Ramanujan J (2018) 47:267–289
https://doi.org/10.1007/s11139-017-9972-8

Romanov type problems

Christian Elsholtz[1] · Florian Luca[2,3] · Stefan Planitzer[1]

Received: 27 January 2017 / Accepted: 13 November 2017 / Published online: 8 February 2018

Abstract Romanov proved that the proportion of positive integers which can be represented as a sum of a prime and a power of 2 is positive. We establish similar results for integers of the form $n = p + 2^{2^k} + m!$ and $n = p + 2^{2^k} + 2^q$ where $m, k \in \mathbb{N}$ and p, q are primes. In the opposite direction, Erdős constructed a full arithmetic progression of odd integers none of which is the sum of a prime and a power of two. While we also exhibit in both cases full arithmetic progressions which do not contain any integers of the two forms, respectively, we prove a much better result for the proportion of integers not of these forms: (1) The proportion of positive integers not of the form $p + 2^{2^k} + m!$ is larger than $\frac{3}{4}$. (2) The proportion of positive integers not of the form $p + 2^{2^k} + 2^q$ is at least $\frac{2}{3}$.

Keywords Romanov's theorem · Smooth numbers · Diophantine equation · Sumsets

The first and the third authors are supported by the Austrian Science Fund (FWF): W1230, Doctoral Program 'Discrete Mathematics'. The second author is supported by Grant No. 17-02804S of the Czech Grant Agency.

✉ Stefan Planitzer
planitzer@math.tugraz.at

Christian Elsholtz
elsholtz@math.tugraz.at

Florian Luca
Florian.Luca@wits.ac.za

1 Institute of Analysis and Number Theory, Graz University of Technology, Kopernikusgasse 24, 8010 Graz, Austria

2 School of Mathematics, Wits University, Private Bag X3, Wits, Johannesburg 2050, South Africa

3 Department of Mathematics, Faculty of Sciences, University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic

Mathematics Subject Classification Primary 11P32 · Secondary 11A07 , 11A41 , 11N36

1 Introduction

An old result of Romanov [16] states that a positive proportion of the positive integers can be written in the form $p+g^k$, where p is a prime and $g \geq 2$ is a positive integer. As there are about $x/\log x$ primes $p \leq x$ and $\lfloor \log x/\log g \rfloor$ powers $g^k \leq x$, this result implicitly gives some information about the number $r(n)$ of representations of $n = p+g^k$. There are not too many integers $n \leq x$ with a very large number of representations and on average $r(n)$ is bounded. The most prominent special case of Romanov's result is the one concerning sums of primes and powers of 2. Euler [9] observed in a letter to Goldbach that 959 can not be written as the sum of a prime and a power of two. Euler's letter was also mentioned by de Polignac [3] and provides a counter example to a conjecture of de Polignac himself, stating that any odd positive integer is the sum of a prime and a power of 2. In 1950, Erdős [5] and van der Corput [18] independently proved that also the lower density of odd integers not of the form $p + 2^k$ is positive. Here and in the following the lower density of a set $\mathcal{A} \subset \mathbb{N}$ is defined to be

$$\liminf_{x\to\infty} \frac{|\{a \in \mathcal{A} : a \leq x\}|}{x}.$$

Replacing lim inf with lim sup leads to what we call upper density and if lower and upper density coincide we speak of the density of the set $\mathcal{A}$.

Concerning Romanov's theorem one may ask how this result can be generalized. One way would be by replacing the sequence of powers of g with another sequence $(a_n)_{n\geq 1}$. Generalizing a result of Lee [13] who replaced the powers of g by the Fibonacci sequence, Ballot and Luca [1] proved an analogue of Romanov's theorem for the case when $(a_n)_{n\geq 1}$ is a linearly recurrent sequence with certain additional properties. For certain quadratic recurrences $(a_n)_{n\geq 1}$ this was done by Dubickas [4].

We would expect that for many sets $\mathcal{A} \subset \mathbb{N}$, with $|\mathcal{A} \cap [1, x]| \geq c \log x$ for some positive constant c, one can write a positive proportion of integers $n \leq x$ as $n = p+a$, p prime and $a \in \mathcal{A}$. In this paper we study sets $\mathcal{A}$ with $|\mathcal{A} \cap [1, x]| \sim c_{\mathcal{A}} \log x$ but of a quite different nature compared to previous ones. In particular, we study

$$\mathcal{A}_1 = \{2^{2^k} + m! : k, m \in \mathbb{N}_0\},$$
$$\mathcal{A}_2 = \{2^{2^k} + 2^q : k \in \mathbb{N}_0, q \text{ prime}\}.$$

Using the machinery of Romanov [16], we prove the following two theorems.

Theorem 1 *The lower density of integers of the form* $p + 2^{2^k} + m!$ *for* $k, m \in \mathbb{N}_0$ *and* p *prime is positive.*

Theorem 2 *The lower density of integers of the form* $p + 2^{2^k} + 2^q$ *for* $k \in \mathbb{N}_0$ *and* p, q *prime is positive.*

Concerning integers not of the form $p+2^{2^k}+m!$ we consider two different questions: The first one is finding a large set, in the sense of lower density, of odd positive integers not of this form.

The second question is if there is a full arithmetic progression of odd positive integers not of the form $p+2^{2^k}+m!$. The positive answer to this question is given in Theorem 4. Note that, the density of the set constructed in the proof of Theorem 4 is considerably less than the density of the set used in the proof of Theorem 3.

Theorem 3 *The lower density of odd positive integers not of the form* $p+2^{2^k}+m!$ *for* $k, m \in \mathbb{N}_0$ *and* p *prime is at least* $615850829669273873/2459565876494606882 > 1/4$. *The lower density of all positive integers without a representation of the form* $p+2^{2^k}+m!$ *is therefore larger than* $3/4$.

Theorem 4 *There exists a full arithmetic progression of odd positive integers not of the form* $p+2^{2^k}+m!$ *for* $k, m \in \mathbb{N}_0$ *and* p *prime.*

Finally, we prove analogous results for integers not of the form $p+2^{2^k}+2^q$.

Theorem 5 *There exists a subset of the odd positive integers not of the form* $p+2^{2^k}+2^q$, *for* $k \in \mathbb{N}$ *and* p, q *prime, with lower density* $1/6$. *The lower density of all positive integers without a representation of the form* $p+2^{2^k}+2^q$ *is therefore larger than* $2/3$.

Furthermore, there exists a full arithmetic progression of odd positive integers not of the form $p+2^{2^k}+2^q$.

Concerning the last result, we recall that Erdős conjectured that the lower density of the set of positive odd integers not of the form $p+2^k+2^m$ is positive for $k, m \in \mathbb{N}_0$, p prime (see for example [10, Sect. A19]).

For the proofs of Theorems 1 and 2, we apply the method of Romanov [16]. This means that we start with the Cauchy–Schwarz inequality in the form

$$\left(\sum_{\substack{n\le x\\ r_i(n)>0}} 1\right)\left(\sum_{n\le x} r_i(n)^2\right) \ge \left(\sum_{n\le x} r_i(n)\right)^2 \tag{1}$$

for $i \in \{1, 2\}$, where $r_1(n)$ denotes the number of representations of n in the form $p+2^{2^k}+m!$, and $r_2(n)$ counts the number of representations of n in the form $p+2^{2^k}+2^q$. Note that the first sum on the left-hand side of Eq. (1) equals the number of integers less than x having a representation of the required form. It thus suffices to check that

$$\sum_{n\le x} r_i(n) \gg x \text{ and } \sum_{n\le x} r_i(n)^2 \ll x$$

for both $i = 1,\ 2$ in order to get positive lower density for the sets of those integers. The estimates $\sum_{n\le x} r_1(n) \gg x$ and $\sum_{n\le x} r_1(n)^2 \ll x$ are proved in Sect. 3, Lemmas

3 and 4, respectively. The analogous results for $r_2(n)$ are proved in Sect. 4, Lemmas 5 and 6, respectively. Theorems 3 and 4 are proved at the end of Sect. 3 and Theorem 5 at the end of Sect. 4.

2 Notation

Let $\mathbb{N}$, as usual, denote the set of positive integers, $\mathbb{N}_0$ the set of non-negative integers and let $\mathbb{P}$ denote the set of primes. The variables p and q will always denote prime numbers. For any prime $p \in \mathbb{P}$ and any positive integer $n \in \mathbb{N}$, let $\nu_p(n)$ denote the p-adic valuation of n, i.e. $\nu_p(n) = k$ where p^k is the highest power of p dividing n. For an integer n, $P(n)$ denotes its largest prime factor. For any set $S \subset \mathbb{N}$ let $S(x) = |S \cap [1, x]|$ denote the counting function of S. As usual φ denotes Euler's totient function and μ the Möbius function. Furthermore, for an odd positive integer n we denote by $t(n)$ the order of 2 mod n. We use the symbols $\ll$, $\gg$, $\mathcal{O}$ and o within the context of the well-known Vinogradov and Landau notation.

3 Integers of the form $p + 2^{2^k} + m!$

Before proving Lemmas 3 and 4, we establish and collect several results needed in due course. The following is a classical result due to Legendre (see for example Theorems 2.6.1 and 2.6.4 in [14]).

Lemma A *For any prime $p \in \mathbb{P}$ and any positive integer $n \in \mathbb{N}$, we have that*

$$\nu_p(n!) = \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor.$$

Furthermore, if $\sigma_p(n)$ denotes the sum of base p digits of n, then

$$\nu_p(n!) = \frac{n - \sigma_p(n)}{p - 1}.$$

Theorem 6 *The equation $2^{x_1} + y_1! = 2^{x_2} + y_2!$ has only four non-negative integer solutions (x_1, y_1, x_2, y_2) with $x_1 > x_2$ where either $x_2 \le 52$ or $y_2 \le 8$. These solutions are*

$$(x_1, y_1, x_2, y_2) \in \{(1, 0, 0, 2), (1, 1, 0, 2), (3, 2, 2, 3), (7, 4, 5, 5)\}.$$

Proof Suppose that $x_2 \le 52$ and note that $y_1 = 0$ either implies that $y_2 \in \{0, 1\}$ if $x_2 > 0$, which leads to a solution where $x_1 = x_2$, which is excluded, or implies that $x_2 = 0$, whence $x_1 = 1$ and $y_2 = 2$. Hence, the only solution where $y_1 = 0$ is $(x_1, y_1, x_2, y_2) = (1, 0, 0, 2)$. From now on, we may suppose that $y_1 \ge 1$. In this case, from Lemma A, we get that $\nu_2(y_1!) \ge \frac{y_1}{2} - 1$. This yields $\frac{y_1}{2} - 1 \le x_2$ and thus $y_1 \le 106$. Since

$$2^{x_2} - y_1! = 2^{x_1} - y_2!,$$

we have $\nu_2(2^{x_2} - y_1!) = \nu_2(2^{x_1} - y_2!)$. Certainly $|2^{x_2} - y_1!| \le 2^{52} + 106!$ which implies that $\nu_2(2^{x_2} - y_1!) \le \frac{\log(2^{52}+106!)}{\log 2} < 816$. If $x_1 \ge 816$ and $y_2 \ge 822$, then $\nu_2(2^{x_1} - y_2!) \ge 816$, a contradiction. The cases where either $x_1 \le 815$ or $y_2 \le 821$ can be checked by a computer search which leads to the solutions

$$(x_1, y_1, x_2, y_2) \in \{(1, 0, 0, 2), (1, 1, 0, 2), (3, 2, 2, 3), (7, 4, 5, 5)\}.$$

Now suppose that $y_2 \le 8$ and consider

$$0 < 2^{x_1} - 2^{x_2} = y_2! - y_1!,$$

which implies that $y_1 \le y_2 \le 8$. In particular, $|y_2! - y_1!| \le 2 \cdot 8!$ and thus

$$\nu_2(y_2! - y_1!) \le \frac{\log(2 \cdot 8!)}{\log 2} < 17.$$

Since $\nu_2(2^{x_1} - 2^{x_2}) = x_2$, we have that $x_2 < 17$ which is included in the case $x_2 \le 52$ treated above. □

Theorem 7 *If we exclude solutions arising from interchanging* (x_1, y_1) *and* (x_2, y_2)*, the equation* $2^{x_1} + y_1! = 2^{x_2} + y_2!$ *has only four non-negative integer solutions* (x_1, y_1, x_2, y_2) *with* $(x_1, y_1) \ne (x_2, y_2)$ *and* $(y_1, y_2) \notin \{(1, 0), (0, 1)\}$ *if* $x_1 = x_2$*. These are the solutions presented in Theorem 6.*

Proof We compare the 2-adic and 3-adic valuation of both sides of equivalent forms of the equation $2^{x_1} + y_1! = 2^{x_2} + y_2!$ to get information about the size of the parameters x_1, x_2, y_1 and y_2.

If $x_1 = x_2$ we have that $y_1! = y_2!$ and hence either $y_1 = y_2$ or $(y_1, y_2) \in \{(1, 0), (0, 1)\}$ which leads to the excluded trivial solutions. Therefore, w.l.o.g., we may suppose that $x_1 > x_2$ and write

$$2^{x_2}(2^{x_1-x_2} - 1) = y_1!((y_1 + 1) \cdots y_2 - 1). \tag{2}$$

Next we compute the 2-adic valuation of both sides of the last equality. For the left-hand side we simply have $\nu_2(2^{x_2}(2^{x_1-x_2} - 1)) = x_2$ while for the right-hand side we use that the factor $((y_1 + 1) \cdots y_2 - 1)$ is odd as soon as $y_2 \ge y_1 + 2$ which yields

$$\nu_2(y_1!((y_1 + 1) \cdots y_2 - 1)) = \begin{cases} \nu_2(y_1!), & \text{if } y_2 \ge y_1 + 2, \\ \nu_2(y_1!) + \nu_2(y_1), & \text{if } y_2 = y_1 + 1. \end{cases}$$

From this, Lemma A and the fact that $1 \le \sigma_2(y_1) \le \frac{\log y_1}{\log 2} + 1$ (note that as in the proof of Theorem 6, $y_1 \in \{0, 1\}$ leads to a single non-trivial solution listed there), we get the following two inequalities:

$$x_2 = \nu_2(2^{x_2}(2^{x_1-x_2}-1)) = \nu_2(y_1!((y_1+1)\cdots y_2-1)) \le \nu_2(y_1!) + \nu_2(y_1) < y_1 + \frac{\log y_1}{\log 2}, \tag{3}$$

$$x_2 = \nu_2(2^{x_2}(2^{x_1-x_2}-1)) = \nu_2(y_1!((y_1+1)\cdots y_2-1)) \ge \nu_2(y_1!) \ge y_1 - \left(\frac{\log y_1}{\log 2}+1\right). \tag{4}$$

By Theorem 6, we may suppose that $x_2 \ge 5$ without loosing solutions. In this case, the last inequality implies $y_1 \le 2x_2$.

Next, we look at

$$2^{x_1} = 2^{x_2} + y_2! - y_1!.$$

Since $2^{x_2} \le 2^{x_1-1} = \frac{2^{x_1}}{2}$, we have that $y_2! > \frac{2^{x_1}}{2}$, whence we get

$$y_2^{y_2} \ge y_2! > \frac{2^{x_1}}{2},$$

and thus

$$y_2 \log y_2 > (x_1-1)\log 2 \text{ and } y_2 > \frac{(x_1-1)\log 2}{\log y_2}.$$

To get the last inequality we used that by Theorem 6 we may suppose that $y_2 \ge 9$ whence $\log y_2 > 0$. Now $x_2 \ge 5$ implies that $x_1 \ge 6$. If we would have that $y_2 \le x_1$ the last inequality would imply that

$$y_2 > \frac{\log 2}{2}\left(\frac{x_1}{\log y_2}\right) > \frac{1}{4}\left(\frac{x_1}{\log x_1}\right). \tag{5}$$

In order to prove (5), it therefore suffices to prove that $y_2 \le x_1$ for $x_1 \ge 6$. In order to do so, we consider the equation

$$2^{x_1} = y_1!((y_1+1)\cdots y_2 - 1) + 2^{x_2}$$

from which we readily deduce that $y_1! < 2^{x_1}$. This together with $2^{x_1} = y_2! - y_1! + 2^{x_2}$ implies that

$$y_2! < 2 \cdot 2^{x_1}.$$

This implies that $y_2 \le x_1$, since otherwise $(x_1+1)! \le 2^{x_1+1}$ which is true for $x_1 \le 2$ only. By Theorem 6 again, we may suppose that $y_2 \ge 9$. In this case, applying Lemma A, we obtain

$$\nu_3(y_2!) \ge \left\lfloor \frac{y_2}{3} \right\rfloor + \left\lfloor \frac{y_2}{9} \right\rfloor \ge \frac{y_2}{3} > \frac{1}{12}\left(\frac{x_1}{\log x_1}\right), \tag{6}$$

where the last inequality follows by (5). Now we compute the 3-adic valuation of both sides of Eq. (2). By inequality (3) and Lemma A for the right-hand side, we get

$$k = \nu_3(y_1!((y_1+1)\cdots y_2 - 1)) \geq \nu_3(y_1!) = \frac{y_1 - \sigma_3(y_1)}{2} \geq \frac{y_1}{2} - \frac{\log y_1}{\log 3} - 1$$
$$\geq \frac{x_2}{2} - \log(y_1)\left(\frac{1}{2\log 2} + \frac{1}{\log 3}\right) - 1.$$

Since for the left-hand side of (2) we have $3^k | 2^{x_1 - x_2} - 1$, we deduce that $\varphi(3^k) = 2 \cdot 3^{k-1} | x_1 - x_2$. Here we used that 2 is a primitive root modulo any power of 3. This is a direct consequence of Jacobi's observation [12, p. XXXV] that a primitive root modulo p^2 is also a primitive root modulo any higher power of p. Using the above bound for k and the fact that $y_1 \leq 2x_2$, we get

$$x_1 \geq x_2 + 2\cdot 3^{k-1} \geq x_2 + \frac{2}{9}3^{x_2/2 - \log(y_1)(1/2\log 2 + 1/\log 3)} \geq x_2 + \frac{2\cdot 3^{x_2/2}}{36x_2^2} \geq \frac{3^{x_2/2}}{18x_2^2}. \quad (7)$$

Next we find an upper bound for x_1 in terms of x_2. Consider the equation

$$2^{x_1} - y_2! = 2^{x_2} - y_1!.$$

Equation (5) yields that $y_2 > \frac{1}{4}\frac{x_1}{\log x_1} > \frac{1}{4}\sqrt{x_1}$. Thus, by Lemma A, $\nu_2(y_2!) > \frac{\sqrt{x_1}}{8} - 1$ and hence $\nu_2(2^{x_1} - y_2!) \geq \frac{\sqrt{x_1}}{8} - 1$.

On the other hand, $|2^{x_2} - y_1!| \leq 2^{x_2} + y_1! \leq 2^{x_2} + (2x_2)^{2x_2} \leq 2\cdot(2x_2)^{2x_2}$. Now $\nu_2(2^{x_2} - y_1!)$ is certainly bounded from above by the highest power of 2 less than $2\cdot(2x_2)^{2x_2}$:

$$2^a \leq 2\cdot(2x_2)^{2x_2} \Leftrightarrow a \leq \frac{2x_2\log(2x_2)}{\log 2} + 1.$$

We therefore have that $\nu_2(2^{x_2} - y_1!) \leq 4x_2\log(2x_2) + 1$ and putting everything together, we get

$$\frac{\sqrt{x_1}}{8} - 1 \leq \nu_2(2^{x_1} - y_2!) = \nu_2(2^{x_2} - y_1!) \leq 4x_2\log(2x_2) + 1,$$

which implies that $x_1 \leq (32x_2\log(2x_2) + 16)^2$. Combining this with (7), we finally arrive at

$$3^{x_2/2} \leq 18x_2^2(32x_2\log(2x_2) + 16)^2.$$

This inequality is valid only for $x_2 \leq 52$ and the solutions satisfying this restriction are given in Theorem 6. $\square$

Lemma 1 *Let $m_1, m_2, m_3, m_4 \in \mathbb{N}$ such that $m_1 > m_2$, $m_3 > m_4$ and*

$$m_1! - m_2! = m_3! - m_4!. \tag{8}$$

Then $(m_1, m_2) = (m_3, m_4)$ or $m_1 = m_3$ and $(m_2, m_4) \in \{(0, 1), (1, 0)\}$.

Proof We start with the case where either $m_1 = m_2 + 1$ or $m_3 = m_4 + 1$ and w.l.o.g suppose that $m_1 = m_2 + 1$. If furthermore $m_2 \le m_4$, we get from Eq. (8)

$$m_2!\, m_2 = m_4!((m_4 + 1)\cdots m_3 - 1) \ge m_2! m_4,$$

which leads to $m_2 \ge m_4$ and thus $m_2 = m_4$ which implies $m_1 = m_3$. On the other hand, if $m_1 = m_2 + 1$ and $m_2 > m_4$ Eq. (8) implies that

$$m_2(m_4 + 1)\cdots m_2 = (m_4 + 1)\cdots m_3 - 1, \tag{9}$$

and therefore $m_4 + 1|1$ if $m_3 > m_4 + 1$ and $m_4 + 1|m_4$ otherwise, whence $m_4 = 0$ in both cases. Now $m_3 = 1$ implies that $(m_1, m_2) = (1, 0)$ and we are done. Otherwise, if $m_3 \ne 1$, then the right-hand side of (9) is odd. In order for the left-hand side to be odd we need $m_2 = 1$, which implies that $m_1 = m_3$.

It remains to consider the case where $m_1 \ge m_2 + 2$ and $m_3 \ge m_4 + 2$ and w.l.o.g. we suppose that $m_2 > m_4$. We look at Eq. (8) in the form

$$m_2!((m_2 + 1)\cdots m_1 - 1) = m_4!((m_4 + 1)\cdots m_3 - 1). \tag{10}$$

By assumption, we have that $\nu_2(m_2!) = \nu_2(m_4!)$ which implies that m_4 is even and $m_2 = m_4 + 1$. We hence may rewrite Eq. (10) to get

$$(m_4 + 1)\cdots m_1 - m_4 = (m_4 + 1)\cdots m_3.$$

It follows that $m_4 + 1|m_4$ which implies that $m_4 = 0$. This leads to $m_2 = 1$ and $m_1 = m_3$. □

Lemma 2 *For odd positive n, let $t(n)$ be the order of 2 mod n and $t(n) = 2^{a(n)}b(n)$ such that $b(n)$ is odd. Then the series*

$$\sum_{\substack{2\nmid n\\ \mu^2(n)=1}} \frac{1}{nt(b(n))}$$

converges.

Proof Recall that $P(n)$ denotes the largest prime factor of n and observe that if $u|v$ then $t(u)|t(v)$, thus $b(u)|b(v)$ and further $t(b(u))|t(b(v))$. From this and Mertens' formula in the weak form

$$\prod_{p\le x}\left(1 + \frac{1}{p}\right) \ll \log x,$$

we get

$$\sum_{\substack{2\nmid n\\ \mu^2(n)=1}} \frac{1}{nt(b(n))} \le \sum_{\substack{p\ge 3\\ p\in\mathbb{P}}} \frac{1}{pt(b(p))} \sum_{\substack{2\nmid m\\ \mu(m)^2=1\\ P(m)<p}} \frac{1}{m} = \sum_{\substack{p\ge 3\\ p\in\mathbb{P}}} \frac{1}{pt(b(p))} \prod_{\substack{q<p\\ q\in\mathbb{P}}} \left(1+\frac{1}{q}\right)$$
$$\ll \sum_{\substack{p\ge 3\\ p\in\mathbb{P}}} \frac{\log p}{pt(b(p))}. \tag{11}$$

We split the primes into two subsets $\mathcal{P}$ and $\mathcal{Q}$ and consider the contribution of these sets separately. We set $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3 \cup \mathcal{P}_4$ where

$$\begin{aligned}
\mathcal{P}_1 &:= \left\{p \in \mathbb{P} : t(p) < p^{1/3}\right\},\\
\mathcal{P}_2 &:= \left\{p \in \mathbb{P} : P(t(p)) < p^{1/\log\log p},\ p \notin \mathcal{P}_1\right\},\\
\mathcal{P}_3 &:= \{p \in \mathbb{P} : P(t(p)) \in \mathcal{P}_1,\ p \notin \mathcal{P}_1 \cup \mathcal{P}_2\},\\
\mathcal{P}_4 &:= \{p \in \mathbb{P} : p \le p_0\},
\end{aligned}$$

for some fixed p_0 to be chosen later. The set $\mathcal{Q}$ is then defined to be $\mathbb{P}\backslash(\mathcal{P} \cup \{2\})$. We start by showing that

$$\mathcal{P}(x) \ll \frac{x}{(\log x)^3}. \tag{12}$$

For $\mathcal{P}_1$, applying an idea of Erdős and Murty [6], we use that $p|2^k - 1$ where $k = t(p)$, whence we have that

$$\prod_{\substack{p\le x\\ p\in\mathcal{P}_1}} p \ \Big| \prod_{k\le x^{1/3}} (2^k - 1).$$

From this, we get

$$2^{\mathcal{P}_1(x)} \le \prod_{\substack{p\le x\\ p\in\mathcal{P}_1}} p \le \prod_{k\le x^{1/3}} (2^k - 1) \le 2^{\sum_{k\le x^{1/3}} k} \le 2^{x^{2/3}},$$

which shows that

$$\mathcal{P}_1(x) \ll x^{2/3} = o\left(\frac{x}{(\log x)^3}\right). \tag{13}$$

To deal with the contribution of the set $\mathcal{P}_2$, we set

$$\Psi(x, y) := |\{n \le x : P(n) \le y\}|.$$

By known results on smooth numbers (in particular, a result of Canfield, Erdős and Pomerance from [2, Corollary p.15]), we have for $y > (\log x)^2$,

$$\Psi(x,y)=\frac{x}{\exp((1+o(1))u\log u)},\quad \text{where}\quad u=\frac{\log x}{\log y}, \tag{14}$$

as both y and u tend to infinity. For $p\in\mathcal{P}_2$ we may suppose that $p>x^{1/2}$ since there are at most $\mathcal{O}(\pi(x^{1/2}))=\mathcal{O}(x^{1/2}/\log x)=o(x/(\log x)^3)$ primes in $\mathcal{P}_2$ less than $\sqrt{x}$. If $p>x^{1/2}$, then $\log\log p>\log\log x/2$ for sufficiently large x, and hence for $x^{1/2}<p<x$ in $\mathcal{P}_2$, we have

$$P(t(p))<p^{1/\log\log p}<x^{2/\log\log x}.$$

Put $y:=x^{2/\log\log x}$. Thus, $p-1$ is a number which is at most x, having a divisor $t(p)>p^{1/3}>x^{1/6}$, whose largest prime factor is at most y. It follows that $p-1\le x$ is a multiple of some number $d>x^{1/6}$ with $P(d)\le y$. For a fixed d, the number of such p is at most $\lfloor x/d\rfloor\le x/d$. Summing over d, we get that

$$\begin{aligned}\mathcal{P}_2(x)&\ll\sum_{\substack{x^{1/6}<d<x\\ P(d)<y}}\frac{x}{d}=x\int_{x^{1/6}}^{x}\frac{1}{t}d\Psi(t,y)\\ &=x\left(\left(\frac{\Psi(t,y)}{t}\right)\Big|_{t=x^{1/6}}^{t=x}+\int_{x^{1/6}}^{x}\frac{1}{t^2}\Psi(t,y)dt\right)\\ &\ll x\left(\frac{\Psi(x,y)}{x}+\int_{x^{1/6}}^{x}\frac{\Psi(t,y)}{t^2}dt\right).\end{aligned}$$

Putting $u_0:=\log x^{1/6}/\log y=(1/12)\log\log x$, we get that $u=\log t/\log y\ge u_0$ for all $t\in[x^{1/6},x]$, and

$$(1+o(1))u_0\log u_0=\left(\frac{1}{12}+o(1)\right)\log\log x\log\log\log x>4\log\log x \tag{15}$$

for large x. Using (14) and (15), we thus get that

$$\mathcal{P}_2(x)\ll\frac{x+x\log x}{\exp((1+o(1))u_0\log u_0)}\ll\frac{x}{(\log x)^3}.$$

Next we consider the contribution of $\mathcal{P}_3$. This set contains primes p such that $p-1$ is divisible by some prime $q>p^{1/\log\log p}$ but $q\in\mathcal{P}_1$. We may assume again that $p>x^{1/2}$, then $q>p^{1/\log\log p}>y^{1/4}$, where as before $y=x^{2/\log\log x}$. Fixing q, the number of primes $p\le x$ such that $p-1$ is a multiple of q is at most x/q. Summing up over $q\in\mathcal{P}_1$ and using (13), we get that

$$\begin{aligned}\mathcal{P}_3(x)&\le\sum_{\substack{y^{1/4}<q<x\\ q\in\mathcal{P}_1}}\frac{x}{q}\ll x\int_{y^{1/4}}^{x}\frac{\mathrm{d}\mathcal{P}_1(t)}{t}=x\left(\left(\frac{\mathcal{P}_1(t)}{t}\right)\Big|_{t=y^{1/4}}^{x}+\int_{y^{1/4}}^{x}\frac{\mathcal{P}_1(t)}{t^2}\mathrm{d}t\right)\\ &\ll x\left(\frac{1}{x^{1/3}}+\int_{y^{1/4}}^{x}\frac{\mathrm{d}t}{t^{4/3}}\right)\ll\frac{x}{y^{1/12}}\ll\frac{x}{(\log x)^3}.\end{aligned}$$

Finally, choose p_0 such that for $p > p_0$ we have that $p^{1/3 \log\log p} > (\log p)^3$ and get

$$\mathcal{P}_3(x) \ll 1 \ll \frac{x}{(\log x)^3}.$$

We are now ready to prove that the sum on the right-hand side of (11) converges. For the contribution of primes $p \in \mathcal{P}$, we use the Abel summation formula as well as (12) and get

$$\begin{aligned}\sum_{\substack{p\le x\\ p\in\mathcal{P}}} \frac{\log p}{pt(b(p))} &\le \sum_{\substack{p\le x\\ p\in\mathcal{P}}} \frac{\log p}{p} = \int_3^x \frac{\log t}{t}\mathrm{d}\mathcal{P}(t)\\ &= \frac{\mathcal{P}(t)\log t}{t}\Big|_{t=3}^{x} - \int_3^x \frac{1-\log t}{t^2}\mathcal{P}(t)\mathrm{d}t\\ &\ll 1 + \int_3^x \frac{\log t}{t^2}\frac{t}{(\log t)^3}\mathrm{d}t = 1 + \int_3^x \frac{\mathrm{d}t}{t(\log t)^2} \ll 1.\end{aligned}$$

By the definition of $\mathcal{Q}$ for $p \in \mathcal{Q}$ we have that $q = P(t(p)) > p^{1/\log\log p}$ which implies that $q|b(p)$ for large p. Furthermore, $q \notin \mathcal{P}_1$ so $t(q) > q^{1/3} > p^{1/3 \log\log p}$. By the choice of the constant p_0 in the definition of $\mathcal{P}_4$ this implies that $t(b(p)) \ge t(q) > (\log p)^3$. Finally, this implies that

$$\sum_{p\in\mathcal{Q}} \frac{\log p}{pt(b(p))} \le \sum_{n\in\mathbb{N}} \frac{1}{n(\log n)^2} \ll 1,$$

which finishes the proof of the lemma. □

Lemma 3 *The following estimate holds:*

$$\sum_{n\le x} r_1(n) \gg x.$$

Proof We certainly have that

$$\sum_{n\le x} r_1(n) \ge \left(\sum_{p\le x/3} 1\right)\left(\sum_{2^{2^k}\le x/3} 1\right)\left(\sum_{m!\le x/3} 1\right).$$

By the Prime Number Theorem

$$\sum_{p\le x/3} 1 \sim \frac{x}{3\log(x/3)} \gg \frac{x}{\log x}, \tag{16}$$

and $2^{2^k} \leq x/3$ implies that $k \leq \frac{\log(\log(x/3)) - \log 2}{\log 2}$ and hence

$$\sum_{2^{2^k} \leq x/3} 1 \gg \log \log x. \tag{17}$$

We use that $m! \leq m^m$ and that $m^m \leq x/3$ for $m \leq \log x / 2 \log \log x$ and sufficiently large x. This implies that

$$\sum_{m! \leq x/3} 1 \gg \frac{\log x}{\log \log x}. \tag{18}$$

The bounds in (16), (17) and (18) show that

$$\sum_{n \leq x} r_1(n) \gg x.$$

□

Lemma 4 *The following estimate holds:*

$$\sum_{n \leq x} r_1(n)^2 \ll x.$$

Proof We begin with the observation that the sum counts exactly the number of solutions of the equation

$$p_1 + 2^{2^{k_1}} + m_1! = p_2 + 2^{2^{k_2}} + m_2!$$

in p_1, p_2, k_1, k_2, m_1 and m_2 where $p_1 + 2^{2^{k_1}} + m_1! \leq x$. For fixed k_1, k_2, m_1 and m_2 this amounts to counting pairs of primes (p_1, p_2) such that $p_2 = p_1 + h$, where

$$h := 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2!.$$

If $h = 0$, then we apply Theorem 7 to get that either $(k_1, m_1) = (k_2, m_2)$ or $k_1 = k_2$ and $(m_1, m_2) \in \{(1, 0), (0, 1)\}$[1]. The number of choices of the form $(p_1, p_2, k_1, k_2, m_1, m_2)$ in this case is

$$\mathcal{O}\left(\frac{x}{\log x}\left(\log \log x \frac{\log x}{\log \log x} + \log \log x\right)\right) = \mathcal{O}(x).$$

If h is odd, then one of the primes p_1 and p_2 equals 2 and any choice of (k_1, k_2, m_1, m_2) fixes the other prime. There are

$$\mathcal{O}\left((\log \log x)^2 \left(\frac{\log x}{\log \log x}\right)^2\right) = o(x)$$

[1] Note that x_1 and x_2 in the non-trivial solutions in Theorem 7 are never both powers of 2.

choices for $(p_1, p_2, k_1, k_2, m_1, m_2)$ in this case. To deal with the remaining even $h \neq 0$, we use a classical sieve bound (cf. for example [15, Theorem 7.3]). In this case, the number of pairs (p_1, p_2) of primes such that $p_2 = p_1 + h$ is

$$\mathcal{O}\left(\frac{x}{(\log x)^2}\prod_{p|h}\left(1+\frac{1}{p}\right)\right).$$

Summing over all choices (k_1, k_2, m_1, m_2) such that $h \neq 0$ is even (this range of summation is indicated by the dash in the superscript of the sum below), we hence need to show that

$$\frac{x}{(\log x)^2}\sideset{}{'}\sum_{(k_1,k_2,m_1,m_2)}\prod_{p|h}\left(1+\frac{1}{p}\right) \ll x. \tag{19}$$

Observing that the prime $p = 2$ contributes just a constant factor, this amounts to showing that

$$\sideset{}{'}\sum_{(k_1,k_2,m_1,m_2)}\prod_{\substack{p|h\\p>2}}\left(1+\frac{1}{p}\right) \ll (\log x)^2,$$

which we do in what follows. We now rewrite the left-hand side of the last inequality above as

$$\begin{aligned}\sideset{}{'}\sum_{(k_1,k_2,m_1,m_2)}\prod_{\substack{p|h\\p>2}}\left(1+\frac{1}{p}\right) &= \sideset{}{'}\sum_{(k_1,k_2,m_1,m_2)}\sum_{\substack{d|h\\d\text{ odd}}}\frac{\mu(d)^2}{d}\\ &= \sideset{}{'}\sum_{\substack{d\text{ odd}\\ \mu(d)^2=1}}\frac{|\{(k_1,k_2,m_1,m_2) : d|h\}|}{d}.\end{aligned}$$

Therefore we need to study, for a given odd square-free d, the cardinality of the set

$$S_d := \{(k_1, k_2, m_1, m_2) : d|h, h \neq 0, 2 \nmid h\}.$$

We start with the subset $S_{1,d} \subset S_d$ where

$$S_{1,d} := \{(k_1, k_2, m_1, m_2) \in S_d : m_1 = m_2 \text{ or } \{m_1, m_2\} = \{0, 1\}\}. \tag{20}$$

We thus first deal with

$$\sideset{}{'}\sum_{\substack{d\text{ odd}\\ \mu(d)^2=1}}\frac{|S_{1,d}|}{d}.$$

By (20), (m_1, m_2) is chosen in at most $O(\log x/\log\log x)$ ways. As for (k_1, k_2), we have $2^{2^{k_1}} \equiv 2^{2^{k_2}} \pmod d$. Since d is odd this implies that $2^{2^{k_1}-2^{k_2}} \equiv 1 \pmod d$.

Recall that $t(d)$ is the order of 2 modulo d. The above congruence makes $2^{k_1} \equiv 2^{k_2} \pmod{t(d)}$. As above we write $t(d) = 2^{a(d)}b(d)$, where $b(d)$ is odd and $a(d)$ is some non-negative integer. This implies that $2^{k_1-k_2} \equiv 1 \pmod{b(d)}$. The above cancellation again is justified since $b(d)$ is odd. Hence, for k_2 fixed, k_1 is in a fixed arithmetic progression modulo $t(b(d))$. The number of such k_1 with $2^{2^{k_1}} \le x$ is of order (up to a constant) at most

$$\left\lfloor \frac{\log\log x}{t(b(d))} \right\rfloor + 1.$$

Since k_2 is chosen in $\mathcal{O}(\log\log x)$ ways, we have

$$\begin{aligned}\sideset{}{'}\sum_{\substack{d \text{ odd}\\ \mu(d)^2=1}} \frac{|S_{1,d}|}{d} &\ll \left(\frac{\log x}{\log\log x}\right)\log\log x\left(\log\log x \sum_{\substack{d \text{ odd}\\ \mu(d)^2=1}} \frac{1}{dt(b(d))} + \sum_{\substack{d\le x\\ d \text{ odd}\\ \mu(d)^2=1}} \frac{1}{d}\right)\\ &\ll (\log x)^2,\end{aligned}$$

where we used Lemma 2 and the fact that

$$\sum_{\substack{d\le x\\ d \text{ odd}\\ \mu(d)^2=1}} \frac{1}{d} \ll \log x.$$

From now on, we deal with $S_d \backslash S_{1,d}$. Any quadruple (k_1, k_2, m_1, m_2) in the above set gives $m_1! - m_2! \ne 0$ and we assume that $m_1 > m_2$. We partition the numbers d in the range of summation into two different sets A and B. We set

$$A := \left\{ d \in \mathbb{N}\colon \begin{array}{c} 2 \nmid d,\ \mu(d)^2 = 1,\ \forall\{(k_1,k_2,m_1,m_2),(k_3,k_4,m_3,m_4)\} \in (S_d\backslash S_{1,d})^2 : \\ 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2! = 2^{2^{k_3}} + m_3! - 2^{2^{k_4}} - m_4! = h \end{array} \right\},$$

$$B := \left\{ d \in \mathbb{N}\colon \begin{array}{c} 2 \nmid d,\ \mu(d)^2 = 1,\ \exists\{(k_1,k_2,m_1,m_2),(k_3,k_4,m_3,m_4)\} \in (S_d\backslash S_{1,d})^2 : \\ 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2! \ne 2^{2^{k_3}} + m_3! - 2^{2^{k_4}} - m_4! \end{array} \right\}.$$

In the set A we thus collect all d for which all solutions in $S_d \backslash S_{1,d}$ give the same h and the set B contains all other d. For $d \in A$ we fix k_1 and k_2 for solutions in $S_d \backslash S_{1,d}$ and get

$$m_1! - m_2! = h - 2^{2^{k_1}} + 2^{2^{k_2}}.$$

The existence of some other element $(k_1, k_2, m_3, m_4) \in S_d \backslash S_{1,d}$ with $m_3 > m_4$ would imply that $m_1! - m_2! = m_3! - m_4!$ which by Lemma 1 leads to $(m_1, m_2) = (m_3, m_4)$. Hence, for $d \in A$ and for $(k_1, k_2, m_1, m_2) \in S_d \backslash S_{1,d}$ with $m_1 > m_2$, the last two coordinates are uniquely determined by the first two whence for $d \in A$ we have

$$|(S_d \backslash S_{1,d})| \ll (\log\log x)^2.$$

We thus get that

$$\sum_{d \in A} \frac{|(S_d \backslash S_{1,d})|}{d} \ll (\log\log x)^2 \sum_{d \le x} \frac{1}{d} \ll (\log x)(\log\log x)^2 = o((\log x)^2).$$

Finally, we deal with the contribution of $d \in B$. By definition, we may find two quadruples (k_1, k_2, m_1, m_2) with $m_1 > m_2$ and (k_3, k_4, m_3, m_4) with $m_3 > m_4$ both in $S_d \backslash S_{1,d}$ such that

$$h := 2^{2^{k_1}} + m_1! - 2^{2^{k_2}} - m_2! \neq 2^{2^{k_3}} + m_3! - 2^{2^{k_4}} - m_4! =: h'. \tag{21}$$

Let $\mathcal{P}$ be the set of possible prime factors of $d \in B$ which exceed $\log x$. We shall prove that $|\mathcal{P}| = \mathcal{O}((\log x)^5)$. For h, h' in (21) we have that they are both divisible by d and thus $d | h - h'$. Every prime factor of d (in particular the ones larger than $\log x$) divides

$$\prod_{k_i, m_i}{}' \left((2^{2^{k_1}} - 2^{2^{k_2}} + m_1! - m_2!) - (2^{2^{k_3}} - 2^{2^{k_4}} + m_3! - m_4!) \right),$$

where the product is taken over all m_i with $m_i! \le x$ and all k_i with $2^{2^{k_i}} \le x$ for $i = 1, 2, 3, 4$. The dash indicates that the product is to be taken over the non-zero factors only. Since each factor in this product is of size $\mathcal{O}(x)$ any of these factors has at most $\mathcal{O}(\log x)$ prime factors. Furthermore, for the octuple $(k_1, k_2, k_3, k_4, m_1, m_2, m_3, m_4)$ we have $\mathcal{O}((\log\log x)^4 (\log x/\log\log x)^4) = \mathcal{O}((\log x)^4)$ choices and altogether we have that $|\mathcal{P}| = \mathcal{O}((\log x)^5)$. Write $d = u_d v_d$, where u_d is divisible by primes $p \le \log x$ only. Hence the factor v_d is divisible only by primes in $\mathcal{P}$. Then

$$\sum_{d \in B} \frac{|(S_d \backslash S_{1,d})|}{d} \le \left(\sum_{\substack{u \text{ odd} \\ \mu(u)^2 = 1 \\ P(u) < \log x}} \frac{|(S_u \backslash S_{1,u})|}{u} \right) \left(\sum_{\substack{v \text{ odd} \\ \mu(v)^2 = 1 \\ p | v \Rightarrow p \in \mathcal{P}}} \frac{1}{v} \right),$$

where we used that $S_d \backslash S_{1,d} \subset S_u \backslash S_{1,u}$ if $u \mid d$. For the second sum we have

$$\sum_{\substack{v \text{ odd} \\ \mu(v)^2 = 1 \\ p | v \Rightarrow p \in \mathcal{P}}} \frac{1}{v} = \prod_{p \in \mathcal{P}} \left(1 + \frac{1}{p} \right) = \mathcal{O}(1),$$

which follows from partial summation and the fact that $\mathcal{P}$ has $\mathcal{O}((\log x)^5)$ elements all larger than $\log x$. It thus remains to bound

$$\sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{|(S_u \backslash S_{1,u})|}{u}.$$

For this, we fix (m_1, m_2) with $m_1 > m_2$ not both in $\{0, 1\}$. Then putting $M_{1,2} = m_2! - m_1!$, we need to count the number of (k_1, k_2) such that $2^{2^{k_1}} - 2^{2^{k_2}} \equiv M_{1,2} \pmod{u}$. Analogously as before, for fixed k_2, this puts k_1 into a fixed arithmetic progression modulo $t(b(u))$. The number of k_1 with $2^{2^{k_1}} \le x$ in this progression is of order $O(\log\log x/t(b(u)) + 1)$. Thus, we have

$$\sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{|(S_u \backslash S_{1,u})|}{u} \ll \left(\frac{\log x}{\log\log x}\right)^2 (\log\log x)$$
$$\times \left(\log\log x \sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{1}{ut(b(u))} + \sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{1}{u} \right) \ll (\log x)^2.$$

Here, we used Lemma 2 and Mertens' formula, which yields

$$\sum_{\substack{u \text{ odd} \\ \mu(u)^2=1 \\ P(u)<\log x}} \frac{1}{u} = \prod_{3 \le p \le \log x} \left(1 + \frac{1}{p}\right) \ll \log\log x.$$

□

Proof of Theorem 3 Since the density of integers of the form $p + 2^{2^k} + m!$, $p \in \mathbb{P}$, $m, k \in \mathbb{N}$ and $m < 2^{2^6} - 1$ is zero, we may suppose that $m \ge 2^{2^6} - 1$. In this case, we have $m! \equiv 0 \bmod 2^{2^6} - 1$, and for $k \ge 6$, we have that $2^{2^k} \equiv 1 \bmod 2^{2^6} - 1$. If $n \equiv a + 1 \bmod 2^{2^6} - 1$, where a is a residue class $\bmod\ 2^{2^6} - 1$ with $(a, 2^{2^6} - 1) > 1$, then $(n - 2^{2^k} - m!, 2^{2^6} - 1) > 1$ which leaves only finitely many choices for the prime $p = n - 2^{2^k} - m!$. This implies that the proportion of such n with a representation of the form $n = p + 2^{2^k} + m!$ is zero. We have $2^{2^6} - 1 - \varphi(2^{2^6} - 1)$ choices for the residue class a and half of the integers in these residue classes are odd which yields a density of

$$\frac{2^{2^6} - 1 - \varphi(2^{2^6} - 1)}{2 \cdot (2^{2^6} - 1)} = \frac{615850829669273873}{2459565876494606882}.$$

We note that a more refined version of the above argument was used by Habsieger and Roblot [11, Sect. 3] to prove an upper bound on the proportion of odd integers not of the form $p + 2^k$. □

Proof of Theorem 4 We will show that none of the integers n satisfying the following system of congruences is of the form $p + 2^{2^k} + m!$:

1 mod 2	1 mod 3	3 mod 5
2 mod 7	6 mod 11	3 mod 17
7 mod 19	9 mod 23.	

By the Chinese Remainder Theorem, the arithmetic progressions above intersect in a unique arithmetic progression. Let n be an element of this progression and suppose that $n = p + 2^{2^k} + m!$.

If $m \geq 3$, then $n = p + 2^{2^k} + m! \equiv p + 2^{2^k} \bmod 3$. All primes except for 3 are in the residue classes $1, 2 \bmod 3$ and $2^{2^k} \equiv 1 \bmod 3$ for $k \geq 1$. Thus, for $m \geq 3$ and $k \geq 1$ we have that $n = p + 2^{2^k} + m! \equiv 1 \bmod 3$; hence, the only possible choice for p is $p = 3$.

Next, we show that if $p = 3$, then $m < 5$. To do so, we use that $2^{2^k} \equiv 1 \bmod 5$ for $k \geq 2$; hence for $m \geq 5$ we are left with $n = 3 + 2^{2^k} + m! \equiv \{0, 2, 4\} \bmod 5$, a contradiction to $n \equiv 3 \bmod 5$.

In the case that $k = 0$, we will show that $m \geq 3$ implies $m < 7$. Let $n = p+2+m!$ and $m \geq 3$. Then $n \equiv 1 \bmod 3$ implies that $p \equiv 2 \bmod 3$. If additionally $m \geq 7$, then $n = p + 2 + m! \equiv p + 2 \bmod 7$. Since $n \equiv 2 \bmod 7$, the only possible choice for p is $p = 7$, which contradicts $p \equiv 2 \bmod 3$.

Using the above observations, the only cases we need to consider are those of $m = 0, m = 1, m = 2, m = 3, 4$ and $k = 0$ or $p = 3$ and $m = 5, 6$ and $k = 0$.

If $m \in \{0, 1\}$ and we additionally have that p is odd, then $n = p + 2^{2^k} + 1$ is even, a contradiction to $n \equiv 1 \bmod 2$. It remains to deal with the case when $p = 2$. Then we have $n = 2 + 2^{2^k} + 1$ and we get a contradiction from $n \equiv 3 \bmod 5$ which would imply that $2^{2^k} \equiv 0 \bmod 5$.

For the case $m = 2$, we use that $2^{2^k} \equiv 1 \bmod 17$ for $k \geq 3$. Hence, for $m = 2$ and $k \geq 3$, we have that $n = p + 2^{2^k} + 2 \equiv p + 3 \bmod 17$ which together with $n \equiv 3 \bmod 17$ leaves us with $p = 17$. We use that $n = 17+2^{2^k}+2 \equiv 2 \bmod 3$ to get a contradiction to $n \equiv 1 \bmod 3$. Since $m = 2$ and $k = 0$ imply $n = p+4 \equiv p+1 \bmod 3$, the only possible choice for p in this case is $p = 3$ but $n = 7 \not\equiv 3 \bmod 5$. If $m = 2$ and $k = 1$, then $n = p + 6$ and $n \equiv 6 \bmod 11$ implies that $p = 11$. This contradicts $n \equiv 1 \bmod 3$. Last we need to deal with $m = 2$ and $k = 2$. In this case, $n = p+18 \equiv p+3 \bmod 5$, and hence, $n \equiv 3 \bmod 5$ implies that $p = 5$. Now $n = 23$ does not satisfy the congruence $n \equiv 1 \bmod 3$.

If $m = 3$ and $p = 3$ we have that $n = 9+2^{2^k} \equiv 8, 10, 11, 13 \bmod 17$ contradicting $n \equiv 3 \bmod 17$. On the other hand, if $m = 3$ and $k = 0$, then $n = p+8 \equiv p+3 \bmod 5$ and we get a contradiction as shown above.

For $m = 4$ and $p = 3$ we get $n = 27 + 2^{2^k} \equiv \{9, 11, 12, 14\} \bmod 17$, a contradiction to $n \equiv 3 \bmod 17$. If $m = 4$ and $k = 0$, it follows that $n = p + 26 \equiv p + 7 \bmod 19$ which implies $p = 19$ and $n = 45$. This contradicts $n \equiv 3 \bmod 5$.

In the case when $m = 5$ and $k = 0$, we have that $n = p + 122 \equiv p + 3 \bmod 17$. Together with $n \equiv 3 \bmod 17$ this only leaves $p = 17$ which contradicts $n \equiv 3 \bmod 5$.

Finally, if $m = 6$ and $k = 0$, then $n = p + 722 \equiv p + 9 \bmod 23$. Together with $n \equiv 9 \bmod 23$, this only leaves $p = 23$ which yields a contradiction to $n \equiv 3 \bmod 5$. □

4 Integers of the form $p + 2^{2^k} + 2^q$

Lemma 5 *The following estimate holds:*

$$\sum_{n \le x} r_2(n) \gg x.$$

Proof The lemma follows from

$$\sum_{n \le x} r_2(n) \ge \left(\sum_{\substack{p \le x/3 \\ p \in \mathbb{P}}} 1 \right) \left(\sum_{2^{2^k} \le x/3} 1 \right) \left(\sum_{\substack{q \le \log x/3 \\ q \in \mathbb{P}}} 1 \right).$$

By the Prime Number Theorem, we have

$$\sum_{\substack{p \le x/3 \\ p \in \mathbb{P}}} 1 \gg \frac{x}{\log x} \quad \text{and} \quad \sum_{\substack{q \le \log x/3 \\ q \in \mathbb{P}}} 1 \gg \frac{\log x}{\log \log x}.$$

Together with

$$\sum_{2^{2^k} \le x/3} 1 \gg \log \log x,$$

this finishes the proof of the lemma. □

Lemma 6 *The following estimate holds:*

$$\sum_{n \le x} r_2(n)^2 \ll x.$$

Proof Again $r_2(n)^2$ counts the number of solutions of the equation

$$p_1 + 2^{2^{k_1}} + 2^{q_1} = p_2 + 2^{2^{k_2}} + 2^{q_2}$$

in p_1, p_2, k_1, k_2, q_1 and q_2 where $p_1 + 2^{2^{k_1}} + 2^{q_1} \leq x$. This means counting pairs of primes (p_1, p_2) such that $p_2 = p_1 + h$, where

$$h := 2^{2^{k_1}} + 2^{q_1} - 2^{2^{k_2}} - 2^{q_2}.$$

If $h = 0$ then either $(k_1, q_1) = (k_2, q_2)$ or w.l.o.g. $k_1 > k_2$ and

$$2^{2^{k_2}} \left(2^{2^{k_1}-2^{k_2}} - 1\right) = 2^{q_1}\left(2^{q_2-q_1} - 1\right).$$

Since $2^{2^{k_1}-2^{k_2}} - 1$ and $2^{q_2-q_1} - 1$ are odd, we have that $2^{k_2} = q_1$ and hence $k_2 = 1$ and $q_1 = 2$. This leads to $2^{k_1} = q_2$ and hence to $k_1 = 1$ and $q_2 = 2$ a contradiction to $k_1 > k_2$. If $h = 0$ we thus have that $(k_1, q_1) = (k_2, q_2)$ and p_2 is fixed by a choice of p_1, k_1 and q_1. The last three parameters may be chosen in $\mathcal{O}(x)$ ways and we can deal with the contribution of solutions of the equation $p_2 = p_1 + h$ where $h \neq 0$. Since h is even, we may directly use the sieve bound from [15, Theorem 7.3] which, after summing over all h, yields an upper bound of order

$$\frac{x}{(\log x)^2} \sideset{}{'}\sum_{(k_1,q_1,k_2,q_2)} \prod_{p|h} \left(1 + \frac{1}{p}\right) \tag{22}$$

for the sum in the lemma, where the dash indicates that $(k_1, q_1) \neq (k_2, q_2)$. Noting that the contribution of the prime 2 is just a constant factor, we disregard it. Furthermore $h \leq x$ by definition, and a very crude upper bound for the number of prime factors of h, in particular for those larger than $\log x$, is given by $\log x/\log 2$. We thus get

$$\begin{aligned}
\sideset{}{'}\sum_{(k_1,q_1,k_2,q_2)} \prod_{\substack{p|h\\p>2}} \left(1 + \frac{1}{p}\right) &\ll \sideset{}{'}\sum_{(k_1,q_1,k_2,q_2)} \underbrace{\left(1 + \frac{1}{\log x}\right)^{\log x/\log 2}}_{\leq e^{1/\log 2}} \prod_{\substack{p|h\\2<p\leq \log x}} \left(1 + \frac{1}{p}\right)\\
&\ll \sideset{}{'}\sum_{(k_1,q_1,k_2,q_2)} \sum_{\substack{d|h\\d \text{ odd}\\P(d)\leq \log x}} \frac{\mu(d)^2}{d}\\
&= \sum_{\substack{d\leq x\\d \text{ odd}\\P(d)\leq \log x}} \frac{\mu(d)^2}{d} \sideset{}{'}\sum_{\substack{(k_1,q_1,k_2,q_2)\\d|h}} 1.
\end{aligned} \tag{23}$$

If we fix k_1, q_1 and k_2, then the fact that $d \mid h$ implies

$$2^{q_2} \equiv 2^{2^{k_1}} + 2^{q_1} - 2^{2^{k_2}} =: l \bmod d,$$

where l is a fixed residue class mod d. This puts q_2 in a fixed residue class mod $t(d)$. Since we are counting representations of integers $n \leq x$, we have $q_2 \leq \log x/\log 2$.

Hence if $t(d) > \log x$ there are at most two choices for q_2. If $t(d) \le \log x$, the Brun–Titchmarsh inequality yields an upper bound of

$$\mathcal{O}\left(\frac{\log x/\log 2}{\varphi(t(d)) \log\left(\log x/t(d) \log 2\right)}\right)$$

for the number of choices of q_2. We thus get an upper bound of the following order for (23)

$$\log x \log\log x \left(\sum_{\substack{d \text{ odd} \\ P(d) \le \log x \\ t(d) \le \log x}} \frac{\mu(d)^2 \left(\log x/\log 2\right)}{d\varphi(t(d)) \log\left(\log x/t(d) \log 2\right)} + \sum_{\substack{d \text{ odd} \\ P(d) \le \log x \\ t(d) > \log x}} \frac{\mu(d)^2}{d} \right). \quad (24)$$

As earlier, by Mertens' formula

$$\sum_{\substack{d \text{ odd} \\ P(d) \le \log x}} \frac{\mu(d)^2}{d} \ll \log\log x.$$

To deal with the first sum in (24), we use $\varphi(m) \gg m/\log\log m$ (see [17, Theorem 15]) and split the range of summation in two parts and get

$$\sum_{\substack{d \text{ odd} \\ P(d) \le \log x \\ t(d) \le \log x}} \frac{\mu(d)^2 \left(\log x/\log 2\right)}{d\varphi(t(d)) \log\left(\log x/t(d) \log 2\right)} \ll \frac{\log x}{\log\log x} \sum_{\substack{d \text{ odd} \\ P(d) \le \log x \\ t(d) \le \sqrt{\log x}}} \frac{\mu(d)^2 \log\log t(d)}{dt(d)}$$
$$+ (\log x)^{3/4} \sum_{\substack{d \text{ odd} \\ P(d) \le \log x \\ \sqrt{\log x} < t(d) \le \log x}} \frac{\mu(d)^2 \log\log t(d)}{d\sqrt{t(d)}}.$$

By a result of Erdős and Turán [7,8], the sums

$$\sum_{d \text{ odd}} \frac{\log\log t(d)}{dt(d)} \text{ and } \sum_{d \text{ odd}} \frac{\log\log t(d)}{d\sqrt{t(d)}}$$

converge which altogether proves an upper bound of order $\mathcal{O}((\log x)^2)$ for (23) and hence an upper bound of order $\mathcal{O}(x)$ for (22). □

Proof of Theorem 5 We prove the theorem by showing that the subset of positive integers in the residue class 3 mod 6 having a representation of the form $p + 2^{2^k} + 2^q$ has density 0.

If $k > 0$, then $2^{2^k} = 4^{2^{k-1}}$. The fact that $4^2 \equiv 4 \bmod 6$ puts the term 2^{2^k} into the residue class 4 mod 6 if $k > 0$. Using the same fact again, we get for $q = 2l + 1$

$$2^q = 2^{2l+1} = 2 \cdot 4^l \equiv 2 \bmod 6.$$

Furthermore, all primes except 2 and 3 are in the residue classes $\{1, 5\} \bmod 6$. Thus if n is in none of the sets

$$\begin{aligned} S_1 &:= \{p + 2 + 2^q : p, q \in \mathbb{P}\}, \\ S_2 &:= \{p + 2^{2^k} + 4 : p \in \mathbb{P}, k \in \mathbb{N}\}, \\ S_3 &:= \{2 + 2^{2^k} + 2^q : k \in \mathbb{N}, q \in \mathbb{P}\}, \\ S_4 &:= \{3 + 2^{2^k} + 2^q : k \in \mathbb{N}, q \in \mathbb{P}\}, \end{aligned}$$

all of which have density 0, and if n has a representation of the form $n = p + 2^{2^k} + 2^q$, then n is in one of the residue classes

$$\{1, 5\} + \{4\} + \{2\} = \{1, 5\} \bmod 6.$$

The set

$$S = \{n \in \mathbb{N} : n \equiv 3 \bmod 6\} \backslash (S_1 \cup S_2 \cup S_3 \cup S_4)$$

has density $1/6$, consists of odd integers only and none of its members is of the form $p + 2^{2^k} + 2^q$. This proves the first part of the Theorem.

To find a full arithmetic progression of integers not of the form $p + 2^{2^k} + 2^q$, we will add additional congruences ruling out the integers in the sets S_1, S_2, S_3 and S_4. We claim that none of the integers n satisfying the congruences

3 mod 6	4 mod 5	4 mod 7
9 mod 13	5 mod 17	8 mod 19
20 mod 23	2 mod 29	3 mod 31
10 mod 37		

is of the form $p + 2^{2^k} + 2^q$. By the above considerations, it suffices to check that none of the integers in the sets S_1, S_2, S_3 and S_4 is contained in this arithmetic progression.

We start with the integers in S_1. Take $n = p + 2 + 2^q \in S_1$ and suppose that n is in the arithmetic progression constructed above. We use that except for $q \in \{2, 3\}$, we have that $q \equiv \{1, 5, 7, 11\} \bmod 12$ and that for any $l \in \mathbb{N}_0$ we have that

$$2^{12l+1} \equiv 2^{12l+5} \equiv 2 \bmod 5,\ 2^{12l+7} \equiv 2 \bmod 7,\ 2^{12l+11} \equiv 7 \bmod 13.$$

If $q \equiv \{1, 5\} \bmod 12$, then $n = p + 2 + 2^q \equiv p + 4 \bmod 5$. Since $n \equiv 4 \bmod 5$, this implies that $p = 5$. Now $7 + 2^{12l+1} \equiv 2 \bmod 7$ and $7 + 2^{12l+5} \equiv 0 \bmod 13$, contradiction to $n \equiv 4 \bmod 7$ and $n \equiv 9 \bmod 13$. In the case of $q = 12l + 7$, we get $n = p + 2 + 2^{12l+7} \equiv p + 4 \bmod 7$ and the only possible choice for p is $p = 7$. Then $9 + 2^{12l+7} \equiv 2 \bmod 5$, a contradiction to $n \equiv 4 \bmod 5$. Finally if $q = 12l + 11$, then $n = p + 2 + 2^{12l+11} \equiv p + 9 \bmod 13$ and from $n \equiv 9 \bmod 13$ we get $p = 13$. Since $n = 15 + 2^{12l+11} \equiv 3 \bmod 5$, we again get a contradiction to $n \equiv 4 \bmod 5$. To finish off the integers in the set S_1, it remains to deal with $q \in \{2, 3\}$. If $q = 2$ we have $n = p + 6 \equiv p \bmod 6$. Since $n \equiv 3 \bmod 6$, we are left with $p = 3$ and $n = 9$ which contradicts to $n \equiv 4 \bmod 7$. If $q = 3$ then $n = p + 10$ and from $n \equiv 10 \bmod 37$, we need to have that $p = 37$, and hence $n = 47$. This is impossible since it contradicts to $n \equiv 4 \bmod 5$.

Next, we deal with the integers in S_2 and we use that $2^{2^k} \equiv 1 \bmod 17$ for $k \geq 3$. Thus, for $k \geq 3$ and $n = p + 2^{2^k} + 4 \in S_2$ we have that $n = p + 2^{2^k} + 4 \equiv p + 5 \bmod 17$. From $n \equiv 5 \bmod 17$, we see that the only admissible choice for p is $p = 17$, and hence, $n = 21 + 2^{2^k}$. As above we use that $2^{2^k} \equiv \{2, 4\} \bmod 6$ and thus $21 + 2^{2^k} \equiv \{1, 5\} \bmod 6$ a contradiction to $n \equiv 3 \bmod 6$. We are left with $k \in \{0, 1, 2\}$. For $k = 0$, we get $n = p+6$ which was ruled out when we dealt with the integers in S_1. If $k = 1$ we have $n = p + 8$ and from $n \equiv 8 \bmod 19$, the only possible choice for p is $p = 19$ and thus $n = 27$. This contradicts to $n \equiv 4 \bmod 5$. Finally, if $k = 2$ we have $n = p + 20$ and from $n \equiv 20 \bmod 23$ we again are left with a single possible choice for p, namely $p = 23$. Now $n = 43$, contradicting to $n \equiv 4 \bmod 5$.

For integers n in the set S_3, we have $n = 2+2^{2^k}+2^q$. If $q = 2$ we have $n \equiv 2^{2^k} \bmod 6$ and again using that $2^{2^k} \in \{2, 4\} \bmod 6$, we get a contradiction to $n \equiv 3 \bmod 6$. If q is odd, then $2^q \equiv 2 \bmod 6$. If furthermore $k = 0$, then $n = 4 + 2^q \equiv 0 \bmod 6$, and if $k = 1$, we get $n = 6 + 2^q \equiv 2 \bmod 6$. In both cases this yields a contradiction to $n \equiv 3 \bmod 6$. For $k \geq 2$ and q odd, we have that $2^{2^k} \equiv \{16, 24, 25\} \bmod 29$ and $2^q \equiv \{2, 3, 8, 10, 11, 12, 14, 15, 17, 18, 19, 21, 26, 27\} \bmod 29$. For $k \geq 2$ and q odd, it is thus true that $2^{2^k} + 2^q \not\equiv 0 \bmod 29$ and thus $n = 2 + 2^{2^k} + 2^q \equiv 2 \bmod 29$ yields a contradiction in this case.

Finally, for integers in the set S_4 we apply a similar argument as for integers in the set S_3. For any prime q we have that $2^q \equiv \{1, 2, 4, 8, 16\} \bmod 31$, and for all $k \in \mathbb{N}_0$ we get $2^{2^k} \equiv \{2, 4, 8, 16\} \bmod 31$. Again $2^{2^k} + 2^q \not\equiv 0 \bmod 31$ for any prime q and any non-negative integer k. Thus, $n = 3+2^{2^k}+2^q \equiv 3 \bmod 31$ yields a contradiction. $\square$

Acknowledgements Open access funding provided by Austrian Science Fund (FWF). Parts of this research work were done when the second author was visiting the Institute of Analysis and Number Theory at Graz University of Technology and the Max Planck Institute for Mathematics in Bonn, Germany. He wants to thank these institutions and Professor Robert Tichy. Furthermore, we would like to thank the referee for carefully reading the manuscript.

References

1. Ballot, C., Luca, F.: On the sumset of the primes and a linear recurrence. Acta Arith. **161**(1), 33–46 (2013)
2. Canfield, E.R., Erdős, P., Pomerance, C.: On a problem of Oppenheim concerning "factorisatio numerorum". J. Number Theory **17**(1), 1–28 (1983)
3. de Polignac, A.: Recherches nouvelles sur les nombres premiers. C. R. Hebd. Séances Acad. Sci. **29**, 397–401 and 738–739 (1849)
4. Dubickas, A.: Sums of primes and quadratic linear recurrence sequences. Acta Math. Sin. (Engl. Ser.) **29**(12), 2251–2260 (2013)
5. Erdős, P.: On integers of the form $2^k + p$ and some related problems. Summa Brasil. Math. **2**, 113–123 (1950)
6. Erdős, P., Murty, M.R.: On the order of $a \pmod p$. CRM Proceedings & Lecture Notes, vol. 19. American Mathematical Society, Providence, RI (1999)
7. Erdős, P., Turán, P.: Ein zahlentheoretischer Satz. Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk **1**, 101–103 (1935)
8. Erdős, P., Turán, P.: Über die Vereinfachung eines Landauschen Satzes. Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk **1**, 144–147 (1935)
9. Euler, L.: Letter to Goldbach, 16.12.1752. http://eulerarchive.maa.org/correspondence/letters/OO0879.pdf
10. Guy, R.K.: Unsolved Problems in Number Theory. Problem Books in Mathematics, 3rd edn. Springer, New York (2004)
11. Habsieger, L., Roblot, X.F.: On integers of the form $p + 2^k$. Acta Arith. **122**(1), 45–50 (2006)
12. Jacobi, C.G.J.: Canon arithmeticus sive tabulae quibus exhibentur pro singulis numeris primis vel primorum potestatibus infra 1000 numeri ad datos indices et indices ad datos numeros pertinentes. Berolinum (1839)
13. Lee, K.S.E.: On the sum of a prime and a Fibonacci number. Int. J. Number Theory **6**(7), 1669–1676 (2010)
14. Moll, V.H.: Numbers and Functions: From a Classical-Experimental Mathematician's Point of View. American Mathematical Society, Providence, RI (2012)
15. Nathanson, M.B.: Additive Number Theory—The Classical Bases. Graduate Texts in Mathematics, vol. 164. Springer, New York (1996)
16. Romanoff, N.P.: Über einige Sätze der additiven Zahlentheorie. Math. Ann. **109**(1), 668–678 (1934)
17. Rosser, J.B., Schoenfeld, L.: Approximate formulas for some functions of prime numbers. Ill. J. Math. **6**, 64–94 (1962)
18. van der Corput, J.G.: Over het vermoeden van de Polignac. Simon Stevin **27**, 99–105 (1950)