x) = O\left(\frac{x\ln_6 x}{\ln_2^{\varepsilon} x}\right) + O\left(\frac{x\ln_3 x\ln_6 x}{\ln_2^{1-\varepsilon} x}\right) + O\left(\frac{x}{\ln_4 x}\right)$$

which is $\ll x\ln_4^{-1} x$ when $x\geqslant x_\varepsilon$, for some x_ε.

Let $t = x/p$ with $p\leqslant x^{1/4}$. Then $x^{3/4}\leqslant t\leqslant x$. Let $\varepsilon\leqslant 1/2$ so that our q and k satisfy $\ln_2^{1/2} x\leqslant q^k\leqslant \ln_2^{3/2} x$. Thus we have $\ln_2^{1/2} t\leqslant q^k\leqslant \ln_2^2 t$ when x is sufficiently large. By Lemmas 4.3 and 4.4, with $\mathscr{P} = \{p\colon p\equiv 1 \bmod q^k\}$, we have

$$\sum_{\substack{m\leqslant t\\ (m,\, P(t))=1}} 1 = t\left[\, W(z) + O\left(\frac{\ln_3 t}{q^k}\, W(z) + \frac{1}{q^k\ln_2^2 t}\right)\right], \tag{15}$$

where $z = \exp(\ln t/\ln_2^2 t)$. Recall that $W(z) = \prod_{p<z,\, p\in\mathscr{P}}(1-p^{-1})$. Using Lemma 4.1 and the facts that $\ln_2^{1/2} x\leqslant q^k\leqslant \ln_2^{3/2} x$ and $\ln_2 t = \ln_2 x + O(1)$, we have the following result:

$$\begin{aligned} W(z) &= \exp\left(-\frac{\ln_2 z}{\phi(q^k)} + O\left(\frac{\ln q^k}{q^k}\right)\right) = \exp\left(-\frac{\ln_2 x}{\phi(q^k)} + O\left(\frac{\ln q^k}{q^k}\right)\right)\\ &= \left(1 + O\left(\frac{\ln q^k}{q^k}\right)\right)\exp\left(-\frac{\ln_2 x}{\phi(q^k)}\right). \end{aligned}$$

The first application of the formula is to simplify the error within the big-O term of (15) as follows

$$\frac{\ln_3 t}{q^k}\, W(z) + \frac{1}{q^k\ln_2^2 t} \ll \frac{q^k\ln_3 x}{\ln_2^2 x}.$$

Here we have used the estimate $\exp(-\ln_2 x/\phi(q^k)) \ll q^{2k}/\ln_2^2 x$ and $\ln_2 t = \ln_2 x + O(1)$. Secondly we can write $W(z)$ as

$$W(z) = \exp\left(-\frac{\ln_2 x}{\phi(q^k)}\right) + O\left(\frac{q^k \ln_3 x}{\ln_2^2 x}\right),$$

by the same estimate for $\exp(-\ln_2 x/\phi(q^k))$ and $\ln q^k = O(\ln_3 x)$. Thus, by (15), we have

$$\sum_{\substack{m \leqslant x/p \\ (m,\, P_q k(x/p))=1}} 1 = \frac{x}{p}\left(\exp\left(-\frac{\ln_2 x}{\phi(q^k)}\right) + O\left(\frac{q^k \ln_3 x}{\ln_2^2 x}\right)\right).$$

When we put this formula in (14), the error generated is

$$= O\left(\sum_{q \leqslant \ln_4 x} \frac{1}{q} \sum_k{}' \sum_{\substack{p \leqslant x^{1/4} \\ q^k \parallel p-1}} \frac{x}{p} \cdot \frac{q^k \ln_3 x}{\ln_2^2 x}\right) \ll x \sum_{q \leqslant \ln_4 x} \frac{1}{q} \sum_k{}' \frac{\ln_2 x}{q^k} \cdot \frac{q^k \ln_3 x}{\ln_2^2 x}$$

$$\leqslant x \sum_{q \leqslant \ln_4 x} \frac{1}{q} \cdot \frac{2\varepsilon \ln_3 x}{\ln q} \cdot \frac{\ln_3 x}{\ln_2 x} \ll \frac{x \ln^2{}_3 x}{\ln_2 x}$$

by Lemmas 4.1.

Therefore we can rewrite (14) as

$$\sum_{n \leqslant x} \tilde{f}(n) = x \sum_{q \leqslant \ln_4 x} \frac{1}{q} \sum_k{}' \sum_{\substack{p \leqslant x^{1/4} \\ q^k \parallel p-1}} \frac{1}{p} \exp\left(-\frac{\ln_2 x}{\phi(q^k)}\right) + O\left(\frac{x}{\ln_4 x}\right).$$

Applying Lemma 4.1 to $\sum_{p \leqslant x^{1/4},\, q^k \parallel p-1} \frac{1}{p}$, we obtain the main term exactly as in Lemma 3.1 while we can put the generated error term in $O(x/\ln_4 x)$.

7. PROOF OF LEMMA 3.2

By definition we have

$$\sum_{n \leqslant x} \tilde{f}^2(n) = \sum_{n \leqslant x} \sum_{\substack{q_i \leqslant \ln_4 n \\ \Delta_{q_i}(n)=1}} \frac{1}{q_1 q_2}$$

$$= \sum_{\substack{i=1,2 \\ q_i \leqslant \ln_4 x}} \frac{1}{q_1 q_2} \sum_{\substack{n \leqslant x \\ \Delta_{q_i}(n)=1}} 1 + O\left(\frac{x \ln_6 x}{\ln_4 x}\right)$$

since the difference between the two sums of the second equality is bounded by

$$2 \sum_{n \leqslant x} \sum_{\substack{q_1 \leqslant \ln_4 x \\ \ln_4 n \leqslant q_2 \leqslant \ln_4 x}} \frac{1}{q_1 q_2} \ll \frac{x \ln_6 x}{\ln_4 x}.$$

When $q_1 = q_2$ we have $\sum_{q \leqslant \ln_4 x} \frac{1}{q^2} \sum_{n \leqslant x} 1 = O(x)$. Hence

$$\sum_{n \leqslant x} \tilde{f}^2(n) = \sum_{\substack{q_i \leqslant \ln_4 x \\ q_1 \neq q_2}} \frac{1}{q_1 q_2} \sum_{\substack{n \leqslant x \\ \Delta_{q_i}(n) = 1}} 1 + O(x). \tag{16}$$

In the remaining part of this section we will show that, for $x \geqslant x_0$ with x_0 being a constant, the inner sum of (16) is equal to

$$\sum_{\substack{n \leqslant x \\ \Delta_{q_1}(n) = 1 \\ \Delta_{q_2}(n) = 1}} 1 = x \sum_{k_1, k_2}' \frac{\ln_2^2 x}{q_1^{k_1} q_2^{k_2}} \exp\left(- \sum_{i=1,2} \frac{\ln_2 x}{\phi(q_i^{k_i})} \right) + O\left(\frac{x}{\ln_2^{\varepsilon} x} \right), \tag{17}$$

where $\varepsilon \in (0, 1/4]$ is fixed and the symbol $'$ in $\sum'_{k_1, k_2}$ indicates that k_1 and k_2 are subject to the condition $|k_i - \frac{\ln_3 x}{\ln q_i}| \leqslant \frac{\varepsilon \ln_3 x}{\ln q_i}$. Lemma 3.2 follows immediately from (16) and (17).

Our strategy to prove (17) is to write the inner sum of (16) explicitly in terms of sums. Then simplify it. For briefness let $P_i(t) := \prod_{p \leqslant t,\, q_i^{k_i} \| p-1} p$.

Step 1. An explicit formula for $\sum_{n \leqslant x,\, \Delta_{q_i}(n) = 1} 1$.

Since $\ln_4 x < \ln_2^{\varepsilon} x$ for x sufficiently large, by Theorem 5.1 the sum in question is equal to

$$\#\{n \leqslant x : \Delta_{q_i}(n) = 1 \ \&\ H(q_1, q_2)\} + O\left(\frac{x}{\ln_2^{\varepsilon} x} \right)$$

where the number $\varepsilon \in (0, 1/2)$ will be chosen later, and the condition

$$H(q_1, q_2) : \left| v_{q_i}(\lambda(n)) - \frac{\ln_3 x}{\ln q_i} \right| \leqslant \frac{\varepsilon \ln_3 x}{\ln q_i} \qquad \text{for} \quad i = 1, 2.$$

By analyzing the main term of the above formula, we claim that

$$\sum_{\substack{n \leqslant x \\ \Delta_{q_1}(n) = \Delta_{q_2}(n) = 1}} 1 = \sum_{k_1, k_2}' \sum_{\substack{p_1,\, p_2 \leqslant x \\ q_i^{k_i} \| p_i - 1}} \sum_{r_1, r_2 \geqslant 1} \sum_{\substack{m \leqslant x / p_1^{r_1} p_2^{r_2} \\ (m,\, P_1(x/p_1^{r_1} p_2^{r_2})\, P_2(x/p_1^{r_1} p_2^{r_2})) = 1}} 1$$

$$+ \sum_{k_1, k_2}' \sum_{\substack{p \leqslant x \\ q_i^{k_i} \| p - 1}} \sum_{\substack{m \leqslant x/p \\ (m,\, P_1(x/p)\, P_2(x/p)) = 1}} 1 + O\left(\frac{x}{\ln_2^{\varepsilon} x} \right). \tag{18}$$

The second triple sum on the right of (18) counts the number of $n \leqslant x$, subject to $H(q_1, q_2)$, such that $p_1 = p_2 = p$ and $p \parallel n$.

Let q be a prime and let $k = v_q(\lambda(n))$. By definition of $\Delta_q(n)$ (see (1)), if $\Delta_q(n) = 1$, then either $q^{k+1} \mid n$ or n contains only one prime factor p such that $q^k \mid p-1$. Conversely if n contains only one prime factor p such that $q^k \mid p-1$, then either $\Delta_q(n) = 1$ or $q^{k+1} \mid n$.

For q_1 fixed, the number of n up to x subject to the conditions that $|k_1 - \frac{\ln_3 x}{\ln q_1}| \leqslant \varepsilon \frac{\ln_3 x}{\ln q_1}$ and $q_1^{k_1+1} \mid n$ is bounded by $O(x/(q_1 \ln_2^{1-\varepsilon} x)) = O(x/\ln_2^{\varepsilon} x)$ since we assume that $\varepsilon \in (0, 1/2)$. The same bound holds if we switch the role of q_1 and q_2. The number of n up to x, such that n has only one prime p_1 with $v_{q_1}(p_1 - 1) = v_{q_1}(\lambda(n))$ and only one prime p_2 with $v_{q_2}(p_2 - 1) = v_{q_2}(\lambda(n))$ where both $v_{q_i}(\lambda(n))$ are subject to the condition $|v_{q_i}(\lambda(n)) - \frac{\ln_3 x}{\ln q_i}| \leqslant \varepsilon \frac{\ln_3 x}{\ln q_i}$, is counted by the four fold sum and the three fold sum in (18). The numbers n counted by the above two cases exhaust all the numbers n counted by the main term of (17). Thus we have proved (18).

We are going to simplify the nasty four fold sum and the three fold sum in a few steps. Also we want to trace the error terms that pop out. Remember that the integers k_i always fall in the range $[\frac{(1-\varepsilon)\ln_3 x}{\ln q_i}, \frac{(1+\varepsilon)\ln_3 x}{\ln q_i}]$.

Step 2. The three-fold sum in (18) is equal to $O(\frac{x}{\ln_2^{1-2\varepsilon} x})$.

By Lemma 4.1 the three fold sum is

$$\leqslant \sum_{k_1, k_2}{}' \sum_{\substack{p \leqslant x \\ q_1^{k_1} q_2^{k_2} \mid p-1}} \frac{x}{p} = x \sum_{k_1, k_2}{}' \frac{\ln_2 x + O(k_1 \ln q_1 + k_2 \ln q_2)}{\phi(q_1^{k_1} q_2^{k_2})}$$

$$\leqslant x \frac{\ln_2 x + O(\ln_3 x)}{\ln_2^{2-2\varepsilon} x} = O\left(\frac{x}{\ln_2^{1-2\varepsilon} x}\right).$$

Step 3. The contribution of the terms with $r_1 > 1$ or $r_2 > 1$ in (18) is $O(\frac{x}{\ln_2^{2-3\varepsilon} x})$.

For sake of simplicity let us use $\sum_m''$ to denote the sum over m subject to the condition $(m, P_1(\frac{x}{p_1^{r_1} p_2^{r_2}})\, P_2(\frac{x}{p_1^{r_1} p_2^{r_2}})) = 1$. Write

$$\sum_{r_1, r_2 \geqslant 1} \sum_{m \leqslant x/p_1^{r_1} p_2^{r_2}}{}'' 1 = \sum_{m \leqslant x/p_1 p_2}{}'' 1 + \sum_{r_1 + r_2 \geqslant 3} \sum_{m \leqslant x/p_1^{r_1} p_2^{r_2}}{}'' 1.$$

Let E denote the second sum. Then

$$E \leqslant \sum_{r_1 + r_2 \geqslant 3} \frac{x}{p_1^{r_1} p_2^{r_2}} \ll \frac{x}{p_1 p_2^2} + \frac{x}{p_1^2 p_2}.$$

But

$$\sum_{k_1,k_2}' \sum_{\substack{p_1,p_2\leqslant x\\ q_i^{k_i}\parallel p_i-1}} \frac{x}{p_1 p_2^2} \ll x \sum_{k_i}' \frac{\ln_2 x}{q_1^{k_1} q_2^{2k_2}} \ll \frac{x}{\ln_2^{2-3\varepsilon} x}.$$

Thus $\sum'_{k_1,k_2} \sum_{p_1,p_2\leqslant x, q_i^{k_i}\parallel p_i-1} E = O(x/\ln_2^{2-3\varepsilon} x)$ and (18) can be written as

$$\sum_{\substack{n\leqslant x\\ \Delta_{q_1}(n)=\Delta_{q_2}(n)=1}} 1 = \sum_{k_i}' \sum_{\substack{p_i\leqslant x\\ q_i^{k_i}\parallel p_i-1}} \sum_{m\leqslant x/p_1 p_2}'' 1 + O\left(\frac{x}{\ln_2^{\varepsilon} x}\right), \tag{19}$$

since the error generated by E and the error in Step 2 can be taken in the above error if we choose $\varepsilon < 1/3$.

Step 4. The contribution from the terms in (19) with p_1 or $p_2 > x^{1/4}$ is at most $O(x/\ln_2^{1-2\varepsilon} x)$. We have, by Lemmas 4.1 and 4.4, that

$$\sum_{k_i}' \sum_{\substack{x^{1/4}<p_1\leqslant x\\ p_2\leqslant x\\ q_i^{k_i}\parallel p_i-1}} \sum_{m\leqslant x/p_1 p_2}'' 1 \leqslant \sum_{k_i}' \sum_{\substack{x^{1/4}<p_1\leqslant x\\ p_2\leqslant x\\ q_i^{k_i}\parallel p_i-1}} \frac{x}{p_1 p_2}$$

$$\ll x \sum_{k_i}' \frac{\ln_2 x}{q_1^{k_1} q_2^{k_2}} \ll \frac{x}{\ln_2^{1-2\varepsilon} x}.$$

This bound holds when we switch the role of p_1 and p_2. Substituting the estimate in (19), we have

$$\sum_{\substack{n\leqslant x\\ \Delta_{q_1}(n)=\Delta_{q_2}(n)=1}} 1 = \sum_{k_i}' \sum_{\substack{p_i\leqslant x^{1/4}\\ q_i^{k_i}\parallel p_i-1}} \sum_{m\leqslant x/p_1 p_2}'' 1 + O\left(\frac{x}{\ln_2^{\varepsilon} x}\right). \tag{20}$$

Step 5. Simplification of the main term in (20).

Let $t = x/(p_1 p_2)$ with $p_i \leqslant x^{1/4}$. Then we have $x^{1/2} \leqslant t \leqslant x$. Let us choose ε in the interval $(0,1/5]$ so that $\ln_2^{4/5} x \leqslant q_i^{k_i} \leqslant \ln_2^{6/5} x$ and so $\ln_2^{4/5} t \leqslant q_i^{k_i} \leqslant 2\ln_2^{6/5} t$ if x is sufficiently large.

We can apply Lemmas 4.3 and 4.4 to the innermost sum in (20) to get

$$\sum_{m\leqslant x/p_1 p_2}'' 1 = tW(z) + O\left(t \ln_3 t \cdot W(z) \cdot \sum_{i=1}^{2} \frac{1}{q_i^{k_i}}\right) + O\left(\frac{t}{\ln_2^2 t} \sum_{i=1}^{2} \frac{1}{q_i^{k_i}}\right)$$

where $z = \exp(\ln t/\ln_2^2 t)$ and $W(z) = \prod_{l \leqslant z,\, q_1^{k_1} \text{ or } q_2^{k_2} \mid l-1}(1-1/l)$, the letter l running over primes. The above expression can be written as

$$tW(z) + O\left(\frac{t \ln_3 x}{\ln_2^{1-\varepsilon} x} W(z)\right) + O\left(\frac{t}{\ln_2^{3-\varepsilon} x}\right).$$

To compute $W(z)$ let us note that, by Lemma 4.1 and the relations among $q_i^{k_i}$, z, t and x, we have

$$\begin{aligned}\sum_{\substack{p \leqslant z \\ q_1^{k_1} \text{ or } q_2^{k_2} \mid p-1}} \frac{1}{p} &= \sum_{i=1,2} \frac{\ln_2 z + O(\ln q_i^{k_i})}{\phi(q_i^{k_i})} + O\left(\frac{\ln_2 z + O(\ln q_1^{k_1} q_2^{k_2})}{\phi(q_1^{k_1} q_2^{k_2})}\right) \\ &= \sum_{i=1,2} \frac{\ln_2 t}{\phi(q_i^{k_i})} + O\left(\frac{1}{\ln_2^{1-2\varepsilon} t}\right) = \sum_{i=1,2} \frac{\ln_2 x}{\phi(q_i^{k_i})} + O\left(\frac{1}{\ln_2^{1-2\varepsilon} x}\right).\end{aligned}$$

Hence

$$W(z) = \exp\left(-\sum_{i=1,2} \frac{\ln_2 x}{\phi(q_i^{k_i})}\right)\left(1 + O\left(\frac{1}{\ln_2^{1-2\varepsilon} x}\right)\right).$$

Put the above estimates together. We have proved that, for x sufficiently large,

$$\sum_{m \leqslant x/p_1 p_2}'' 1 = t \exp\left(-\sum_{i=1,2} \frac{\ln_2 x}{\phi(q_i^{k_i})}\right) + O\left(\frac{t}{\ln_2^{1-2\varepsilon} x}\right) \tag{21}$$

Put (21) into (20). The first term provides the main term for $\sum_{n \leqslant x,\, \Delta_{q_1}(n) = \Delta_{q_2}(n) = 1} 1$, which is

$$x \sum_{k_i}' \sum_{\substack{p_i \leqslant x^{1/4} \\ q_i^{k_i} \parallel p_i - 1}} \frac{1}{p_1 p_2} \exp\left(-\sum_{i=1,2} \frac{\ln_2 x}{\phi(q_i^{k_i})}\right). \tag{22}$$

By Lemma 4.1, we have

$$\sum_{\substack{p_i \leqslant x^{1/4} \\ q_i^{k_i} \parallel p_i - 1}} \frac{1}{p_1 p_2} = \prod_{i=1,2} \frac{\ln_2 x + O(k_i \ln q_i)}{q_i^{k_i}} = \frac{\ln_2^2 x + O(\ln_2 x \ln_3 x)}{q_1^{k_1} q_2^{k_2}}.$$

Summing over k_1, k_2, the error in this estimate is $O(x \ln_3 x/\ln_2^{1-2\varepsilon} x)$.

What is left is to estimate accumulation of the error of (21) in (20). It is bounded by

$$\frac{x}{\ln_2^{1-2\varepsilon} x} \sum_{k_i}' \sum_{\substack{p_i \leqslant x^{1/4} \\ q_i^{k_i} \| p_i - 1}} \frac{1}{p_1 p_2} \ll \frac{x}{\ln_2^{1-2\varepsilon} x} \sum_{k_i}' \frac{\ln_2^2 x}{q_1^{k_1} q_2^{k_2}} \ll \frac{x}{\ln_2^{1-4\varepsilon} x}.$$

by Lemma 4.1.

Assuming $\varepsilon \in (0, 1/5]$, the above estimates of the error, and (22) yield that

$$\sum_{\substack{n \leqslant x \\ \Delta_{q_1}(n) = \Delta_{q_2}(n) = 1}} 1 = x \sum_{k_1, k_2}' \frac{\ln_2^2 x}{q_1^{k_1} q_2^{k_2}} \exp\left(- \sum_{i=1,2} \frac{\ln_2 x}{\phi(q_i^{k_i})} \right) + O\left(\frac{x}{\ln_2^{\varepsilon} x} \right).$$

Our lemma then follows by substituting this formula into (16).

REFERENCES

1. R. D. Carmichael, "The Theory of Numbers," Wiley, New York, 1914.
2. H. Davenport, "Multiplicative Number Theory," second ed., Springer-Verlag, New York, 1980.
3. P. D. T. A. Elliot, "Probabilistic Number Theory," Vol. II, Springer-Verlag, New York, 1980.
4. P. Erdös, A. Granville, G. Pomerance, and C. Spiro, On the normal behavior of the iterates of some arithmetic functions, *in* "Analytic Number Theory, Allerton Park, IL, 1989," Progr. Math., Vol. 85, pp. 165–204, Birkhäuser, Boston, 1990.
5. P. Erdős and C. Pomerance, On the normal number of prime factors of $\phi(n)$, *Rocky Mountain Math. J.* **15** (1985), 343–352.
6. P. Erdős, C. Pomerance, and E. Schmutz, Carmichael's lambda function, *Acta Arith.* **58** (1991), 363–385.
7. M. Goldfeld, Artin's conjecture on the average, *Mathematika* **15** (1968), 223–226.
8. H. Halberstam, and H. E. Richert, "Sieve Methods," Academic Press, New York, 1974.
9. C. Hooley, On Artin's conjecture, *J. Reine Angew. Math.* **225** (1967), 209–220.
10. W. J. LeVeque, "Topics in Number Theory," Vol. I, Addison–Wesley, Reading, MA, 1956.
11. S. Li, On the number of elements with maximal order in the multiplicative group modulo n, *Acta Arith.* **86** (1998), 113–132.
12. S. Li, "On Artin's Conjecture for Composite Moduli," Doctorate dissertation, University of Georgia, 1998.
13. G. Martin, The least prime primitive root and the shifted sieve, *Acta Arith.* **80** (1997), 277–288.
14. H. L. Montgomery and R. C. Vaughan, The large sieve, *Mathematika* **20** (1973), 119–134.
15. K. K. Norton, On the number of restricted prime factors of an integer I, *Illinois J. Math.* **20** (1976), 681–705.
16. C. Pomerance, On the distribution of amicable numbers, *J. Reine Angew. Math.* **293/294** (1977), 217–222.

17. I. J. Schoenberg, On asymptotic distributions of arithmetical functions, *Trans. Amer. Math. Soc.* **39** (1936), 315–330.
18. P. J. Stephens, An average result for Artin's conjecture, *Mathematika* **16** (1969), 178–188.
19. A. E. Western and J. C. P. Miller, "Tables of Indices and Primitive Roots," Royal Society Mathematical Tables, Vol. 9, pp. xxxviii, Cambridge Univ. Press, Cambridge, UK, 1968.