THE ITERATED CARMICHAEL λ-FUNCTION AND THE NUMBER OF CYCLES OF THE POWER GENERATOR

GREG MARTIN AND CARL POMERANCE

1. INTRODUCTION

A common pseudorandom number generator is the power generator: $x \mapsto x^\ell \pmod n$. Here, ℓ, n are fixed integers at least 2, and one constructs a pseudorandom sequence by starting at some residue mod n and iterating this ℓth power map. (Because it is the easiest to compute, one often takes $\ell = 2$; this case is known as the BBS generator, for Blum, Blum, and Shub.) To be a good generator, the period should be large. Of course, the period depends somewhat on the number chosen for the initial value. However, a universal upper bound for this period is $\lambda(\lambda(n))$ where λ is Carmichael's function. Here, $\lambda(m)$ is defined as the order of the largest cyclic subgroup of the multiplicative group $(\mathbb{Z}/m\mathbb{Z})^\times$. It may be computed via the identity $\lambda(\mathrm{lcm}\{a,b\}) = \mathrm{lcm}\{\lambda(a), \lambda(b)\}$ and its values at prime powers: with ϕ being Euler's function, $\lambda(p^a) = \phi(p^a) = (p-1)p^{a-1}$ for every odd prime power p^a and for 2 and 4, and $\lambda(2^a) = \phi(2^a)/2 = 2^{a-2}$ for $a \geq 3$.

Statistical properties of $\lambda(n)$ were studied by Erdős, Schmutz, and the second author in [7], and in particular, they showed that $\lambda(n) = n/\exp((1+o(1))\log\log n \log\log\log n)$ as $n \to \infty$ through a certain set of integers of asymptotic density 1. This does not quite pinpoint the normal order of $\lambda(n)$ (even the sharper version of this theorem from [7] falls short in this regard), but it is certainly a step in this direction, and does give the normal order of the function $\log(n/\lambda(n))$.

In this paper we prove a result of similar quality for the function $\lambda(\lambda(n))$, which we have seen arises in connection with the period of the power generator. We obtain the same expression as with $\lambda(n)$, except that the $\log\log n$ is squared. That is, $\lambda(\lambda(n)) = n/\exp((1+o(1))(\log\log n)^2 \log\log\log n)$ almost always.

We are able to use this result to say something nontrivial about the number of cycles for the power generator. This problem has been considered in several papers, including [3], [4], and [15]. We show that for almost all integers n, the number of cycles for the ℓth power map modulo n is at least $\exp((1+o(1))(\log\log n)^2 \log\log\log n)$, and we conjecture that this lower bound is actually the truth. Under the assumption of the Generalized Riemann Hypothesis (GRH), and using a new result of Kurlberg and the second author [12], we prove our conjecture. (By the GRH, we mean the Riemann Hypothesis for Kummerian fields as used by Hooley in his celebrated conditional proof of the Artin conjecture.)

For an arithmetic function $f(n)$ whose values are in the natural numbers, let $f_k(n)$ denote the kth iterate of f evaluated at n. One might ask about the normal behavior of $\lambda_k(n)$ for $k \geq 3$. Here we make a conjecture for each fixed k. We also briefly consider the function $L(n)$ defined as the least k such that $\lambda_k(n) = 1$. A similar undertaking was made

G.M. is supported in part by the National Sciences and Engineering Research Council of Canada. C.P. is supported in part by the National Science Foundation.

by Erdős, Granville, Spiro, and the second author in [5] for the function $F(n)$ defined as the least k with $\phi_k(n) = 1$. Though λ is very similar to ϕ, the behavior of $L(n)$ and $F(n)$ seem markedly different. We know that $F(n)$ is always of order of magnitude $\log n$, and it is shown in [5], assuming the Elliott–Halberstam conjecture on the average distribution of primes in arithmetic progressions with large moduli, that in fact $F(n) \sim \alpha \log n$ on a set of asymptotic density 1 for a particular positive constant α. We know far less about $L(n)$, not even its typical order of magnitude. We raise the possibility that it is normally of order $\log\log n$ and show that it is bounded by this order infinitely often.

A more formal statement of our results follows.

Theorem 1. *The normal order of* $\log\big(n/\lambda(\lambda(n))\big)$ *is* $(\log\log n)^2\log\log\log n$. *That is,*

$$\lambda(\lambda(n)) = n\exp\big(-(1+o(1))(\log\log n)^2\log\log\log n\big)$$

as $n \to \infty$ *through a set of integers of asymptotic density* 1.

We actually prove the slightly stronger result: given any function $\psi(n)$ going to infinity arbitrarily slowly, we have

$$\lambda(\lambda(n)) = n\exp\big(-(\log\log n)^2(\log\log\log n + O(\psi(n)))\big)$$

for almost all n.

Given integers $\ell, n \ge 2$, let $C(\ell, n)$ denote the number of cycles when iterating the modular power map $x \mapsto x^\ell \pmod n$.

Theorem 2. *Given any fixed integer* $\ell \ge 2$, *there is a set of integers of asymptotic density* 1 *such that as* $n \to \infty$ *through this set,*

$$C(\ell, n) \ge \exp\big((1+o(1))(\log\log n)^2\log\log\log n\big). \qquad (1)$$

Further, if $\varepsilon(n)$ *tends to* 0 *arbitrarily slowly, we have* $C(\ell, n) \le n^{1/2-\varepsilon(n)}$ *for almost all* n. *Moreover, for a positive proportion of integers* n *we have* $C(\ell, n) \le n^{.409}$. *Finally, if the Generalized Riemann Hypothesis (GRH) is true, we have equality in* (1) *on a set of integers* n *of asymptotic density* 1.

Conjecture 3. *The normal order of* $\log(n/\lambda_k(n))$ *is* $(1/(k-1)!)(\log\log n)^k\log\log\log n$. *That is, for each fixed integer* $k \ge 1$,

$$\lambda_k(n) = n\exp\bigg(-\bigg(\frac{1}{(k-1)!} + o(1)\bigg)(\log\log n)^k(\log\log\log n)\bigg)$$

for almost all n.

Define $L(n)$ to be the number of iterations of λ required to take n to 1, that is, $L(n)$ equals the smallest nonnegative integer k such that $\lambda_k(n) = 1$.

Theorem 4. *There are infinitely many integers* n *such that* $L(n) < (1/\log 2 + o(1))\log\log n$.

2. Notation, strategy, and preliminaries

The proof of Theorem 1, our principal result, proceeds by comparing the prime divisors of $\lambda(\lambda(n))$ with those of $\phi(\phi(n))$. The primes dividing $\phi(m)$ and $\lambda(m)$ are always the same. However, this is not always true for $\phi(\phi(m))$ and $\lambda(\lambda(m))$. The prime 2 clearly causes problems; for example, we have $\phi(\phi(8)) = 2$ but $\lambda(\lambda(8)) = 1$. However this problem also arises from the interaction between different primes, for example, $\phi(\phi(91)) = 24$ but $\lambda(\lambda(91)) = 2$.

We shall use the following notation throughout the paper. The letters p, q, r will always denote primes. Let $v_q(n)$ denote the exponent on q in the prime factorization of n, so that

$$n = \prod_q q^{v_q(n)}$$

for every positive integer n. We let $\mathcal{P}_n = \{p \colon p \equiv 1 \pmod n\}$. We let $x > e^{e^e}$ be a real number and $y = y(x) = \log\log x$. By $\psi(x)$ we denote a function tending to infinity but more slowly than $\log\log\log x = \log y$. In Sections 2–5, the phrase "for almost all n" always means "for all but $O(x/\psi(x))$ integers $n \le x$".

First we argue that the "large" prime divisors typically do not contribute significantly:

Proposition 5. *For almost all $n \le x$, the prime divisors of $\phi(\phi(n))$ and $\lambda(\lambda(n))$ that exceed y^2 are identical.*

Proposition 6. *For almost all $n \le x$,*

$$\sum_{\substack{q > y^2 \\ v_q(\phi(\phi(n))) \ge 2}} v_q(\phi(\phi(n))) \log q \ll y^2 \psi(x). \tag{2}$$

Next we argue that the contribution of "small" primes to $\lambda(\lambda(n))$ is typically small:

Proposition 7. *For almost all $n \le x$, we have*

$$\sum_{q \le y^2} v_q(\lambda(\lambda(n))) \log q \ll y^2 \psi(x).$$

Finally, we develop an understanding of the typical contribution of small primes to $\phi(\phi(n))$ by comparing it to the additive function $h(n)$ defined by

$$h(n) = \sum_{p \mid n} \sum_{r \mid p-1} \sum_{q \le y^2} v_q(r-1) \log q. \tag{3}$$

Proposition 8. *For almost all $n \le x$,*

$$\sum_{q \le y^2} v_q(\phi(\phi(n))) \log q = h(n) + O(y \log y \cdot \psi(x)).$$

Proposition 9. *For almost all $n \le x$, we have $h(n) = y^2 \log y + O(y^2)$.*

Proof of Theorem 1. Let x be a sufficiently large real number. For any positive integer $n \le x$ we may write

$$\log \frac{n}{\lambda(\lambda(n))} = \log \frac{n}{\phi(n)} + \log \frac{\phi(n)}{\phi(\phi(n))} + \log \frac{\phi(\phi(n))}{\lambda(\lambda(n))}.$$

Recall that $n/\phi(n) \ll \log\log n$, and so the first two terms are both $O(\log\log\log x)$. Thus, it suffices to show that

$$\log \frac{\phi(\phi(n))}{\lambda(\lambda(n))} = (\log\log x)^2(\log\log\log x + O(\psi(x))) = y^2 \log y + O(y^2 \psi(x)) \tag{4}$$

for almost all $n \le x$. We write

$$\begin{aligned}\log\frac{\phi(\phi(n))}{\lambda(\lambda(n))} &= \sum_q \big(v_q(\phi(\phi(n))) - v_q(\lambda(\lambda(n)))\big)\log q \\ &= \sum_{q\le y^2} v_q(\phi(\phi(n)))\log q - \sum_{q\le y^2} v_q(\lambda(\lambda(n)))\log q \qquad (5) \\ &\quad + \sum_{q>y^2} \big(v_q(\phi(\phi(n))) - v_q(\lambda(\lambda(n)))\big)\log q.\end{aligned}$$

Since $\lambda(\lambda(n))$ always divides $\phi(\phi(n))$, the coefficients of $\log q$ in this last sum are all nonnegative. On the other hand, Proposition 5 tells us that for almost all $n \le x$, whenever $v_q(\phi(\phi(n))) > 0$ we have $v_q(\lambda(\lambda(n))) > 0$ as well. Therefore the primes q for which $v_q(\phi(\phi(n))) \le 1$ do not contribute to this last sum at all, that is,

$$\begin{aligned}0 &\le \sum_{q>y^2} \big(v_q(\phi(\phi(n))) - v_q(\lambda(\lambda(n)))\big)\log q \\ &= \sum_{\substack{q>y^2 \\ v_q(\phi(\phi(n)))\ge 2}} \big(v_q(\phi(\phi(n))) - v_q(\lambda(\lambda(n)))\big)\log q \\ &\le \sum_{\substack{q>y^2 \\ v_q(\phi(\phi(n)))\ge 2}} v_q(\phi(\phi(n)))\log q \ll y^2\psi(x)\end{aligned}$$

for almost all $n \le x$ by Propositions 5 and 6. Moreover, Proposition 7 tells us that the second sum on the right-hand side of equation (5) is $O(y^2\psi(x))$ for almost all $n \le x$. Therefore equation (5) becomes

$$\log\frac{\phi(\phi(n))}{\lambda(\lambda(n))} = \sum_{q\le y^2} v_q(\phi(\phi(n)))\log q + O(y^2\psi(x))$$

for almost all $n \le x$. By Proposition 8, the sum on the right-hand side can be replaced by $h(n)$ for almost all $n \le x$, the error $O(y\log y\cdot\psi(x))$ in that proposition being absorbed into the existing error $O(y^2\psi(x))$. Finally, Proposition 9 tells us that $h(n) = y^2\log y + O(y^2)$ for almost all $n \le x$. We conclude that equation (4) is satisfied for almost all $n \le x$, which establishes the theorem. □

Given integers a and n, recall that $\pi(t; n, a)$ denotes the number of primes up to t that are congruent to a (mod n). The Brun–Titchmarsh inequality (see [10, Theorem 3.7]) states that

$$\pi(t; n, a) \ll \frac{t}{\phi(n)\log(t/n)} \qquad (6)$$

for all $t > n$. We use repeatedly a weak form of this inequality, valid for all $t > e^e$,

$$\sum_{\substack{p\le t \\ p\in\mathcal{P}_n}} \frac{1}{p} \ll \frac{\log\log t}{\phi(n)}, \qquad (7)$$

which follows from the estimate (6) with $a = 1$ by partial summation. When $n/\phi(n)$ is bounded, this estimate simplifies to

$$\sum_{\substack{p\le t\\ p\in\mathcal{P}_n}} \frac{1}{p} \ll \frac{\log\log t}{n}. \tag{8}$$

For example, we shall employ this last estimate when n is a prime or a prime power and when n is the product of two primes or prime powers; in these cases we have $n/\phi(n) \le 3$. We also quote the fact (see Norton [13] or the paper [14] of the second author) that

$$\sum_{\substack{p\in\mathcal{P}_n\\ p\le t}} \frac{1}{p} = \frac{\log\log t}{\phi(n)} + O\Big(\frac{\log n}{\phi(n)}\Big). \tag{9}$$

This readily implies that

$$\sum_{\substack{p\in\mathcal{P}_n\\ p\le t}} \frac{1}{p-1} = \frac{\log\log t}{\phi(n)} + O\Big(\frac{\log n}{\phi(n)}\Big) \tag{10}$$

as well, since (noting that the smallest possible term in the sum is $p = n+1$) the difference equals

$$\sum_{\substack{p\in\mathcal{P}_n\\ p\le t}} \frac{1}{(p-1)p} \le \sum_{i=1}^{\infty} \frac{1}{in(in+1)} \ll \frac{1}{n^2}.$$

We occasionally use the Chebyshev upper bound

$$\sum_{p\le z} \log p \le \sum_{n\le z} \Lambda(n) \ll z, \tag{11}$$

where $\Lambda(n)$ is the von Mangoldt function, as well as the weaker versions

$$\sum_{p\le z} \frac{\log p}{p} \ll \log z, \qquad \sum_{p\le z} \frac{\log^2 p}{p} \ll \log^2 z \tag{12}$$

and the tail estimates

$$\sum_{p>z} \frac{\log p}{p^2} \ll \frac{1}{z}, \qquad \sum_{p>z} \frac{1}{p^2} \ll \frac{1}{z\log z}, \tag{13}$$

each of which can be derived from the estimate (11) by partial summation. We shall also need at one point a weak form of the asymptotic formula of Mertens,

$$\sum_{p\le z} \frac{\log p}{p} = \log z + O(1). \tag{14}$$

For any polynomial $P(x)$, we also note the series estimate

$$\sum_{a=0}^{\infty} \frac{P(a)}{m^a} \ll_P 1$$

uniformly for $m \ge 2$, valid since the series $\sum_{a=0}^{\infty} P(a)z^a$ converges uniformly for $|z| \le \frac{1}{2}$. The estimates

$$\sum_{a\in\mathbb{N}} \frac{P(a)}{m^a} \ll_P \frac{1}{m}, \qquad \sum_{\substack{a\in\mathbb{N} \\ m^a>z}} \frac{P(a)}{m^a} \ll_P \frac{1}{z}, \tag{15}$$

valid uniformly for any integer $m \ge 2$, follow easily by factoring out the first denominator occurring in each sum.

3. Large primes dividing $\phi(\phi(n))$ and $\lambda(\lambda(n))$

Proof of Proposition 5. If q is any prime, then q divides $\phi(\phi(n))$ if and only if at least one of the following criteria holds:

- $q^3 \mid n$,
- there exists $p \in \mathcal{P}_{q^2}$ with $p \mid n$,
- there exists $p \in \mathcal{P}_q$ with $p^2 \mid n$,
- there exist $r \in \mathcal{P}_q$ and $p \in \mathcal{P}_r$ with $p \mid n$,
- $q^2 \mid n$ and there exists $p \in \mathcal{P}_q$ with $p \mid n$,
- there exist distinct $p_1, p_2 \in \mathcal{P}_q$ with $p_1p_2 \mid n$.

In the first four of these six cases, it is easily checked that $q \mid \lambda(\lambda(n))$ as well. (This is not quite true for $q = 2$, but in this proof we shall only consider primes $q > y^2$.) Therefore we can estimate the number of integers $n \le x$ for which q divides $\phi(\phi(n))$ but not $\lambda(\lambda(n))$ as follows:

$$\sum_{\substack{n\le x \\ q\mid\phi(\phi(n)) \\ q\nmid\lambda(\lambda(n))}} 1 \le \sum_{p\in\mathcal{P}_q} \sum_{\substack{n\le x \\ q^2p\mid n}} 1 + \sum_{p_1\in\mathcal{P}_q} \sum_{\substack{p_2\in\mathcal{P}_q \\ p_2\ne p_1}} \sum_{\substack{n\le x \\ p_1p_2\mid n}} 1 \le \sum_{p\in\mathcal{P}_q} \frac{x}{q^2p} + \sum_{p_1\in\mathcal{P}_q} \sum_{p_2\in\mathcal{P}_q} \frac{x}{p_1p_2}.$$

Using three applications of the Brun–Titchmarsh inequality (8), we conclude that for any odd prime q,

$$\sum_{\substack{n\le x \\ q\mid\phi(\phi(n)) \\ q\nmid\lambda(\lambda(n))}} 1 \ll \frac{xy}{q^3} + \frac{xy^2}{q^2} \ll \frac{xy^2}{q^2}.$$

Consequently, by the tail estimate (13) and the condition $\psi(x) = o(\log y)$,

$$\sum_{q>y^2} \sum_{\substack{n\le x \\ q\mid\phi(\phi(n)) \\ q\nmid\lambda(\lambda(n))}} 1 \ll xy^2 \sum_{q>y^2} \frac{1}{q^2} \ll \frac{xy^2}{y^2\log y^2} < \frac{x}{\log y} \ll \frac{x}{\psi(x)}.$$

Therefore for almost all $n \le x$, every prime $q > y^2$ dividing $\phi(\phi(n))$ also divides $\lambda(\lambda(n))$, as asserted. □

Lemma 10. *Given a real number $x \ge 3$ and a prime $q > y^2$, define $S_q = S_q(x)$ to be the set of all integers $n \le x$ for which at least one of the following criteria holds:*

- $q^2 \mid n$,
- *there exists $p \in \mathcal{P}_{q^2}$ with $p \mid n$,*
- *there exist distinct $p_1, p_2 \in \mathcal{P}_q$ with $p_1p_2 \mid n$,*
- *there exist $r \in \mathcal{P}_{q^2}$ and $p \in \mathcal{P}_r$ with $p \mid n$,*

- *there exist distinct $r_1, r_2, r_3 \in \mathcal{P}_q$ and $p \in \mathcal{P}_{r_1 r_2 r_3}$ with $p \mid n$,*
- *there exist distinct $r_1, r_2, r_3, r_4 \in \mathcal{P}_q$, $p_1 \in \mathcal{P}_{r_1 r_2}$, and $p_2 \in \mathcal{P}_{r_3 r_4}$ with $p_1 p_2 \mid n$.*

Then the cardinality of S_q is $O(xy^2/q^2)$.

Note that if $q^2 \mid \phi(n)$, then at least one of the first three of the six conditions in the statement of the lemma must be satisfied.

Proof. The number of integers up to x for which any particular one of the six criteria holds is easily shown to be $O(xy^2/q^2)$. For the sake of conciseness, we show the details of this calculation only for the last criterion, which is the most complicated. The number of integers n up to x for which there exist distinct $r_1, r_2, r_3, r_4 \in \mathcal{P}_q$, $p_1 \in \mathcal{P}_{r_1 r_2}$, and $p_2 \in \mathcal{P}_{r_3 r_4}$ with $p_1 p_2 \mid n$ is at most

$$\sum_{r_1,r_2,r_3,r_4 \in \mathcal{P}_q} \sum_{\substack{p_1 \in \mathcal{P}_{r_1 r_2} \\ p_2 \in \mathcal{P}_{r_3 r_4}}} \sum_{\substack{n \le x \\ p_1 p_2 \mid n}} 1 \le \sum_{r_1,r_2,r_3,r_4 \in \mathcal{P}_q} \sum_{\substack{p_1 \in \mathcal{P}_{r_1 r_2} \\ p_2 \in \mathcal{P}_{r_3 r_4}}} \frac{x}{p_1 p_2}.$$

Using six applications of the Brun–Titchmarsh estimate (8), we have

$$\sum_{r_1,r_2,r_3,r_4 \in \mathcal{P}_q} \sum_{\substack{p_1 \in \mathcal{P}_{r_1 r_2} \\ p_2 \in \mathcal{P}_{r_3 r_4}}} \frac{x}{p_1 p_2} \ll \sum_{r_1,r_2,r_3,r_4 \in \mathcal{P}_q} \frac{xy^2}{r_1 r_2 r_3 r_4} \ll \frac{xy^6}{q^4} < \frac{xy^2}{q^2},$$

the last inequality being valid due to the hypothesis $q > y^2$. □

Proof of Proposition 6. Define $S = S(x)$ to be the union of S_q over all primes $q > y^2$, where S_q is defined as in the statement of Lemma 10. Using #A to denote the cardinality of a set A, Lemma 10 implies that

$$\#S \le \sum_{q > y^2} \#S_q \ll \sum_{q > y^2} \frac{xy^2}{q^2} \ll \frac{xy^2}{y^2 \log y^2} \ll \frac{x}{\psi(x)}$$

by the tail estimate (13) and the condition $\psi(x) = o(\log y)$. Therefore to prove that the estimate (2) holds for almost all integers $n \le x$, it suffices to prove that it holds for almost all integers $n \le x$ that are not in the set S. This in turn is implied by the upper bound

$$\sum_{\substack{n \le x \\ n \notin S}} \sum_{\substack{q > y^2 \\ v_q(\phi(\phi(n))) \ge 2}} v_q(\phi(\phi(n))) \log q \ll xy^2, \tag{16}$$

which we proceed now to establish.

Fix a prime $q > y^2$ and an integer $a \ge 2$ for the moment. In general, there are many ways in which q^a could divide $\phi(\phi(n))$, depending on the power to which q divides n itself, the power to which q divides numbers of the form $p - 1$ with $p \mid n$, and so forth. However, for integers $n \notin S$, most of these various possibilities are ruled out by one of the six criteria defining the sets S_q. In fact, for $n \notin S$, there are only two ways for q^a to divide $\phi(\phi(n))$:

- there are distinct $r_1, \dots, r_a \subset \mathcal{P}_q$ and distinct $p_1 \in \mathcal{P}_{r_1}, \dots, p_a \in \mathcal{P}_{r_a}$ with $p_1 \dots p_a | n$,
- there are distinct $r_1, \dots, r_a \subset \mathcal{P}_q$, distinct $p_1 \in \mathcal{P}_{r_1}, \dots, p_{a-2} \in \mathcal{P}_{r_{a-2}}$, and $p \in \mathcal{P}_{r_{a-1} r_a}$ with $p_1 \dots p_{a-2} p | n$.

(We refer to the former case as the "supersquarefree" case.)

Still considering q and a fixed, the number of integers n up to x satisfying each of these two conditions is at most

$$\sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{a!} \sum_{\substack{p_1\in\mathcal{P}_{r_1}\\ \vdots\\ p_a\in\mathcal{P}_{r_a}}} \sum_{\substack{n\le x\\ p_1\dots p_a|n}} 1 \;\le\; \sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{a!} \sum_{\substack{p_1\in\mathcal{P}_{r_1}\\ \vdots\\ p_a\in\mathcal{P}_{r_a}}} \frac{x}{p_1\dots p_a}$$

and

$$\sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{2!(a-2)!} \sum_{\substack{p_1\in\mathcal{P}_{r_1}\\ \vdots\\ p_{a-2}\in\mathcal{P}_{r_{a-2}}\\ p\in\mathcal{P}_{r_{a-1}r_a}}} \sum_{\substack{n\le x\\ p_1\dots p_{a-2}p|n}} 1 \;\le\; \sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{(a-2)!} \sum_{\substack{p_1\in\mathcal{P}_{r_1}\\ \vdots\\ p_{a-2}\in\mathcal{P}_{r_{a-2}}\\ p\in\mathcal{P}_{r_{a-1}r_a}}} \frac{x}{p_1\dots p_{a-2}p},$$

respectively, the factors $1/a!$ and $1/2!(a-2)!$ coming from the various possible permutations of the primes r_i. Letting $c\ge 1$ be the constant implied in the Brun–Titchmarsh inequality (8) as applied to moduli n that are divisible by at most two distinct primes, we see that

$$\sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{a!} \sum_{\substack{p_1\in\mathcal{P}_{r_1}\\ \vdots\\ p_a\in\mathcal{P}_{r_a}}} \frac{x}{p_1\dots p_a} \;\le\; \sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{a!}\frac{x(cy)^a}{r_1\dots r_a} \;\le\; \frac{x(cy)^{2a}}{a!q^a}$$

and

$$\sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{(a-2)!} \sum_{\substack{p_1\in\mathcal{P}_{r_1}\\ \vdots\\ p_{a-2}\in\mathcal{P}_{r_{a-2}}\\ p\in\mathcal{P}_{r_{a-1}r_a}}} \frac{x}{p_1\dots p_{a-2}p} \;\le\; \sum_{r_1,\dots,r_a\in\mathcal{P}_q} \frac{1}{(a-2)!}\frac{x(cy)^{a-1}}{r_1\dots r_a} \;\le\; \frac{x(cy)^{2a-1}}{(a-2)!q^a}.$$

Therefore the number of integers $n\le x$ such that $n\notin S$ and $q^a \mid \phi(\phi(n))$ is

$$\le\; \frac{x(cy)^{2a}}{a!q^a} + \frac{x(cy)^{2a-1}}{(a-2)!q^a} \;<\; \frac{c^{2a}xy^4}{(a-2)!q^2}, \tag{17}$$

where we have used the assumption $q>y^2$.

We now establish the estimate (16). Note that

$$\begin{aligned}\sum_{\substack{n\le x\\ n\notin S}} \sum_{\substack{q>y^2\\ v_q(\phi(\phi(n)))\ge 2}} v_q(\phi(\phi(n)))\log q \;&\le\; 2\sum_{\substack{n\le x\\ n\notin S}} \sum_{\substack{q>y^2\\ v_q(\phi(\phi(n)))\ge 2}} \big(v_q(\phi(\phi(n)))-1\big)\log q\\ &=\; 2\sum_{q>y^2}\log q\sum_{a\ge 2}\sum_{\substack{n\le x\\ n\notin S\\ q^a|\phi(\phi(n))}} 1.\end{aligned}$$

Therefore, using the bound (17) for each pair q and a,

$$\begin{aligned}\sum_{\substack{n\le x\\ n\notin S}} \sum_{\substack{q>y^2\\ v_q(\phi(\phi(n)))\ge 2}} v_q(\phi(\phi(n)))\log q \;&\le\; 2\sum_{q>y^2}\log q\sum_{a\ge 2}\frac{c^{2a}xy^4}{(a-2)!q^2}\\ &=\; 2c^4e^{c^2}xy^4\sum_{q>y^2}\frac{\log q}{q^2} \;\ll\; \frac{xy^4}{y^2} \;=\; xy^2\end{aligned}$$

by the tail estimate (13). This establishes the estimate (16) and hence the proposition. □

4. Small primes and the reduction to $h(n)$

Lemma 11. *For any prime power q^a, the number of positive integers $n \le x$ for which q^a divides $\lambda(\lambda(n))$ is $O(xy^2/q^a)$.*

Proof. When q is an odd prime, the prime power q^a divides $\lambda(\lambda(n))$ if and only if at least one of the following criteria holds:

- $q^{a+2} \mid n$,
- there exists $p \in \mathcal{P}_{q^{a+1}}$ with $p \mid n$,
- there exists $p \in \mathcal{P}_{q^a}$ with $p^2 \mid n$,
- there exist $r \in \mathcal{P}_{q^a}$ and $p \in \mathcal{P}_r$ with $p \mid n$.

Even when $q = 2$, at least one of these four conditions must hold for q^a to divide $\lambda(\lambda(n))$, although they are not quite sufficient. In either case, we still have the upper bound

$$\begin{aligned}\sum_{\substack{n\le x\\ q^a\mid\lambda(\lambda(n))}} 1 &\le \sum_{\substack{n\le x\\ q^{a+2}\mid n}} 1 + \sum_{p\in\mathcal{P}_{q^{a+1}}} \sum_{\substack{n\le x\\ p\mid n}} 1 + \sum_{p\in\mathcal{P}_{q^a}} \sum_{\substack{n\le x\\ p^2\mid n}} 1 + \sum_{r\in\mathcal{P}_{q^a}} \sum_{p\in\mathcal{P}_r} \sum_{\substack{n\le x\\ p\mid n}} 1 \\ &\le \frac{x}{q^{a+2}} + \sum_{\substack{p\in\mathcal{P}_{q^{a+1}}\\ p\le x}} \frac{x}{p} + \sum_{\substack{p\in\mathcal{P}_{q^a}\\ p\le\sqrt{x}}} \frac{x}{p^2} + \sum_{r\in\mathcal{P}_{q^a}} \sum_{\substack{p\in\mathcal{P}_r\\ p\le x}} \frac{x}{p}. \end{aligned} \tag{18}$$

In the second of these three sums, it is sufficient to notice that any $p \in \mathcal{P}_{q^a}$ must exceed q^a, which leads to the estimate

$$\sum_{\substack{p\in\mathcal{P}_{q^a}\\ p\le\sqrt{x}}} \frac{x}{p^2} < \sum_{m>q^a} \frac{x}{m^2} < \frac{x}{q^a}.$$

To bound the first and third sums in (18), we invoke the Brun–Titchmarsh estimate (8) a total of three times:

$$\sum_{\substack{p\in\mathcal{P}_{q^{a+1}}\\ p\le x}} \frac{x}{p} \ll \frac{xy}{q^{a+1}}$$

$$\sum_{r\in\mathcal{P}_{q^a}} \sum_{\substack{p\in\mathcal{P}_r\\ p\le x}} \frac{x}{p} \ll \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le x}} \frac{xy}{r} \ll \frac{xy^2}{q^a}.$$

Using these three estimates, (18) gives

$$\sum_{\substack{n\le x\\ q^a\mid\lambda(\lambda(n))}} 1 \ll \frac{x}{q^{a+2}} + \frac{xy}{q^{a+1}} + \frac{x}{q^a} + \frac{xy^2}{q^a} \ll \frac{xy^2}{q^a},$$

which establishes the lemma. □

Proof of Proposition 7. We have

$$\sum_{q\le y^2} v_q(\lambda(\lambda(n))) \log q = \sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a\mid\lambda(\lambda(n))}} 1 \le \sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le y^2}} 1 + \sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a>y^2\\ q^a\mid\lambda(\lambda(n))}} 1.$$

Since the first sum is simply

$$\sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le y^2}} 1 \;=\; \sum_{m\le y^2} \Lambda(m) \;\ll\; y^2$$

by the Chebyshev estimate (11), we have uniformly for $n \le x$,

$$\sum_{q\le y^2} v_q(\lambda(\lambda(n))) \log q \;\ll\; y^2 + \sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a>y^2\\ q^a\mid\lambda(\lambda(n))}} 1. \tag{19}$$

To show that this quantity is usually small, we sum this last double sum over n and apply Lemma 11, yielding

$$\sum_{n\le x}\sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a>y^2\\ q^a\mid\lambda(\lambda(n))}} 1 \;=\; \sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a>y^2}} \sum_{\substack{n\le x\\ q^a\mid\lambda(\lambda(n))}} 1 \;\ll\; \sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a>y^2}} \frac{xy^2}{q^a}.$$

Using the geometric series sum (15) and the Chebyshev estimate (11), this becomes

$$\sum_{n\le x}\sum_{q\le y^2} \log q \sum_{\substack{a\in\mathbb{N}\\ q^a>y^2\\ q^a\mid\lambda(\lambda(n))}} 1 \;\ll\; \sum_{q\le y^2} \log q \cdot \frac{xy^2}{y^2} \;\ll\; xy^2.$$

Therefore if we sum both sides of (19) over n, we obtain

$$\sum_{n\le x}\sum_{q\le y^2} v_q(\lambda(\lambda(n))) \log q \;\ll\; xy^2.$$

This implies that for almost all $n \le x$, we have

$$\sum_{q\le y^2} v_q(\lambda(\lambda(n))) \log q \;\ll\; y^2\psi(x),$$

as desired. □

Proof of Proposition 8. Fix a prime q for the moment. For any positive integer m, the usual formula for $\phi(m)$ readily implies

$$v_q(\phi(m)) \;=\; \max\{0, v_q(m)-1\} + \sum_{p\mid m} v_q(p-1),$$

which we use in the form

$$\sum_{p\mid m} v_q(p-1) \;\le\; v_q(\phi(m)) \;\le\; \sum_{p\mid m} v_q(p-1) + v_q(m).$$

Using these inequalities twice, first with $m=\phi(n)$ and then with $m=n$, we see that

$$\begin{aligned}\sum_{p\mid\phi(n)} v_q(p-1) \;\le\; v_q(\phi(\phi(n))) \;&\le\; \sum_{p\mid\phi(n)} v_q(p-1) + v_q(\phi(n))\\ &\le\; \sum_{p\mid\phi(n)} v_q(p-1) + \sum_{p\mid n} v_q(p-1) + v_q(n).\end{aligned} \tag{20}$$

Now a prime r divides $\phi(n)$ if and only if either $r^2 \mid n$ or there exists a prime $p \mid n$ such that $r \mid p-1$. Therefore

$$\sum_{p\mid n}\sum_{r\mid p-1} v_q(r-1) \;\le\; \sum_{r\mid \phi(n)} v_q(r-1) \;\le\; \sum_{p\mid n}\sum_{r\mid p-1} v_q(r-1) + \sum_{r\colon r^2\mid n} v_q(r-1),$$

the latter inequality accounting for the possibility that both criteria hold for some prime r. When we combine these inequalities with those in equation (20) and subtract the double sum over p and r throughout, we obtain

$$\begin{aligned} 0 \;\le\; v_q(\phi(\phi(n))) - \sum_{p\mid n}\sum_{r\mid p-1} v_q(r-1) \;&\le\; \sum_{r\colon r^2\mid n} v_q(r-1) + \sum_{p\mid n} v_q(p-1) + v_q(n) \\ &\le\; 2\sum_{p\mid n} v_q(p-1) + v_q(n). \end{aligned}$$

Now we multiply through by $\log q$ and sum over all primes $q \le y^2$ to conclude that for any positive integer n,

$$0 \;\le\; \sum_{q\le y^2} v_q(\phi(\phi(n)))\log q - h(n) \;\le\; 2\sum_{q\le y^2}\sum_{p\mid n} v_q(p-1)\log q + \sum_{q\le y^2} v_q(n)\log q.$$

It remains to show that the right-hand side of this last inequality is $O(y\log y\cdot\psi(x))$ for almost all $n \le x$, which we accomplish by establishing the estimate

$$\sum_{n\le x}\sum_{q\le y^2}\sum_{p\mid n} v_q(p-1)\log q + \sum_{n\le x}\sum_{q\le y^2} v_q(n)\log q \;\ll\; xy\log y. \tag{21}$$

We may rewrite the first term on the left-hand side as

$$\begin{aligned} \sum_{n\le x}\sum_{q\le y^2}\sum_{p\mid n} v_q(p-1)\log q \;&=\; \sum_{n\le x}\sum_{q\le y^2}\sum_{p\mid n}\sum_{\substack{a\in\mathbb{N}\\ q^a\mid p-1}} \log q \\ &=\; \sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\sum_{p\in\mathcal{P}_{q^a}}\sum_{\substack{n\le x\\ p\mid n}} 1 \;\le\; \sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\sum_{p\in\mathcal{P}_{q^a}} \frac{x}{p}. \end{aligned}$$

Using the Brun–Titchmarsh inequality (8) and the geometric series estimate (15), we obtain

$$\sum_{n\le x}\sum_{q\le y^2}\sum_{p\mid n} v_q(p-1)\log q \;\ll\; x\sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\frac{y}{q^a} \;\ll\; xy\sum_{q\le y^2}\frac{\log q}{q} \;\ll\; xy\log y^2.$$

The second term on the left-hand side of (21) is even simpler: we have

$$\sum_{n\le x}\sum_{q\le y^2} v_q(n)\log q \;=\; \sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\sum_{\substack{n\le x\\ q^a\mid n}} 1 \;\le\; \sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\frac{x}{q^a},$$

and using the geometric series bound (15) and the weak Chebyshev estimate (12) yields

$$\sum_{n\le x}\sum_{q\le y^2} v_q(n)\log q \;\ll\; x\sum_{q\le y^2}\frac{\log q}{q} \;\ll\; x\log y^2.$$

The last two estimates therefore establish (21) and hence the proposition. □

5. The normal order of $h(n)$

Recall the definition (3): $h(n) = \sum_{p|n} \sum_{r|p-1} \sum_{q \le y^2} v_q(r-1) \log q$. We now calculate the normal order of the additive function $h(n)$ via the Turán–Kubilius inequality (see [11], Lemma 3.1). If we define

$$M_1(x) = \sum_{p \le x} \frac{h(p)}{p}, \qquad M_2(x) = \sum_{p \le x} \frac{h(p)^2}{p},$$

then the Turán-Kubilius inequality asserts that

$$\sum_{n \le x} (h(n) - M_1(x))^2 \ll x M_2(x). \tag{22}$$

Proposition 12. *We have $M_1(x) = y^2 \log y + O(y^2)$ for all $x > e^{e^e}$.*

Proposition 13. *We have $M_2(x) \ll y^3 \log^2 y$ for all $x > e^{e^e}$.*

Proof of Proposition 9. Let N denote the number of $n \le x$ for which $|h(n) - M_1(x)| > y^2$. The contribution of such n to the sum in (22) is at least $y^4 N$. Thus, Proposition 13 implies that $N \ll x(\log y)^2/y$. Hence, Proposition 12 implies that $h(n) = y^2 \log y + O(y^2)$ for all $n \le x$ but for a set of size $O(x(\log y)^2)/y)$. This proves Proposition 9. □

To calculate $M_1(x)$ and $M_2(x)$ we shall first calculate $\sum_{p \le t} h(p)$ and $\sum_{p \le t} h(p)^2$ and then account for the weights $1/p$ using partial summation. We begin the evaluation of $\sum_{p \le t} h(p)$ with a lemma.

Lemma 14. *Let b be a positive integer and $t > e^e$ a real number.*

(a) *If $b > t^{1/4}$ then*

$$\sum_{r \in \mathcal{P}_b} \pi(t; r, 1) \ll \frac{t \log t}{b}.$$

(b) *If $b \le t^{1/4}$ then*

$$\sum_{\substack{r \in \mathcal{P}_b \\ r > t^{1/3}}} \pi(t; r, 1) \ll \frac{bt}{\phi(b)^2 \log t}.$$

and

$$\sum_{r \in \mathcal{P}_b} \pi(t; r, 1) \ll \frac{t \log \log t}{\phi(b) \log t}$$

Remark. The exponents $\frac14$ and $\frac13$ are rather arbitrary and chosen only for simplicity; any two exponents $0 < \alpha < \beta < \frac12$ would do equally well.

Proof. Notice that in all three sums, the only contributing terms are those with $r > b$ and $r < t$. If $b > t^{1/4}$, then the trivial bound $\pi(t; r, 1) \le t/r$ gives

$$\sum_{r \in \mathcal{P}_b} \pi(t; r, 1) \le \sum_{\substack{r \in \mathcal{P}_b \\ t^{1/4} < r \le t}} \frac{t}{r} \le \sum_{\substack{m \equiv 1 \ (\mathrm{mod}\ b) \\ t^{1/4} < m \le t}} \frac{t}{m} \ll \frac{t \log t}{b},$$

proving part (a) of the lemma.

We now assume $b \le t^{1/4}$. We have

$$\begin{aligned}\sum_{\substack{r\in\mathcal{P}_b\\ r>t^{1/3}}} \pi(t;r,1) &= \#\{(m,r)\colon r\equiv 1 \ (\mathrm{mod}\ b),\ r>t^{1/3},\ mr+1\le t,\ mr+1 \text{ and } r \text{ both prime}\}\\ &\le \sum_{m<t^{2/3}} \#\{r<\tfrac{t}{m}\colon r\equiv 1\ (\mathrm{mod}\ b),\ mr+1 \text{ and } r \text{ both prime}\}\\ &\ll \sum_{m<t^{2/3}} \frac{bt}{\phi(mb)\phi(b)\log^2\frac{t}{mb}}\end{aligned}$$

by Brun's sieve method (see [10, Corollary 2.4.1]). We have $\frac{t}{mb}\ge t^{1/12}$ and so $\log\frac{t}{mb}\gg \log t$. We also have $\phi(mb)\ge\phi(m)\phi(b)$ and the standard estimate

$$\sum_{m\le z}\frac{1}{\phi(m)} \ll \log z. \tag{23}$$

Therefore

$$\sum_{\substack{r\in\mathcal{P}_b\\ r>t^{1/3}}} \pi(t;r,1) \ll \sum_{m<t^{2/3}} \frac{bt}{\phi(m)\phi(b)^2\log^2 t} \ll \frac{bt\log t^{2/3}}{\phi(b)^2\log^2 t} \le \frac{bt}{\phi(b)^2\log t},$$

establishing the first estimate in part (b). Finally, by the Brun–Titchmarsh inequalities (6) and (8),

$$\sum_{\substack{r\in\mathcal{P}_b\\ r\le t^{1/3}}} \pi(t;r,1) \ll \sum_{\substack{r\in\mathcal{P}_b\\ r\le t^{1/3}}} \frac{t}{\phi(r)\log\frac{t}{r}} \ll \sum_{\substack{r\in\mathcal{P}_b\\ r\le t^{1/3}}} \frac{t}{r\log t} \ll \frac{t\log\log t}{\phi(b)\log t}.$$

Combining this estimate with the first half of part (b) and the standard estimate $b/\phi(b)\ll \log\log b$ establishes the second half. □

Lemma 15. *For all real numbers $x>e^{e^e}$ and $t>e^e$, we have*

$$\sum_{p\le t} h(p) = \frac{2t\log\log t\log y}{\log t} + O\Big(\frac{t\log\log t}{\log t} + \frac{t\log^2 y}{\log t} + t^{3/4}\log t\cdot y^2\Big).$$

Remark. In particular, we have $\sum_{p\le x} h(p) \ll x\log\log x\log y/\log x = xy\log y/\log x$.

Proof. We may rewrite

$$\begin{aligned}\sum_{p\le t} h(p) &= \sum_{p\le t}\sum_{r\mid p-1}\sum_{q\le y^2} v_q(r-1)\log q = \sum_{p\le t}\sum_{r\mid p-1}\sum_{q\le y^2}\sum_{\substack{a\in\mathbb{N}\\ q^a\mid r-1}} \log q\\ &= \sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\sum_{r\colon q^a\mid r-1}\sum_{\substack{p\le t\\ r\mid p-1}} 1 = \sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\sum_{r\in\mathcal{P}_{q^a}} \pi(t;r,1).\end{aligned} \tag{24}$$

The main contribution to this triple sum comes from the terms with $q^a\le t^{1/4}$ and $r\le t^{1/3}$. In fact, using Lemma 14(a) we can bound the contribution from the terms with q^a large by

$$\sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a>t^{1/4}}} \sum_{r\in\mathcal{P}_{q^a}} \pi(t;r,1) \ll \sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a>t^{1/4}}} \frac{t\log t}{q^a}$$
$$\ll t\log t \sum_{q\le y^2} \frac{\log q}{t^{1/4}} \ll t^{3/4}\log t\cdot y^2,$$

where the last two estimates are due to the geometric series bound (15) and the Chebyshev bound (11). Similarly, using the first half of Lemma 14(b) we can bound the contribution from the terms with q^a small and r large by

$$\sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r>t^{1/3}}} \pi(t;r,1) \ll \sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \frac{t}{q^a\log t} \ll \frac{t}{\log t}\sum_{q\le y^2}\frac{\log q}{q} \ll \frac{t\log y}{\log t},$$

where again the last two estimates are due to the geometric series bound (15) and the weak Chebyshev bound (12). In light of these two estimates, equation (24) becomes

$$\sum_{p\le t} h(p) = \sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}} \pi(t;r,1) + O\Big(t^{3/4}\log t\cdot y^2 + \frac{t\log y}{\log t}\Big). \tag{25}$$

Define $E(t;r,1) = \pi(t;r,1) - \mathrm{li}(t)/(r-1)$. We have

$$\sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}} \pi(t;r,1) = \sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}} \Big(\frac{\mathrm{li}(t)}{r-1} + E(t;r,1)\Big)$$
$$= \sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}} \frac{\mathrm{li}(t)}{r-1} + O\Big(\sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}} |E(t;r,1)|\Big). \tag{26}$$

Let $\Omega(m)$ denote the number of divisors of m that are primes or prime powers. Using the estimate $\Omega(m)\ll\log m$, we quickly dispose of

$$\sum_{q\le y^2}\log q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}} |E(t;r,1)| \le 2\log y \sum_{r\le t^{1/3}} |E(t;r,1)| \sum_{q\le y^2} \sum_{\substack{a\in\mathbb{N}\\ q^a\mid r-1}} 1$$
$$\le 2\log y \sum_{r\le t^{1/3}} |E(t;r,1)|\,\Omega(r-1)$$
$$\ll \log y\log t \sum_{r\le t^{1/3}} |E(t;r,1)| \ll \frac{t\log y}{\log t}$$

by the Bombieri–Vinogradov theorem (we could equally well put any power of $\log t$ in the denominator of the final expression if we needed). Inserting this estimate into equation (26), we see that equation (25) becomes

$$\sum_{p\le t} h(p) = \mathrm{li}(t)\sum_{q\le y^2}\log q \sum_{a\in\mathbb{N}} \sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}} \frac{1}{r-1} + O\Big(t^{3/4}\log t\cdot y^2 + \frac{t\log y}{\log t}\Big). \tag{27}$$

We have by equation (10)

$$\begin{aligned}
\sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}}\frac{1}{r-1} &= \sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\Big(\frac{\log\log t^{1/3}}{\phi(q^a)}+O\Big(\frac{\log q^a}{q^a}\Big)\Big)\\
&= (\log\log t+O(1))\sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\Big(\frac{1}{q^a}+O\Big(\frac{1}{q^{a+1}}\Big)\Big)+O\Big(\sum_{q\le y^2}\log^2 q\sum_{a\in\mathbb{N}}\frac{a}{q^a}\Big)\\
&= (\log\log t+O(1))\sum_{q\le y^2}\Big(\frac{\log q}{q}+O\Big(\frac{\log q}{q^2}\Big)\Big)+O\Big(\sum_{q\le y^2}\frac{\log^2 q}{q}\Big),
\end{aligned}$$

using the geometric series estimate (15). Using the Mertens formula (14) to evaluate the main term and the weak Chebyshev estimates (12) to bound the error terms, we see that

$$\sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\sum_{\substack{r\in\mathcal{P}_{q^a}\\ r\le t^{1/3}}}\frac{1}{r-1} = \log\log t\log y^2+O(\log y+\log\log t+\log^2 y).$$

We conclude from equation (27) and the fact that $\operatorname{li}(t)=t/\log t+O(t/\log^2 t)$ that

$$\begin{aligned}
\sum_{p\le t}h(p) &= \operatorname{li}(t)\big(\log\log t\log y^2+O(\log y+\log\log t+\log^2 y)\big)\\
&\qquad +O\Big(t^{3/4}\log t\cdot y^2+\frac{t\log y}{\log t}\Big)\\
&= \frac{2t\log\log t\log y}{\log t}+O\Big(\frac{t\log\log t}{\log t}+\frac{t\log^2 y}{\log t}+t^{3/4}\log t\cdot y^2\Big),
\end{aligned}$$

as asserted. □

Proof of Proposition 12. In an explicit example of the technique of partial summation, we write

$$\begin{aligned}
M_1(x) &= \sum_{p\le x}\frac{h(p)}{p} = \sum_{p\le e^e}\frac{h(p)}{p}+\sum_{e^e<p\le x}h(p)\Big(\frac{1}{x}+\int_p^x\frac{dt}{t^2}\Big)\\
&= O(1)+\frac{1}{x}\sum_{e^e<p\le x}h(p)+\int_{e^e}^x\frac{dt}{t^2}\sum_{e^e<p\le t}h(p).
\end{aligned}$$

The quantity $\sum_{p\le t}h(p)$ has been evaluated asymptotically in Lemma 15, and the quantity $\sum_{e^e<p\le t}h(p)$ differs by only $O(1)$. Therefore we may use Lemma 15 and the remark

following its statement to write

$$\begin{aligned} M_1(x) &= O(1) + \frac{1}{x}O\Big(\frac{xy\log y}{\log x}\Big) \\ &\quad + \int_{e^e}^x \frac{dt}{t^2}\Big(\frac{2t\log\log t\log y}{\log t} + O\Big(\frac{t\log\log t}{\log t} + \frac{t\log^2 y}{\log t} + t^{3/4}\log t\cdot y^2\Big)\Big) \\ &= O\Big(\frac{y\log y}{\log x}\Big) + \log y\int_{e^e}^x \frac{2\log\log t}{t\log t}\,dt \\ &\quad + O\bigg(\int_{e^e}^x \frac{\log\log t}{t\log t}\,dt + \log^2 y\int_{e^e}^x \frac{dt}{t\log t} + y^2\int_{e^e}^x \frac{dt}{t^{5/4}}\bigg). \end{aligned}$$

Each of these integrals can be explicitly evaluated, resulting in the asymptotic formula

$$\begin{aligned} M_1(x) &= \log y\big((\log\log x)^2 - 1\big) + O\Big(\frac{y\log y}{\log x} + (\log\log x)^2 + \log^2 y\cdot\log\log x + y^2\Big) \\ &= y^2\log y + O(y^2), \end{aligned}$$

as claimed. □

Now we turn our attention to $M_2(x)$, beginning with some preliminary lemmas.

Lemma 16. *For all real numbers $x > e^{e^e}$ and $t > e^e$, we have*

$$\sum_{q_1,q_2\le y^2} \log q_1\log q_2 \sum_{a_1,a_2\in\mathbb{N}} \sum_{r\in\mathcal{P}_{q_1^{a_1}}\cap\mathcal{P}_{q_2^{a_2}}} \sum_{\substack{p\le t\\ p\equiv 1\ (\mathrm{mod}\ r)}} 1 \ll t^{7/8}\log t\cdot y^2\log y + \frac{t\log\log t\cdot\log^2 y}{\log t}.$$

Proof. Since the exact form of $\mathcal{P}_{q_1^{a_1}}\cap\mathcal{P}_{q_2^{a_2}}$ depends on whether or not $q_1 = q_2$, we split the expression in question into two separate sums:

$$\begin{aligned} &\sum_{q_1,q_2\le y^2} \log q_1\log q_2 \sum_{a_1,a_2\in\mathbb{N}} \sum_{r\in\mathcal{P}_{q_1^{a_1}}\cap\mathcal{P}_{q_2^{a_2}}} \sum_{\substack{p\le t\\ p\equiv 1\ (\mathrm{mod}\ r)}} 1 \qquad (28) \\ &\quad = \sum_{q\le y^2}\log^2 q \sum_{a_1,a_2\in\mathbb{N}} \sum_{r\in\mathcal{P}_{q^{\max\{a_1,a_2\}}}} \pi(t;r,1) + \sum_{\substack{q_1,q_2\le y^2\\ q_1\ne q_2}} \log q_1\log q_2 \sum_{a_1,a_2\in\mathbb{N}} \sum_{r\in\mathcal{P}_{q_1^{a_1}q_2^{a_2}}} \pi(t;r,1). \end{aligned}$$

Noting that there are exactly $2a-1$ ordered pairs (a_1,a_2) for which $\max\{a_1,a_2\} = a$, we have

$$\begin{aligned} \sum_{q\le y^2}\log^2 q \sum_{a_1,a_2\in\mathbb{N}} \sum_{r\in\mathcal{P}_{q^{\max\{a_1,a_2\}}}} \pi(t;r,1) &= \sum_{q\le y^2}\log^2 q\sum_{a\in\mathbb{N}}(2a-1)\sum_{r\in\mathcal{P}_{q^a}}\pi(t;r,1) \\ &\ll \sum_{q\le y^2}\log^2 q \sum_{\substack{a\in\mathbb{N}\\ q^a>t^{1/4}}} \frac{at\log t}{q^a} + \sum_{q\le y^2}\log^2 q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \frac{at\log\log t}{q^a\log t} \end{aligned}$$

by Lemma 14. Since

$$\sum_{q\le y^2}\log^2 q \sum_{\substack{a\in\mathbb{N}\\ q^a>t^{1/4}}} \frac{at\log t}{q^a} \ll t\log t\log y^2\sum_{q\le y^2}\frac{\log q}{t^{1/4}} \ll t^{3/4}\log t\cdot y^2\log y$$

by the Chebyshev bound (11), and

$$\sum_{q\le y^2}\log^2 q \sum_{\substack{a\in\mathbb{N}\\ q^a\le t^{1/4}}} \frac{at\log\log t}{q^a\log t} \ll \frac{t\log\log t}{\log t}\sum_{q\le y^2}\frac{\log^2 q}{q} \ll \frac{t\log\log t\cdot\log^2 y}{\log t}$$

by (11) and its weaker version (12), the first term on the right-hand side of equation (28) is bounded by the estimate asserted in the statement of the lemma.

It remains to satisfactorily bound the second term on the right-hand side of equation (28). Again dividing the sum so that Lemma 14 can be applied, we have

$$\begin{aligned}\sum_{\substack{q_1,q_2\le y^2\\ q_1\ne q_2}}\log q_1\log q_2 \sum_{a_1,a_2\in\mathbb{N}}\sum_{r\in\mathcal{P}_{q_1^{a_1}q_2^{a_2}}}\pi(t;r,1) &\ll \sum_{q_1,q_2\le y^2}\log q_1\log q_2 \sum_{\substack{a_1,a_2\in\mathbb{N}\\ q_1^{a_1}q_2^{a_2}>t^{1/4}}}\frac{t\log t}{q_1^{a_1}q_2^{a_2}}\\ &\quad+\sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{\substack{a_1,a_2\in\mathbb{N}\\ q_1^{a_1}q_2^{a_2}\le t^{1/4}}}\frac{t\log\log t}{q_1^{a_1}q_2^{a_2}\log t}.\end{aligned}$$

In the first of these two terms, at least one of the $q_i^{a_i}$ must exceed $t^{1/8}$, and so using the estimates (15), (11), and (12) we see that

$$\begin{aligned}\sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{\substack{a_1,a_2\in\mathbb{N}\\ q_1^{a_1}q_2^{a_2}>t^{1/4}}}\frac{t\log t}{q_1^{a_1}q_2^{a_2}} &\le 2t\log t\sum_{q_1\le y^2}\log q_1\sum_{\substack{a_1\in\mathbb{N}\\ q_1^{a_1}>t^{1/8}}}\frac{1}{q_1^{a_1}}\sum_{q_2\le y^2}\log q_2\sum_{a_2\in\mathbb{N}}\frac{1}{q_2^{a_2}}\\ &\ll t\log t\sum_{q_1\le y^2}\frac{\log q_1}{t^{1/8}}\sum_{q_2\le y^2}\frac{\log q_2}{q_2}\\ &\ll t^{7/8}\log t\cdot y^2\log y.\end{aligned}$$

In the second, we simply ignore the restriction $q_1^{a_1}q_2^{a_2}\le t^{1/4}$ and use the estimates (15) and (12), obtaining

$$\begin{aligned}\sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{a_1,a_2\in\mathbb{N}}\frac{t\log\log t}{q_1^{a_1}q_2^{a_2}\log t} &= \frac{t\log\log t}{\log t}\bigg(\sum_{q\le y^2}\log q\sum_{a\in\mathbb{N}}\frac{1}{q^a}\bigg)^2\\ &\ll \frac{t\log\log t}{\log t}\bigg(\sum_{q\le y^2}\frac{\log q}{q}\bigg)^2\\ &\ll \frac{t\log\log t\cdot\log^2 y}{\log t}.\end{aligned}$$

This concludes the proof of the lemma. □

The following lemma is similar in spirit to Lemma 14 but is a bit more complicated to state and prove.

Lemma 17. *Let b_1 and b_2 be positive integers and $t > e^e$ a real number.*

(a) *If $b_1 > t^{1/8}$ or $b_2 > t^{1/8}$ then*

$$\sum_{r_1\in\mathcal{P}_{b_1}}\sum_{r_2\in\mathcal{P}_{b_2}}\pi(t;r_1r_2,1) \ll \frac{t\log^2 t}{b_1b_2}.$$

(b) *If neither b_1 nor b_2 exceeds $t^{1/8}$ then*

$$\sum_{r_1\in\mathcal{P}_{b_1}} \sum_{\substack{r_2\in\mathcal{P}_{b_2} \\ r_1r_2>t^{1/3}}} \pi(t;r_1r_2,1) \ll \frac{b_2t\log\log t}{\phi(b_1)\phi(b_2)^2\log t}$$

and

$$\sum_{r_1\in\mathcal{P}_{b_1}} \sum_{r_2\in\mathcal{P}_{b_2}} \pi(t;r_1r_2,1) \ll \frac{t(\log\log t)^2}{\phi(b_1)\phi(b_2)\log t}.$$

Remark. Again, the values 1/8 and 1/3 for the exponents are rather arbitrary.

Proof. The bound in part (a) follows from the trivial estimate $\pi(t;r_1r_2,1) \ll t/r_1r_2$, just as in the proof of Lemma 14(a). For the first estimate in part (b), we my assume that $r_1 \le r_2$ by symmetry. We use Brun's method again:

$$\begin{aligned}
&\sum_{r_1\in\mathcal{P}_{b_1}} \sum_{\substack{r_2\in\mathcal{P}_{b_2} \\ r_1\le r_2 \\ r_1r_2>t^{1/3}}} \pi(t;r_1r_2,1) \\
&\qquad = \#\{(m,r_1,r_2)\colon r_1\equiv 1 \pmod{b_1},\ r_2\equiv 1 \pmod{b_2},\ r_1\le r_2,\ r_1r_2>t^{1/3}, \\
&\qquad\qquad mr_1r_2+1\le t, \text{ and } r_1, r_2, \text{ and } mr_1r_2+1 \text{ are all prime}\} \\
&\qquad \le \sum_{m<t^{2/3}} \sum_{\substack{r_1<\sqrt{t/m} \\ r_1\in\mathcal{P}_{b_1}}} \sum_{\substack{r_2<t/mr_1 \\ r_2\in\mathcal{P}_{b_2} \\ mr_1r_2+1 \text{ prime}}} 1 \\
&\qquad \ll \sum_{m<t^{2/3}} \sum_{\substack{r_1<\sqrt{t/m} \\ r_1\in\mathcal{P}_{b_1}}} \frac{mr_1b_2}{\phi(b_2)\phi(mr_1b_2)} \cdot \frac{t/mr_1}{\log^2(t/mr_1b_2)}.
\end{aligned}$$

Notice that $t/mr_1b_2 > (\sqrt{t/m})/b_2 > t^{1/6}/t^{1/8} = t^{1/24}$, and so

$$\begin{aligned}
\sum_{r_1\in\mathcal{P}_{b_1}} \sum_{\substack{r_2\in\mathcal{P}_{b_2} \\ r_1\le r_2 \\ r_1r_2>t^{1/3}}} \pi(t;r_1r_2,1) &\ll \frac{t}{\log^2 t} \sum_{m<t^{2/3}} \sum_{\substack{r_1<\sqrt{t/m} \\ r_1\in\mathcal{P}_{b_1}}} \frac{b_2}{\phi(b_2)^2\phi(m)\phi(r_1)} \\
&\ll \frac{b_2t\log\log t}{\phi(b_1)\phi(b_2)^2\log^2 t} \sum_{m<t^{2/3}} \frac{1}{\phi(m)} \\
&\ll \frac{b_2t\log\log t}{\phi(b_1)\phi(b_2)^2\log t}.
\end{aligned}$$

by the estimates (7) and (23) as desired. The second estimate of part (b) is a consequence of the first estimate and

$$\sum_{r_1\in\mathcal{P}_{b_1}} \sum_{\substack{r_2\in\mathcal{P}_{b_2} \\ r_1r_2\le t^{1/3}}} \pi(t;r_1r_2,1) \ll \frac{t(\log\log t)^2}{\phi(b_1)\phi(b_2)\log t},$$

which follows from the Brun–Titchmarsh inequality just as in the proof of Lemma 14(b). □

Proof of Proposition 13. We may rewrite

$$\begin{aligned}\sum_{p\le t} h(p)^2 &= \sum_{p\le t}\Bigg(\sum_{r\mid p-1}\sum_{q\le y^2}\sum_{\substack{a\in\mathbb{N}\\ q^a\mid r-1}}\log q\Bigg)^2\\ &= \sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{a_1,a_2\in\mathbb{N}}\sum_{\substack{r_1\in\mathcal{P}_{q_1^{a_1}}\\ r_2\in\mathcal{P}_{q_2^{a_2}}}}\sum_{\substack{p\le t\\ p\equiv 1 \ (\mathrm{mod}\ r_1)\\ p\equiv 1 \ (\mathrm{mod}\ r_2)}} 1\\ &= \sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{a_1,a_2\in\mathbb{N}}\sum_{\substack{r_1\in\mathcal{P}_{q_1^{a_1}}\\ r_2\in\mathcal{P}_{q_2^{a_2}}\\ r_1\ne r_2}}\sum_{\substack{p\le t\\ p\equiv 1 \ (\mathrm{mod}\ r_1)\\ p\equiv 1 \ (\mathrm{mod}\ r_2)}} 1\\ &\qquad + O\Big(t^{7/8}\log t\cdot y^2\log y+\frac{t\log\log t\cdot\log^2 y}{\log t}\Big),\end{aligned}$$

the last step due to Lemma 16. Since r_1 and r_2 are distinct primes, the innermost sum is simply $\pi(t;r_1r_2,1)$, and thus

$$\begin{aligned}\sum_{p\le t} h(p)^2 \le \sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{a_1,a_2\in\mathbb{N}}\sum_{\substack{r_1\in\mathcal{P}_{q_1^{a_1}}\\ r_2\in\mathcal{P}_{q_2^{a_2}}}}\pi(t;r_1r_2,1)&\\ + O\Big(t^{7/8}\log t\cdot y^2\log y+\frac{t\log\log t\cdot\log^2 y}{\log t}\Big).&\end{aligned}\tag{29}$$

The contribution to the sum on the right-hand side of equation (29) from those terms for which $q_1^{a_1}>t^{1/8}$ is

$$\begin{aligned}\sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{\substack{a_1,a_2\in\mathbb{N}\\ q_1^{a_1}>t^{1/8}}}\sum_{\substack{r_1\in\mathcal{P}_{q_1^{a_1}}\\ r_2\in\mathcal{P}_{q_2^{a_2}}}}\pi(t;r_1r_2,1)&\\ \ll \sum_{q_1,q_2\le y^2}\log q_1\log q_2\sum_{\substack{a_1,a_2\in\mathbb{N}\\ q_1^{a_1}>t^{1/8}}}\frac{t\log^2 t}{q_1^{a_1}q_2^{a_2}}&\\ \ll t\log^2 t\sum_{q_1\le y^2}\sum_{\substack{a_1\in\mathbb{N}\\ q_1^{a_1}>t^{1/8}}}\frac{\log q_1}{q_1^{a_1}}\sum_{q_2\le y^2}\sum_{a_2\in\mathbb{N}}\frac{\log q_2}{q_2^{a_2}}&\\ \ll t\log^2 t\sum_{q_1\le y^2}\frac{\log q_1}{t^{1/8}}\sum_{q_2\le y^2}\frac{\log q_2}{q_2}&\\ \ll t^{7/8}\log^2 t\cdot y^2\log y&\end{aligned}$$

by Lemma 17(a) and the estimates (15), (11), and (12); the contribution from the terms for which $q_2^{a_2} > t^{1/8}$ is bounded likewise. The remaining contribution is

$$\begin{aligned}\sum_{q_1,q_2\le y^2} \log q_1 \log q_2 \sum_{\substack{a_1,a_2\in\mathbb{N}\\ q_1^{a_1},q_2^{a_2}\le t^{1/8}}} \sum_{\substack{r_1\in\mathcal{P}_{q_1^{a_1}}\\ r_2\in\mathcal{P}_{q_2^{a_2}}}} \pi(t;r_1r_2,1) \qquad\qquad & \\ \ll \sum_{q_1,q_2\le y^2} \log q_1 \log q_2 \sum_{\substack{a_1,a_2\in\mathbb{N}\\ q_1^{a_1},q_2^{a_2}\le t^{1/8}}} \frac{t(\log\log t)^2}{q_1^{a_1}q_2^{a_2}\log t} & \\ \ll \frac{t(\log\log t)^2}{\log t}\biggl(\sum_{q\le y^2}\sum_{a\in\mathbb{N}}\frac{\log q}{q^a}\biggr)^2 & \\ \ll \frac{t(\log\log t)^2\log^2 y}{\log t} &\end{aligned}$$

by Lemma 17(b) and the estimates (15) and (12). Using both these bounds in equation (29), we conclude that

$$\sum_{p\le t} h(p)^2 \ll t^{7/8}\log t\cdot y^2\log y + \frac{t(\log\log t)^2\log^2 y}{\log t}.$$

We now evaluate $M_2(x)$ using partial summation. We have

$$\begin{aligned} M_2(x) = \sum_{p\le x}\frac{h(p)^2}{p} &= \sum_{p\le e^e}\frac{h(p)^2}{p} + \frac{1}{x}\sum_{e^e<p\le x} h(p)^2 + \int_{e^e}^x \frac{dt}{t^2}\sum_{e^e<p\le t} h(p)^2 \\ &\ll 1 + \frac{1}{x}\cdot\frac{x(\log\log x)^2\log y}{\log x} \\ &\qquad + \int_{e^e}^x \frac{dt}{t^2}\Bigl(t^{7/8}\log t\cdot y^2\log y + \frac{t(\log\log t)^2\log^2 y}{\log t}\Bigr) \\ &\ll \frac{y^2\log y}{\log x} + y^2\log y\int_{e^e}^x \frac{\log t\,dt}{t^{9/8}} + \log^2 y\int_{e^e}^x \frac{(\log\log t)^2}{t\log t}\,dt. \end{aligned}$$

Evaluating these two integrals explicitly, we obtain

$$M_2(x) \ll \frac{y^2\log y}{\log x} + y^2\log y + \log^2 y\cdot(\log\log x)^3 \ll y^3\log^2 y$$

as claimed. □

6. Normal number of cycles for the power generator

If $(u,n) = 1$, then the sequence $u^i \pmod n$ for $i = 1,2,\dots$ is purely periodic. We denote the length of the period by $\mathrm{ord}(u,n)$, which of course is the multiplicative order of u in $(\mathbb{Z}/n\mathbb{Z})^\times$. Even when $(u,n) > 1$, the sequence $u^i \pmod n$ is eventually periodic, and we denote the length of the eventual cycle by $\mathrm{ord}^*(u,n)$. So, letting $n_{(u)}$ denote the largest divisor of n coprime to u, we have $\mathrm{ord}^*(u,n) = \mathrm{ord}(u,n_{(u)})$. For example, let $u = 2$, $n = 24$. The sequence $u^i \pmod n$ is $2,4,8,16,8,16,\dots$ with cycle length 2, and so $\mathrm{ord}^*(2,24) = \mathrm{ord}(2,3) = 2$.

When iterating the ℓth power map modulo n, the length of the eventual cycle starting with $x = u$ is given by $\mathrm{ord}^*(\ell, \mathrm{ord}^*(u,n))$. We would like to have a criterion for when a residue is part of some cycle, that is, for when a residue is eventually sent back to itself when iterating $x \mapsto x^\ell \pmod n$.

Lemma 18. *A residue u is part of some cycle under iteration of the map $x \mapsto x^\ell \pmod n$ if and only if $(\ell, \mathrm{ord}^*(u,n)) = 1$ and, with $d = (u,n)$, we have $(d, n/d) = 1$.*

Proof. If $(u,n) = d$, then high powers of u will be $\equiv 0 \pmod{n/n_{(d)}}$. Thus, for u to be in a cycle it is necessary that $n/n_{(d)} = d$, that is, $(d, n/d) = 1$. Further, it is necessary that $(\ell, \mathrm{ord}^*(u,n)) = 1$. Indeed, if $\sigma = \mathrm{ord}^*(u,n)$, we would need $\ell^i \pmod \sigma$ to be purely periodic, which is equivalent to $(\ell,\sigma) = 1$. This proves the necessity of the condition. For the sufficiency, we have just noted that $(\ell,\sigma) = 1$ implies that $\ell^i \pmod \sigma$ is purely periodic. This implies in turn that the sequence $u^{\ell^i} \pmod{n_{(u)}}$ is purely periodic. But the condition $(d, n/d) = 1$ implies that $n_{(u)} = n/d$, and as each $u^{\ell^i} \equiv 0 \pmod d$, we have that $u^{\ell^i} \pmod n$ is purely periodic. □

For $d|n$ with $(d, n/d) = 1$, let $C_d(\ell, n)$ denote the number of cycles in the ℓth power map mod n that involve residues u with $(u,n) = d$. For the lower bound in Theorem 2 we shall deal only with $C_1(\ell,n)$, that is, cycles involving numbers coprime to n.

Lemma 19. *We have $C_1(\ell,n) \geq \phi(n)_{(\ell)}/\lambda(\lambda(n))$.*

Proof. It is easy to see that the subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$ of residues u with $(\ell, \mathrm{ord}(u,n)) = 1$ has size $\phi(n)_{(\ell)}$. (In fact, this is true for any finite abelian group G: the size of the subgroup of elements with order coprime to ℓ is $|G|_{(\ell)}$.) As the length of *any* cycle in the ℓth power map is bounded above by $\lambda(\lambda(n))$, the lemma follows immediately. □

To investigate the normal size of $\phi(n)_{(\ell)}$, we introduce the function

$$f_\ell(n) = \sum_{p|\ell} v_p(\phi(n)) \log p.$$

We also make use of the notation $q^a \| n$, which means that q^a is the exact power of q dividing n, that is, q^a divides n but q^{a+1} does not.

Proposition 20. *For any fixed ℓ, we have $f_\ell(n) \leq (\log\log n)^2$ for almost all n, in fact for all but $O_\ell(x/\log\log x)$ integers $n \leq x$.*

Proof. We have

$$\begin{aligned}\sum_{n\leq x} f_\ell(n) &= \sum_{p|\ell}\sum_{n\leq x}\sum_{q^a\|n} v_p(\phi(q^a))\log p \;\leq\; x\sum_{p|\ell}\log p\sum_{q^a\leq x}\frac{v_p(\phi(q^a))}{q^a}\\ &\leq\; x\sum_{p|\ell}\log p\sum_{p^a\leq x}\frac{a-1}{p^a} + x\sum_{p|\ell}\log p\sum_{q\leq x}\frac{v_p(q-1)}{q}.\end{aligned}$$

Now

$$x\sum_{p|\ell}\log p\sum_{p^a\leq x}\frac{a-1}{p^a} \;\ll_\ell\; x$$

and, by (8),

$$x\sum_{p\mid\ell}\log p\sum_{q\le x}\frac{v_p(q-1)}{q} = x\sum_{p\mid\ell}\log p\sum_{a\ge1}\sum_{q\in\mathcal{P}_{p^a},\,q\le x}\frac{1}{q} \ll x\sum_{p\mid\ell}\log p\sum_{a\ge1}\frac{\log\log x}{p^a} \ll_\ell x\log\log x.$$

Hence,

$$\sum_{n\le x} f_\ell(n) \ll_\ell x\log\log x,$$

so that the number of $n \le x$ with $f_\ell(n) > (\log\log n)^2$ is $O_\ell(x/\log\log x)$. □

It is interesting that one can prove an Erdős–Kac theorem for $f_\ell(n)$ using as a tool the criterion of Kubilius–Shapiro (see [11], [16]).

Proof of the lower bound in Theorem 2. Noting that $\phi(n)_{(\ell)} = \phi(n)/e^{f_\ell(n)}$, we have $\phi(n)_{(\ell)} \ge \phi(n)/\exp((\log\log n)^2)$ for almost all n by Proposition 20. Of course, $n \ge \phi(n) \gg n/\log\log n$ for all $n \ge 3$. Therefore, using Lemma 19 and Theorem 2, we have

$$\begin{aligned} C(\ell,n) \ge C_1(\ell,n) \ge \frac{\phi(n)_{(\ell)}}{\lambda(\lambda(n))} &\ge \frac{\phi(n)}{\exp((\log\log n)^2)\lambda(\lambda(n))} \\ &= \frac{\phi(n)/n}{\exp((\log\log n)^2)}\frac{n}{\lambda(\lambda(n))} \\ &= \exp((1+o(1))(\log\log n)^2\log\log\log n) \end{aligned}$$

for almost all n. This completes the proof of the lower bound in Theorem 2. □

We now consider the upper bounds in Theorem 2, first establishing a lemma.

Lemma 21. *Suppose m is a positive integer and $(d,m)=1$. For any integer $j \mid \lambda(m)$, the number of integers $u \in [1,m]$ with $(u,m)=1$ and $\operatorname{ord}(du,m) \mid \lambda(m)/j$ is at most $\phi(m)/j$.*

Proof. In fact, we prove a more general statement for any finite abelian group G: let $\lambda(G)$ denote the exponent of G, that is, the order of the largest cyclic subgroup of G, or equivalently the least common multiple of the orders of the elements of G. Then for any $d \in G$ and any $j \mid \lambda(G)$, the number of elements $u \in G$ for which the order of du divides $\lambda(G)/j$ is at most $\#G/j$. It is clear that the lemma follows immediately from this statement upon taking G to be $(\mathbb{Z}/m\mathbb{Z})^\times$. It is also clear that in this statement, the element d plays no role whatsoever except to shuffle the elements of G around, and so we assume without loss of generality that d is the identity of G.

Let p be any prime dividing $\lambda(G)$, and choose $a \le b$ so that $p^a \| j$ and $p^b \| \lambda(G)$. When we write G canonically as isomorphic to the direct product of cyclic groups of prime-power order, at least one of the factors must be isomorphic to $\mathbb{Z}/p^b\mathbb{Z}$. In every such factor, only one out of every p^a elements has order dividing $\lambda(G)/j$, since all but p^{b-a} elements of the factor have order divisible by p^{b-a+1}. Since there is at least one such factor for every $p^a \| j$, we conclude that at most one out of every j elements of G has order dividing $\lambda(G)/j$, as claimed. □

Note that this result in the case $d = 1$ is Lemma 1 in [9]. The above proof, while similar in spirit to the proof in [9], is simpler.

Let $\tau(m)$ denote the number of positive divisors of m.

Proposition 22. *For any integers $\ell, n \geq 2$ we have $C(\ell, n) \leq n\tau(\lambda(n))\tau(n)/\operatorname{ord}^*(\ell, \lambda(n))$.*

Proof. It is sufficient to show that for each $\ell, n \geq 2$ and each $d \mid n$ with $(d, n/d) = 1$, we have

$$C_d(\ell, n) \leq \frac{n\tau(\lambda(n))}{\operatorname{ord}^*(\ell, \lambda(n))}. \tag{30}$$

Let $d \mid n$ with $(d, n/d) = 1$. We have seen in Lemma 18 that for a residue u (mod n) with $(u, n) = d$ to be involved in a cycle, it is necessary and sufficient that $(\ell, \operatorname{ord}(u, n/d)) = 1$. For each integer $j \mid \lambda(n/d)$, let $C_{d,j}(\ell, n)$ denote the number of cycles corresponding to residues u with $(u, n) = d$ and $\operatorname{ord}(u, n/d) = \lambda(n/d)/j$. Writing such a residue u as du_1, we have $u_1 \in [1, n/d]$ and $(u_1, n/d) = 1$. Thus, by Lemma 21, we have that the number of such residues u is at most $\phi(n/d)/j \leq n/dj$. Hence we have

$$C_{d,j}(\ell, n) \leq \frac{n/dj}{\operatorname{ord}(\ell, \lambda(n/d)/j)}.$$

Now $\lambda(n/d) = \lambda(n)/d_1$ for some integer $d_1 \leq d$. It is shown in (15) of [12] that for $k \mid m$ we have $\operatorname{ord}^*(a, m/k) \geq \operatorname{ord}^*(a, m)/k$ for any nonzero integer a. Hence

$$\operatorname{ord}(\ell, \lambda(n/d)/j) = \operatorname{ord}(\ell, \lambda(n)/d_1 j) \geq \operatorname{ord}^*(\ell, \lambda(n))/d_1 j,$$

so that

$$C_{d,j}(\ell, n) \leq \frac{n/dj}{\operatorname{ord}^*(\ell, \lambda(n))/d_1 j} \leq \frac{n}{\operatorname{ord}^*(\ell, \lambda(n))}.$$

Letting j range over all divisors of $\lambda(n/d)$, we get that

$$C_d(\ell, n) \leq \frac{n\tau(\lambda(n/d))}{\operatorname{ord}^*(\ell, \lambda(n))},$$

which immediately gives (30). □

Proof of the upper bounds in Theorem 2. Note that from [6, Theorem 4.1], we have $\tau(\lambda(n)) < \exp((\log\log n)^2)$ for almost all n. Furthermore, letting $\Omega(n)$ denote the number of prime factors of n counted with multiplicity, we know that the normal order of $\Omega(n)$ is $\log\log n$; in particular, we have $\Omega(n) < \log\log n/\log 2$ for almost all n. Since the inequality $\tau(n) \leq 2^{\Omega(n)}$ is elementary, this implies that $\tau(n) < \log n$ for almost all n. We conclude from Proposition 22 that

$$C(\ell, n) < n\exp(2(\log\log n)^2)/\operatorname{ord}^*(\ell, \lambda(n))$$

for almost all n.

The three upper bounds in Theorem 2 therefore follow respectively from three results in the new paper of Kurlberg and the second author [12]: Theorem 4 (1), which states that for any function $\varepsilon(n) \to 0$, we have $\operatorname{ord}^*(\ell, \lambda(n)) \geq n^{1/2+\varepsilon(n)}$ almost always; Theorem 22, which states that a positive proportion of integers n have $\operatorname{ord}^*(\ell, \lambda(n)) \geq n^{.592}$; and Theorem 28, which states that if the GRH is true, then

$$\operatorname{ord}^*(\ell, \lambda(n)) = n/\exp((1 + o(1))(\log\log n)^2 \log\log\log n)$$

on a set of asymptotic density 1. (Note that the proof of this result uses Theorem 1 of the current paper.) □

7. Higher iterates

Here we sketch what we believe to be a viable strategy for establishing an analogue of Theorem 1 for the higher iterates λ_k where $k \geq 3$. As in the case of $k = 2$, we have generally that

$$\frac{n}{\lambda_k(n)} = \frac{n}{\phi_k(n)} \frac{\phi_k(n)}{\lambda_k(n)}.$$

We always have $n/\phi_k(n) \leq (c \log\log n)^k$, which is already a good enough estimate for our purposes. Even better, however, it is known [5] that for each fixed k, we have $n/\phi_k(n) \ll (\log\log\log n)^k$ for almost all n. The problem therefore reduces to comparing $\lambda_k(n)$ to $\phi_k(n)$. Probably it is not hard to get analogs of Propositions 5 and 6, where we replace y^2 with y^k. The problem comes in with the proliferation of cases needed to deal with small prime factors. As with the second iterate, we expect the main contribution to come from the "supersquarefree" case. In particular, let

$$h_k(n) = \sum_{p_1 | n} \sum_{p_2 | p_1 - 1} \cdots \sum_{p_k | p_{k-1} - 1} \sum_{q \leq y^k} v_q(p_k - 1) \log q.$$

We expect $h_k(n)$ to be the dominant contribution to $\log(\phi_k(n)/\lambda_k(n))$ almost always. But it seems hard not only to prove this in general but also to establish the normal order of $h_k(n)$.

It would seem useful in this endeavor to have a uniform estimate of the shape

$$\sum_{p \in \mathcal{P}_m,\, p \leq x} \frac{1}{p} \sim \frac{\log\log x - \log\log m}{\phi(m)} \quad \text{for } x \geq m^{1+\varepsilon}. \tag{31}$$

Even under the assumption of the Riemann Hypothesis for Dirichlet L-functions, (31) seems difficult, and maybe it is false. It implies with $x = m^2$ that the sum is $\ll 1/\phi(m)$, when all we seem to be able to prove, via sieve methods, is that it is $\ll (\log\log m)/\phi(m)$.

Assuming uniformity in (31), it seems that on average

$$h_k(n) \sim \frac{1}{(k-1)!} (\log\log n)^k \log\log\log n,$$

supporting Conjecture 3. It would be a worthwhile enterprise to try to verify or disprove the Conjecture in the case $k = 3$, which may be tractable.

Going out even further on a limb, it may be instructive to think of what Conjecture 3 has to say about the normal order of $L(n)$, the minimum value of k with $\lambda_k(n) = 1$. The expression $(1/(k-1)!)(\log\log n)^k \log\log\log n$ reaches its maximum value when $k \approx \log\log n$. Is this formula then trying to tell us that we have $L(n) \ll \log\log n$ almost always? Perhaps so.

There is a second argument supporting the thought that $L(n) \ll \log\log n$ almost always. Let $P(n)$ denote the largest prime factor of an integer $n > 1$, and let $\ell(n) = P(n) - 1$ for $n > 1$, $\ell(1) = 1$. Clearly, $\ell(n) \mid \lambda(n)$ for all n, so that if $L_0(n)$ is the least k with $\ell_k(n) = 1$, then $L_0(n) \leq L(n)$. It may be that the difference $L(n) - L_0(n)$ is usually not large. In any event, it seems safe to conjecture that $L_0(n)$ is usually of order of magnitude $\log\log n$, due to the following argument. For an odd prime p, consider the quantity

$\log \ell(p)/\log p \approx \log P(p-1)/\log(p-1)$. It may be that this quantity is distributed as p varies through the primes in the same way that $\log P(n)/\log n$ is distributed as n varies through the integers, namely the Dickman distribution. Such a conjecture has been made in various papers. If so, it may be that the sequence

$$\frac{\log \ell(p)}{\log p}, \frac{\log \ell_2(p)}{\log \ell(p)}, \dots$$

behaves like a sequence of independent random variables, each with the Dickman distribution. And if so, it may then be reasonable to assume that almost always we get down to small numbers and terminate in about $\log\log n$ steps. A similar probabilistic model is considered in [1], but for the simpler experiment of finding the joint distribution of logarithmic sizes of the various prime factors of a given number n.

At the very least, we can prove that $L(n) \ll \log\log n$ infinitely often.

Proof of Theorem 4. Notice that the definition of $\lambda(n)$ as a least common multiple, together with the fact that $\lambda(p^a) \mid \lambda(p^{a+1})$ always, implies that

$$\lambda\big(\operatorname{lcm}\{m_1,\dots,m_j\}\big) = \operatorname{lcm}\big\{\lambda(m_1),\dots,\lambda(m_j)\big\}$$

for any positive integers $m_1,\dots,m_j$. A trivial induction then shows that

$$\lambda_k\big(\operatorname{lcm}\{m_1,\dots,m_j\}\big) = \operatorname{lcm}\big\{\lambda_k(m_1),\dots,\lambda_k(m_j)\big\}$$

for any $k \ge 0$. Since the least common multiple of a set of numbers equals 1 precisely when each number in the set equals 1, we deduce that

$$L\big(\operatorname{lcm}\{m_1,\dots,m_j\}\big) = \max\big\{L(m_1),\dots,L(m_j)\big\}.$$

We apply this identity with $m_i = i$. Let $n_j = \operatorname{lcm}\{1,2,\dots,j\}$. We have $\log n_j = \sum_{i\le j}\Lambda(i)$, which is asymptotic to j by the prime number theorem. On the other hand, it is trivial that for any number n we have $L(n) \le 1 + (1/\log 2)\log n$, as $\lambda_{i+1}(n) \le (1/2)\lambda_i(n)$ for $1 \le i < L(n)$. Therefore

$$\begin{aligned} L(n_j) &= \max\{L(1),\dots,L(j)\} \le 1 + \max\left\{\frac{\log 1}{\log 2},\dots,\frac{\log j}{\log 2}\right\} \\ &= 1 + \frac{\log j}{\log 2} = \left(\frac{1}{\log 2} + o(1)\right)\log\log n_j. \end{aligned}$$

□

We can improve on the estimate in Theorem 4, but not by much. Say we let N_j be the product of all primes $p \le j^{3.29}$ with $p-1 \mid n_j$, with n_j as in the above proof. It follows from Friedlander [8] that a positive proportion of the primes $p \le j^{3.29}$ have the required property. Thus, $N_j > \exp(cj^{3.29})$ for some positive constant c and all sufficiently large values of j. But $\lambda(N_j) \mid n_j$, so that $L(N_j) \le 2 + j/\log 2$. Hence $L(N_j) < .439\log\log N_j$ for j sufficiently large. (This result can be improved by a very small margin using a more recent result of Baker and Harman [2], but the argument is a bit more difficult, since they do not get a positive proportion of the primes with the required property.) It is likely that $L(n) \ll \log\log\log n$ infinitely often, possibly even that $L(n) \ll_k \log_k n$ infinitely often for arbitrary k-fold-iterated logarithms.

One may also study the maximal order of $L(n)$. The analogous problem for the iterated ϕ-function is relatively trivial, but not so for λ. If there can exist very long "Sophie Germain chains", that is, sequences of primes $p_1, p_2, \dots, p_k$ where each $p_i = 2p_{i-1} + 1$, for $i > 1$, then we might have $L(p_k) \sim (1/\log 2) \log p_k$. We might even perturb such a chain by a small amount and keep the asymptotic relation, say by occasionally having $p_i = 4p_{i-1} + 1$. It seems hard to prove that long enough chains to get the the asymptotic for $L(p_k)$ do not exist, but probably they don't on probabilistic grounds. We can at least say that $L(n) \geq 1 + (1/\log 3) \log n$ infinitely often, since this inequality is attained when n is a power of 3.

References

[1] E. Bach, Analytic methods in the analysis and design of number-theoretic algorithms, MIT Press, Cambridge, MA, 1985.

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

[3] E. Blanton, S. Hurd, J. McCranie, On the digraph defined by squaring mod m, when m has primitive roots, Cong. Numerantium **82** (1992), 167–177.

[4] J. J. Brennan and B. Geist, Analysis of iterated modular exponentiation: the orbit of x^α mod N, Designs, Codes, and Cryptography **13** (1998), 229–245.

[5] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, On the normal behavior of the iterates of some arithmetic functions, in Analytic number theory (Allerton Park, IL, 1989), 165–204, Progr. Math., 85, Birkhäuser Boston, Boston, MA, 1990.

[6] P. Erdős and C. Pomerance, On the normal number of prime factors of $\varphi(n)$, Rocky Mountain J. Math., **15** (1985), 343–352. Corrigendum in [5].

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

[8] J. B. Friedlander, Shifted primes without large prime factors, in Number theory and applications (Banff, AB, 1988), 393–401, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 265, Kluwer Acad. Publ., Dordrecht, 1989.

[9] J. B. Friedlander, C. Pomerance, and I. E. Shparlinski, Period of the power generator and small values of Carmichael's function, Math. Comp., **70** (2001), 1591–1605. Corrigendum. Math. Comp., **71** (2002), 1803–1806.

[10] H. Halberstam and H.-E. Richert, Sieve methods, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974. London Mathematical Society Monographs, No. 4.

[11] J. P. Kubilius, Probabilistic methods in the theory of numbers, Translations of Mathematical Monographs, Vol. 11, American Math. Soc., Providence, 1964.

[12] P. Kurlberg and C. Pomerance, On the period of the linear congruential and power generators, to appear.

[13] K. K. Norton, On the number of restricted prime factors of an integer. I, Illinois J. Math., **20** (1976), 681–705.

[14] C. Pomerance, On the distribution of amicable numbers, J. Reine Angew. Math., 293/294 (1977), 217–222.

[15] T. D. Rogers, The graph of the square mapping on the prime fields, Discrete Math., **148** (1996), 317–324.

[16] H. N. Shapiro, Distribution functions of additive arithmetic functions, Proc. Nat. Acad. Sci. USA, **42** (1956), 426–430.

Department of Mathematics, University of British Columbia, Room 121, 1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada
E-mail address: gerg@math.ubc.ca

Mathematics Department, Dartmouth College, Hanover, NH 03755-3551, U.S.A.
E-mail address: carlp@math.dartmouth.edu