The Curling Number Conjecture

Benjamin Chaffin,
Intel Processor Architecture,
2111 NE 25th Avenue,
Hillsboro, OR 97124, USA,
Email: chaffin@gmail.com
and
N. J. A. Sloane[(a)],
AT&T Shannon Labs,
180 Park Avenue,
Florham Park, NJ 07932-0971, USA,
Email: njas@research.att.com

[(a)] Corresponding author.

December 11, 2009; revised February 13, 2010

Abstract

Given a finite nonempty sequence of integers S, by grouping adjacent terms it is always possible to write it, possibly in many ways, as $S = X\,Y^k$, where X and Y are sequences and Y is nonempty. Choose the version which maximizes the value of k: this k is the *curling number* of S. The *curling number conjecture* is that if one starts with *any* initial sequence S, and extends it by repeatedly appending the curling number of the current sequence, the sequence will eventually reach 1. The conjecture remains open, but we will report on some numerical results in the case when S consists of only 2's and 3's.

Keywords: curling number, Gijswijt sequence, recurrences, conjectures
AMS 2000 Classification: Primary 11B37

1. The curling number conjecture

Let S be a finite nonempty sequence of integers. By grouping adjacent terms, it is always possible to write it as $S = X\,Y\,Y\,\cdots\,Y = X\,Y^k$, where X and Y are sequences of integers and Y is nonempty. There may be several ways to do this: choose the one that maximizes the value of k: this k is the *curling number* of S, denoted by $k(S)$.

For example, if $S = 0\,1\,2\,2\,1\,2\,2\,1\,2\,2$, we could write it as $X\,Y^2$, where $X = 0\,1\,2\,2\,1\,2\,2\,1$ and $Y = 2$, or as $X\,Y^3$, where $X = 0$ and $Y = 1\,2\,2$. The latter representation is to be preferred, since it has $k = 3$, and as $k = 4$ is impossible, the curling number of this S is 3.

The *curling number conjecture* is that if one starts with any initial sequence S, and extends it by repeatedly appending the curling number of the current sequence, the sequence will eventually reach 1. In other words, if $S_0 = S$ is any finite nonempty sequence of integers, and we define

$$S_{n+1} = S_n\ k(S_n) \ \text{ for } n \geq 0\,, \tag{1}$$

then the conjecture is that for some $m \geq 0$ we will have $k(S_m) = 1$.

For example, suppose we start with $S_0 = 2\,3\,2\,3$. By taking $X = \emptyset$, $Y = 2\,3$, we have $S_0 = Y^2$, so $k(S_0) = 2$, and we get $S_1 = 2\,3\,2\,3\,2$. By taking $X = 2$, $Y = 3\,2$ we get $k(S_1) = 2$, $S_2 = 2\,3\,2\,3\,2\,2$. By taking $X = 2\,3\,2\,3$, $Y = 2$ we get $k(S_2) = 2$, $S_3 = 2\,3\,2\,3\,2\,2\,2$. Again taking $X = 2\,3\,2\,3$, $Y = 2$

we get $k(S_3) = 3$, $S_4 = 2\,3\,2\,3\,2\,2\,2\,3$. Now, unfortunately, it is impossible to write $S_4 = X\,Y^k$ with $k > 1$, so $k(S_4) = 1$, $S_5 = 2\,3\,2\,3\,2\,2\,2\,3\,1$, and we have reached a 1, as predicted by the conjecture. (If we continue the sequence from this point, it appears to join Gijswijt's sequence, discussed in Section 3.)

The conjecture is stated in [1], [3], [4], and is mentioned in several entries in [2]. Some of the proofs in [1] could be shortened if the conjecture were known to be true. All the available evidence suggests that the conjecture is true, but it has so far resisted all attempts to prove it.

Notation. Y^k means $Y\,Y\,\cdots\,Y$, where Y is repeated k times. $\emptyset$ denotes the empty sequence. We usually separate the parts of a sequence by small spaces.

2. Initial sequences of 2's and 3's

One way to approach the problem is to consider initial sequences S_0 that contain only 2's and 3's, and to see how far such a sequence will extend using the rule (1) before reaching a 1.

Let $\mu(n)$ denote the maximal length that can be achieved before a 1 appears, for any starting sequence consisting of n 2's and 3's. The Curling Number Conjecture implies that $\mu(n) < \infty$ for all n. Reference [1] gave $\mu(n)$ for $1 \le n \le 30$. Since then we have computed the values of $\mu(n)$ for all $n \le 52$, and we have lower bounds, which are probably the exact values, for $53 \le n \le 80$. The results are shown in Table 1 and Figure 1. The values of $\mu(n)$ also form sequence A094004 in [2].

n	1	2	3	4	5	6	7	8	9	10	11	12
$\mu(n)$	1	4	5	8	9	14	15	66	68	70	123	124
n	13	14	15	16	17	18	19	20	21	22	23	24
$\mu(n)$	125	132	133	134	135	136	138	139	140	142	143	144
n	25	26	27	28	29	30	31	32	33	34	35	36
$\mu(n)$	145	146	147	148	149	150	151	152	153	154	155	156
n	37	38	39	40	41	42	43	44	45	46	47	48
$\mu(n)$	157	158	159	160	161	162	163	164	165	166	167	179
n	49	50	51	52	53	54	55	56	57	58	59	60
$\mu(n)$	180	181	182	183	$184^\dagger$	$185^\dagger$	$186^\dagger$	$187^\dagger$	$188^\dagger$	$189^\dagger$	$190^\dagger$	$191^\dagger$
n	61	62	63	64	65	66	67	68	69	70	71	72
$\mu(n)$	$192^\dagger$	$193^\dagger$	$194^\dagger$	$195^\dagger$	$196^\dagger$	$197^\dagger$	$198^\dagger$	$200^\dagger$	$201^\dagger$	$202^\dagger$	$203^\dagger$	$204^\dagger$
n	73	74	75	76	77	78	79	80				
$\mu(n)$	$205^\dagger$	$206^\dagger$	$207^\dagger$	$209^\dagger$	$250^\dagger$	$251^\dagger$	$252^\dagger$	$253^\dagger$				

Table 1: $\mu(n)$, the record for a starting sequence of n 2's and 3's. Entries marked with a dagger (†) are only lower bounds but are conjectured to be exact.

As can be seen from Fig. 1, up to $n = 52$, $\mu(n)$ increases in a piecewise linear manner. At the values $n = 1, 2, 4, 6, 8, 9, 10, 11, 14, 19, 22, 48$ (entry A160766 in [2]) there is a jump, but at the other values of n up to 52, $\mu(n)$ is simply $\mu(n-1)+1$. Table 2 gives the starting sequences where $\mu(n) > \mu(n-1)+1$. Up to $n = 52$, such sequences are always unique and start with a 2 (except for $n = 1$).

Note from Table 1 that

$$\mu(n) = n + 120 \text{ for } 22 \le n \le 47\,. \qquad (2)$$

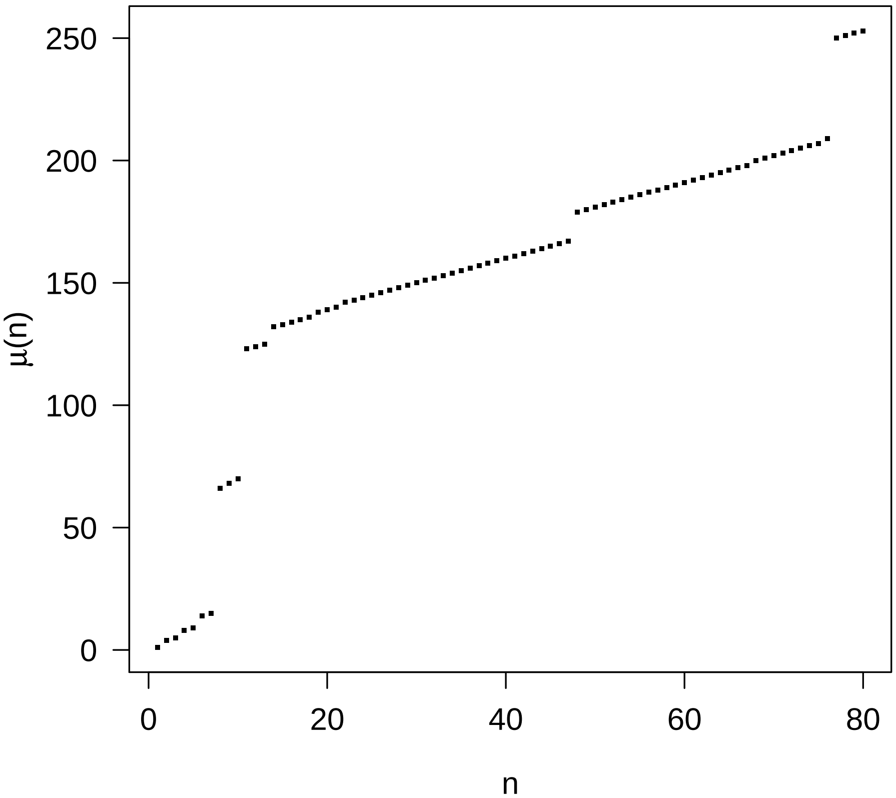

Figure 1: Known values of $\mu(n)$, the maximal length of a sequence produced by any starting sequence of n 2's and 3's, stopping when first 1 is reached. The values for $53 \le n \le 80$ are only lower bounds, but are conjectured to be exact.

That is, in this range one cannot do any better than taking the starting sequence for $n = 22$ and prefixing it by an irrelevant sequence of $47 - n$ 2's and 3's. However, at $n = 48$ a new record-holder appears, and we have

$$\mu(n) = n + 131 \text{ for } 48 \le n \le 52\,. \tag{3}$$

The main reason we were able to extend the search as far as $n = 52$ is the following lemma.

Lemma 1. *If $\mu(n) > \mu(n-1) + 1$ then no starting sequence of n 2's and 3's that achieves $\mu(n)$ can contain a subsequence W^4 for a nonempty sequence W.*

Proof. Let S be a starting sequence of length n that achieves $\mu(n) > \mu(n-1) + 1$, say $S = s_1\, s_2\, \ldots\, s_n$, and let $T = s_1\, s_2\, \ldots\, s_n\, \ldots\, s_{\mu(n)}$ (with $s_i \in \{2, 3\}$ for $1 \le i \le \mu(n)$, $s_{\mu(n)+1} \notin \{2, 3\}$) be its extension. Call S "weak" if s_1 is never used in T, that is, if $s_1\, s_2\, \ldots\, s_r$ can be written as $XY^{s_{r+1}}, X \ne \emptyset, Y \ne \emptyset$, for all r satisfying $n \le r \le \mu(n) - 1$. If S is weak, then $\mu(n) \le \mu(n-1) + 1$, since T is simply s_1 followed by the extension of $s_2\, \ldots\, s_n$. Hence S is not weak, and so for some r with $n \le r \le \mu(n) - 1$ we have $s_1\, s_2\, \ldots\, s_r = Y^k$, where $k = s_{r+1} \ge 2$ and Y is a sequence that

n	Starting sequence
1	2
2	2 2
4	2 3 2 3
6	2 2 2 3 2 2
8	2 3 2 2 2 3 2 3
9	2 2 3 2 2 2 3 2 3
10	2 3 2 3 2 2 2 3 2 2
11	2 2 3 2 3 2 2 2 3 2 2
14	2 2 3 2 3 2 2 2 3 2 2 3 2 3
19	2 2 3 2 2 3 2 3 2 2 2 3 2 2 3 2 2 3 2
22	2 3 2 2 3 2 2 3 2 3 2 2 2 3 2 3 2 2 3 2 2 3
48	2 2 3 2 2 3 2 3 2 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3 2 3 2 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3 2 3

Table 2: Starting sequences of n 2's and 3's for which $\mu(n) > \mu(n-1)+1$, complete for $1 \le n \le 52$.

begins with S. If S were to contain a sequence of the form W^4, we would have a contradiction unless W^4 were at the end of Y, for when growing the second copy of Y, after W^4 has appeared, the next term is ≥ 4, whereas S only contains 2's and 3's. On the other hand, if W^4 is at the end of Y, the next term is ≥ 4, so $k = 1$, again a contradiction. ∎

We also made use of a number of more obvious shortcuts, such as not considering a starting sequence $s_1 \dots s_n$ if $k(s_1 \dots s_{n-1}) = s_n$, since we may assume that we have already considered all starting sequences of length $n-1$.

The lemma cuts down the number of starting sequences of 2's and 3's that must be considered. Even if we simply exclude sequences that contain four consecutive 2's or four consecutive 3's, the number of length n (see entry A135491 in [2]) drops from 2^n to constant$\cdot\alpha^n$, where $\alpha = 1.839\dots$.

Inspection of the best starting sequences in Table 2 suggests they must satisfy another condition, which however we have not been able to prove: namely that they do not contain two consecutive 3's. This is true for all the best starting sequences of lengths $n \le 52$.

Making the assumption (as yet unjustified) that we need only consider starting sequences with at most three consecutive 2's and with no pair of adjacent 3's, we were able extend the search to $n = 78$. This produced three further jumps, at $n = 68$, 76 and 77, establishing that $\mu(n) \ge n+132$ for $68 \le n \le 75$, $\mu(n) \ge n+133$ for $n = 76$ and $\mu(n) \ge n+173$ for $77 \le n \le 80$. The corresponding starting sequences are shown in Table 3.

n	Starting sequence
68	2 2 3 2 2 3 2 2 2 3 2 2 3 2 3 2 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3 2 3 2 2 2 3 2 2
	2 3 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3 2
76	2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3 2 3 2 2 2 3 2 2 2 3 2 2 3 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3 2 3
	2 2 2 3 2 2 2 3 2 2 3 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 3
77	2 2 3 2 2 2 3 2 3 2 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 2 3 2 3 2 2 2 3 2 2 2 3 2 2 3 2 2 2 3 2 2 2
	3 2 3 2 2 2 3 2 2 2 3 2 2 3 2 3 2 2 2 3 2 2 3 2 2 2 3 2 3

Table 3: Starting sequences of n 2's and 3's for which $\mu(n) > \mu(n-1)+1$, conjectured to be complete for $53 \le n \le 80$.

We have not succeeded in finding any algebraic constructions for good starting sequences.

However, one simple construction enables us to obtain lower bounds on $\mu(n)$ for some larger values of n. Let S be a sequence of length n that achieves $\mu(n)$, and let T be the sequence of length $\mu(n)$ that it generates (up to just before the first 1 appears). Then in some cases the starting sequence $T\,S$ will extend to $T\,T\,2$ and beyond before reaching a 1. For example, taking S to be the length-48 sequence in Table 2, the sequence $T\,S$ has length $179 + 48 = 227$ and extends to a total length of 596 before reaching a 1, showing that $\mu(227) \geq 596$.

It would be nice to have some further exact values of $\mu(n)$, even though they will require extensive computations. Can the especially good starting sequences shown in Tables 2 and 3 (especially at lengths 22, 48 and 77) be generalized? What makes them so special? And above all, what is the asymptotic behavior of $\mu(n)$?

3. Gijswijt's sequence

If we simply start with $S_0 = 1$, and generate an infinite sequence by continually appending the curling number of the current sequence, as in (1), we obtain

$$1, 1, 2, 1, 1, 2, 2, 2, 3, 1, 1, 2, 1, 1, 2, 2, 2, 3, 2, 1, 1, 2, 1, 1, 2, 2, 2, 3, 1, 1, 2, 1, 1, \ldots .$$

This is *Gijswijt's sequence*, invented by Dion Gijswijt when he was a graduate student at the University of Amsterdam, and analyzed in [1]. It is entry A090822 in [2].

The first time a 4 appears is at term 220. One can calculate quite a few million terms without finding a 5 (as the authors of [1] discovered!), but in [1] it was shown that a 5 eventually appears for the first time at about term

$$10^{10^{23}} .$$

Reference [1] also shows that the sequence is in fact unbounded, and conjectures that the first time that a number $m \geq 6$ appears is at about term number

$$2^{2^{3^{4^{\cdot^{\cdot^{\cdot^{m-1}}}}}}} ,$$

a tower of height $m - 1$.

There is another question, also still open, which relates Gijswijt's sequence to the discussion in the previous section. If we start with an initial sequence S of 2's and 3's, extend it until we reach the first 1, say at the $(k+1)$st step, and then keep going, it *appears* that the result is always simply the first k terms of the extension of S, followed by Gijswijt's sequence. In other words, there is never any interaction between the first k terms of the extension of S and an initial segment of Gijswijt's sequence when computing curling numbers after the kth step. This seems plausible, but we would not be surprised if there was a counterexample. It would be nice to have this question settled one way or the other.

One final remark: To avoid 1's in the sequence, we might define $h(S) = \max\{k(S), 2\}$, and replace the recurrence (1) by

$$S_{n+1} = S_n \; h(S_n) \quad \text{for } n \geq 0 , \tag{4}$$

If we start with $S_0 = 1$ and use the rule (4) to extend it, the resulting sequence (A091787) is again unbounded, and now it is possible to compute exactly when the first 5 appears, which is at step

$$777094043884153701608292469323456921 80 .$$

See [1] for further information.

References

[1] F. J. van der Bult, D. C. Gijswijt, J. P. Linderman, N. J. A. Sloane and A. R. Wilks, A slow-growing sequence defined by an unusual recurrence, *J. Integer Sequences*, **10** (2007), #07.1.2 [arXiv:math.NT/0602498].

[2] N. J. A. Sloane, editor, *The On-Line Encyclopedia of Integer Sequences*, published electronically at www.research.att.com/∼njas/sequences/, 1996–2009, and at oeis.org from 2010 onwards.

[3] N. J. A. Sloane, *Seven staggering sequences*, in *Homage to a Pied Puzzler*, E. Pegg Jr., A. H. Schoen and T. Rodgers (editors), A. K. Peters, Wellesley, MA, 2009, pp. 93-110 [arXiv:0912.2394].

[4] N. J. A. Sloane, *Eight hateful sequences*, to appear, 2010 [arXiv:0805.2128].