ACTA ARITHMETICA
120.3 (2005)

Roughly squarefree values of the Euler and Carmichael functions

by

William D. Banks (Columbia, MO) and Florian Luca (Morelia)

1. Introduction. Let φ denote the Euler function, whose value at an integer $n \ge 1$ is given by

$$\varphi(n) = \prod_{p^a \,\|\, n} p^{a-1}(p-1). \tag{1}$$

Recall that an integer m is said to be *squarefree* if $p^2 \nmid m$ for any prime p. Using (1), it is easy to see that if $m = \varphi(n)$ is squarefree, then the following properties hold:

- If a prime p divides n, then $p-1$ is squarefree.
- $p^3 \nmid n$ for any prime p.
- If $4 \mid n$, then $p \nmid n$ for any odd prime p (and thus, $n = 4$).
- If $4 \nmid n$, then $p \mid n$ for at most one odd prime p.

These properties imply that $n \in \{2, 4, p, 2p, p^2, 2p^2\}$ for some prime $p > 2$ such that $p-1$ is squarefree. Hence, the problem of estimating the number of integers $n \le x$ for which $\varphi(n)$ is squarefree reduces to that of estimating the number of primes $p \le x$ for which $p-1$ is squarefree. These questions have been previously investigated in [9], where it is shown that for any constant $A > 0$, the asymptotic relation

$$\#\{p \le x : p-1 \text{ is squarefree}\} = \alpha\, \pi(x) + O\left(\frac{x}{\log^A x}\right) \tag{2}$$

holds (see also [8]), and consequently,

$$\#\{n \le x : \varphi(n) \text{ is squarefree}\} = \frac{3\alpha}{2}\, \pi(x) + O\left(\frac{x}{\log^A x}\right). \tag{3}$$

2000 *Mathematics Subject Classification*: Primary 11N37.

Here, α is the *Artin constant* (see, for example, [3, 7]):

$$\alpha = \prod_p \left(1 - \frac{1}{p(p-1)}\right) = 0.373956\ldots.$$

As is clear from the analysis above, the prime $p = 2$ plays a crucial role in the proof of (3) by limiting the number of distinct odd primes that can divide any integer n for which $\varphi(n)$ is squarefree. A similar idea has been exploited in [2] to establish an asymptotic expression for the number of positive integers $n \le x$ for which $\varphi(n)$ is free of kth powers.

Now consider the problem of estimating the number of positive integers $n \le x$ for which the *odd part* of $\varphi(n)$ is squarefree (in this case, we say that $m = \varphi(n)$ is *oddly squarefree*). This problem is clearly more complicated in that, by disregarding the power of 2 that divides $\varphi(n)$, one can no longer control the number of distinct odd primes dividing n.

More generally, for a real number $y > 0$, let $\mathcal{N}(y)$ denote the set of natural numbers n with the property that $p^2 \nmid n$ for any prime $p > y$. We say that n is *y-squarefree* if $n \in \mathcal{N}(y)$. In particular, $\mathcal{N}(1)$ is the set of squarefree natural numbers, and $\mathcal{N}(2)$ is the set of oddly squarefree natural numbers. It is easy to see that the set of y-squarefree numbers has an asymptotic density equal to $\prod_{p>y}(1 - 1/p^2)$. Our goal in this paper is to derive estimates for the cardinality of the set

$$\mathcal{F}_y(x) = \{n \le x : \varphi(n) \in \mathcal{N}(y)\}.$$

We also consider the problem of estimating the cardinality of the set

$$\mathcal{L}_y(x) = \{n \le x : \lambda(n) \in \mathcal{N}(y)\}.$$

Here, $\lambda(n)$ denotes the Carmichael function, which is defined for an integer $n \ge 1$ as the largest possible order of any element in the multiplicative group of integers modulo n. More explicitly, for a prime power p^a, one has

$$\lambda(p^a) = \begin{cases} p^{a-1}(p-1) & \text{if } p \ge 3 \text{ or } a \le 2, \\ 2^{a-2} & \text{if } p = 2 \text{ and } a \ge 3, \end{cases}$$

and for an arbitrary integer $n \ge 2$ with prime factorization $n = p_1^{a_1} \cdots p_k^{a_k}$, one has

$$\lambda(n) = \operatorname{lcm}[\lambda(p_1^{a_1}), \ldots, \lambda(p_k^{a_k})].$$

Clearly, $\lambda(1) = 1$.

In what follows, we use the Landau symbols O, o, and $\asymp$, and the Vinogradov symbols $\ll$ and $\gg$ with their usual meanings. Recall that, for positive functions F and G, the notations $F \ll G$, $F \gg G$ and $F = O(G)$ are all equivalent, and $F \asymp G$ is equivalent to $F \ll G \ll F$.

For an integer $k \ge 1$ and a real number $x > 0$, we write $\log_k x$ for the recursively defined function given by $\log_1 x = \max\{\ln x, 1\}$ and $\log_k x =$

$\max\{\ln(\log_{k-1} x), 1\}$ for $k \geq 2$, where $\ln x$ denotes the natural logarithm. When $k = 1$, we omit the subscript with the understanding that $\log x \geq 1$ for all $x > 0$.

The letters p and q are always used to denote prime numbers. As usual, we denote by $\pi(x)$ the number of primes $p \leq x$, and for coprime integers $l, k \geq 1$ we denote by $\pi(x; k, l)$ the number of primes $p \leq x$ that satisfy the congruence $p \equiv l \pmod{k}$.

Acknowledgements. We thank the anonymous referee for remarks that improved the quality of this paper and for suggesting the questions and problems that appear in the last section of this paper. Most of this work was done during a visit by W. B. to the Universidad Nacional Autónoma de México, and during visits by both authors to Macquarie University; the hospitality and support of both of these institutions are gratefully acknowledged. During the preparation of this paper, W. B. was supported in part by NSF grant DMS-0070628, and F. L. was supported in part by grant PAPIIT IN104505.

2. y-Squarefree values of $\varphi(n)$. As in the introduction, we define

$$\mathcal{F}_y(x) = \{n \leq x : \varphi(n) \in \mathcal{N}(y)\},$$

where $\mathcal{N}(y)$ is the set of natural numbers n such that $p^2 \nmid n$ for any prime $p > y$. Let

$$r(x, y) = \log_2 x \prod_{y < p \leq \log_2 x} \left(1 - \frac{1}{p-1}\right).$$

Here, and in what follows, an empty product is taken to be 1, as usual. Since the estimate

$$\prod_{p \leq t} \left(1 - \frac{1}{p-1}\right) = \frac{c}{\log t}\left(1 + O\left(\frac{1}{\log t}\right)\right),$$

holds as $t \to \infty$ for some positive constant c, it follows that

$$r(x, y) = \frac{\log_2 x \log y}{\log_3 x}\left(1 + O\left(\frac{1}{y} + \frac{1}{\log_3 x}\right)\right)$$

uniformly for $2 \leq y \leq \log_2 x$.

The main result of this section is the following:

Theorem 1. *Uniformly for x and $y \geq 2$, we have*

$$\#\mathcal{F}_y(x) = \frac{x}{\log x} \exp\left(r(x, y)\left(1 + O\left(\frac{\log_4 x}{\log_3 x}\right)\right)\right).$$

Our proofs of both the upper and the lower bound are rather intricate and rely on standard results from multiplicative number theory, including the study of shifted primes free of prime factors from certain intervals, the use of sieves, and various averaging techniques. Several of the arguments presented here use variations of techniques that are already present in the literature, e.g. in [4], where similar techniques are used to study the average value of the Carmichael function. However, as we did not find specific arguments in the literature which can be directly applied to our problem, we develop these ideas here in some detail.

We begin with the following lemma:

LEMMA 1. *Let* $\pi_1(x,y,z)$ *be the number of primes* $p \le x$ *with the property that if a prime* q *divides* $p-1$, *then either* $q \le y$, *or* $q > z$ *and* $q^2 \nmid p-1$. *Then, uniformly for* $\max\{y,z\} \le \frac{1}{3}\log x$ *and* $z \to \infty$, *the estimate*

$$\pi_1(x,y,z) = f(y,z)\pi(x) + O\biggl(\frac{x}{z\log z\log x}\biggr)$$

holds, where

$$f(y,z) = \prod_{y<p\le z}\biggl(1 - \frac{1}{p-1}\biggr).$$

Proof. For each $d \ge 1$, let $\mathcal{A}_d = \{p \le x : p \equiv 1 \pmod d\}$. If $\mathcal{B}$ is the set of primes $p \le x$ such that $p-1$ is coprime to $R = \prod_{y<q\le z} q$, then

$$\begin{aligned}
\#\mathcal{B} &= \sum_{d\mid R}\mu(d)\#\mathcal{A}_d = \sum_{d\mid R}\mu(d)\pi(x;d,1)\\
&= \sum_{d\mid R}\mu(d)\,\frac{\pi(x)}{\varphi(d)} + O\biggl(\sum_{d\mid R}\biggl|\pi(x;d,1) - \frac{\pi(x)}{\varphi(d)}\biggr|\biggr)\\
&= f(y,z)\pi(x) + O\biggl(\frac{x}{\log^3 x}\biggr) = f(y,z)\pi(x) + O\biggl(\frac{x}{z\log z\log x}\biggr),
\end{aligned}$$

where we have used the Bombieri–Vinogradov Theorem together with the fact that

$$R \le \prod_{q\le z} q = \exp(z(1+o(1))) \le x^{1/3+o(1)} \le x^{2/5}$$

when x is sufficiently large. On the other hand, if $\mathcal{C}$ is the set of primes $p \in \mathcal{B}$ such that $q^2 \mid p-1$ for some $q > z$, then using the Brun–Titchmarsh Theorem, we have

$$\begin{aligned}
\#\mathcal{C} &\le \sum_{z<q\le x^{1/2}} \pi(x;q^2,1) \ll \sum_{z<q\le x^{1/2}} \frac{x}{q^2\log(2x/q^2)}\\
&\ll \frac{x}{\log x}\sum_{z<q\le x^{1/3}}\frac{1}{q^2} + x^{1/3}\sum_{x^{1/3}<q\le x^{1/2}} 1 \ll \frac{x}{z\log z\log x}.
\end{aligned}$$

Since $\pi_1(x,y,z) = \#\mathcal{B} - \#\mathcal{C}$, we obtain the stated bound. ∎

Proof of Theorem 1. The range of y. We first note that it suffices to assume that $y \le \log_2^2 x$. Indeed, if $y > \log_2^2 x$, then the bound asserted by Theorem 1 is

$$x \exp\left(O\left(\frac{\log_2 x \log_4 x}{\log_3 x}\right)\right) = x^{1+o(1)}.$$

On the other hand, it is easy to see that $\#\mathcal{F}_y(x) = (1+o(1))x$. Indeed, let us count the complement of $\mathcal{F}_y(x)$ in $[1,x]$, that is, the set consisting of those positive integers $n \le x$ such that $p^2 \mid \varphi(n)$ for some $p > y$. Clearly, every such integer n must be of one of the following types:

- $p^3 \mid n$ for some $p > y$. The number of such $n \le x$ is at most
$$\sum_{p>y} \frac{x}{p^3} \ll \frac{x}{y^2} = o(x).$$
- $p^2 \mid n$ and $p \mid q-1$ for some $q \mid n$, where $p > y$. The number of such $n \le x$ is at most
$$\sum_{p>y} \sum_{\substack{q \le x/p^2 \\ p \mid q-1}} \frac{x}{p^2 q} \le x \sum_{p>y} \frac{1}{p^2} \sum_{\substack{q<x \\ q \equiv 1 \pmod p}} \frac{1}{q}$$
$$\ll x \log_2 x \sum_{p>y} \frac{1}{p^3} \ll \frac{x \log_2 x}{y^2} = o(x).$$
- $p^2 \mid q-1$ for some $q \mid n$, where $p > y$. The number of such $n \le x$ is at most
$$\sum_{p>y} \sum_{\substack{q \le x \\ p^2 \mid q-1}} \frac{x}{q} \ll x \log_2 x \sum_{p>y} \frac{1}{p^2} \ll \frac{x \log_2 x}{y} = o(x).$$
- There exist two distinct prime factors q_1 and q_2 of n with $q_1 \equiv q_2 \equiv 1 \pmod p$ for some $p > y$. In this last and most numerous case, the number of such $n \le x$ is bounded by
$$\sum_{p>y} \sum_{\substack{q_1 \equiv q_2 \equiv 1 \pmod p \\ q_1 < q_2 < x}} \frac{x}{q_1 q_2} \ll x \sum_{p>y} \Bigg(\sum_{\substack{q<x \\ q \equiv 1 \pmod p}} \frac{1}{q} \Bigg)^2$$
$$\ll x \log_2^2 x \sum_{p>y} \frac{1}{p^2} \ll \frac{x \log_2^2 x}{y \log y} = o(x).$$

Hence, from now on, we can assume that $y \le \log_2^2 x$.

Lower bound. Let x be a large real number, put $z = \log_2 x \log_3^5 x$ and $k = \lfloor f(y,z) \log_2 x \rfloor$. Note that $f(y,z) = 1$ if $y \in [z, \log_2^2 x]$, and that

$$\frac{\log_2 x}{\log_3 x} \ll k \le \log_2 x$$

for all y in our range. Put $w = \exp(\log_2^2 x)$, $v = x^{1/(6k)}$, and let $\mathcal{I}$ be the closed interval $\mathcal{I} = [w, v]$.

Let $\mathcal{P}$ be the set of primes $p \in \mathcal{I}$ with the property that if a prime $q > y$ divides $p - 1$, then $q > z$ and $q^2 \nmid p - 1$. Since $z \le \frac{1}{3}\log w$ if x is sufficiently large, by Lemma 1, it follows that

$$\pi_1(t, y, z) = f(y, z)\pi(t) + O\left(\frac{t}{z\log z\log t}\right)$$

uniformly for all $t \in \mathcal{I}$. Using partial summation, we derive that

$$\begin{aligned}\sum_{p\in\mathcal{P}}\frac{1}{p} &= \int_w^v \frac{d\pi_1(t,y,z)}{t}\\ &= \frac{\pi_1(t,y,z)}{t}\bigg|_{t=w}^{t=v} + f(y,z)\int_w^v \frac{\pi(t)}{t^2}\,dt + O\left(\frac{1}{z\log z}\int_w^v \frac{1}{t\log t}\,dt\right)\\ &= f(y,z)\left(\log_2 v - \log_2 w + O\left(\frac{1}{\log w}\right)\right) + O\left(\frac{1}{\log w} + \frac{\log_2 v}{z\log z}\right)\\ &= f(y,z)\log_2 v\left(1 + O\left(\frac{\log_2 w}{\log_2 v} + \frac{\log z}{\log w\log_2 v} + \frac{1}{z}\right)\right)\\ &= f(y,z)\log_2 x\left(1 + O\left(\frac{\log_3 x}{\log_2 x}\right)\right),\end{aligned}$$

where we used the fact that $f(y, z) \gg 1/\log z$.

Let $\mathcal{Q}$ be the subset of $\mathcal{P}$ obtained by removing from $\mathcal{P}$ those primes p for which $p-1$ has more than $\log_2^2 x$ distinct prime factors. Let $K = \lfloor \log_2^2 x\rfloor + 1$. Since

$$\sum_{\substack{p\le x\\ \omega(p-1)>\log_2^2 x}} \frac{1}{p} \le \sum_{j\ge K}\frac{1}{j!}\left(\sum_{q\le x}\frac{1}{q}\right)^j \ll \left(\frac{e\log_2 x + O(1)}{\log_2^2 x}\right)^K = o\left(\frac{1}{\log x}\right),$$

while $f(y, z) \gg 1/\log z \gg 1/\log_3 x$, it follows that

$$\sum_{p\in\mathcal{Q}}\frac{1}{p} = f(y,z)\log_2 x\left(1 + O\left(\frac{\log_3 x}{\log_2 x}\right)\right). \tag{4}$$

Let also $\widetilde{\mathcal{Q}}$ be the set of powers of primes from $\mathcal{Q}$. Clearly,

$$\begin{aligned}\sum_{p^a\in\widetilde{\mathcal{Q}}}\frac{1}{p^a} &= \sum_{p\in\mathcal{Q}}\frac{1}{p} + O\left(\sum_{p\ge w}\frac{1}{p^2}\right) = \sum_{p\in\mathcal{Q}}\frac{1}{p} + O\left(\frac{1}{w}\right)\\ &= f(y,z)\log_2 x\left(1 + O\left(\frac{\log_3 x}{\log_2 x}\right)\right).\end{aligned} \tag{5}$$

Now let $\mathcal{M}$ be the set of squarefree natural numbers m with precisely k prime factors, each one lying in $\mathcal{Q}$, with the property that $\varphi(m) \in \mathcal{N}(z^2)$.

Note that if $m \in \mathcal{M}$, and $p^2 \mid \varphi(m)$ for some prime $p > y$, then $p \in [z, z^2]$. For every positive integer $m \in \mathcal{M}$ we write $d(m)$ for the largest divisor of m such that $\varphi(d(m))$ lies in $\mathcal{N}(y)$; clearly, $\varphi(d(m)) \in \mathcal{N}(z)$. Let $\mathcal{D}$ be the set of all numbers d such that $d = d(m)$ for some $m \in \mathcal{M}$.

Let d be a fixed element of $\mathcal{D}$; observe that $d = d(m) \le m \le x^{1/6}$ for some $m \in \mathcal{M}$, and therefore $d < x^{1/4} < (x/d)^{1/3}$. Let $\mathcal{P}_d$ be the set of primes P with the properties:

- $x^{1/4} < P \le x/d$.
- If a prime $q > y$ divides $P - 1$, then $q > z$ and $q^2 \nmid P - 1$.
- If a prime q divides $\gcd(P - 1, \varphi(d))$, then $q \le y$.

Now let n be an integer of the form $n = dP$, where $d \in \mathcal{D}$ and $P \in \mathcal{P}_d$. Note that $n \le x$. Since $P > d$, it follows that P is the largest prime factor of n. This shows that d and P are uniquely determined by n; hence, the integers $n \le x$ constructed in this way are pairwise distinct. Since $\varphi(n) = \varphi(d)(P - 1)$, the conditions on P guarantee that $\varphi(n) \in \mathcal{N}(y)$; therefore,

$$\#\mathcal{F}_y(x) \ge \sum_{d \in \mathcal{D}} \#\mathcal{P}_d. \tag{6}$$

To estimate $\#\mathcal{P}_d$, let us first observe that the number of primes $P \le x/d$ such that either $P \le x^{1/4}$, or $q^2 \mid P - 1$ for some $q > z$, is bounded above by

$$\begin{aligned}\pi(x^{1/4}) + \sum_{q > z} \pi(x/d; q^2, 1) &\ll \pi(x^{1/4}) + \frac{x}{d \log x} \sum_{z < q < (x/d)^{1/3}} \frac{1}{q^2} \\ &\quad + \frac{x}{d} \sum_{q \ge (x/d)^{1/3}} \frac{1}{q^2} \\ &\ll \pi(x^{1/4}) + \frac{x}{dz \log z \log x} + \left(\frac{x}{d}\right)^{2/3} \\ &\ll \frac{x}{dz \log z \log x},\end{aligned}$$

where we used the fact that $x/d \ge x^{5/6} > \pi(x^{1/4}) z \log z \log x$ and also $(x/d)^{1/3} \ge x^{5/18} > z \log z \log x$, if x is large enough. Thus, writing

$$R_d = \prod_{\substack{q > y \\ q \le z \text{ or } q \mid \varphi(d)}} q,$$

we see that

$$\#\mathcal{P}_d = \sum_{d_1 \mid R_d} \mu(d_1) \pi(x/d; d_1, 1) + O\left(\frac{x}{dz \log z \log x}\right) \tag{7}$$

$$
\begin{aligned}
\text{(7) [cont.]}\qquad &= \sum_{d_1 | R_d} \mu(d_1)\, \frac{\pi(x/d)}{\varphi(d_1)} \\
&\quad + O\biggl(\sum_{d_1 | R_d} \biggl| \pi(x/d; d_1, 1) - \frac{\pi(x/d)}{\varphi(d_1)} \biggr| + \frac{x}{dz \log z \log x} \biggr) \\
&= g(d)\pi(x/d) + O\biggl(\frac{x}{dz \log z \log x} \biggr),
\end{aligned}
$$

where

$$
g(d) = \prod_{\substack{q>y \\ q\le z \text{ or } q|\varphi(d)}} \biggl(1 - \frac{1}{q-1}\biggr).
$$

Here, we have used the Bombieri–Vinogradov Theorem together with the fact that $R_d \le \varphi(d) < d < (x/d)^{1/3}$.

We now remark that $g(d) \gg 1/\log z$. Indeed, $\varphi(d)$ has no more than $k \log_2^2 x \le \log_2^3 x$ distinct prime factors larger than y, and every such prime is larger than z by construction. Since $z > \log_2 x$, from the Prime Number Theorem, it follows that the number of prime factors of $\varphi(d)$ that are larger than z cannot exceed the number of primes in the interval $[z, z^4]$ if x is sufficiently large. Thus,

$$
\begin{aligned}
\prod_{\substack{q>y \\ q|\varphi(d)}} \biggl(1 - \frac{1}{q-1}\biggr) &\gg \exp\biggl(- \sum_{z \le p \le z^4} \frac{1}{p} + O\biggl(\sum_{q>z} \frac{1}{q^2} \biggr) \biggr) \\
&= \exp(-\log 4 + o(1)) \ge 0.2,
\end{aligned}
$$

and therefore $g(d) \gg f(y,z) \gg 1/\log z$.

Since

$$
\pi(x/d) \gg \frac{x}{d \log(x/d)} \gg \frac{x}{d \log x},
$$

from (7) we deduce that

$$
\#\mathcal{P}_d \gg \frac{x}{d \log z \log x}.
$$

Using this estimate in (6) and summing over all $d \in \mathcal{D}$, we obtain

$$
\#\mathcal{F}_y(x) \gg \frac{x}{\log z \log x} \sum_{d \in \mathcal{D}} \frac{1}{d}. \tag{8}
$$

To complete the proof of the lower bound in the theorem, it suffices to find a suitable lower bound for the sum

$$
S_{\mathcal{D}} = \sum_{d \in \mathcal{D}} \frac{1}{d}.
$$

To do this, we begin by showing that the following estimate holds:

$$S_{\mathcal{M}} = \sum_{m\in\mathcal{M}} \frac{1}{m} \gg S, \tag{9}$$

where

$$S = \frac{1}{k!}\biggl(\sum_{p\in\mathcal{Q}} \frac{1}{p}\biggr)^{k}.$$

Using the multinomial formula, it is easy to see that (9) follows from the two estimates

$$\frac{1}{(k-2)!}\biggl(\sum_{p^a\in\widetilde{\mathcal{Q}}} \frac{1}{p^a}\biggr)^{k-2} \sum_{p\in\mathcal{Q}} \frac{1}{p^2} = o(S) \tag{10}$$

and

$$\frac{1}{(k-2)!}\biggl(\sum_{p^a\in\widetilde{\mathcal{Q}}} \frac{1}{p^a}\biggr)^{k-2} \sum_{q>z^2} \sum_{\substack{p_1,p_2\in\mathcal{Q}\\ p_1\equiv p_2\equiv 1 \ (\mathrm{mod}\ q)}} \frac{1}{p_1p_2} = o(S). \tag{11}$$

Indeed, the estimate (10) implies that the main contribution to S comes from the sum S^* of the reciprocals of *squarefree* numbers composed of k primes from the set $\mathcal{Q}$, while the estimate (11) implies that the main contribution to S^* comes from integers m lying in $\mathcal{M}$ rather than integers m for which $\varphi(m) \notin \mathcal{N}(z^2)$. Concerning (10), using (4) and (5), we obtain

$$\begin{aligned}\frac{1}{(k-2)!}\biggl(\sum_{p^\alpha\in\widetilde{\mathcal{Q}}} \frac{1}{p^\alpha}\biggr)^{k-2} \sum_{p\in\mathcal{Q}} \frac{1}{p^2} &\ll Sk^2\biggl(\frac{1}{f(y,z)\log_2 x}\biggr)^2 \frac{1}{w\log w} \\ &\ll \biggl(\frac{k}{f(y,z)\log_2 x}\biggr)^2 \frac{S}{w\log w} \ll \frac{S}{w\log w} = o(S),\end{aligned}$$

where we have used the fact that $k \asymp f(y,z)\log_2 x$. Concerning (11), if we combine the same argument with Mertens' Theorem, it follows that

$$\begin{aligned}&\frac{1}{(k-2)!}\biggl(\sum_{p^\alpha\in\widetilde{\mathcal{Q}}} \frac{1}{p^\alpha}\biggr)^{k-2} \sum_{q>z^2} \sum_{\substack{p_1,p_2\in\mathcal{Q}\\ p_1\equiv p_2\equiv 1 \ (\mathrm{mod}\ q)}} \frac{1}{p_1p_2} \\ &\qquad \ll S\biggl(\frac{k}{f(y,z)\log_2 x}\biggr)^2 \sum_{q>z^2} \biggl(\sum_{\substack{p\in\mathcal{Q}\\ p\equiv 1 \ (\mathrm{mod}\ q)}} \frac{1}{p}\biggr)^2 \\ &\qquad \ll S\biggl(\frac{k}{f(y,z)\log_2 x}\biggr)^2 \sum_{q>z^2} \frac{\log_2^2 x}{q^2} \ll S\,\frac{\log_2^2 x}{z^2\log z} = o(S).\end{aligned}$$

Thus, we obtain (9).

We now turn to the lower bound for $S_{\mathcal{D}}$.

Let $\mathcal{M}_1$ be the set of integers $m \in \mathcal{M}$ with the property that there exist two primes $q_1, q_2 \in [z, z^2]$ and two prime factors p_1 and p_2 of m such that $p_1 \equiv p_2 \equiv 1 \pmod{q_1 q_2}$. By arguments similar to those above, we have (since $z > \log_2 x$)

$$\begin{aligned}\sum_{m\in\mathcal{M}_1}\frac{1}{m} &\ll \frac{1}{(k-2)!}\Big(\sum_{p^\alpha\in\widetilde{\mathcal{Q}}}\frac{1}{p^\alpha}\Big)^{k-2}\sum_{q_1,q_2\in[z,z^2]}\ \sum_{\substack{p_1,p_2\in\mathcal{Q}\\ p_1\equiv p_2\equiv 1 \pmod{q_1q_2}}}\frac{1}{p_1p_2}\\ &\ll S\Big(\frac{k}{f(y,z)\log_2 x}\Big)^2\sum_{q_1,q_2\ge z}\Big(\sum_{\substack{p\in\mathcal{Q}\\ p\equiv 1 \pmod{q_1q_2}}}\frac{1}{p}\Big)^2\\ &\ll S\sum_{q_1,q_2\ge z}\frac{\log_2^2 x}{(q_1q_2)^2} \ll S\log_2^2 x\Big(\sum_{q\ge z}\frac{1}{q^2}\Big)^2 = S\,\frac{\log_2^2 x}{z^2\log^2 z} = o(S).\end{aligned}$$

Next, let $\mathcal{M}_2$ be the set of integers $m \in \mathcal{M}$ for which there exists a prime $q \in [z, z^2]$ and $L = \lfloor \log_3 x \rfloor$ distinct prime factors p of m with $p \equiv 1 \pmod q$. We have

$$\begin{aligned}\sum_{m\in\mathcal{M}_2}\frac{1}{m} &\le \frac{1}{(k-L)!}\Big(\sum_{p^\alpha\in\widetilde{\mathcal{Q}}}\frac{1}{p^\alpha}\Big)^{k-L}\sum_{q\in[z,z^2]}\ \sum_{\substack{p_1<\cdots<p_L\le x\\ p_i\equiv 1 \pmod q,\, i=1,\dots,L}}\frac{1}{p_1\cdots p_L}\\ &\le S\Big(\frac{2k}{f(y,z)\log_2 x}\Big)^L\sum_{q\ge z}\frac{1}{L!}\Big(\sum_{\substack{p<x\\ p\equiv 1 \pmod q}}\frac{1}{p}\Big)^L\\ &\le S\,\frac{3^L}{L!}\sum_{q\ge z}\Big(\frac{2\log_2 x}{q}\Big)^L = S\,\frac{(6\log_2 x)^L}{L!}\sum_{q\ge z}\frac{1}{q^L}\\ &\ll \frac{S\log_2 x}{L^{5/2}}\Big(\frac{6e\log_2 x}{(L-1)z}\Big)^{L-1} \ll \frac{S}{L^{5/2}} = o(S).\end{aligned}$$

Here, we have used Stirling's formula to approximate $(L-1)!$, together with the fact that

$$\sum_{q\ge z}\frac{1}{q^L} \le \int_z^\infty \frac{dt}{t^L} = \frac{1}{(L-1)z^{L-1}},$$

the estimate

$$\sum_{\substack{p\le x\\ p\equiv 1 \pmod q}}\frac{1}{p} \le \frac{2\log_2 x}{q}, \tag{12}$$

which holds for large x and $q \in [z, z^2]$ (using, for example, the Siegel–Walfisz

Theorem and partial integration), and the inequality

$$\frac{(L-1)z}{6e\log_2 x} > e,$$

which holds when x is large and leads to the estimate

$$\left(\frac{6e\log_2 x}{(L-1)z}\right)^{L-1} \ll e^{-L} \ll \frac{1}{\log_2 x}.$$

Finally, let $\mathcal{M}_3$ be the set of those $m \in \mathcal{M} \setminus (\mathcal{M}_1 \cup \mathcal{M}_2)$ for which there exist at least $T = \lfloor \log_2 x/\log_3^3 x \rfloor$ distinct primes $q \in [z, z^2]$ such that for each prime q, there exist two distinct prime factors $p_{1,q}$ and $p_{2,q}$ of m congruent to 1 modulo q. By arguments similar to those above, we have

$$\begin{aligned}
\sum_{m\in\mathcal{M}_3} \frac{1}{m} &\le \frac{1}{(k-2T)!}\bigg(\sum_{p^\alpha\in\widetilde{\mathcal{Q}}} \frac{1}{p^\alpha}\bigg)^{k-2T} \sum_{\substack{q_1<\cdots<q_T\\ q_i\in[z,z^2]\\ i=1,\dots,T}} \sum_{\substack{p_1,\dots,p_{2T}\le x\\ p_{2i}\equiv p_{2i+1}\equiv 1 \pmod{q_i}\\ i=1,\dots,T}} \frac{1}{p_1\cdots p_{2T}}\\
&\le S\bigg(\frac{2k}{f(y,z)\log_2 x}\bigg)^{2T} \sum_{\substack{q_1<\cdots<q_T\\ q_i\in[z,z^2]\\ i=1,\dots,T}} \prod_{i=1}^{T}\bigg(\sum_{\substack{p\le x\\ p\equiv 1 \pmod{q_i}}} \frac{1}{p}\bigg)^2\\
&\le S\,3^{2T} \sum_{\substack{q_1<\cdots<q_T\\ q_i\in[z,z^2]\\ i=1,\dots,T}} \prod_{i=1}^{T}\bigg(\frac{4\log_2^2 x}{q_i^2}\bigg)\\
&\le S\,\frac{(6\log_2 x)^{2T}}{T!}\bigg(\sum_{q>z}\frac{1}{q^2}\bigg)^{T} \le \frac{S}{T^{1/2}}\bigg(\frac{36e\log_2^2 x}{Tz}\bigg)^{T} = o(S).
\end{aligned}$$

In the above estimates we used, in addition to Stirling's formula for $T!$ and the estimate (12), the fact that the inequality

$$\sum_{q>z}\frac{1}{q^2} \le \frac{1}{z}$$

holds for large z, together with the fact that

$$\frac{36\log_2^2 x}{zT} \le \frac{37}{\log_3^2 x} < 1.$$

Now let $\mathcal{M}_4 = \mathcal{M} \setminus (\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3)$. It follows easily that if $m \in \mathcal{M}_4$, then there exist at most T distinct primes $q \in [z, z^2]$ such that $q^2 \mid \varphi(m)$ (if not, then either there exist two primes p and p' dividing m such that $p-1$ and $p'-1$ have at least two common prime divisors in $[z, z^2]$, which cannot happen since $m \notin \mathcal{M}_1$, or else there exist more than T distinct primes q in $[z, z^2]$, and for each such q there are two prime factors $p_{1,q}$ and $p_{2,q}$

of m such that q divides $p_{i,q}-1$, $i=1,2$, which is again impossible since $m\notin\mathcal{M}_3$). Also, the fact that $m\notin\mathcal{M}_2$ implies that if $q^2\,|\,\varphi(m)$ for some $q>z$, then there exist at most L prime factors p of m such that q divides $p-1$. Thus, if $m=m'd(m)$, then $\omega(m')\le TL\le \log_2 x/\log_3^2 x=o(k)$ since $k\gg \log_2 x/\log_3 x$. From our previous estimates, we immediately obtain

$$\Big(\sum_{d\in\mathcal{D}}\frac{1}{d}\Big)\Big(\sum_{\substack{m'\le x\\ \omega(m')\le TL}}\frac{1}{m'}\Big)\ge\sum_{m\in\mathcal{M}_4}\frac{1}{m}\gg S.$$

Clearly,

$$\sum_{\substack{m'\le x\\ \omega(m')\le TL}}\frac{1}{m'}\le\frac{1}{(TL)!}\Big(\sum_{p\le x}\frac{1}{p}\Big)^{TL}\le\frac{1}{(TL)^{1/2}}\Big(\frac{e\log_2 x+O(1)}{TL}\Big)^{TL}$$
$$=\exp\Big(O\Big(\frac{\log_2 x\log_4 x}{\log_3^2 x}\Big)\Big)=\exp\Big(O\Big(\frac{f(y,z)\log_2 x\log_4 x}{\log_3 x}\Big)\Big).$$

Thus,

$$S_{\mathcal{D}}=\sum_{d\in\mathcal{D}}\frac{1}{d}\ge S\exp\Big(O\Big(\frac{f(y,z)\log_2 x\log_4 x}{\log_3 x}\Big)\Big). \tag{13}$$

By our choice of k, the definition of S, and the formula (4), we have

$$\begin{aligned} S&\gg\frac{1}{k^{1/2}}\Big(\frac{ef(y,z)\log_2 x}{k}\Big(1+O\Big(\frac{\log_3 x}{\log_2 x}\Big)\Big)\Big)^k \\ &=\exp\Big(f(y,z)\log_2 x\Big(1+O\Big(\frac{\log_3 x}{\log_2 x}+\frac{\log k}{f(y,z)\log_2 x}\Big)\Big)\Big) \\ &=\exp\Big(f(y,z)\log_2 x\Big(1+O\Big(\frac{\log_3^2 x}{\log_2 x}\Big)\Big)\Big). \end{aligned} \tag{14}$$

The lower bound of Theorem 1 now follows from the estimates (8), (13) and (14), together with the observation that

$$\begin{aligned} f(\log_2 x,z)&=\prod_{\log_2 x<p\le z}\Big(1-\frac{1}{p-1}\Big) \\ &=\exp\Big(-\sum_{\log_2 x<p\le z}\frac{1}{p}+O\Big(\sum_{p>\log_2 x}\frac{1}{p^2}\Big)\Big) \\ &=\exp\Big(\log\Big(1+O\Big(\frac{\log_4 x}{\log_3 x}\Big)\Big)+O\Big(\frac{1}{\log_2 x}\Big)\Big)=1+O\Big(\frac{\log_4 x}{\log_3 x}\Big). \end{aligned}$$

Upper bound. Since the bound in the statement of Theorem 1 is $x^{1+o(1)}$ for $y>\log_2 x$, we may assume that $y\le\log_2 x$ for our proof of the upper bound. Let $z = \log_2 x/\log_3^2 x$. Since

$$f(z, \log_2 x) = \prod_{z<p\le\log_2 x}\left(1 - \frac{1}{p-1}\right) = \exp\left(-\sum_{z<p\le\log_2 x}\frac{1}{p} + O\left(\sum_{p>z}\frac{1}{p^2}\right)\right)$$
$$= \exp\left(\log_2(\log_2 x) - \log_2 z + O\left(\frac{1}{\log_2 x}\right)\right) = 1 + O\left(\frac{\log_4 x}{\log_3 x}\right),$$

we may further assume that $y \le z$.

Let $\mathcal{A}_y(x)$ be the subset of integers $n \in \mathcal{F}_y(x)$ that are squarefree. Our first goal is to establish the following upper bound:

$$\#\mathcal{A}_y(x) \le \frac{x}{\log x}\exp\left(f(y,z)\log_2 x\left(1 + O\left(\frac{\log_4 x}{\log_3 x}\right)\right)\right). \tag{15}$$

For any positive integer k, let $\pi_k(x)$ be the number of positive integers $n \le x$ such that $\omega(n) = k$. By a well known result of Hardy and Ramanujan (see [5]), the following estimate holds:

$$\pi_k(x) \ll \frac{x}{\log x}\frac{1}{(k-1)!}(\log_2 x + O(1))^{k-1}. \tag{16}$$

Using Stirling's formula, we get

$$\pi_k(x) \ll \frac{x}{\log x}\left(\frac{e\log_2 x + O(1)}{k-1}\right)^{k-1}. \tag{17}$$

Since the function appearing on the right hand side of (17) is increasing for $k \le \frac{1}{2}\log_2 x$ once x is large enough, if we put $K_1 = \lfloor z\rfloor$, it follows that

$$\sum_{k\le K_1}\pi_k(x) \ll \frac{xz}{\log x}(O(\log_3^2 x))^z = \frac{x}{\log x}\exp\left(O\left(\frac{\log_2 x\log_4 x}{\log_3^2 x}\right)\right)$$
$$= \frac{x}{\log x}\exp\left(O\left(\frac{f(y,z)\log_2 x\log_4 x}{\log_3 x}\right)\right).$$

Using again the estimate (16), we note that if $k \ge K_2 = \lfloor 3e\log_2 x\rfloor + 1$, then the inequality

$$\pi_k(x) \ll \frac{x}{\log x}\left(\frac{e\log_2 x + O(1)}{k}\right)^k \le \frac{x}{\log x}\left(\frac{1}{3} + o(1)\right)^k < \frac{x}{\log x}\left(\frac{1}{2^k}\right)$$

holds uniformly for such k provided that x is large enough. Therefore,

$$\sum_{k\ge K_2}\pi_k(x) \ll \frac{x}{\log x}\sum_k\frac{1}{2^k} \ll \frac{x}{\log x}.$$

Thus, to prove (15), it suffices to bound the number of integers $n \in \mathcal{A}_y(x)$ for which $\omega(n)$ lies in the interval $[K_1, K_2]$; let $\mathcal{A}_y^*(x)$ denote the set of such integers n.

Fix $k \in [K_1, K_2]$ and $n \in \mathcal{A}_y^*(x)$ with $\omega(n) = k$. Let us write $n = n_1n_2$, where n_2 is the largest divisor of n with the property that if a prime $q \mid \varphi(n_2)$, then $q \notin [y, z]$. Notice that if $q \in [y, z]$ is a prime dividing $\varphi(n)$, then (since $n \in \mathcal{F}_y(x)$) there exists a unique prime $p \,|\, n$ such that $q \,|\, p - 1$; by the maximal property defining n_2, it follows that n_1 is the product of all such primes p. Since there are only $\pi(z) \ll \log_2 x / \log_3^3 x$ primes $q \le z$, we see that n_2 has at least $k - \pi(z) = k(1 + o(1))$ distinct prime factors.

Let $\mathcal{P}_{y,z}$ denote the set of all primes $p \le x$ such that $p - 1$ is free of primes in the interval $[y, z]$ and such that $q^2 \nmid p - 1$ for any prime $q > z$. Suppose that $n = n_1 n_2$ (as above), where n_1 has precisely $t \le \pi(z)$ prime factors, and n_2 has $k - t$ prime factors, each of which necessarily lies in $\mathcal{P}_{y,z}$. For fixed t, the number of such $n \in \mathcal{A}_y^*(x)$ is bounded by a constant times

$$\frac{x \log_2^2 x}{\log x} \, \frac{1}{t!} \Big(\sum_{p \le x} \frac{1}{p} \Big)^t \frac{1}{(k-t)!} \Big(\sum_{p \in \mathcal{P}_{y,z}} \frac{1}{p} \Big)^{k-t}. \tag{18}$$

To prove this, let $P = P(n)$ be the largest prime factor of one such n, and write $n = Pm$. Using well known results about the distribution of *smooth numbers* (see, for example, [6]), we have

$$\begin{aligned} \#\{n \le x : P(n) \le \exp(\log x / \log_2 x)\} &= x \, \exp((1 + o(1)) \log_2 x \log_3 x) \\ &= o(x / \log x); \end{aligned}$$

hence, we may assume that $P \ge \exp(\log x / \log_2 x)$. For a fixed value of m, it follows that P can be selected in at most

$$\pi(x/m) \ll \frac{x \log_2 x}{m \log x}$$

different ways. Summing these contributions over m, we must now consider whether P divides n_1 or n_2. In either case, using the multinomial formula, we obtain an estimate similar to (18), but in the first case, t has been changed to $t - 1$ in both the factorial and the exponent, whereas in the second case $k - t$ has been changed to $k - t - 1$. At the cost of including an extra factor of $\log_2 x$, we obtain (18) in either case; this follows from the estimates $t \ll \log_2 x$, $k - t \ll \log_2 x$, and

$$\sum_{p \le x} \frac{1}{p} \gg \log_2 x \gg 1, \qquad \sum_{p \in \mathcal{P}_{y,z}} \frac{1}{p} \gg \frac{\log_2 x}{\log z} \gg \frac{\log_2 x}{\log_3 x} \gg 1.$$

Since $t \le \pi(z) \ll \log_2 x / \log_2^3 x$, we have as above

$$\begin{aligned} \frac{1}{t!} \Big(\sum_{p \le x} \frac{1}{p} \Big)^t &\ll \Big(\frac{e \log_2 x + O(1)}{t} \Big)^t = \exp\Big(O\Big(\frac{\log_2 x \log_4 x}{\log_3^3 x} \Big) \Big) \\ &= \exp\Big(O\Big(\frac{f(y,z) \log_2 x \log_4 x}{\log_3 x} \Big) \Big). \end{aligned} \tag{19}$$

We now claim that

$$(20)\qquad \frac{1}{(k-t)!}\Big(\sum_{p\in\mathcal{P}_{y,z}}\frac{1}{p}\Big)^{k-t} \le \exp\Big(f(y,z)\log_2 x\Big(1+O\Big(\frac{\log_4 x}{\log_3 x}\Big)\Big)\Big).$$

To prove this, we apply arguments from our proof of the lower bound to obtain the estimate

$$\sum_{p\in\mathcal{P}_{y,z}}\frac{1}{p} = f(y,z)\log_2 x\Big(1+O\Big(\frac{\log_3 x}{\log_2 x}\Big)\Big).$$

Put $l = k-t$. Then

$$\frac{1}{(k-t)!}\Big(\sum_{p\in\mathcal{P}_{y,z}}\frac{1}{p}\Big)^{k-t} \ll \Big(\frac{ef(y,z)\log_2 x}{l}\Big)^{l}\Big(1+O\Big(\frac{\log_3 x}{\log_2 x}\Big)\Big)^{l}.$$

Since $l\le k\ll \log_2 x$, we have the inequality

$$\Big(1+O\Big(\frac{\log_3 x}{\log_2 x}\Big)\Big)^{l} \ll \exp(O(\log_3 x)) = \exp\Big(O\Big(\frac{f(y,z)\log_2 x\log_4 x}{\log_3 x}\Big)\Big);$$

it therefore suffices to estimate the quantity

$$\Big(\frac{ef(y,z)\log_2 x}{l}\Big)^{l}.$$

The maximum value of this function occurs at $l = f(y,z)\log_2 x$, and for this value we have

$$\Big(\frac{ef(y,z)\log_2 x}{l}\Big)^{l} \le \exp(f(y,z)\log_2 x),$$

and the claim is proved.

Substituting (19) and (20) into inequality (18), and then summing (18) first over all $t\le\pi(z)$, then over all $k\in[K_1,K_2]$, we derive that

$$\begin{aligned}\mathcal{A}_y^*(x) &\ll \frac{x\pi(z)K_2\log_2^2 x}{\log x}\exp\Big(\Big(1+O\Big(\frac{\log_4 x}{\log_3 x}\Big)\Big)f(y,z)\log_2 x\Big)\\ &= \frac{x}{\log x}\exp\Big(\Big(1+O\Big(\frac{\log_4 x}{\log_3 x}\Big)\Big)f(y,z)\log_2 x\Big).\end{aligned}$$

Bearing in mind the contributions to $\mathcal{A}_y(x)$ coming from the values of k outside $[K_1,K_2]$, which have already been discussed, we obtain the desired estimate (15).

Finally, we need to pass from $\mathcal{A}_y(x)$ to the entire set $\mathcal{F}_y(x)$. Suppose that $n = d^2m$ lies in $\mathcal{F}_y(x)$, where m is squarefree. For fixed d, the number of such numbers is at most x/d^2. For those integers with $d>\log x$, we have an overall contribution bounded by

$$x\sum_{d>\log x}\frac{1}{d^2}\ll\frac{x}{\log x},$$

which is sufficient for our upper bound. On the other hand, for integers with $d\le\log x$, by (15) we see that the contribution to $\mathcal{F}_y(x)$ is at most

$$\sum_{d\le\log x}\#\mathcal{A}_y(x/d^2)\le\sum_{d\le\log x}\frac{x_d}{\log x_d}\exp\biggl(f(y,z)\log_2 x_d\biggl(1+O\biggl(\frac{\log_4 x_d}{\log_3 x_d}\biggr)\biggr)\biggr),$$

where $x_d=x/d^2$. Since each $d\le\log x$, we have the estimates

$$\log x_d=\biggl(1+O\biggl(\frac{\log_2 x}{\log x}\biggr)\biggr)\log x,$$

$$\log_2 x_d=\biggl(1+O\biggl(\frac{\log_2 x}{\log x}\biggr)\biggr)\log_2 x,$$

$$\log_3 x_d=(1+o(1))\log_3 x,\qquad \log_4 x_d=(1+o(1))\log_4 x,$$

and we deduce that

$$\sum_{d\le\log x}\#\mathcal{A}_y(x/d^2)\ll\frac{x}{\log x}\exp\biggl(\biggl(1+O\biggl(\frac{\log_4 x}{\log_3 x}\biggr)\biggr)f(y,z)\log_2 x\biggr)\sum_{d\ge1}\frac{1}{d^2}$$

$$\ll\frac{x}{\log x}\exp\biggl(\biggl(1+O\biggl(\frac{\log_4 x}{\log_3 x}\biggr)\biggr)f(y,z)\log_2 x\biggr).$$

This completes the proof of the upper bound and of the theorem. ■

3. y-Squarefree values of $\lambda(n)$. As in the introduction, we define

$$\mathcal{L}_y(x)=\{n\le x:\lambda(n)\in\mathcal{N}(y)\},$$

where λ denotes the Carmichael function. In this section, we follow closely ideas from [9] that were used to establish (2). Our main result is the following analogue of Theorem 1 for the function λ:

THEOREM 2. *For every fixed real number $y\ge2$, there exists a constant $\kappa(y)>0$ such that*

$$\#\mathcal{L}_y(x)=(\kappa(y)+o(1))\frac{x}{(\log x)^{1-\alpha(y)}},$$

where

$$\alpha(y)=\prod_{p>y}\biggl(1-\frac{1}{p(p-1)}\biggr).$$

For historical interest, we remark that positive integers n with the property that $\lambda(n)$ is squarefree have been previously used in the primality test of Adleman, Pomerance and Rumely (see [1]).

Our principal tool for the proof of Theorem 2 is a well known theorem of Wirsing [10], which may be formulated as follows:

Lemma 2. *Suppose that the real-valued multiplicative function $f(n)$ satisfies the following conditions*:

- $f(n) \geq 0$ *for all positive integers* n.
- *There exist constants* c_1, c_2 *with* $c_2 < 2$ *such that* $f(p^a) \leq c_1 c_2^a$ *for all primes* p *and integers* $a \geq 2$.
- *There exists a constant* $\alpha > 0$ *such that*

$$\sum_{p \leq x} f(p) = (\alpha + o(1)) \frac{x}{\log x}.$$

Then

$$\sum_{n \leq x} f(n) = \left(\frac{1}{e^{\gamma\alpha}\Gamma(\alpha)} + o(1) \right) \frac{x}{\log x} \prod_{p \leq x} \left(\sum_{a=0}^{\infty} \frac{f(p^a)}{p^a} \right),$$

where γ *is the Euler–Mascheroni constant*, *and* $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1}\, dt$.

The next result provides the essential analytic ingredient needed to deduce Theorem 2 from Lemma 2.

Lemma 3. *Let* $y > 0$, *and suppose that* $A > 0$ *is a fixed constant. Then the set of primes*

$$\mathcal{P}(y) = \{p \leq x : p - 1 \in \mathcal{N}(y)\}$$

has cardinality

$$\#\mathcal{P}(y) = \alpha(y)\pi(x) + O(x/(\log x)^A),$$

where $\alpha(y)$ *is the constant of Theorem* 2.

Proof. By standard arguments based on partial summation, it suffices to show that

$$\psi_y(x) = \alpha(y)\psi(x) + O(x/(\log x)^A), \tag{21}$$

where

$$\psi_y(x) = \sum_{\substack{n \leq x \\ n-1 \in \mathcal{N}(y)}} \Lambda(n) \quad \text{and} \quad \psi(x) = \sum_{n \leq x} \Lambda(n),$$

where $\Lambda(n)$ is the von Mangoldt function.

Let $\mu(d)$ denote the Möbius function. Since the characteristic function of the set $\mathcal{N}(y)$ is given by

$$n \mapsto \sum_{\substack{d^2 \mid n \\ p \mid d \Rightarrow p > y}} \mu(d),$$

it follows that

$$\psi_y(x) = \sum_{\substack{d \le x^{1/2} \\ p \mid d \Rightarrow p > y}} \mu(d)\psi(x; d^2, 1), \tag{22}$$

where for integers k, l with $k \ge 1$ and $\gcd(k, l) = 1$,

$$\psi(x; k, l) = \sum_{\substack{n \le x \\ n \equiv l \pmod k}} \Lambda(n).$$

Now let $z = x^{1/2}(\log x)^{-B}$, where $B = A + 5$. By (22), we have

$$\psi_y(x) = x \sum_{\substack{d \le z \\ p \mid d \Rightarrow p > y}} \frac{\mu(d)}{\varphi(d^2)} + O(R_1 + R_2),$$

where

$$R_1 = \sum_{\substack{d \le z \\ p \mid d \Rightarrow p > y}} \left| \psi(x; d^2, 1) - \frac{x}{\varphi(d^2)} \right|, \qquad R_2 = \sum_{\substack{z < d \le x^{1/2} \\ p \mid d \Rightarrow p > y}} \psi(x; d^2, 1).$$

By the Bombieri–Vinogradov Theorem, we have the bound

$$R_1 \le \sum_{k \le z} \left| \psi(x; k, 1) - \frac{x}{\varphi(k)} \right| \ll \frac{x}{(\log x)^A}.$$

Using the trivial bound $\psi(x; k, 1) \le x(\log x)/k$, we also have

$$R_2 \ll \sum_{d > z} \frac{x \log x}{d^2} \ll \frac{x \log x}{z} = x^{1/2}(\log x)^{B+1} \ll \frac{x}{(\log x)^A}.$$

Therefore,

$$\psi_y(x) = x \sum_{\substack{d \le z \\ p \mid d \Rightarrow p > y}} \frac{\mu(d)}{\varphi(d^2)} + O\left(\frac{x}{(\log x)^A} \right).$$

Now

$$\sum_{\substack{d \le z \\ p \mid d \Rightarrow p > y}} \frac{\mu(d)}{\varphi(d^2)} = \sum_{\substack{d \ge 1 \\ p \mid d \Rightarrow p > y}} \frac{\mu(d)}{\varphi(d^2)} + O(R_3),$$

where

$$R_3 = \sum_{d > z} \frac{1}{\varphi(d^2)}.$$

Using the well known bound $\varphi(k) \gg k/\log_2 k$, we obtain

$$R_3 \ll \sum_{d > z} \frac{\log_2(d^2)}{d^2} \ll \sum_{d > z} \frac{1}{d^{3/2}} \ll z^{-1/2} = x^{-1/4}(\log x)^B = O\left(\frac{1}{(\log x)^A} \right).$$

Consequently,

$$\psi_y(x) = x \sum_{\substack{d\ge 1\\ p\mid d\Rightarrow p>y}} \frac{\mu(d)}{\varphi(d^2)} + O\biggl(\frac{x}{(\log x)^A}\biggr).$$

Using the multiplicativity of $\mu(n)$ and $\varphi(n)$ (hence, also of $\varphi(n^2)$), we derive that

$$\sum_{\substack{d\ge 1\\ p\mid d\Rightarrow p>y}} \frac{\mu(d)}{\varphi(d^2)} = \prod_{p>y}\biggl(1-\frac{1}{\varphi(p^2)}\biggr) = \prod_{p>y}\biggl(1-\frac{1}{p(p-1)}\biggr) = \alpha(y),$$

which completes the proof. ∎

Proof of Theorem 2. Let $f(n)$ be the unique multiplicative function such that $f(p^a) = 1$ for every prime $p \le y$ and integer $a \ge 1$, and for any prime $p > y$, $f(p^2) = f(p) = 1$ if $p-1 \in \mathcal{N}(y)$ and $f(p^a) = 0$ if either $a \ge 3$ or $p-1 \notin \mathcal{N}(y)$.

Clearly, $\lambda(n) \in \mathcal{N}(y)$ if and only if $\lambda(p^a) \in \mathcal{N}(y)$ for every prime power p^a dividing n. For any prime $p \le y$, the latter condition holds trivially for all $a \ge 1$, while if $p > y \ge 2$, it is equivalent (since p is odd) to the two conditions $a \le 2$ and $p-1 \in \mathcal{N}(y)$. Therefore, f is the characteristic function of the set of integers n for which $\lambda(n)$ lies in $\mathcal{N}(y)$.

By Lemma 3, we see that all of the conditions of Lemma 2 are satisfied, with $\alpha = \alpha(y)$; thus,

$$\#\mathcal{L}_y(x) = \sum_{n\le x} f(n) = \biggl(\frac{1}{e^{\gamma\alpha(y)}\Gamma(\alpha(y))} + o(1)\biggr)\frac{x}{\log x}\prod_{p\le x}\biggl(\sum_{a=0}^{\infty}\frac{f(p^a)}{p^a}\biggr).$$

To complete the proof, we can apply an analogue of Lemma 4 of [9] to deduce that the estimate

$$\prod_{p\le x}\biggl(\sum_{a=0}^{\infty}\frac{f(p^a)}{p^a}\biggr) = \eta(y)(\log x)^{\alpha(y)} + O((\log x)^{\alpha(y)-1})$$

holds for some absolute constant $\eta(y) > 0$. Taking

$$\kappa(y) = \frac{\eta(y)}{e^{\gamma\alpha(y)}\Gamma(\alpha(y))},$$

we finish the proof. ∎

4. Remarks and open problems. It is clear from the proof of our Theorem 1 that if y is a bit smaller than $(\log_2 x)^2$, then almost all n have the property that $\varphi(n)$ is y-squarefree. It would be interesting to investigate whether there is a threshold, or a distribution. For example, is there a function $y = y(n)$ such that the set of integers n for which $\varphi(n)$ is y-squarefree has asymptotic density 1/2? Or more simply, is there a function $y = y(n)$ such that the set of integers n for which $\varphi(n)$ is y-squarefree has asymptotic density c for some constant c in the interval $(0, 1)$? We leave these questions as open problems for the reader.

References

[1] L. M. Adleman, C. Pomerance and R. S. Rumely, *On distinguishing prime numbers from composite numbers*, Ann. of Math. (2) 117 (1983), 173–206.

[2] W. Banks and F. Pappalardi, *Powerfree values of the Euler function*, preprint, 2003.

[3] R. Crandall and C. Pomerance, *Prime Numbers: A Computational Perspective*, Springer, Berlin, 2001.

[4] P. Erdős, C. Pomerance and E. Schmutz, *Carmichael's lambda function*, Acta Arith. 58 (1991), 363–385.

[5] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of an integer*, Quart. J. Math. Oxford 48 (1917), 76–92.

[6] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors*, J. Théor. Nombres Bordeaux 5 (1993), 411–484.

[7] C. Hooley, *On Artin's conjecture*, J. Reine Angew. Math. 225 (1967), 209–220.

[8] L. Mirsky, *The number of representations of an integer as the sum of a prime and a k-free integer*, Amer. Math. Monthly 56 (1949), 17–19.

[9] F. Pappalardi, F. Saidak and I. E. Shparlinski, *Square-free values of the Carmichael function*, J. Number Theory 103 (2003), 122–131.

[10] E. Wirsing, *Das asymptotische Verhalten von Summen über multiplikative Funktionen*, Math. Ann. 143 (1961), 75–102.

Department of Mathematics
University of Missouri
Columbia, MO 65211, U.S.A.
E-mail: bbanks@math.missouri.edu

Instituto de Matemáticas
Universidad Nacional Autónoma de México
Ap. Postal 61-3 (Xangari)
C.P. 58089
Morelia, Michoacán, México
E-mail: fluca@matmor.unam.mx

Received on 27.5.2004
and in revised form on 3.5.2005 (4774)