

## The Yellowstone Permutation

David L. Applegate  
AT&T  
One AT&T Way  
Bedminster, NJ 07921  
USA

[david@research.att.com](mailto:david@research.att.com)

Hans Havermann  
11 Sykes Ave.  
Weston, ON M9N 1C8  
Canada

[gladhobo@teksavvy.com](mailto:gladhobo@teksavvy.com)

Robert G. Selcoe  
16214 Madewood St.  
Cypress, TX 77429  
USA

[rselcoe@entouchonline.net](mailto:rselcoe@entouchonline.net)

Vladimir Shevelev  
Department of Mathematics  
Ben-Gurion University of the Negev  
Beer-Sheva 84105  
Israel

[shevelev@bgu.ac.il](mailto:shevelev@bgu.ac.il)

N. J. A. Sloane<sup>1</sup>  
The OEIS Foundation Inc.  
11 South Adelaide Ave.  
Highland Park, NJ 08904  
USA

[njasloane@gmail.com](mailto:njasloane@gmail.com)

Reinhard Zumkeller  
Isabellastrasse 13  
D-80798 Munich  
Germany

[reinhard.zumkeller@gmail.com](mailto:reinhard.zumkeller@gmail.com)

March 7, 2015

### Abstract

Define a sequence of positive integers by the rule that  $a(n) = n$  for  $1 \leq n \leq 3$ , and for  $n \geq 4$ ,  $a(n)$  is the smallest number not already in the sequence which has a common factor with  $a(n-2)$

---

<sup>1</sup>To whom correspondence should be addressed.

and is relatively prime to  $a(n-1)$ . We show that this is a permutation of the positive integers. The remarkable graph of this sequence consists of runs of alternating even and odd numbers, interrupted by small downward spikes followed by large upward spikes, suggesting the eruption of geysers in Yellowstone National Park. On a larger scale the points appear to lie on infinitely many distinct curves. There are several unanswered questions concerning the locations of these spikes and the equations for these curves.

## 1 Introduction

Let  $(a(n))_{n \geq 1}$  be defined as in the Abstract. This is sequence [A098550](#)<sup>2</sup> in the *On-Line Encyclopedia of Integer Sequences* [6], contributed by Zumkeller in 2004. Figures 1 and 2 show two different views of its graph, and the first 300 terms are given in Table 1. Figure 1 shows terms  $a(101) = 47$  through  $a(200) = 279$ , with successive points joined by lines. The downward spikes occur when  $a(n)$  is a prime, and the larger upward spikes (the “geysers”, which suggested our name for this sequence) happen two steps later. In the intervals between spikes the sequence alternates between even and odd values in a fairly smooth way.

Figure 2 shows the first 300,000 terms, without lines connecting the points. On this scale the points appear to fall on or close to a number of quite distinct curves. The primes lie on the lowest curve (labeled “p”), and the even terms on the next curve (“E”). The red line is the straight line  $f(x) = x$ , included for reference (it is not part of the graph of the sequence). The heaviest curve (labeled “C”), just above the red line, consists of almost all the odd composite numbers. The higher curves are the relatively sparse “ $\kappa p$ ” points, to be discussed in Section 3 when we study the growth of the sequence more closely. It seems very likely that there are infinitely many curves in the graph, although this, like other properties to be mentioned in Section 3, is at present only a conjecture. We are able to show that every number appears in the sequence, and so  $(a(n))_{n \geq 1}$  is a permutation of the natural numbers.

The definition of this sequence resembles that of the EKG sequence ([A064413](#), [5]), which is that  $b(n) = n$  for  $n = 1$  and 2, and for  $n \geq 3$ ,  $b(n)$  is the smallest number not already in the sequence which has a common factor with  $b(n-1)$ . However, the present sequence seems considerably more complex. (The points of the EKG sequence fall on or near just three curves.) Many other permutations of the natural numbers are discussed in [1, 2, 3, 4].

## 2 Every number appears

**Theorem 1.**  $(a(n))_{n \geq 1}$  is a permutation of the natural numbers.

*Proof.* By definition, there are no repeated terms, so we need only show that every number appears. There are several steps in the argument.

(i) The sequence is certainly infinite, because the term  $pa(n-2)$  is always a candidate for  $a(n)$ , where  $p$  is a prime not dividing any of  $a(1), \dots, a(n-1)$ .

(ii) The set of primes that divide terms of the sequence is infinite. For if not, there is a prime  $p$  such that every term is the product of primes  $< p$ . Using (i), let  $m$  be large enough that  $a(m) > p^2$ , and let  $q$  be a common prime factor of  $a(m-2)$  and  $a(m)$ . Since  $q < p$ ,  $qp < p^2 < a(m)$ , and so  $qp$  would have been a smaller choice for  $a(m)$ , a contradiction.

---

<sup>2</sup>Throughout this article, six-digit numbers prefixed by A refer to entries in [6].

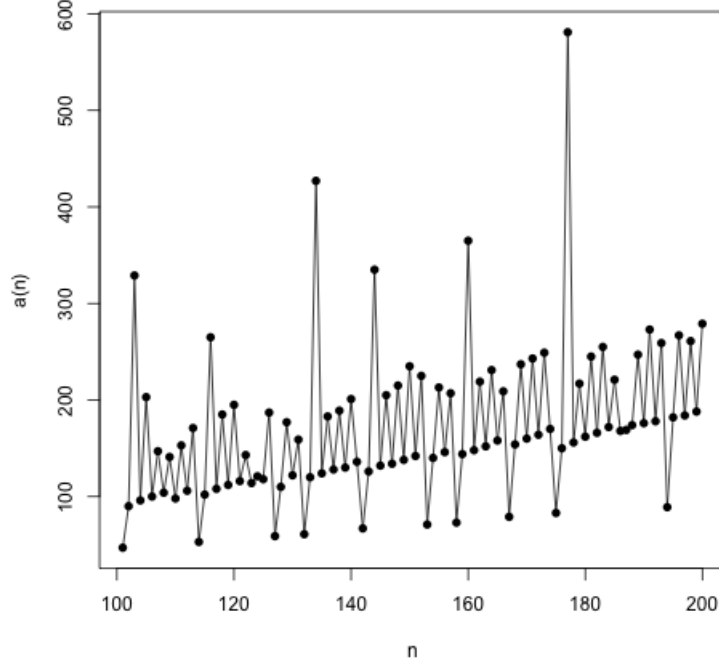


Figure 1: Plot of terms  $a(101)$  through  $a(200)$  of the Yellowstone permutation. The downward spikes occur when  $a(n)$  is a prime, and the larger upward spikes (the “geysers”) happen two steps later.

(iii) For any prime  $p$ , there is a term divisible by  $p$ . For suppose not. Then no prime  $q > p$  can divide any term, for if it did, let  $a(n) = tq$  be the first multiple of  $q$  to appear. But then we could have used  $tp < tq$  instead. So every prime divisor is  $< p$ , contradicting (ii).

(iv) Any prime  $p$  divides infinitely many terms. For suppose not. Let  $N_0$  be such that  $p$  does not divide  $a(n)$  for  $n \geq N_0$ . Choose  $i$  large enough that  $p^i$  does not divide any term in the sequence, and choose a prime  $q > p^i$  which does not divide any of  $a(1), \dots, a(N_0)$ . By (iii), there is some term divisible by  $q$ . Let  $a(m) = tq$  be the first such term. But now  $tp^i < tq$  is a smaller candidate for  $a(m)$ , a contradiction.

(v) For any prime  $p$  there is a term with  $a(n) = p$ . Again suppose not. Using (i), choose  $N_0$  large enough that  $a(n) > p$  for all  $n \geq N_0$ . By (iv), we can find an  $n \geq N_0$  such that  $a(n) = tp$  for some  $t$ . Then  $a(n+2) = p$ , a contradiction.

(vi) All numbers appear. For if not, let  $k$  be the smallest missing number, and choose  $N_0$  so that all of  $1, \dots, k-1$  have occurred in  $a(1), \dots, a(N_0)$ . Let  $p$  be a prime dividing  $k$ . Since, by (iv),  $p$  divides infinitely many terms, there is a number  $N_1 > N_0$  such that  $\gcd(a(N_1), k) > 1$ . This forces

$$\gcd(a(n), k) > 1 \text{ for all } n \geq N_1. \quad (1)$$

(If not, there would be some  $j \geq N_1$  where  $\gcd(a(j), k) > 1$  and  $\gcd(a(j+1), k) = 1$ , which would lead to  $a(j+2) = k$ .) But (1) is impossible, because we know from (v) that infinitely many of the  $a(n)$  are primes.  $\square$

**Remarks.** The same argument, with appropriate modifications, can be applied to many other

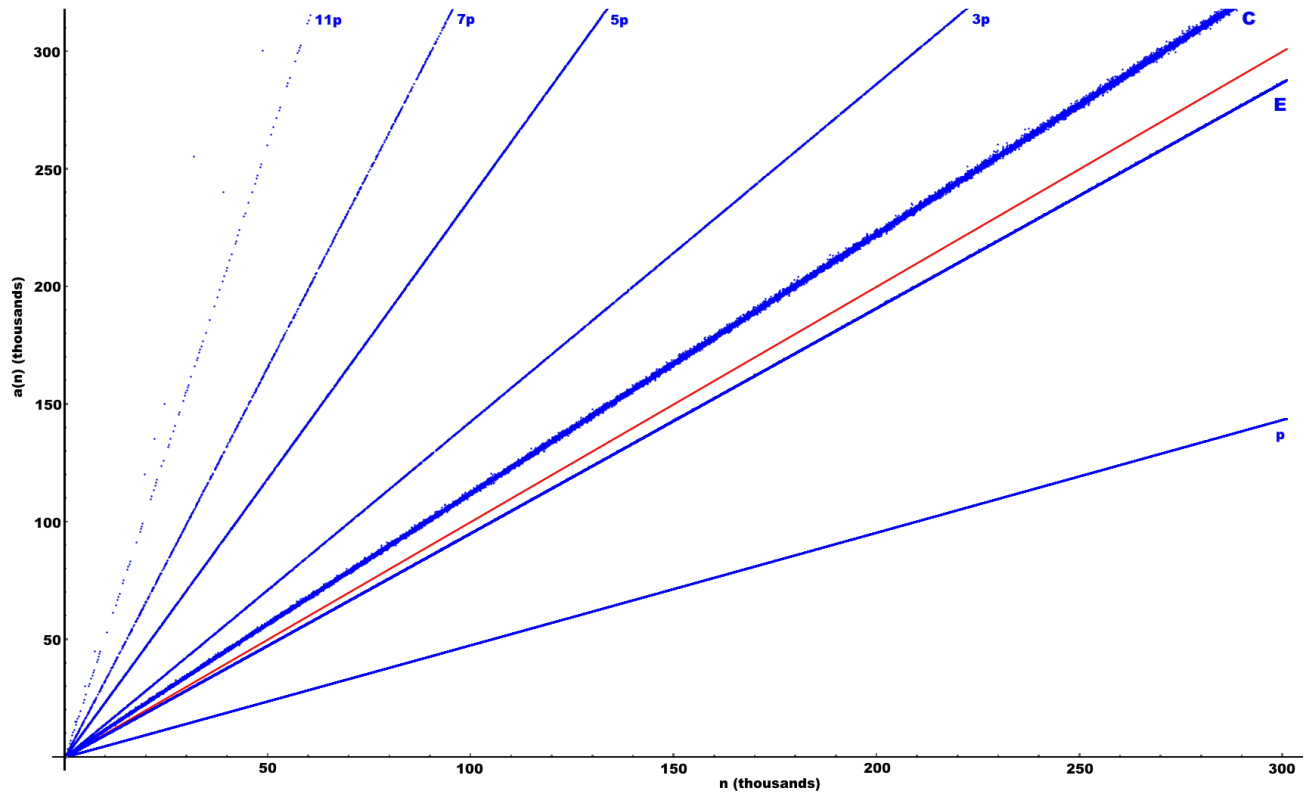


Figure 2: Scatterplot of the first 300,000 terms. The primes lie on the lowest line (labeled “p”), the even numbers on the second line (“E”), the majority of the odd composite numbers on the third line (“C”), and the “ $\kappa p$ ” points on all the higher lines. The lines are not actually straight, except for the red line  $f(x) = x$ , which is included for reference.

sequences. Let  $\Omega$  be a sufficiently large set of positive integers,<sup>3</sup> and define a sequence  $(c(n))_{n \geq 1}$  by specifying that certain members of  $\Omega$  must appear at the start of the sequence (including 1, if  $1 \in \Omega$ ), and that thereafter  $c(n) \in \Omega$  is the smallest number not yet used which satisfies  $\gcd(c(n), c(n-2)) > 1$ ,  $\gcd(c(n-1), c(n)) = 1$ . Then the resulting sequence will be a permutation of  $\Omega$ . We omit the details.

For example, if we take  $\Omega$  to be the odd positive integers, and specify that the sequence begins 1, 3, 5, we obtain [A251413](#).

Or, with  $\Omega$  the positive integers again, we can generalize our main sequence by taking the first three terms to be 1,  $x$ ,  $y$  with  $x > 1$  and  $y > 1$  relatively prime. For example, starting with 1, 3, 2 gives [A251555](#), and starting with 1, 2, 5 gives [A251554](#), neither of which appears to merge with the main sequence, whereas starting with 1, 4, 9 merges with [A098550](#) after five steps. It would be interesting to know more about which of these sequences eventually merge. It follows from Theorem 1 that a necessary and sufficient condition for two sequences  $(c(n))$ ,  $(d(n))$  of this type (that is, beginning 1,  $x$ ,  $y$ ) to merge is that for some  $m$ , terms 1 through  $m-2$  contain the same set of numbers, and  $c(m-1) = d(m-1)$ ,  $c(m) = d(m)$ .

<sup>3</sup>For instance, let  $\mathcal{P}$  be an infinite set of primes, and take  $\Omega$  to consist of the positive numbers all of whose prime factors belong to  $\mathcal{P}$ . We could also exclude any finite subset of numbers from  $\Omega$ . We obtain the Yellowstone permutation by taking  $\mathcal{P}$  to be all primes,  $\Omega$  to be the positive integers, and requiring that the sequence begin with 1, 2, 3.

### 3 Growth of the sequence

Table 1: The first 300 terms  $a(20i+j)$ ,  $0 \leq i \leq 14$ ,  $1 \leq j \leq 20$  of the Yellowstone permutation. The primes (or downward spikes) are shown in red, the “ $\kappa p$ ” points (the upward spikes, or “geysers”) in blue.

| $i \setminus j$ | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0               | 1   | 2   | 3   | 4   | 9   | 8   | 15  | 14  | 5   | 6   | 25  | 12  | 35  | 16  | 7   | 10  | 21  | 20  | 27  | 22  |
| 1               | 39  | 11  | 13  | 33  | 26  | 45  | 28  | 51  | 32  | 17  | 18  | 85  | 24  | 55  | 34  | 65  | 36  | 91  | 30  | 49  |
| 2               | 38  | 63  | 19  | 42  | 95  | 44  | 57  | 40  | 69  | 50  | 23  | 48  | 115 | 52  | 75  | 46  | 81  | 56  | 87  | 62  |
| 3               | 29  | 31  | 58  | 93  | 64  | 99  | 68  | 77  | 54  | 119 | 60  | 133 | 66  | 161 | 72  | 175 | 74  | 105 | 37  | 70  |
| 4               | 111 | 76  | 117 | 80  | 123 | 86  | 41  | 43  | 82  | 129 | 88  | 135 | 92  | 125 | 78  | 145 | 84  | 155 | 94  | 165 |
| 5               | 47  | 90  | 329 | 96  | 203 | 100 | 147 | 104 | 141 | 98  | 153 | 106 | 171 | 53  | 102 | 265 | 108 | 185 | 112 | 195 |
| 6               | 116 | 143 | 114 | 121 | 118 | 187 | 59  | 110 | 177 | 122 | 159 | 61  | 120 | 427 | 124 | 183 | 128 | 189 | 130 | 201 |
| 7               | 136 | 67  | 126 | 335 | 132 | 205 | 134 | 215 | 138 | 235 | 142 | 225 | 71  | 140 | 213 | 146 | 207 | 73  | 144 | 365 |
| 8               | 148 | 219 | 152 | 231 | 158 | 209 | 79  | 154 | 237 | 160 | 243 | 164 | 249 | 170 | 83  | 150 | 581 | 156 | 217 | 162 |
| 9               | 245 | 166 | 255 | 172 | 221 | 168 | 169 | 174 | 247 | 176 | 273 | 178 | 259 | 89  | 182 | 267 | 184 | 261 | 188 | 279 |
| 10              | 190 | 291 | 196 | 97  | 180 | 679 | 186 | 287 | 192 | 301 | 194 | 315 | 202 | 275 | 101 | 198 | 505 | 204 | 295 | 206 |
| 11              | 285 | 103 | 200 | 309 | 208 | 297 | 212 | 253 | 210 | 299 | 214 | 325 | 107 | 220 | 321 | 218 | 303 | 109 | 216 | 545 |
| 12              | 222 | 305 | 224 | 345 | 226 | 327 | 113 | 228 | 565 | 232 | 339 | 230 | 333 | 236 | 351 | 238 | 363 | 244 | 319 | 234 |
| 13              | 341 | 240 | 403 | 242 | 377 | 246 | 455 | 248 | 343 | 250 | 357 | 254 | 289 | 127 | 272 | 381 | 256 | 369 | 260 | 387 |
| 14              | 262 | 375 | 131 | 252 | 655 | 258 | 355 | 264 | 395 | 266 | 405 | 268 | 385 | 274 | 371 | 137 | 280 | 411 | 278 | 393 |

From studying the first 100 million terms of the sequence  $(a(n))$ , we believe we have an accurate model of how the sequence grows. However, at present we have no proofs for any of the following statements. They are merely empirical observations.

The first 212 terms are exceptional (see Table 1). Starting at the 213th term, it appears that the sequence is governed by what we shall call:

**Hypothesis A.** (“A” stands for “alternating”.) The sequence alternates between even and odd composite terms, except that, when an even term is reached which is twice a prime, the alternation of even and odd terms is disrupted, and we see five successive terms of the form

$$2p, 2i + 1, p, 2j, \kappa p, \tag{2}$$

where  $p$  is an odd prime,  $i$  and  $j$  are integers, and  $\kappa < p$  is the least odd prime that does not divide  $j$ . The  $\kappa p$  terms are the “geysers” ([A251544](#)).

For example, terms  $a(213)$  to  $a(217)$  are

$$202 = 2 \cdot 101, 275, 101, 198 = 2 \cdot 3^2 \cdot 11, 505 = 5 \cdot 101.$$

Hypothesis A is only a conjecture, since we cannot rule out the possibility that this behavior breaks down at some much later point in the sequence. It is theoretically possible, for example, that a term that is twice a prime is not followed two steps later by the prime itself (as happens after  $a(8) = 14$ , which is followed two steps later by  $a(10) = 6$  rather than 7). However, as we shall argue later in this section, this is unlikely to happen.

Under Hypothesis A, most of the time the sequence alternates between even and odd composite terms, and the  $n$ th term  $a(n)$  is about  $n$ , to a first approximation. However, the primes appear later than they should, because  $p$  cannot appear until the sequence first reaches  $2p$ , which takes about  $2p$  steps, and so the primes are roughly on the line  $f(x) = x/2$ . On the other hand, the term  $\kappa p$  in (2) appears earlier than it should, and lies roughly on the line  $f(x) = \kappa x/2$ .

Continuing to assume that Hypothesis A holds for terms 213 onwards, we can give heuristic arguments that lead to better asymptotic estimates, as follows. Guided by (2), we divide the terms of the sequence into several types: type  $E$  terms, consisting of all the even terms; type  $p$ , all the odd primes; types  $\kappa p$  for  $\kappa = 3, 5, 7, 11, \dots$ , the terms that appear two steps after a prime; and type  $C$ , all the odd composite terms that are not of type  $\kappa p$  for any  $\kappa$ .

From term 213 onwards, even and odd terms (more precisely, types  $E$  and  $C$ ) alternate, except when the even term is twice a prime, when we see the five-term subsequence (2), containing two  $E$  terms, one  $p$  term, one  $\kappa p$  term for some odd prime  $\kappa$ , and one  $C$  term. Between terms 213 and  $n$ , we will see about  $\lambda$  of these five-term subsequences, where  $\lambda$  is the number of terms in that range that are twice a prime.  $\lambda$  is therefore approximately<sup>4</sup>  $\pi(a(n)/2)$ , where  $\pi(x)$  is the number of primes  $\leq x$ .

There are  $n - 5\lambda$  terms not in the 5-term subsequences, so the total number of even terms out of  $a(1), \dots, a(n)$  is roughly

$$\frac{n - 5\lambda}{2} + 2\lambda = \frac{n - \lambda}{2} \cong \frac{n - \pi(\frac{a(n)}{2})}{2}, \quad (3)$$

where  $\cong$  signifies “is approximately equal to”. Although the even terms do not increase monotonically (compare Table 1), it appears to be a good approximation to assume that, on the average, each even term contributes 2 to the growth of the even subsequence, and so, if  $a(n)$  is an even term, we obtain

$$a(n) \cong n - \pi\left(\frac{a(n)}{2}\right). \quad (4)$$

In other words, the even terms should lie on or close to the curve  $y = f_E(x)$  defined by the functional equation

$$y + \pi\left(\frac{y}{2}\right) = x. \quad (5)$$

The primes then lie on the curve  $f_p(x) = \frac{1}{2}f_E(x)$ , and the  $\kappa p$  terms on the curve  $f_{\kappa p}(x) = \frac{\kappa}{2}f_E(x)$  for  $\kappa = 3, 5, 7, \dots$

Although the reasoning that led us to (5) was far from rigorous, it turns out that (5) is a remarkably good fit to the graph of the even terms, at least for the first  $10^8$  terms. We solved (5) numerically, and computed the residual errors  $a(n) - f_E(n)$ . The fit is very good indeed for the “normal” even terms, those that do not belong to the 5-term subsequences. As can be seen from Fig. 3, up to  $n = 10^7$ , the maximum error is less than 40, in numbers which are around  $10^7$ .

The fit is still good for the even terms in the five-term subsequences, although not so remarkable, as can be seen in Fig. 4. Up to  $n = 10^7$  there are errors as large as 6000, which is on the order of  $\sqrt{n}$ . The errors for the “normal” even terms are shown in this figure in green.

If we use  $\pi(x) \sim x/\log x$  in (5), we obtain

$$f_E(x) = x \left(1 - \frac{1}{2\log x} + o\left(\frac{1}{\log x}\right)\right). \quad (6)$$

However, (5) is a much better fit than just using the first two terms on the right side of (6).

We can study the curve  $f_C(x)$  containing the type  $C$  terms (the odd composite terms not of type  $\kappa p$ ) in a similar manner. This is complicated by the fact that the values of  $\kappa$  are hard to predict. We therefore use a probabilistic model, and let  $\sigma(\kappa)$  denote the probability that the multiplier in

---

<sup>4</sup>We assume  $n$  is very large, and ignore the fact that the first 212 terms are slightly exceptional—asymptotically this makes no difference.

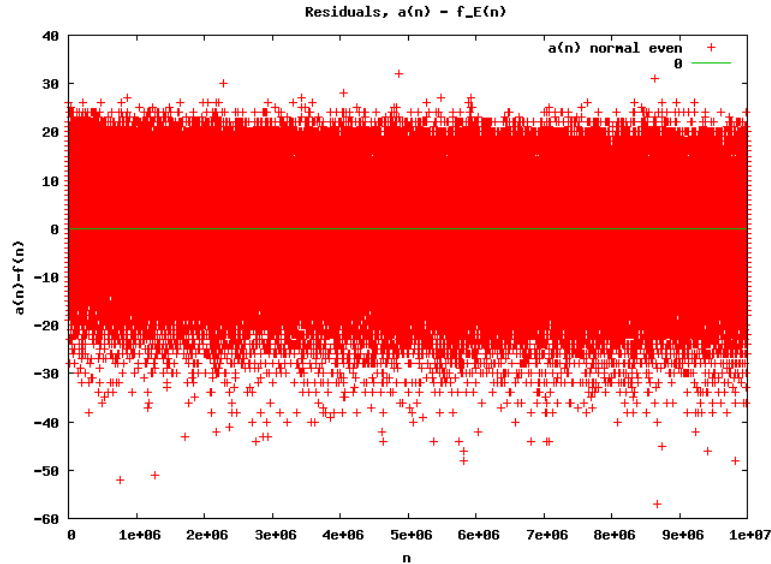


Figure 3: The difference between the “normal” even terms and the approximation  $f_E(n)$  given in (5) is at most 40 for  $n \leq 10^7$ .

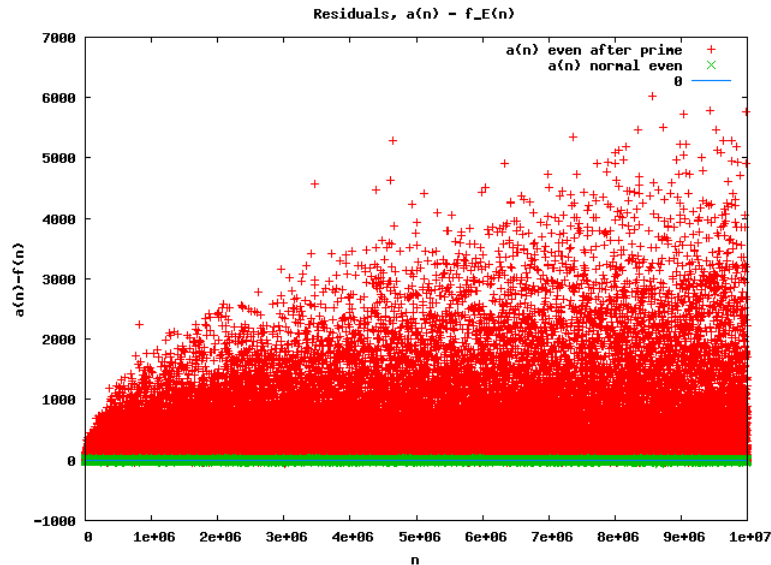


Figure 4: The difference between the even terms that follow a prime and the approximation  $f_E(n)$  given in (5) is at most 6000 for  $n \leq 10^7$ . The green points are the same errors shown in the previous figure, plotted on this scale.

a  $\kappa p$  term is  $\kappa$ . Empirically,  $\sigma(3) \cong 0.334$ ,  $\sigma(5) \cong 0.451$ ,  $\sigma(7) \cong 0.174, \dots$ . The number of type  $C$  terms in the first  $n$  terms is (compare (3))

$$\frac{n - 5\lambda}{2} + \lambda = \frac{n - 3\lambda}{2} \cong \frac{n - 3\pi(\frac{f_E(n)}{2})}{2}. \quad (7)$$

However, type  $C$  terms skip over the primes, and we expect to see  $\pi(f_C(n))$  primes  $\leq f_C(n)$ . Type

$C$  terms also skip over the  $\kappa p$  terms that have already appeared in the sequence. For a given value of  $\kappa$ , terms of type  $\kappa p$  will have been skipped over if  $\kappa p \leq f_C(n)$ , and if that value of  $\kappa$  was chosen, so the number of  $\kappa p$  terms we skip over is

$$\sum_{\text{odd primes } \kappa} \sigma(\kappa) \pi\left(\frac{f_C(n)}{\kappa}\right),$$

where here and in the next two displayed equations the summation ranges over all odd primes  $\kappa \leq \sqrt{f_C(n)}$ . Each of these events contributes 2, on the average, to the growth of the  $C$  terms, so we obtain

$$f_C(n) \cong n - 3\pi\left(\frac{f_E(n)}{2}\right) + 2\pi(f_C(n)) + 2 \sum_{\text{odd primes } \kappa} \sigma(\kappa) \pi\left(\frac{f_C(n)}{\kappa}\right). \quad (8)$$

In other words, the type  $C$  terms should lie on or close to the curve  $y = f_C(x)$  defined by the functional equation

$$y - 2\pi(y) - 2 \sum_{\text{odd primes } \kappa} \sigma(\kappa) \pi\left(\frac{y}{\kappa}\right) = x - 3\pi\left(\frac{f_E(x)}{2}\right). \quad (9)$$

Equation (9) can be solved numerically, using the values of  $f_E(x)$  computed from (5), and gives a good fit to the graph of the type  $C$  terms. As can be seen from Fig. 5, the errors in the first  $10^7$  terms are on the order of  $5\sqrt{n}$ . It is not surprising that the errors are larger for type  $C$  terms than type  $E$  terms, since as can be seen in Fig. 2, the curve with the  $C$  points is much thicker than the  $E$  curve.

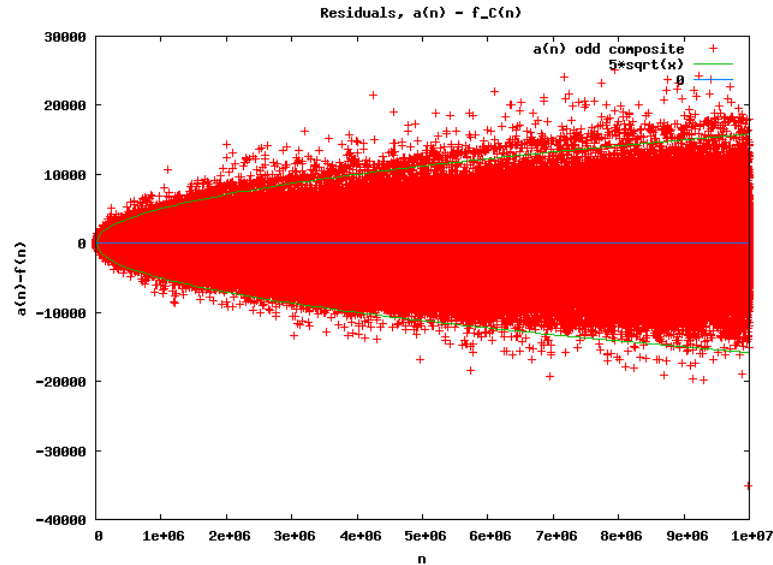


Figure 5: The difference between the odd composite (or type  $C$ ) terms and the approximation  $f_C(n)$  defined implicitly by (9) is on the order of  $5\sqrt{n}$  for  $n \leq 10^7$ .

If we use  $\pi(x) \sim x/\log x$  in (9), we obtain

$$f_C(x) = x \left( 1 + \frac{\alpha}{\log x} + o\left(\frac{1}{\log x}\right) \right), \quad (10)$$



where

$$\alpha = \frac{1}{2} + 2 \sum_{\text{odd primes } \kappa \geq 3} \frac{\sigma(\kappa)}{\kappa} \cong 0.96, \quad (11)$$

and now the summation is over all odd primes  $\kappa$ .

To summarize, our estimates for the curves  $f_E(x)$  and  $f_C(x)$  containing the terms of types  $E$  and  $C$  are given by (5) and (9). Equations (6) and (10) have a simpler form but are less precise. The primes lie on the curve  $f_p(x) = \frac{1}{2}f_E(x)$ , and the  $\kappa p$  terms on the curves  $f_{\kappa p}(x) = \frac{\kappa}{2}f_E(x)$  for  $\kappa = 3, 5, 7, \dots$ . In Fig. 2, reading counterclockwise from the horizontal axis, we see the curves  $f_p(x)$ ,  $f_E(x)$ , the red line  $f(x) = x$ , then  $f_C(x)$ ,  $f_{3p}(x)$ ,  $f_{5p}(x)$ ,  $f_{7p}(x)$ ,  $f_{11p}(x)$ , and a few points from  $f_{\kappa p}(x)$  for  $\kappa \geq 13$ . At this scale, the curves look straight.

To see why Hypothesis A is unlikely to fail, note that when we add an even number to the sequence, most of the time it belongs to the interval  $[m_E, M_E]$  (A251546, A251557), which we call the *even frontier*, where  $m_E$  is the smallest even number that is not yet in the sequence, and  $M_E$  is 2 more than the largest even number that has appeared. The *odd composite frontier*  $[m_C, M_C]$  (A251558, A251559) is defined similarly for the type  $C$  points. For example, when  $n = 10^6$ ,  $a(10^6) = 1094537$ , the even frontier is  $[960004, \dots, 960234]$  and the odd composite frontier is  $[1092467, \dots, 1097887]$ . In fact, at this point, no even number in the range  $960004, \dots, 960230$  is in the sequence. What we see here is typical of the general situation: the length of the even frontier,  $M_E - m_E$ , is much less than the length of the odd composite frontier,  $M_C - m_C$ ; most of the terms in the even frontier are available; and the two frontiers are well separated. As long as this continues, the even and odd frontiers will remain separated, and Hypothesis A will hold. The much larger width of the odd composite frontier is reflected in the greater thickness of the ‘‘C’’ curve in Figure 2.

We know from the proof of Theorem 1 that if  $p < q$  are primes, the first term divisible by  $p$  occurs before the first term divisible by  $q$ . But we do not know that  $p$  itself occurs before  $q$ . This would be a consequence of Hypothesis A, but perhaps it can be proved by arguments similar to those used to prove Theorem 1. Sequences A252837 and A252838 contain additional information related to Hypothesis A.

The OEIS contains a number of other sequences (e.g., A098548, A249167, A251604, A251756, A252868) whose definition has a similar flavor to that of the Yellowstone permutation. Two sequences contributed by Adams-Watters are especially noteworthy: A252865 is an analog of A098550 for square-free numbers, and A252867 is a set-theoretic version.

## 4 Orbits under the permutation

Since the sequence is a permutation of the positive integers, it is natural to study its orbits. It appears that the only fixed points are 1, 2, 3, 4, 12, 50, 86 (A251411). There are certainly no other fixed points below  $10^9$ , and Fig. 2 makes it very plausible that there are no further points on the red line.

At present, 27 finite cycles are known besides the seven fixed points. For example, 6 is in the cycle (6, 8, 14, 16, 10). The finite cycle with the largest minimum term known to date is the cycle of length 45 containing 756023506.

We conjecture that, on the other hand, almost all positive numbers belong to infinite orbits. See Fig. 6 for portions of the conjecturally infinite orbits whose smallest terms are respectively 11 (A251412), 29, 36, 66, and 98 (cf. A251556). The orbits have been displaced sideways so that the conjectured minimal value is positioned at  $x = 0$ . For example, the blue curve to the right of  $x = 0$

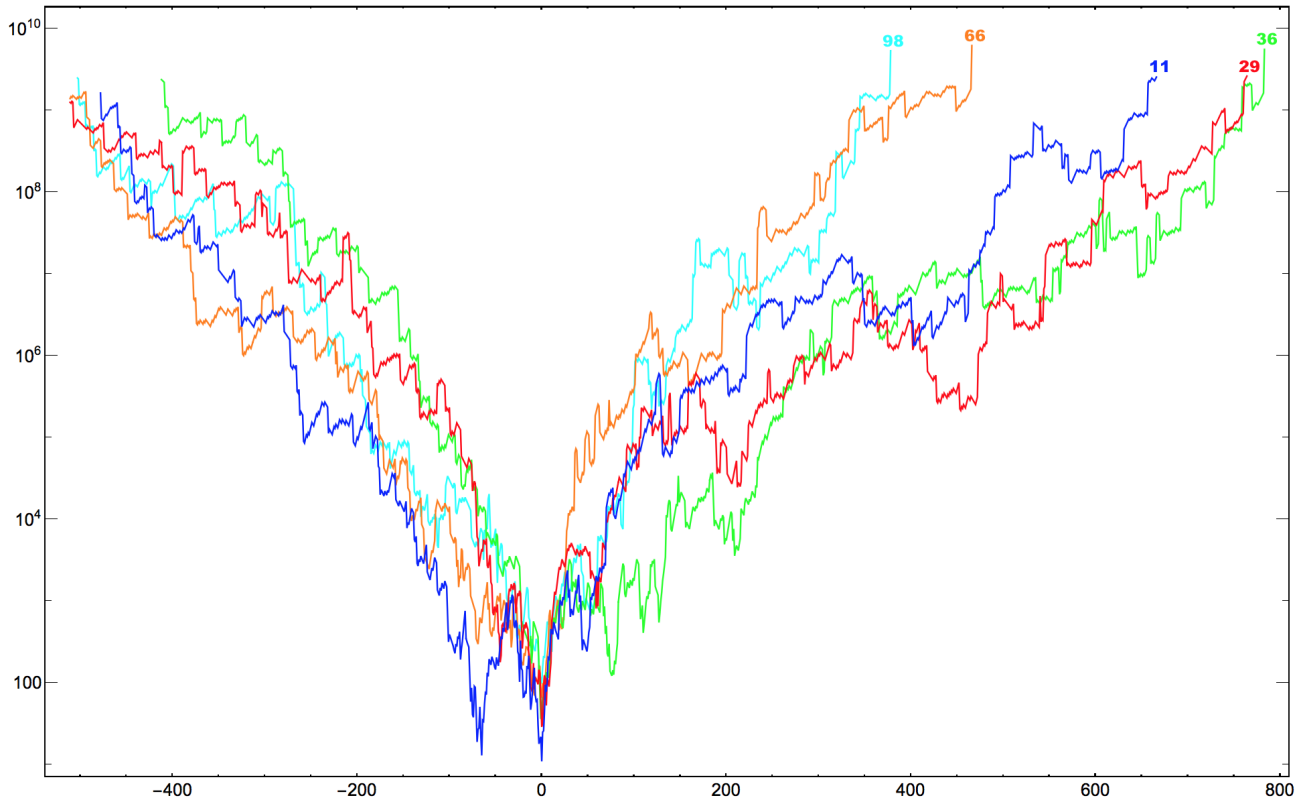


Figure 6: Portions of the conjecturally infinite orbits whose smallest terms are 11 (blue), 29 (red), 36 (green), 66 (orange), 98 (cyan).

shows the initial portion of the trajectory of 11 under repeated applications of the Yellowstone permutation, while the curve to the left shows the trajectory under repeated applications of the inverse permutation. In other words, the blue curve, from upper left to upper right, is a section of the orbit whose minimal value appears to be 11. The inverse trajectory of 11 has near-misses after three steps, when it reaches 18, and after 70 steps, when it reaches 19, but once the numbers get large it seems that there is little chance that the forward and inverse trajectories will ever meet, implying that the orbit is infinite.

However, because of the erratic appearance of the trajectories in Fig. 6, there is perhaps a greater possibility that these paths may eventually close, or merge, compared with the situation for other well-known permutations. For example, in the case of the “amusical permutation” of the nonnegative integers ([A006368](#)) studied by Conway and Guy [1, 2, 4], the empirical evidence that most orbits are infinite is much stronger—compare Figs. 1 and 2 of [2] with our Fig. 6.

## 5 Acknowledgments

We would like to thank Bradley Klee for the key idea of considering terms of the form  $p^i$  in part (iv) of the proof of Theorem 1.

Sequence [A098550](#) was the subject of many discussions on the *Sequence Fans Mailing List* in 2014. Franklin T. Adams-Watters observed that when the first 10000 terms are plotted, the slopes of the various lines in the graph were, surprisingly, not recognizable as rational numbers. Jon E.

Schoenfield noticed that if the third (or “C”) line was ignored, the ratios of the other slopes were consecutive primes. See §3 for our conjectured explanation of these observations. L. Edson Jeffery observed that there appear to be only seven fixed points (see §4).

Other contributors to the discussion include Maximilian F. Hasler, Robert Israel, Benoît Jubin, Jon Perry, Robert G. Wilson v, and Chai Wah Wu.

We thank several of these people, especially Benoît Jubin and Jon E. Schoenfield, for sending comments on the manuscript.

## References

- [1] J. H. Conway, Unpredictable iterations, *Proceedings of the Number Theory Conference (Univ. Colorado, Boulder, CO, 1972)*, pp. 49–52.
- [2] J. H. Conway, On unsettleable arithmetical problems, *Amer. Math. Monthly*, **120** (2013), 192–198.
- [3] P. Erdős, R. Freud, and N. Hegyvári, Arithmetical properties of permutations of integers, *Acta Math. Hungar.*, **41** (1983), 169–176.
- [4] R. K. Guy, *Unsolved Problems in Number Theory*, Springer, 2nd. ed., 1994. See §E17.
- [5] J. C. Lagarias, E. M. Rains, and N. J. A. Sloane, The EKG Sequence, *Experimental Math.*, **11** (2002), 437–446.
- [6] The OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, <http://oeis.org>.

---

2010 *Mathematics Subject Classification*: Primary 11Bxx, 11B83, 11B75.

*Keywords*: Number sequence, EKG sequence, permutation of natural numbers.

---

(Concerned with sequences [A006368](#), [A064413](#), [A098548](#), [A098550](#), [A249167](#), [A249943](#), [A251237](#), [A251411–A251413](#), [A251542–A251547](#), [A251554–A251559](#), [A251604](#), [A251621](#), [A251756](#), [A252837](#), [A252838](#), [A252865](#), [A252867](#), [A252868](#), [A253048](#), [A253049](#). The OEIS entry [A098550](#) contains cross-references to many other related sequences.)

---