SQUARE VALUES OF EULER'S FUNCTION

PAUL POLLACK AND CARL POMERANCE

Abstract. We show that almost all squares are missing from the range of Euler's φ-function.

1. Introduction

Let φ denote Euler's function, let $\mathbf{N}$ denote the set of positive integers, and let $\mathscr{V} = \varphi(\mathbf{N})$, the set of values of φ. Further, let $V(x) = \#\{n \le x : n \in \mathscr{V}\}$. The distribution of $\mathscr{V}$ has been of interest since the 1930s when Erdős showed that $V(x) = x/(\log x)^{1+o(1)}$ as $x \to \infty$. We still do not have an asymptotic for $V(x)$, but after work of Ford [8], we do know the order of magnitude.

For a function $f\colon \mathbf{N} \to \mathbf{N}$, let

$$\mathscr{V}_f = \{n : f(n) \in \mathscr{V}\}, \quad \mathscr{V}^f = \{n : \varphi(n) \in f(\mathbf{N})\},$$

and let $V_f(x), V^f(x)$ be the respective counting functions for $\mathscr{V}_f, \mathscr{V}^f$. The situation when f is a linear polynomial is fairly well-understood. If $f(n) = kn$, where k is a fixed natural number, then $V_f(x) \sim V(kx)$ and $V^f(x) \sim x$ as $x \to \infty$; on the other hand, if $f(n) = kn + j$ with $0 < j < k$, then $V_f(x) = o(V(kx))$ and $V^f(x) = o(x)$. (The V_f-results do not appear to be in the literature, but follow from the method of Ford.) More refined results concerning the cases when $0 < j < k$ can be found in [16, 7, 9]. The case when $f = \sigma$, the sum-of-divisors function, was considered in [10], where some old questions of Erdős were settled (see also [11, 12]). This paper is concerned with the function $f(n) = n^2$, which we denote with the symbol $\square$, so that

$$V_\square(x) = \#\{n \le x : n^2 \in \mathscr{V}\}, \quad V^\square(x) = \#\{n \le x : \varphi(n) = m^2 \text{ for some integer } m\}.$$

It was shown in [2], perhaps counter-intuitively, that $V^\square(x) \ge x^{0.7}$ for all large x, with the conjectured exponent on x allowed to be any number below 1. In that paper it was also shown that $V_\square(x) \ge x^{0.234}$ for all sufficiently large x. This lower bound was considerably improved in [3], where it was shown that $V_\square(x) \gg x/(\log x)^4$ (compare with the case $r = 2$ of [13, Theorem 1.2]).

The paper [2] shows that $V^\square(x) \le x/\exp((1 + o(1))(\log x \log\log\log x)^{1/2})$ as $x \to \infty$, but does not address an upper bound for $V_\square(x)$. It is not immediately clear that $V_\square(x) = o(x)$. In fact, a short computer run shows that $V_\square(10^8) = 26{,}094{,}797$ so that more than half of the even numbers to 10^8 have their squares in the range of φ. In this paper we prove the following results.

Theorem 1. *For all sufficiently large numbers x, we have $V_\square(x) \le x/(\log x)^{0.0063}$.*

Theorem 2. *We have $V_\square(x) \gg x/(\log x \log\log x)^2$.*

In addition, we discuss some heuristics for the estimation of $V_\square(x)$ and we discuss the analogous problems for the sum-of-divisors function.

For the analogous problem with Carmichael's λ-function, one of us (CP) has a heuristic argument that asymptotically all even numbers n have n^2 a λ-value. However, not only have

2010 *Mathematics Subject Classification.* Primary: 11N37, Secondary: 11N25, 11N36.

we failed to prove this, we have not been able to find a proof of a lower bound similar to that of Theorem 2.

Notation. We use the Landau/Bachmann O and o-notation, as well as the associated Vinogradov $\ll$ and $\gg$ notations, with their standard meanings. We write $A \asymp B$ to mean that $A \ll B$ and $B \ll A$. Any dependence of implied constants is noted explicitly, often with a subscript.

The letters p, and ℓ, with or without subscripts, always denote primes. We use $P(n)$ for the largest prime factor of the natural number n, with the convention that $P(1) = 1$. The notation $p^e \parallel n$ means that $p^e \mid n$ but that $p^{e+1} \nmid n$; in this case, we say that p^e *exactly divides* n. As usual, $\Omega(n)$ denotes the number of prime factors of n counted with multiplicity; thus, $\Omega(n) = \sum_{p^k \parallel n} k$. We write $\log_k$ for the k-fold iterate of the natural logarithm.

2. Preparation

2.1. **Anatomy and sieving.** A classical theorem of Hardy and Ramanujan asserts that a typical natural number n has about $\log_2 n$ prime factors, regardless of whether or not the primes are counted with multiplicity. Our first lemma, which may be deduced from the results in Chapter 0 of [15], bounds from above the number of n for which $\Omega(n)$ is atypically large.

Lemma 3. *Let $x \geq 3$, and let $\epsilon > 0$. For $1 \leq \alpha \leq 2 - \epsilon$, the number of $n \leq x$ with $\Omega(n) \geq \alpha \log_2 x$ is $O_\epsilon(x(\log x)^{-Q(\alpha)})$, where we set $Q(\lambda) = \int_1^\lambda \log t\, dt = \lambda \log(\lambda) - \lambda + 1$.*

We now quote two upper bound sieve results, in slightly crude forms that are convenient for our later applications. Both of these follow from the general upper bound O-result appearing as [14, Theorem 2.2].

Lemma 4. *Suppose that $A_1, \dots, A_h$ are positive integers and $B_1, \dots, B_h$ are integers such that*

$$E := \prod_{i=1}^{h} A_i \prod_{1 \leq i < j \leq h} (A_i B_j - A_j B_i) \neq 0.$$

Then for $x \geq 3$,

$$\#\{n \leq x : A_i n + B_i \text{ prime for all } 1 \leq i \leq h\} \ll \frac{x}{(\log x)^h} (\log_2 |3E|)^h,$$

where the implied constant may depend on h.

Lemma 5. *Let A, B, and C be integers with $A > 0$ and $D = B^2 - 4AC$ not a square. Write $D = df^2$, where d is a fundamental discriminant. Then for $x \geq 3$,*

$$\#\{p \leq x : Ap^2 + Bp + C \text{ prime}\} \ll \frac{x}{(\log x)^2} (\log_2 |3ACD|)^3 \prod_{\ell \leq x} \left(1 - \frac{\left(\frac{d}{\ell}\right)}{\ell}\right), \tag{1}$$

where $\left(\frac{d}{\cdot}\right)$ is the Kronecker symbol.

2.2. **Sieving quadratics and short Euler products.** To control the size of the product on ℓ appearing in (1), we appeal to the methods and results of a recent preprint of Chandee, David, Koukoulopoulos, and Smith [5].

Lemma 6. *Let $\epsilon > 0$. Let χ be a nonprincipal real character* mod q. For all real $y \geq 1$, we have

$$\prod_{\ell \leq y} \left(1 - \frac{\chi(\ell)}{\ell}\right) \ll_\epsilon q^\epsilon.$$

Proof. The proof parallels that of [5, Lemma 3.2]. By Mertens' theorem,

$$\prod_{\ell \le \min\{y, \exp(q^\epsilon)\}} \left(1 - \frac{\chi(\ell)}{\ell}\right) \ll q^\epsilon,$$

so we may assume that $y > \exp(q^\epsilon)$ and it suffices to show that

$$\prod_{\exp(q^\epsilon) < \ell \le y} \left(1 - \frac{\chi(\ell)}{\ell}\right) \ll_\epsilon 1.$$

By the classical Siegel–Walfisz estimates (see [6, eq. (3), p. 132]),

$$\sum_{n \le x} \Lambda(n)\chi(n) \ll_\epsilon x/\log x \quad \text{for all} \quad x \ge \exp(q^\epsilon). \tag{2}$$

Recalling that $\log(1-t) = -\sum_{k \ge 1} t^k/k$ (for $|t| < 1$), we find that

$$\begin{aligned} \log \prod_{\exp(q^\epsilon) < \ell \le y} \left(1 - \frac{\chi(\ell)}{\ell}\right) &= - \sum_{n > 1,\ \ell \mid n \Rightarrow \exp(q^\epsilon) < \ell \le y} \frac{\Lambda(n)\chi(n)}{n \log n} \\ &= - \sum_{\exp(q^\epsilon) < n \le y} \frac{\Lambda(n)\chi(n)}{n \log n} + O(1) \ll_\epsilon 1. \end{aligned}$$

where the final estimate is obtained from (2) by partial summation. □

The next lemma is an equivalent form of [5, Lemma 3.3], which the authors of that paper attribute in essence to Elliott.

Lemma 7. *Fix $\delta \in (0, 1]$, and let $Q \ge 3$. We can choose a set $\mathscr{E}_\delta(Q)$ of real, primitive characters, all of conductor bounded by Q, with*

$$\#\mathscr{E}_\delta(Q) \ll_\delta Q^\delta$$

and so that the following holds: If χ is a primitive real character of conductor $q \le Q$ and $\chi \notin \mathscr{E}_\delta(Q)$, then

$$\prod_{y < \ell \le z} \left(1 - \frac{\chi(\ell)}{\ell}\right) \asymp_\delta 1 \quad \textit{uniformly for} \quad z \ge y \ge \log Q.$$

For each nonsquare integer d, let χ_d be the primitive real character of conductor $|D|$ given by the Kronecker symbol $\left(\frac{D}{\cdot}\right)$, where D is the discriminant of $\mathbf{Q}(\sqrt{d})$. It is convenient for us to isolate the following consequence of Lemma 7.

Lemma 8. *Let $\mathscr{D}$ be the set of squarefree $d \ne 1$ for which there exists a real number y with*

$$\prod_{\ell \le y} \left(1 - \frac{\chi_d(\ell)}{\ell}\right) \ge (\log_2 |3d|)^2. \tag{3}$$

For fixed $\delta \in (0, 1]$ and all $x \ge 1$, we have that

$$\#\{d \in \mathscr{D} : |d| \le x\} \ll_\delta x^\delta.$$

Proof. We can assume that x is large. It suffices to prove the stated estimate for $\#\{d \in \mathscr{D} : x^\delta < |d| \le x\}$. Let $\{y_i\}_{i=0}^\infty$ be the sequence of real numbers defined by $y_i = 4^i x^\delta$, and choose j so that $y_j < |d| \le y_{j+1}$. Then the conductor of χ_d is bounded by $4y_{j+1}$, and $4y_{j+1} < 16|d|$.

We claim that if $\chi_d \notin \mathscr{E}_\delta(4y_{j+1})$, then the inequality (3) never holds. Indeed, Lemma 7 (with $Q := 4y_{j+1}$) shows that for every y,

$$\prod_{\ell \le y}\left(1 - \frac{\chi_d(\ell)}{\ell}\right) \ll_\delta \prod_{\ell \le \min\{\log(4y_{j+1}), y\}}\left(1 - \frac{\chi_d(\ell)}{\ell}\right) \ll \log_2 |d|,$$

using Mertens' theorem in the final step. Since $|d| \ge x^\delta$ and x is large, this upper bound is incompatible with (3), proving our claim. Since distinct squarefree d give rise to distinct primitive real characters χ_d, the upper bound for $\#\mathscr{E}_\delta(Q)$ from Lemma 7 yields

$$\#\{d \in \mathscr{D} : x^\delta < |d| \le x\} \le \sum_{\substack{j \ge 0\\ y_j \le x}} \#\mathscr{E}_\delta(4y_{j+1}) \ll_\delta \sum_{0 \le j \le \frac{\log(x^{1-\delta})}{\log 4}} 4^{(j+2)\delta} x^{\delta^2} \ll_\delta x^\delta.$$

This completes the proof of the lemma. □

3. Proof of the upper bound (Theorem 1)

Setup. We assume throughout the argument that x is large. Let $n \le x$ be such that $n^2 = \varphi(m)$ for some integer m. By de Bruijn [4, eq. (1.6)], we can assume that

(i) $P(n) \ge x^{1/\log_2 x}$

since the number of $n \le x$ for which (i) fails is $O(x/\log x)$. We can also assume that

(ii) n is not divisible by any $d \in \mathscr{D}$ with $|d| > \log x$, where $\mathscr{D}$ is the set considered in Lemma 8.

Indeed, since $\#\{d \in \mathscr{D} : |d| \le t\} \ll t^{1/2}$ for all $t \ge 1$, the count of exceptional $n \le x$ is $O(x/(\log x)^{1/2})$ (by partial summation). At the cost of an additional exceptional set of the same order, we can further assume that

(iii) n is not divisible by any square exceeding $\log x$.

Introducing another exceptional set of size $O(x/(\log x)^{1/2})$, we can assume that

(iv) there is no prime p^2 dividing m with $p > \log x$.

Indeed, suppose that $p^2 \mid m$. Setting $r_p = \prod_{\ell^e \| p-1} \ell^{\lceil e/2 \rceil}$, we see that $p \cdot r_p \mid n$. Note that $r_p \ge \sqrt{p-1} \gg \sqrt{p}$. Hence, the number of n with $p^2 \mid m$ for some $p > \log x$ does not exceed

$$\sum_{p > \log x} \frac{x}{p \cdot r_p} \ll x \sum_{p > \log x} \frac{1}{p^{3/2}} \ll x/(\log x)^{1/2}.$$

Let α be a parameter with $1 < \alpha < 2$, which will be chosen later so as to optimize the argument. We assume that

(v) $\Omega(n) \le \alpha \log_2 x$,

noting that Lemma 3 guarantees that the number of exceptions $n \le x$ is

$$\ll_\alpha x/(\log x)^{1-\alpha+\alpha\log\alpha}. \tag{4}$$

Let $p = P(n)$, so that $p^2 \mid n^2 = \varphi(m)$. By (i) and (iv), we have that $p^2 \nmid m$, and so there are only two ways to explain how $p^2 \mid \varphi(m)$:

I. there are two different primes $q_1, q_2 \mid m$ with $q_i \equiv 1 \pmod p$ for $i = 1, 2$,
II. there is a prime $q \mid m$ with $q \equiv 1 \pmod{p^2}$.

Case I. We will assume that the primes q_1, q_2 are not 1 (mod p^2); otherwise $p^3 \mid n^2$, so $p^2 \mid n$, a violation of (i) and (iii). For such a prime q we may write it as $1 + apb^2$, where ap is squarefree. This shows that n may be written in the form

$$n = ua_1a_2a_3b_1b_2p, \quad \text{with} \quad a_1a_2a_3p \text{ squarefree, } 1 + a_1a_3pb_1^2 \text{ prime, } 1 + a_2a_3pb_2^2 \text{ prime.}$$

For each fixed choice of $u, a_1, a_2, a_3, b_1, b_2$ we count primes $p \le x/ua_1a_2a_3b_1b_2$ with the two primality conditions above holding. Using the upper bound sieve in the form of Lemma 4, and recalling that $x/ua_1a_2a_3b_1b_2 \ge p > x^{1/\log_2 x}$, we find that the number of these p is

$$\ll \frac{x}{ua_1a_2a_3b_1b_2(\log x)^3}(\log_2 x)^6. \tag{5}$$

(Explicitly, we apply Lemma 4 with $A_1 = 1$ and $B_1 = 0$, $A_2 = a_1a_3b_1^2$ and $B_2 = 1$, and $A_3 = a_2a_3b_2^2$ and $B_3 = 1$; note that since $q_1 \ne q_2$, we have $E \ne 0$, and $|E| < x^{O(1)}$.) Now we sum our upper bound (5) over the possibilities for $u, a_1, a_2, a_3, b_1, b_2$, keeping in mind that their product is bounded by x and $\Omega(ua_1a_2a_3b_1b_2) \le \alpha \log_2 x$. Here it is helpful to introduce an auxiliary parameter z (Rankin's trick); for $0 < z < 1$,

$$\sum_{\Omega(ua_1a_2a_3b_1b_2) \le \alpha \log_2 x} \frac{1}{ua_1a_2a_3b_1b_2} \le z^{-\alpha \log_2 x} \sum \frac{z^{\Omega(u)} z^{\Omega(a_1)} z^{\Omega(a_2)} z^{\Omega(a_3)} z^{\Omega(b_1)} z^{\Omega(b_2)}}{ua_1a_2a_3b_1b_2}.$$

Keeping only the restriction that $P(ua_1a_2a_3b_1b_2) \le x$, we find that

$$\sum \frac{z^{\Omega(u)} z^{\Omega(a_1)} z^{\Omega(a_2)} z^{\Omega(a_3)} z^{\Omega(b_1)} z^{\Omega(b_2)}}{ua_1a_2a_3b_1b_2} \le \left(\prod_{\ell \le x} (1 - z/\ell)^{-1} \right)^6 \ll (\log x)^{6z}.$$

(The last estimate uses Mertens' theorem.) Comparing the previous two displays, we find that $\sum \frac{1}{ua_1a_2a_3b_1b_2} \ll (\log x)^{6z - \alpha \log z}$. To optimize, we take $z = \alpha/6$ to get an upper bound of $O((\log x)^{\alpha - \alpha \log(\alpha/6)})$ for our reciprocal sum. Referring back to (5), we see that the total count of n in Case I is

$$\ll \frac{x}{(\log x)^{3 - \alpha + \alpha \log(\alpha/6)}}(\log_2 x)^6. \tag{6}$$

Case II. Write $q - 1 = a(bp)^2$ where a is squarefree, so that $n = uabp$ for some integer u. We first consider the sub-case where $P(ua) \le \exp((\log x)^\beta)$, where $0 < \beta < 1$ is to be chosen later. For given values of u, a, b, the number of choices for $p \le x/uab$ satisfying the primality condition is

$$\ll \frac{x}{uab(\log x)^2}(\log_2 x)^5 \prod_{\ell \le x/uab} \left(1 - \frac{\chi_{-a}(\ell)}{\ell}\right). \tag{7}$$

(Here we have applied Lemma 5 with $A = ab^2$, $B = 0$, and $C = 1$, so that $D = -4ab^2$ and d is the discriminant of $\mathbf{Q}(\sqrt{-a})$.) If $-a \notin \mathscr{D}$, then the product appearing in (7) is $O((\log_2 x)^2)$. If $-a \in \mathscr{D}$, our assumption (ii) implies that $a \le \log x$. In that case, Lemma 6 shows that the product in (7) is $O_\epsilon((\log x)^{\epsilon/2})$, for any $\epsilon > 0$. So whether or not $-a \in \mathscr{D}$, the number of choices for p is

$$\ll_\epsilon \frac{x}{uab(\log x)^{2-\epsilon}}. \tag{8}$$

(We have absorbed the power of $\log_2 x$ into the exponent of $\log x$.) We now sum over u, a, b by the method used in Case I, keeping in mind that $P(ua) \le \exp((\log x)^\beta)$. For $0 < z < 1$,

$$\sum \frac{1}{uab} \le z^{-\alpha \log_2 x} \sum \frac{z^{\Omega(u)} z^{\Omega(a)} z^{\Omega(b)}}{uab}$$
$$\le z^{-\alpha \log_2 x} \prod_{\ell_1 \le x} (1 - z/\ell_1)^{-1} \left(\prod_{\ell_2 \le \exp((\log x)^\beta)} (1 - z/\ell_2)^{-1} \right)^2 \ll (\log x)^{-\alpha \log z + (1+2\beta)z}.$$

The optimal choice is $z = \alpha/(1+2\beta)$, which gives $\sum \frac{1}{uab} \ll (\log x)^{\alpha - \alpha \log(\alpha/(1+2\beta))}$. So by (8), the total contribution in this sub-case is

$$\ll_\epsilon \frac{x}{(\log x)^{2-\alpha+\alpha\log(\alpha/(1+2\beta))-\epsilon}}. \tag{9}$$

We divide the remaining sub-case when $P(ua) > \exp((\log x)^\beta)$ into further sub-cases as follows. For each positive integer i, let $\beta_i = \beta + i/\log_2 x$, and let $\mathcal{I}_i$ be the interval

$$\mathcal{I}_i = (\exp((\log x)^{\beta_{i-1}}), \exp((\log x)^{\beta_i})].$$

For each i we consider the sub-case where $p_2 := P(ua) \in \mathcal{I}_i$. Clearly, the number of possible sub-cases is at most $1 + \log_2 x$.

We know that $p_2 \mid ua \mid n$, while (iii) implies that $p_2^2 \nmid n$. Hence, $p_2 \parallel n$. Consequently, $p_2 \nmid bp$ and so $p_2^2 \nmid q-1$. Since $p_2 > \log x$, (iv) gives that $p_2^2 \nmid m$. In conjunction with the relations $p_2^2 \parallel n^2 = \varphi(m)$ and $p_2^2 \nmid q-1$, this shows that there is a prime $q_2 \ne q$ dividing m with $q_2 \equiv 1 \pmod{p_2}$. If $p_2 \mid u$, then either $p_2^2 \parallel q_2 - 1$ or $p_2 \parallel q_2 - 1$ and there is some other prime $q_3 \mid m$ with $p_2 \parallel q_3 - 1$. If $p_2 \mid a$, then $p_2 \parallel q_2 - 1$. We shall sum up these possibilities as $p_2^k \parallel q-1$, $k = 0$ or 1, and $p_2^j \parallel q_2 - 1$, $j = 1$ or 2 and $k + j \le 2$, ignoring the possible existence of a prime q_3.

Set $q_1 = q$, $p_1 = p$, $b_1 = b$. We can select natural numbers a_1, a_2, a_3, b_2 with $a_1a_2a_3p_1p_2$ squarefree and

$$q_1 - 1 = a_1a_3b_1^2p_1^2p_2^k, \quad q_2 - 1 = a_2a_3b_2^2p_2^j.$$

Then n has a decomposition of the form

$$n = u_1a_1a_2a_3b_1b_2p_1p_2.$$

Here, in our old notation, $a = a_1a_3p_2^k$ and $u = u_1a_2b_2p_2^{1-k}$. Thus, $P(u_1a_1a_2a_3b_2) < p_2$. Fixing $u_1, a_1, a_2, a_3, b_1, b_2, p_2$ and using the primality of q_1, we deduce from Lemma 5 (applied with $A = a_1a_3b_1^2p_2^k$, $B = 0$, and $C = 1$) that the number of possible $p_1 \le x/u_1a_1a_2a_3b_1b_2p_2$ is

$$\ll \frac{x}{u_1a_1a_2a_3b_1b_2p_2(\log x)^2}(\log_2 x)^5 \prod_{\ell \le x/u_1a_1a_2a_3b_1b_2p_2} \left(1 - \frac{\chi_{-a_1a_3p_2}(\ell)}{\ell}\right)$$
$$\ll_\epsilon \frac{x}{u_1a_1a_2a_3b_1b_2p_2(\log x)^{2-\epsilon}}. \tag{10}$$

(To estimate the product we use an analysis similar to that in (7).) We now fix u_1, a_1, a_2, a_3, b_1, b_2 and sum on $p_2 \in \mathcal{I}_i$. First assume that $j = 1$. Since p_2 and $a_2a_3b_2^2p_2 + 1$ are both prime, the sieve in the form of Lemma 4 shows that for each $t \ge 3$, the number of possible $p_2 \le t$ is $O(t(\log_2 x)^2/(\log t)^2)$. Now partial summation implies that if we sum (10) over $p_2 \in \mathcal{I}_i$, the result is

$$\ll_\epsilon \frac{x}{u_1a_1a_2a_3b_1b_2(\log x)^{2+\beta_{i-1}-2\epsilon}}. \tag{11}$$

(Indeed, this upper bound holds for the larger sum over all $p_2 \geq \exp((\log x)^{\beta_{i-1}})$.) Now assume $j = 2$. We proceed in the same way, though now we use Lemma 5 and a similar analysis as in (7), getting an estimate of

$$\ll_\epsilon \frac{x}{u_1 a_1 a_2 a_3 b_1 b_2 (\log x)^{2+\beta_{i-1}-3\epsilon}}. \tag{12}$$

Finally, we replace the estimate (11) with the larger bound (12) and sum over u_1, a_1, a_2, a_3, b_1, b_2, keeping in mind that $P(u_1 a_1 a_2 a_3 b_2) \leq \exp((\log x)^{\beta_i})$. For $0 < z < 1$,

$$\begin{aligned} \sum \frac{1}{u_1 a_1 a_2 a_3 b_1 b_2} &\leq z^{-\alpha \log_2 x} \sum \frac{z^{\Omega(u_1)} z^{\Omega(a_1)} z^{\Omega(a_2)} z^{\Omega(a_3)} z^{\Omega(b_1)} z^{\Omega(b_2)}}{u_1 a_1 a_2 a_3 b_1 b_2} \\ &\leq z^{-\alpha \log_2 x} \left(\prod_{\ell_1 \leq \exp((\log x)^{\beta_i})} (1 - z/\ell_1)^{-1} \right)^5 \prod_{\ell_2 \leq x} (1 - z/\ell_2)^{-1} \\ &\ll (\log x)^{-\alpha \log z + (1+5\beta_i) z}. \end{aligned}$$

We select $z = \alpha/(1+5\beta_i)$ and find that $\sum \frac{1}{u_1 a_1 a_2 a_3 b_1 b_2} \ll (\log x)^{\alpha - \alpha \log(\alpha/(1+5\beta_i))}$. Referring back to (12), we deduce that the contribution of the ith sub-case is

$$\ll_\epsilon \frac{x}{(\log x)^{2+\beta_{i-1}-\alpha+\alpha\log(\alpha/(1+5\beta_i))-3\epsilon}}. \tag{13}$$

To continue our analysis, we make the additional assumption that our parameters α and β satisfy

$$0 < \beta \leq \alpha - \frac{1}{5} \leq 1. \tag{14}$$

As $\beta_i - \beta_{i-1} = 1/\log_2 x$, it is straightforward to check that the upper bound in (13) remains valid with the occurrence of β_i replaced by β_{i-1}. Having made this replacement, we now view the exponent of $\log x$ in (13) as a function of β_{i-1}, thinking of α and ϵ as fixed. The minimum value of this function on the closed interval $[\beta, 1]$ occurs when $\beta_{i-1} = \alpha - \frac{1}{5}$, resulting in a contribution of

$$\ll_\epsilon x/(\log x)^{\frac{9}{5}+\alpha\log(\frac{1}{5})-3\epsilon}.$$

Since there are $O(\log_2 x)$ sub-cases, the contribution from all values of i is

$$\ll_\epsilon \frac{x}{(\log x)^{\frac{9}{5}+\alpha\log(\frac{1}{5})-4\epsilon}}. \tag{15}$$

Optimization. We now choose α, β to minimize the size of the total exceptional set obtained by adding the estimates (4), (6), (9), (15). (The other exceptional sets appearing in the argument are of total size $O(x/(\log x)^{1/2})$, which is tiny on the scale we are interested in, so we ignore these.) The optimal choice of α is obtained by setting the exponent $Q(\alpha)$ from (4) equal to the exponent $\frac{9}{5} + \alpha \log(\frac{1}{5})$ from (15), which yields $\alpha = 1.114478\ldots$. This leads to the exponent $Q(\alpha) = 0.006316\ldots$. Choosing $\beta = 0.7$, say, the remaining error terms (6) and (9) are smaller than $x/(\log x)^{Q(\alpha)}$. (Note that (14) is satisfied for these choices of α and β, and that the various choices of the parameter z in the proof all satisfy $0 < z < 1$ as required.) Thus, our count is smaller than $x/(\log x)^{0.0063}$ for all sufficiently large values of x, which completes the proof of Theorem 1.

Remark. Our argument can be modified to show that for each fixed integer $w \leq (\log x)^{1/4}$, the number of integers wn^2 in $[1, x^2]$ which are φ values is uniformly $O(xw^{-1/2}/(\log x)^{0.0063})$. For example, in Case I, let $d = \gcd(\varphi(q_1 q_2), w)$, and write $d = d_1 d_2$, where each $d_i \mid \varphi(q_i)$. Then

$$\frac{q_1 - 1}{d_1} \frac{q_2 - 1}{d_2} \mid n^2.$$

Each factor $\frac{q-1}{d}$ on the left can be written as apb^2. Proceeding as before, we deduce that n has a factorization $n = ua_1a_2a_3b_1b_2p$, where now the primality conditions are that $1 + d_1a_1a_3pb_1^2$ and $1+d_2a_2a_3pb_2^2$ are prime, and where $d_1d_2 \mid w$. One then needs to sum also on the number of possibilities for d_1, d_2, but this is $(\log x)^{o(1)}$ given the small size of w. Other changes are similarly routine.

Using this we claim that the number of squarefull integers in $[1, x^2]$ which belong to $\mathscr{V}$ is $O(x/(\log x)^{0.0063})$. Indeed, all but $O(x/(\log x)^{1/24})$ squarefull numbers in $[1, x^2]$ are of the form m^3n^2 with $m \le (\log x)^{1/12}$. For each such m, the above argument shows that the number of n with $m^3n^2 \in \mathscr{V} \cap [1, x^2]$ is $O(x \cdot m^{-3/2}/(\log x)^{0.0063})$, uniformly in m. Now we sum on m to get the claim.

4. A lower bound and a heuristic

4.1. **Proof of Theorem 2.**

Proof. Let $y = (\log x)^2$. For each prime $p \in [y, 2y]$, let $\mathcal{Q}_p$ denote the set of primes $q \le x$ with $q \equiv 1 \pmod{p^2}$ and let $\mathcal{Q}'_p$ denote the set of those $q \in \mathcal{Q}_p$ such that $(q-1)/p^2$ has no prime factors in $[y, 2y]$. From the Brun–Titchmarsh inequality, it follows that

$$\#(\mathcal{Q}_p \setminus \mathcal{Q}'_p) \ll \sum_{r\in[y,2y]} \frac{x}{p^2 r \log x} \ll \frac{x}{p^2 \log y \log x}.$$

Thus, from the Siegel–Walfisz theorem, we have uniformly for $p \in [y, 2y]$ that

$$\#\mathcal{Q}'_p \sim \#\mathcal{Q}_p \sim \frac{x}{p^2 \log x}, \quad x \to \infty$$

so that

$$\#\mathcal{Q}'_p \asymp \frac{x}{y^2 \log x}. \tag{16}$$

For an integer $a < x/y^2$ free of prime factors from $[y, 2y]$, let $\mathcal{N}(a)$ denote the set of primes $q \le x$ with $q \equiv 1 \pmod a$ and $(q-1)/a = p^2$ for some prime $p \in [y, 2y]$. Thus, $q \in \mathcal{Q}'_p$. If we have two different primes q_1, q_2 in $\mathcal{N}(a)$ with $q_i - 1 = ap_i^2$ for $i = 1, 2$, then

$$\varphi(q_1q_2) = (q_1-1)(q_2-1) = (ap_1p_2)^2, \quad ap_1p_2 < a\max\{p_1^2, p_2^2\} < x.$$

Since a has no prime factors in $[y, 2y]$, an integer $n = ap_1p_2$ constructed in this way determines the value of a and so determines the pair of distinct primes $q_1, q_2 \in \mathcal{N}(a)$. Our strategy then is to count the number of such pairs of distinct primes for all possible values of a.

Let $N(a) = \#\mathcal{N}(a)$ if $\mathcal{N}(a)$ has been defined, with $N(a) = 0$ otherwise. From (16),

$$\sum_{a<x/y^2} N(a) = \sum_{p\in[y,2y]} \#\mathcal{Q}'_p \asymp \frac{y}{\log y} \cdot \frac{x}{y^2 \log x} = \frac{x}{y \log y \log x}.$$

It follows from Cauchy's inequality that

$$\sum_{a<x/y^2} N(a)^2 \ge \frac{y^2}{x} \left(\sum_{a<x/y^2} N(a) \right)^2 \gg \frac{y^2}{x} \cdot \frac{x^2}{y^2(\log y \log x)^2} \gg \frac{x}{(\log x \log\log x)^2}.$$

The last two displays and the choice of y as $(\log x)^2$ imply that

$$\sum_{a<x/y^2} \big(N(a)^2 - N(a)\big) \gg \frac{x}{(\log x \log\log x)^2}.$$

This sum represents the number of pairs of distinct primes in any of the sets $\mathcal{N}(a)$, and as we have seen, it gives a lower bound for $V_\square(x)$. This completes the proof of the theorem. $\square$

Remark. The above argument can be improved by a factor of $\log\log x$ by including the contributions from dyadic intervals $[2^{j-1}y, 2^j y)$ for $2^j \le y^\epsilon$ for a fixed small value of $\epsilon > 0$. In the jth interval we have $\gg x/(\log x \log(2^j y))^2$ solutions, and so summing on j gets us $\gg x/((\log x)^2 \log\log x)$ numbers. To make this work one needs that the parameter a has no prime factors from the interval $[y, y^{1+\epsilon})$, which is easy to arrange if ϵ is small enough.

It is interesting to compare the proof of Theorem 2 with the proof in [3]. The idea there is similar, but instead of taking $y = (\log x)^2$, they take $y = x^{1/6}$ and appeal to the prime number theorem in [1] instead of the Siegel–Walfisz theorem. In addition, instead of using the Cauchy inequality, they use Jensen's inequality to much the same effect.

4.2. **A heuristic.** The above proof gives a lower estimate for the number of squares of the form $\varphi(q_1q_2)$, where q_1, q_2 are distinct primes. One might ask what the "true" answer is, and more generally for the distribution of squares of the form $\varphi(m)$ where m is the product of k distinct odd primes, say $m = q_1 \cdots q_k$. Such a square n^2 has a natural factorization as $(q_1 - 1)\cdots(q_k - 1)$. If $q_i - 1$ is written as $a_i b_i^2$ with a_i squarefree, it follows that $a_1 \cdots a_k$ is a square. For the case $k = 2$, as we have seen in the proof above, this forces $a_1 = a_2$. In the case $k = 3$ we have three numbers A_1, A_2, A_3 with $a_i = A_1A_2A_3/A_i$, for $i = 1, 2, 3$. The situation gets more complicated for 4 or more primes.

Suppose that a number $n \le x$ is divisible by 4, $n/4$ is squarefree, and $\Omega(n/4) \ge \alpha \log_2 x$, where we fix a real number $\alpha > 1$. The number of ordered factorizations of n as $A_1A_2A_3b_1b_2b_3$ with at least 2 of A_1, A_2, A_3 even is at least $6^{\Omega(n/4)} \ge (\log x)^{\alpha \log 6}$. The "chance" that each of $1 + b_i^2 A_1A_2A_3/A_i$ is prime for $i = 1, 2, 3$ "should be" about $(\log x)^{-3}$. So, if $\alpha \log 6 > 3$, i.e., $\alpha > 3/\log 6$, there should be at least one such factorization. Thus, most numbers $n \le x$ with $n/4$ squarefree and $\Omega(n/4) > \alpha \log_2 x$ with α a fixed real larger than $3/\log 6$ should have $n^2 \in \mathscr{V}$. It should then follow that $V_\square(x) \gg x/(\log x)^{Q(\alpha)}$. Since $Q(3/\log 6) = 0.18864255\ldots$, we thus should have $V_\square(x) \ge x/(\log x)^{0.189}$ for all sufficiently large values of x. Note that repeating this argument with products of 2 or 4 primes gives a worse result.

5. Square values of the sum-of-divisors function

Both Theorems 1 and 2 remain true with σ replacing φ. When porting over the proofs, the main idea is to replace every occurrence of $\varphi(q) = q - 1$ with $\sigma(q) = q + 1$. This works without much fuss for Theorem 2, and we leave the details to the reader. For Theorem 1, we meet additional difficulties owing to the more complicated behavior of σ on prime powers. In this section, we sketch a way around these roadblocks.

5.1. **Outline.** Assume that $n \le x$ is such that $n^2 = \sigma(m)$. We can assume all of our previous conditions (i)–(v) on n and m, with the exception of (iv), which we replace with

(iv′) m has no prime power divisor $q^e > \exp((\log x)^{1/2})$ with $e \ge 2$.

We leave the justification of (iv′) to the end of this section, where it is shown (Lemma 9) that this assumption introduces an exceptional set of size $O(x/(\log x)^{1/4})$. For the rest of the argument, we fix the values of α and β to the constants we found above. Thus, $\alpha = 1.114478\ldots$ and $\beta = 0.7$.

With $p = P(n)$, we have $p^2 \mid n^2 = \sigma(m)$. It cannot be the case that $p \mid \sigma(q^e)$ for a prime power $q^e \parallel m$ having $e \ge 2$, for then $2q^e > q^e + q^{e-1} + \cdots + 1 = \sigma(q^e) \ge p$, forcing $q^e > \frac{p}{2} > \frac{1}{2}x^{1/2\log_2 x}$ and contradicting (iv′). This leaves only two possibilities:

I′. there are two different primes $q_1, q_2 \parallel m$ with $q_i \equiv -1 \pmod p$ for $i = 1, 2$,

II′. there is a prime $q \parallel m$ with $q \equiv -1 \pmod{p^2}$.

Case I′. This is handled exactly as Case I above, replacing $q - 1$ with $q + 1$ throughout the argument. We find that the total count of n in Case I′ satisfies our earlier upper bound (6).

Case II′. We start by writing $q+1 = a(bp)^2$, so that $n = uabp$ for some integer u. Our first sub-case, when $P(ua) \le \exp((\log x)^\beta)$, is handled exactly as was the first sub-case of Case II. Note that in the analogue of the sieve bound (7), the character χ_a appears in place of χ_{-a}. (We do not have to worry that a is a square, as that would imply $q = a(bp)^2 - 1$ factors.) This sub-case makes a total contribution of size (9).

In the remaining sub-cases, $P(ua) > \exp((\log x)^\beta)$. We again partition these according to the interval $\mathcal{I}_i$ to which $p_2 := P(ua)$ belongs. Reasoning as in our treatment of Case II, we find that $p_2 \parallel n$; moreover, if we choose k so that $p_2^k \parallel q+1$, then $k = 0$ or 1 according to whether or not $p_2 \mid a$. Hence,

$$p_2 \mid \frac{n^2}{q+1} = \sigma(m/q).$$

Thus, there is a prime power $q_2^e \parallel m/q$ for which p_2 divides $\sigma(q_2^e)$. Note that $q_2^e > \frac{1}{2}p_2 > \frac{1}{2}\exp((\log x)^\beta)$, so that if $e \ge 2$, we obtain a contradiction with (iv′). So $e = 1$ and $p_2 \mid q_2+1$. We choose j so that $p_2^j \parallel q_2+1$. Then $j = 1$ or $j = 2$, and $k + j \le 2$. We now set $q_1 = q$, $p_1 = p$, $b_1 = b$, and continue to mimick our earlier arguments. We find that the contribution from all of the possible sub-cases of this sort satisfies (15).

Combining our estimates as before, we obtain the σ-analogue of Theorem 1 with the same exponent 0.0063.

5.2. **Proof that we can assume (iv′).**

Lemma 9. *The count of $n \le x$ with $n^2 = \sigma(m)$ for some m failing (iv′) is $O(x/(\log x)^{1/4})$.*

Proof. We continue to assume that x is large. For the duration of the argument, we let $y = \exp((\log x)^{1/2})$. Suppose that $q^e \parallel m$. Then $\sigma(q^e) \mid \sigma(m) = n^2$, and so $r_{q^e} := \prod_{\ell^f \parallel \sigma(q^e)} \ell^{\lceil f/2 \rceil}$ is a divisor of n. Thus,

$$\frac{1}{x}\#\{n \le x : n^2 = \sigma(m) \text{ for an } m \text{ where (iv}'\text{) fails}\} \le \sum\nolimits^{(1)} + \sum\nolimits^{(2)} + \sum\nolimits^{(3)}, \tag{17}$$

where

$$\sum\nolimits^{(1)} := \sum_{\substack{q^e > y \\ e \ge 3}} \frac{1}{r_{q^e}}, \qquad \sum\nolimits^{(2)} := \sum_{\substack{q > \sqrt{y} \\ r_{q^2} > q\log q}} \frac{1}{r_{q^2}}, \qquad \text{and} \qquad \sum\nolimits^{(3)} := \sum_{\substack{q > \sqrt{y} \\ r_{q^2} \le q\log q}} \frac{1}{r_{q^2}}.$$

Since $r_{q^e} \ge (\sigma(q^e))^{1/2} > q^{e/2}$, we have $\sum^{(1)} \le \sum_{q^e > y,\ e\ge 3} q^{-e/2} \le \sum_{\text{cubefull } c > y} c^{-1/2} \ll y^{-1/6}$, using in the final step that the count of cubefull numbers up to height t is $O(t^{1/3})$. By partial summation and the prime number theorem, $\sum^{(2)} \le \sum_{q > \sqrt{y}} (q\log q)^{-1} \ll (\log y)^{-1}$. It remains to estimate $\sum^{(3)}$.

Let us show that $\mathcal{Q} := \{q : r_{q^2} \le q\log q\}$ is a sparse set of primes. We begin with a simple observation: If q^2+q+1 has an exact prime divisor $\ell_0 > (\log q)^2$, then

$$r_{q^2} = \ell_0 \prod_{\substack{\ell^f \parallel q^2+q+1 \\ \ell \ne \ell_0}} \ell^{\lceil f/2 \rceil} \ge \ell_0 \sqrt{\frac{q^2+q+1}{\ell_0}} > q\sqrt{\ell_0} > q\log q,$$

and thus $q \notin \mathcal{Q}$. So if we suppose that $q \in \mathcal{Q} \cap (t/2, t]$ for a large real number t, then $q \in \mathcal{Q}_1 \cup \mathcal{Q}_2$, where

$$\mathcal{Q}_1 := \{q \in (t/2, t] : q^2+q+1 \text{ has no prime divisors in } ((\log t)^2, t^{1/10}]\},$$
$$\mathcal{Q}_2 := \{q \in (t/2, t] : \ell^2 \mid q^2+q+1 \text{ for some } \ell \in ((\log t)^2, t^{1/10}]\}.$$

Let $\varrho(r)$ be the number of roots modulo r of the polynomial X^2+X+1. For primes $\ell > 3$, we have $\varrho(\ell) = 2$ when $\ell \equiv 1 \pmod 3$ and $\varrho(\ell) = 0$ otherwise. By the upper bound sieve (for instance, in the form of [14, Theorem 4.2, p. 134]),

$$\#\mathcal{Q}_1 \ll \frac{t}{\log t} \prod_{(\log t)^2 < \ell \le t^{1/10}} \left(1 - \frac{\varrho(\ell)}{\ell}\right) \ll \frac{t}{(\log t)^2} \log_2 t \ll \frac{t}{(\log t)^{3/2}}.$$

(To estimate the product, we used a version of Mertens's theorem for primes congruent to 1 modulo 3.) We estimate $\#\mathcal{Q}_2$ crudely. Observing that $\varrho(\ell^2) \le 2$ for all primes $\ell > 3$ (for instance, by Hensel's lemma), we obtain immediately that

$$\#\mathcal{Q}_2 \le \sum_{(\log t)^2 < \ell \le t^{1/10}} \left(\frac{2t}{\ell^2} + 2\right) \ll \frac{t}{(\log t)^2}.$$

Hence, $\#\mathcal{Q} \cap (t/2, t] \le \#\mathcal{Q}_1 + \#\mathcal{Q}_2 \ll t/(\log t)^{3/2}$. Summing dyadically, we find that $\#\mathcal{Q} \cap [1, t] \ll t/(\log t)^{3/2}$ for all $t \ge 3$.

We now return to the problem of estimating $\sum^{(3)}$. Using the lower bound $r_{q^2} > q$, we find that $\sum^{(3)} \le \sum_{q > \sqrt{y},\ q \in \mathcal{Q}} q^{-1} \ll (\log y)^{-1/2}$, by partial summation. Lemma 9 now follows from (17) and our estimates for $\sum^{(1)}, \sum^{(2)}$, and $\sum^{(3)}$. $\square$

Acknowledgments

We thank the referee for some helpful comments. The second author was supported in part by NSF grant DMS-1001180.

References

[1] W. R. Alford, A. Granville, and C. Pomerance, *There are infinitely many Carmichael numbers*, Ann. of Math. (2) **139** (1994), 703–722.

[2] W. D. Banks, J. B. Friedlander, C. Pomerance, and I. E. Shparlinski, *Multiplicative structure of values of the Euler function*, in High primes and misdemeanours: Lectures in honour of the sixtieth birthday of Hugh Cowie Williams, A. J. van der Poorten, ed., Fields Inst. Comm. **41** (2004), 29–47.

[3] W. D. Banks and F. Luca, *Power totients with almost primes*, Integers **11** (2011), 307–313.

[4] N. G. de Bruijn, *On the number of positive integers $\le x$ and free of prime factors $> y$*, Indag. Math. **13** (1951), 50–60.

[5] V. Chandee, C. David, D. Koukoulopoulos, and E. Smith, *Group structures of elliptic curves over finite fields*, Int. Math. Res. Notices, to appear.

[6] H. Davenport, *Multiplicative number theory*, third ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000.

[7] T. Dence and C. Pomerance, *Euler's function in residue classes*, Ramanujan J. **2** (1988), 7–20.

[8] K. Ford, *The distribution of totients*, Ramanujan J. **2** (1998), 67–151, updated version available as `arXiv:1104.3264 [math.NT]`.

[9] K. Ford, S. V. Konyagin, and C. Pomerance, *Residue classes free of values of Euler's function*, in Number theory in progress, K. Gyory, H. Iwaniec, and J. Urbanowicz, eds., vol. 2, de Gruyter, Berlin and New York, 1999, pp. 805–812.

[10] K. Ford, F. Luca, and C. Pomerance, *Common values of the arithmetic functions ϕ and σ*, Bull. London Math. Soc. **42** (2010), 478–488.

[11] K. Ford and P. Pollack, *On common values of $\varphi(n)$ and $\sigma(m)$, I*, Acta Math. Hungarica **133** (2011), 251–271.

[12] ———, *On common values of $\varphi(n)$ and $\sigma(m)$, II*, Algebra and Number Theory **6** (2013), 1669–1696.

[13] T. Freiberg, *Products of shifted primes simultaneously taking perfect power values*, J. Aust. Math. Soc. (special issue dedicated to Alf van der Poorten) **92** (2012), 145–154.

[14] H. Halberstam and H.-E. Richert, *Sieve methods*, London Mathematical Society Monographs, vol. 4, Academic Press, London-New York, 1974.

[15] R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics, vol. 90, Cambridge University Press, Cambridge, 1988.

[16] W. Narkiewicz, *On distribution of values of multiplicative functions in residue classes*, Acta Arith. **12** (1966/1967), 269–279.

University of Georgia, Department of Mathematics, Athens, GA 30602, USA
E-mail address: pollack@uga.edu

Dartmouth College, Department of Mathematics, Hanover, NH 03755, USA
E-mail address: carlp@math.dartmouth.edu