

On the sum of the square of a prime and a square-free number

Adrian W. Dudek and David J. Platt

ABSTRACT

We prove that every integer $n \geq 10$ such that $n \not\equiv 1 \pmod{4}$ can be written as the sum of the square of a prime and a square-free number. This makes explicit a theorem of Erdős that every sufficiently large integer of this type may be written in such a way. Our proof requires us to construct new explicit results for primes in arithmetic progressions. As such, we use the second author's numerical computation regarding the generalised Riemann hypothesis to extend the explicit bounds of Ramaré–Rumely.

1. Introduction

We say that a positive integer is square-free if it is not divisible by the square of any prime number. It was proven by Erdős [7] in 1935 that every sufficiently large integer $n \not\equiv 1 \pmod{4}$ may be written as the sum of the square of a prime and a square-free number. The congruence condition here is sensible. If $n \equiv 1 \pmod{4}$, then $4|(n-p^2)$ for any odd prime p . This only leaves the case $p = 2$, but $n - 4$ fails to be square-free infinitely often[†].

It is the objective of this paper to make explicit the proof provided by Erdős, to the end of proving the following theorem.

THEOREM 1. *Let $n \geq 10$ be an integer such that $n \not\equiv 1 \pmod{4}$. Then there exist a prime p and a square-free number k such that $n = p^2 + k$.*

In a recent paper [5], the first author proved that every integer greater than two can be written as the sum of a prime and a square-free number. One can think of such a result as a weak-but-explicit form of Goldbach's conjecture. Theorem 1 is significantly stronger than this, for the sequence of squares of primes is far more sparse than the sequence of the primes. To prove Theorem 1, we combine modern explicit results on primes in arithmetic progressions and computation.

The proof may be outlined as follows. For any integer n satisfying the conditions of the above theorem, we want to show that there exists a prime $p < \sqrt{n}$ such that $n - p^2$ is square-free. That is, we require some prime p such that

$$n - p^2 \not\equiv 0 \pmod{q^2}$$

for all odd primes $q < \sqrt{n}$. The idea is to consider, for some large n and each odd prime $q < \sqrt{n}$, those *mischievous* primes p that satisfy the congruence

$$n \equiv p^2 \pmod{q^2}.$$

Then, for each q , we explicitly bound from above (with logarithmic weights) the number of primes p which satisfy the above congruence. Summing over all moduli q gives us an upper

Received 12 April 2015; revised 14 October 2015.

2010 Mathematics Subject Classification 11N13, 11P32 (primary).

[†]For example, one can consider the congruence class $13 \pmod{36}$.

bound for the weighted count of the so-called mischievous primes

$$\sum_{q < \sqrt{n}} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p.$$

It is then straightforward to show that for large enough n , the above sum is less than the weighted count of *all* primes less than \sqrt{n} , and therefore there must exist a prime $p < \sqrt{n}$ such that $n - p^2$ is not divisible by the square of any prime.

This method works well, and allows us to prove Theorem 1 for all integers $n \geq 2.5 \cdot 10^{14}$ that satisfy the congruence condition. We eliminate the remaining cases by direct computation to complete the proof.

2. Theorem 1 for large integers

2.1. Case 1

We start by considering integers in the range $n \geq 2.5 \cdot 10^{14}$ such that $n \not\equiv 1 \pmod{4}$. As usual, we define

$$\theta(x; k, l) = \sum_{\substack{p \leq x \\ p \equiv l \pmod{k}}} \log p,$$

where p denotes a prime number.

The paper of Ramaré and Rumely [10] provides us with bounds of the form

$$\left| \theta(x; k, l) - \frac{x}{\varphi(k)} \right| < \epsilon(k, x_0) \frac{x}{\varphi(k)}$$

and

$$\left| \theta(x; k, l) - \frac{x}{\varphi(k)} \right| < \omega(k, x_1) \sqrt{x}$$

for various ranges of $x \geq x_0$ and $x \leq x_1$, respectively. These computations were in turn based on Rumely’s numerical verification of the generalised Riemann hypothesis (GRH) [12] for various moduli and to certain heights. Since then, the second author has verified the GRH for a wider range of moduli and to greater heights [9]. For our purposes, we rely only on the following result.

LEMMA 2. *Let q be a prime satisfying $17 \leq q \leq 97$. All non-trivial zeros ρ of Dirichlet L -functions derived from characters of modulus q^2 with $\Im \rho \leq 1000$ have $\Re \rho = 1/2$.*

Proof. See [9, Theorem 10.1]. □

We can therefore extend the results of Ramaré–Rumely with the following lemma.

LEMMA 3. *For $x > 10^{10}$, we have*

$$\left| \theta(x; q^2, l) - \frac{x}{\varphi(q^2)} \right| < \epsilon(q^2, 10^{10}) \frac{x}{\varphi(q^2)}$$

for the values of q and $\epsilon(q^2, 10^{10})$ in Table 1.

Proof. We refer to [10]. The values for $q \in \{3, 5, 7, 11, 13\}$ are from Table 1 of that paper. For the other entries, we use Theorem 5.1.1 with $H_\chi = 1000$ and $C_1(\chi, H_\chi) = 9.14$ (see display 4.2). We set $m = 10$ for $q \leq 23$, $m = 12$ for $q \geq 47$ and $m = 11$ otherwise. We use $\delta = 2e/H_\chi$ and, for \tilde{A}_χ , we use the upper bound of Lemma 4.2.1. Finally, for \tilde{E}_χ , we rely on Lemma 4.1.2 and we note that $2 \cdot 9.645908801 \cdot \log^2(1000/9.14) \geq \log 10^{10}$, as required. \square

LEMMA 4. We have

$$\begin{aligned} \omega(3^2, 10^{10}) &= 1.109042, \\ \omega(5^2, 10^{10}) &= 0.821891, \\ \omega(7^2, 10^{10}) &= 0.744132, \\ \omega(11^2, 10^{10}) &= 0.711433 \end{aligned}$$

and

$$\omega(13^2, 10^{10}) = 0.718525.$$

If q is a prime such that $17 \leq q \leq 97$, we have

$$\omega(q^2, 10^{10}) = \frac{\log 7 - 7/\varphi(q^2)}{\sqrt{7}}.$$

Proof. The results for $\{3^2, 5^2, 7^2, 11^2, 13^2\}$ are from [10, Table 2] with a slight correction to the entry for 5^2 . A short computation shows that the maximum occurs for all of the other q when $x = 7$ and $a = 7$. \square

LEMMA 5. Let $T = \sqrt{2.5 \cdot 10^{14}}$. Then, for $x \geq T$ and $q \leq 97$ an odd prime, we have

$$\left| \theta(x; q^2, l) - \frac{x}{\varphi(q^2)} \right| < \epsilon(q^2, T) \frac{x}{\varphi(q^2)},$$

where the values of $\epsilon(q^2, T)$ are given in Table 2.

Proof. Using $\omega(q^2, 10^{10})$, we have

$$\left| \theta(T; q^2, l) - \frac{T}{\varphi(q^2)} \right| < \omega(q^2, 10^{10}) \sqrt{T}$$

and so, for $x \in [T, 10^{10}]$, we have

$$\left| \theta(x; q^2, l) - \frac{x}{\varphi(q^2)} \right| < \frac{\omega(q^2, 10^{10}) \varphi(q^2)}{\sqrt{T}} \frac{x}{\varphi(q^2)}$$

TABLE 1. Values for $\epsilon(q^2, 10^{10})$.

q	$\epsilon(q^2, 10^{10})$						
3	0.003228	19	0.17641	43	0.95757	71	2.82639
5	0.012214	23	0.25779	47	1.15923	73	3.00162
7	0.017015	29	0.41474	53	1.50179	79	3.56158
11	0.031939	31	0.47695	59	1.89334	83	3.96363
13	0.042497	37	0.69397	61	2.03488	89	4.61023
17	0.14271	41	0.86446	67	2.49293	97	5.55434

and so we can take

$$\epsilon(q^2, T) = \max\left(\epsilon(q^2, 10^{10}), \frac{\omega(q^2, 10^{10})\varphi(q^2)}{\sqrt{T}}\right). \quad \square$$

Let $n \geq 2.5 \cdot 10^{14}$ be such that $n \not\equiv 1 \pmod 4$ and consider the case where q is an odd prime such that $q \leq 97$. We want to bound from above the number of primes $p < \sqrt{n}$ satisfying

$$n \equiv p^2 \pmod{q^2}. \tag{1}$$

Clearly, p can belong to at most two arithmetic progressions modulo q^2 . Therefore, by Lemma 5, we can estimate the weighted count of such primes as follows:

$$\sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p \leq \theta(\sqrt{n}; q^2, l) + \theta(\sqrt{n}; q^2, l') < \frac{2(1 + \epsilon(q^2, T))}{q(q-1)} \sqrt{n},$$

where l and l' are the possible congruence classes for p and $\epsilon(q^2, T)$ is given in Table 2. Summing this over all 24 values of q gives us the contribution

$$\sum_{q \in \{3, \dots, 97\}} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < 0.568\sqrt{n}. \tag{2}$$

2.2. Case 2

We now consider the case where $97 < q \leq n^c$ and $c \in (0, 1/4)$ is to be chosen later to achieve an optimal result. Montgomery and Vaughan’s [8] explicit version of the Brun–Titchmarsh theorem gives us that

$$\pi(x; k, l) \leq \frac{2x}{\varphi(k) \log(x/k)}$$

for all $x > q$. Trivially, one has that

$$\theta(\sqrt{n}; q^2, l) \leq \frac{\sqrt{n}}{q(q-1)} \frac{\log n}{\log(\sqrt{n}/q^2)}.$$

As $q < n^c$, it follows that

$$\sum_{97 < q \leq n^c} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < \frac{\sqrt{n}}{1/4 - c} \sum_{97 < q \leq n^c} \frac{1}{q(q-1)}. \tag{3}$$

TABLE 2. Values for $\epsilon(q^2, T)$ for Lemma 5.

q	$\epsilon(q^2, T)$						
3	0.00323	19	0.17641	43	0.95757	71	2.82639
5	0.01222	23	0.25779	47	1.15923	73	3.00162
7	0.01702	29	0.41474	53	1.50179	79	3.56158
11	0.03194	31	0.47695	59	1.89334	83	3.96363
13	0.04250	37	0.69397	61	2.03488	89	4.61023
17	0.14271	41	0.86446	67	2.49293	97	5.55434

We can bound the sum as follows:

$$\begin{aligned} \sum_{97 < q \leq n^c} \frac{1}{q(q-1)} &< \sum_{97 < q < 1000001} \frac{1}{q(q-1)} + \sum_{n \geq 1000001} \frac{1}{n(n-1)} \\ &= \sum_{97 < q < 1000001} \frac{1}{q(q-1)} + \frac{1}{1000000} < 0.00183. \end{aligned}$$

Substituting this into (3) gives us that

$$\sum_{97 < q \leq n^c} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < \frac{0.00183\sqrt{n}}{1/4 - c}. \tag{4}$$

2.3. Case 3

Let q be an odd prime such that $n^c < q < A\sqrt{n}$ and $A \in (0, 1)$ is to be chosen later for optimisation. Since there are at most two possible residue classes modulo q^2 for p , the number of primes p such that $n \equiv p^2 \pmod{q^2}$ is trivially less than

$$2\left(\frac{\sqrt{n}}{q^2} + 1\right).$$

Clearly, including our logarithmic weights, one has that

$$\sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < \left(\frac{\sqrt{n}}{q^2} + 1\right) \log n$$

and so

$$\sum_{n^c < q < A\sqrt{n}} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < \sqrt{n} \log n \sum_{m > n^c} \frac{1}{m^2} + \pi(A\sqrt{n}) \log(n),$$

where $\pi(x)$ denotes the number of primes not exceeding x . The sum can be estimated in a straightforward way by

$$\sum_{m > n^c} \frac{1}{m^2} < \frac{1}{n^{2c}} + \int_{n^c}^{\infty} \frac{1}{t^2} dt = \frac{1}{n^{2c}} + \frac{1}{n^c}$$

and Dusart [6, Theorem 6.9] gives us that

$$\pi(A\sqrt{n}) < \frac{A\sqrt{n}}{\log(A\sqrt{n})} \left(1 + \frac{1.2762}{\log(A\sqrt{n})}\right).$$

Therefore, putting this all together, we have

$$\sum_{n^c < q < A\sqrt{n}} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < \sqrt{n}(n^{-2c} + n^{-c}) \log n + \frac{A\sqrt{n} \log n}{\log(A\sqrt{n})} \left(1 + \frac{1.2762}{\log(A\sqrt{n})}\right). \tag{5}$$

2.4. Case 4

Finally, we consider the range $A\sqrt{n} \leq q < \sqrt{n}$. If $n - p^2$ is divisible by q^2 , then

$$n = p^2 + Bq^2 \tag{6}$$

for some positive integer $B < A^{-2}$. We will need some preliminary results here. First, it is known by the theory of quadratic forms (see Davenport [4, Chapter 6]) that the equation $ax^2 + by^2 = n$, where a, b and n are given positive integers, has at most $w2^{\omega(n)}$ proper solutions, that is, solutions with $\gcd(x, y) = 1$. Note that w denotes the number of automorphs of the above form and $\omega(n)$ denotes the number of different prime factors of n . The number of automorphs is directly related to the discriminant of the form; specifically, $w = 4$ for the case $B = 1$ and $w = 2$ for $B > 1$. Moreover, we are only interested in the case where x and y are both positive, and so it follows that equation (6) has at most $w2^{\omega(n)-2}$ proper solutions. Finally, noting that there will be at most one improper solution to (6), namely $p = q$, we can bound the overall number of solutions to (6) by $w2^{\omega(n)-2} + 1$.

Furthermore, Robin [11, Theorem 11] gives us the explicit bound

$$\omega(n) \leq 1.3841 \frac{\log n}{\log \log n}$$

for all $n \geq 3$. Thus, for fixed n and B , it is easy to bound explicitly from above the number of solutions to (6). It remains to sum this bound over all valid values of B . However, we should note that given an integer n , there are not too many good choices of B , and this will allow us to make a further saving.

This comes from the observation that every prime $p > 3$ satisfies $p^2 \equiv 1 \pmod{24}$. For, with $p > 3$ and $q > 3$, equation (6) becomes

$$B \equiv n - 1 \pmod{24},$$

and this confines B to the integers in a single residue class modulo 24.

Formally and explicitly, we argue as follows. Consider first the case where B is an integer in the range

$$\frac{n - 9}{A^2n} \leq B < \frac{1}{A^2}.$$

The left-most inequality above keeps $p \leq 3$. Here, there are clearly at most

$$\frac{9}{A^2n} + 1$$

integer values for B . We now consider the case where $p > 3$, and it follows that $B \equiv n - 1 \pmod{24}$. Clearly, then, there are at most

$$\frac{1}{24A^2} + 1$$

values for B in this range. Therefore, in total, there are at most

$$2 + \frac{1}{24A^2} + \frac{9}{A^2n}$$

values of B for which we need to sum the solution counts to equation (6). Also, we must also consider that $w = 4$ for $B = 1$. Therefore, we have that the number of solutions to equation (6) summed over B is bounded above by

$$2^{\omega(n)-1} \left(3 + \frac{1}{24A^2} + \frac{9}{A^2n} \right).$$

Therefore, the number of primes p (including weights) which satisfy (6) is at most

$$\sum_{A\sqrt{n} \leq q < \sqrt{n}} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < 2^{1.3841 \log n / \log \log n} \left(\frac{3}{2} + \frac{1}{48A^2} + \frac{9}{2A^2n} \right) \log n. \tag{7}$$

2.5. *Collecting terms*

Now, collecting together (2), (4), (5) and (7), we have that the weighted count over all the so-called mischievous primes can be bounded thus:

$$\begin{aligned} \sum_{q < \sqrt{n}} \sum_{\substack{p < \sqrt{n} \\ n \equiv p^2 \pmod{q^2}}} \log p < & \left(0.568 + \frac{0.00183}{1/4 - c} + (n^{-2c} + n^{-c}) \log n \right) \sqrt{n} \\ & + \frac{A\sqrt{n} \log n}{\log(A\sqrt{n})} \left(1 + \frac{1.2762}{\log(A\sqrt{n})} \right) \\ & + 2^{1.3841 \log n / \log \log n} \left(\frac{3}{2} + \frac{1}{48A^2} + \frac{9}{2A^2n} \right) \log n. \end{aligned}$$

As expected, however, the weighted count over all primes exceeds this for large enough n and good choices of c and A . Dusart [6] gives us that

$$\theta(x) \geq x - 0.2 \frac{x}{\log^2 x}$$

for all $x \geq 3594641$, and thus it follows that

$$\theta(\sqrt{n}) \geq \sqrt{n} - 0.8 \frac{\sqrt{n}}{\log^2 n}$$

for all $n \geq 10^{14}$. Therefore, if we denote by $R(n)$ the (weighted) count of primes p such that $n - p^2$ is square-free, it follows that

$$\begin{aligned} R(n) > & \left(1 - 0.568 - \frac{0.00183}{1/4 - c} - \frac{0.8}{\log^2 n} - (n^{-2c} + n^{-c}) \log n \right) \sqrt{n} \\ & - \frac{A\sqrt{n} \log n}{\log(A\sqrt{n})} \left(1 + \frac{1.2762}{\log(A\sqrt{n})} \right) \\ & - 2^{1.3841 \log n / \log \log n} \left(\frac{3}{2} + \frac{1}{48A^2} + \frac{9}{2A^2n} \right) \log n. \end{aligned}$$

It is now straightforward to check that choosing $c = 0.209$ and $A = 0.0685$ gives $R(n) > 0$ for all $n \geq 2.5 \times 10^{14}$.

3. *Numerical verification for ‘small’ n*

We now describe a computation undertaken to confirm that all $n \not\equiv 1 \pmod{4}$, $10 \leq n \leq 4\,000\,023\,301\,851\,135$, can be written as the sum of a prime squared and a square-free number[†]. We will first describe the algorithm used, and then say a few words about its implementation.

[†]This is a factor of 16 further than we actually needed to check, but we did not expect our analytic approach to fare as well as it did.

3.1. The algorithm

We aim to test $3 \cdot 10^{15}$ different n . We quickly conclude that we cannot afford to individually test each candidate $n - p^2$ to see if they are square-free. There is an analytic algorithm [3] that is conjectured to be able to test a number of size n in time $\mathcal{O}(\exp([\log n]^{2/3+o(1)}))$, but this is contingent on the GRH. We would be left needing to factor each $n - p^2$, which would be prohibitively expensive.

We proceed instead by choosing a largest prime P and a sieve width W . To check all the integers in $[N, N + W)$, we first sieve all the integers in $[N - P^2, N + W - 4)$ by crossing out any that are divisible by a prime square p^2 with $p < \sqrt{(N + W - 5)/2}$. Now, for each $n \in [N, N + W)$, $n \not\equiv 1 \pmod{4}$, we look up in our sieve to see if $n - 4$ is square-free[†]. If not, we try $n - 9$, then $n - 25$ and so on until $n - p^2$ is square-free. If it fails all these tests up to and including $n - P^2$, we output n for later checking.

3.2. The implementation

Numbers of this size fit comfortably in the 64 bit native word size of modern CPUs and we implemented the algorithm in C++. We use a character array for the sieve[‡] and choose a sieve width $W = 2^{31}$, as this allows us to run 16 such sieves in parallel in the memory available. We set the prime limit $P = 43$, as this was found to reduce the number of failures to a manageable level (see below). To generate the primes used to sieve the character array, we used Walisch's PrimeSieve [13].

We were able to run 16 threads on a node of the University of Bristol's BlueCrystal Phase III cluster [1] and in total we required 5400 core hours of CPU time to check all $n \in [2048, 4\,000\,023\,301\,851\,135]$. A total of 4915 n were rejected as none of $n - p^2$ with $p \leq 43$ were square-free. We checked these 4915 cases in seconds using PARI [2] and found that $p = 47$ eliminated 4290 of them, 53 does for a further 538, 59 for 14 more, 61 for 61 (!), 67 does not help (!), 71 kills off 11 more and the last one standing, $n = 1\,623\,364\,493\,706\,484$, falls away with $p = 73$. Finally, we use PARI again to check $n \in [10, 2047]$ with $n \not\equiv 1 \pmod{4}$ and we are done.

It is interesting to consider the efficiency of the main part of this algorithm. The CPUs on the compute nodes of Phase III are 2.6 GHz Intel Xeon processors and we checked $3 \cdot 10^{15}$ individual n in 5400 hours. This averages less than 17 clock ticks per n , which suggests that the implementation must have made good use of cache.

References

1. ACRC, University of Bristol, BlueCrystal Phase 3 User Guide, 2014, <https://www.acrc.bris.ac.uk/pdf/bc-user-guide.pdf>.
2. C. BATUT, K. BELABAS, D. BERNARDI, H. COHEN and M. OLIVIER, User's Guide to PARI-GP, 2000, <http://pari.math.u-bordeaux.fr/pub/pari/manuals/2.3.3/users.pdf>.
3. A. R. BOOKER, G. A. HIARY and J. P. KEATING, 'Detecting squarefree numbers', *Duke Math. J.* 164 (2015) no. 2, 235–275.
4. H. DAVENPORT, *The higher arithmetic: an introduction to the theory of numbers*, 8th edn (Cambridge University Press, 2008).
5. A. W. DUDEK, 'On the sum of a prime and a square-free number', *Ramanujan J.* (2015), to appear.
6. P. DUSART, 'Inégalités explicites pour $\psi(x)$, $\theta(x)$, $\pi(x)$ et les nombres premiers', *C. R. Math. Rep. Acad. Sci. Can.* 21 (1999) no. 1, 53–59.
7. P. ERDŐS, 'The representation of an integer as the sum of the square of a prime and of a square-free integer', *J. Lond. Math. Soc.* 4 (1935) 243–245.
8. H. L. MONTGOMERY and R. C. VAUGHAN, 'The large sieve', *Mathematika* 20 (1973) 119–134.

[†]Unless $n \equiv 0 \pmod{4}$.

[‡]We considered using each byte to represent eight or more n , but the cost of the necessary bit twiddling proved too heavy.

9. D. J. PLATT, 'Numerical computations concerning the GRH', *Math. Comp.* (2015), to appear.
10. O. RAMARÉ and R. RUMELY, 'Primes in arithmetic progressions', *Math. Comp.* 65 (1996) no. 213, 397–425.
11. G. ROBIN, 'Estimation de la fonction de Tchebychef θ sur le k -ième nombre premier et grandes valeurs de la fonction $\omega(n)$ nombre de diviseurs premiers de n ', *Acta Arith.* 42 (1983) no. 4, 367–389.
12. R. RUMELY, 'Numerical computations concerning the ERH', *Math. Comp.* 61 (1993) no. 203, 415–440.
13. K. WALISCH, PrimeSieve, <http://primesieve.org/>.

Adrian W. Dudek
Mathematical Sciences Institute
The Australian National University
Acton ACT 2601
Australia
adrian.dudek@anu.edu.au

David J. Platt
Heilbronn Institute for Mathematical
Research
University of Bristol
Bristol BS8 1SN
United Kingdom
dave.platt@bris.ac.uk