Bit-Position Dynamics and a Lower Bound for Collatz Cycle Length

Jochen Kiemes

April 21, 2025

Abstract

We present a novel reformulation of the Collatz conjecture by leveraging the binary structure of positive integers, focusing on the sequence of odd terms. Through an analysis of leading and trailing bit-position dynamics, we derive a substantial lower bound of at least 17,026,679,261 steps for any hypothetical non-trivial cycle, offering new insights into its structural constraints.

1 Introduction

The Collatz conjecture, also known as the $3n+1$ problem, asserts a universal property of sequences generated by a simple iterative rule applied to positive integers. For any $n \in \mathbb{N}$, define the sequence $\{c_i\}$ with $c_0 = n$ and

$$c_{i+1} = \begin{cases} \frac{c_i}{2} & \text{if } c_i \equiv 0 \pmod 2, \\ 3c_i + 1 & \text{if } c_i \equiv 1 \pmod 2. \end{cases} \tag{1}$$

The conjecture posits that for every n, there exists a finite k such that $c_k = 1$, after which the sequence cycles as $1, 4, 2, 1$. Despite its elementary formulation, the conjecture remains open, resisting resolution despite extensive study via diverse approaches. There have been extensive work about the lower bound for the minimal cycle length of an existing non-trivial cycle [Eliahou, 1993], [Simons and de Weger, 2010] and [Hercher, 2018].

This paper uses a reformulation that shifts focus to the binary structure of the sequence terms as introduced in [Kiemes, 2025]. We define a sequence $\{a_i\}$ based on odd parts, where a_0 is the odd part of n, and each subsequent term is computed as $a_{i+1} = 3a_i + 2^{t(a_i)}$, with $t(a_i)$ being the position of the least significant 1-bit in a_i. This allows to write a closed formula for the elements a_i. Based on this closed form, we derive a lower bound for the cycle length of any non-trivial cycle in the Collatz sequence.

2 Preliminaries

To analyze the Collatz conjecture through its binary structure, we define a sequence based on odd parts and their bit positions. This section introduces the essential notation and concepts used in our reformulation, with additional definitions provided as needed in later sections.

2.1 Binary Representation and Odd Part

For a positive integer $n \in \mathbb{N}$, let its binary form be $n = \sum_{i=0}^{k} b_i 2^i$, with $b_i \in \{0, 1\}$ and $b_k = 1$. Define the *trailing bit position* $t(n) = \min\{i \geq 0 : b_i = 1\}$, the index of the least significant 1-bit. The *odd part* of n is:

$$m(n) = \frac{n}{2^{t(n)}}, \tag{2}$$

which is odd and satisfies $m(n) = m(n \cdot 2^i)$ for $i \in \mathbb{N}_0$. For example, if $n = 40 = 101000_2$, then $t(n) = 3$, and $m(n) = 40/2^3 = 5$. This function captures the number of trailing zeros in the binary representation of n, a key element in our sequence's dynamics.

2.2 Leading Bit Position

The leading bit position $l(n)$ corresponds to the highest 1-bit in n's binary form, defined as:

$$l(n) = \lfloor \log_2 n \rfloor. \tag{3}$$

This satisfies $2^{l(n)} \leq n < 2^{l(n)+1}$. For $n = 40$, $l(n) = \lfloor \log_2 40 \rfloor \approx \lfloor 5.322 \rfloor = 5$, since $2^5 = 32 \leq 40 < 64 = 2^6$.

3 Reformulation of the Collatz Conjecture

We propose a reformulation of the Collatz conjecture that highlights the odd terms and their binary structure, focusing on transitions between odd numbers to simplify the sequence's dynamics. Recall the Collatz sequence $\{c_i\}$ with $c_0 = n$ and

$$c_{i+1} = \begin{cases} \dfrac{c_i}{2} & \text{if } c_i \equiv 0 \pmod 2, \\ 3c_i + 1 & \text{if } c_i \equiv 1 \pmod 2. \end{cases}$$

Each c_i even step divides by 2 until an odd number is reached, while c_i odd steps apply $3c_i + 1$, which is always even. For our analysis, we introduce a refined sequence that omits trivial even steps and focuses solely on the odd components:

Corollary 1. *For any $n \in \mathbb{N}$, let $a_0 = m(n)$ and*

$$a_{i+1} = 3a_i + 2^{t(a_i)}, \tag{4}$$

where a_i is the odd part of c_i. The sequence $\{a_i\}$ is defined for all $i \in \mathbb{N}_0$ and satisfies: where $t(a_i)$ is the trailing bit position from Section 2. The sequence $\{a_i\}$ is defined for all $i \in \mathbb{N}_0$, and the Collatz conjecture holds if there exists some i such that $m(a_i) = 1$.

Proof. Given $a_i = m(a_i) \cdot 2^{t(a_i)}$, we express the Collatz sequence term as:

$$c_i = a_i = m(a_i) \cdot 2^{t(a_i)} \tag{5}$$

The trivial Collatz steps (division by 2) eliminate the factor $2^{t(a_i)}$ in $t(a_i)$ iterations, yielding:

$$c_{i+t(a_i)} = m(a_i) \tag{6}$$

Applying the non-trivial Collatz step:

$$c_{i+t(a_i)+1} = 3 * m(a_i) + 1 \tag{7}$$

To obtain the next term in the $\{a_i\}$ sequence, we reintroduce the trailing bit factor:

$$\begin{aligned} a_{i+1} &= c_{i+t(a_i)+1} \cdot 2^{t(a_i)} && (8) \\ &= (3 * m(a_i) + 1) \cdot 2^{t(a_i)} && (9) \\ &= 3 * m(a_i) \cdot 2^{t(a_i)} + 2^{t(a_i)} && (10) \\ &= 3 * a_i + 2^{t(a_i)} && (11) \end{aligned}$$

which matches the recurrence in (4). If $m(a_i) = 1$, then $a_i = 2^{t(a_i)}$, a power of 2. For subsequent terms, $m(a_{i+1}) = 1$. In the Collatz sequence, powers of 2 divide repeatedly to 1, entering the cycle $1, 4, 2, 1$. Thus, $m(a_i) = 1$ ensures convergence to the trivial cycle. ☐

Remark 1. *If $a_k = 2^n$ for some k, then $a_{k+1} = 3 \cdot 2^n + 2^n = 2^{n+2}$, which is again a power of 2. Consequently, the trivial Collatz cycle $1, 4, 2, 1$ in the sequence $\{c_i\}$ corresponds to a sequence of powers of 2 in $\{a_i\}$: $2^n, 2^{n+2}, 2^{n+4}, \ldots$, with $m(a_i) = 1$ for all $i \geq k$.*

This reformulation bypasses even steps, enabling a focused study of the binary dynamics of a_i through the leading and trailing bit positions $l(a_i)$ and $t(a_i)$ in later sections.

4 Closed Form of the Reformulation

Our reformulation enables a closed-form expression for a_k:

Lemma 1. *For $k \in \mathbb{N}_0$, the sequence $\{a_k\}$ satisfies:*

$$a_k = 3^k a_0 + \sum_{i=0}^{k-1} 3^{k-1-i} \cdot 2^{t(a_i)}, \tag{12}$$

where $2^{t(a_i)} = a_i / m(a_i)$.

Proof. We verify the base case for $k = 1$:

$$a_1 = 3a_0 + 2^{t(a_0)},$$

and from (12):

$$a_1 = 3^1 a_0 + \sum_{i=0}^{0} 3^{1-1-i} \cdot 2^{t(a_i)} = 3a_0 + 3^0 2^{t(a_0)} = 3a_0 + 2^{t(a_0)},$$

which matches. For the general step, assume (12) holds for k. Then:

$$a_{k+1} = 3a_k + 2^{t(a_k)} = 3\left(3^k a_0 + \sum_{i=0}^{k-1} 3^{k-1-i} \cdot 2^{t(a_i)}\right) + 2^{t(a_k)}.$$

Distribute and adjust indices:

$$\begin{aligned} a_{k+1} &= 3^{k+1} a_0 + 3\sum_{i=0}^{k-1} 3^{k-1-i} \cdot 2^{t(a_i)} + 2^{t(a_k)} \\ &= 3^{k+1} a_0 + \sum_{i=0}^{k-1} 3^{k+1-1-i} \cdot 2^{t(a_i)} + 3^{k+1-1-k} \cdot 2^{t(a_k)} \\ &= 3^{k+1} a_0 + \sum_{i=0}^{(k+1)-1} 3^{(k+1)-1-i} \cdot 2^{t(a_i)}, \end{aligned}$$

which is (12) for $k + 1$. □

5 Basis for Lower Bounds

Assume there exists a sequence $\{a_i\}$ with $a_0 \in \mathbb{O}$ (the set of positive odd integers) such that $m(a_i) \neq 1$ for all $i \in \mathbb{N}_0$, where $a_{i+1} = 3a_i + 2^{t(a_i)}$. Denote the set of all elements in this sequence as $\mathbb{V} = \{a_i : i \in \mathbb{N}_0\}$. This implies $d(a_i) = l(a_i) - t(a_i) > 0$ for all $a_i \in \mathbb{V}$, since $d(a_i) = 0$ would yield $a_i = 2^{t(a_i)}$ and $m(a_i) = 1$.

If such a $\mathbb{V}$ exists, we select an element $a_0 \in \mathbb{V}$ with the property:

$$d(a_0) \leq d(a) \text{ for all } a \in \mathbb{V}, \tag{13}$$

defining $d_0 = d(a_0)$ as the minimal bit distance in $\mathbb{V}$.

In case a_0 forms a non-trivial cycle, then the minimal distance allows to define a lower bound for the cycle.

6 Bounds on Leading Bit Position

Under the assumption of a non-converging sequence $\{a_i\}$ with $d(a_i) \geq d_0 > 0$ for all i (Section 5, (13)), we derive upper and lower bounds for a_i and $l(a_i)$ as functions of i. Recall the recurrence:

$$a_{i+1} = 3a_i + 2^{t(a_i)}.$$

Since $d(a_i) = l(a_i) - t(a_i) \geq d_0$, we have $2^{t(a_i)} = 2^{l(a_i)-d(a_i)} \leq 2^{l(a_i)}/2^{d_0}$. For $i = 0$:

$$a_1 = 3a_0 + 2^{t(a_0)} \leq 3a_0 + \frac{2^{l(a_0)}}{2^{d_0}} \leq 3a_0 + \frac{a_0}{2^{d_0}} = a_0\left(3 + \frac{1}{2^{d_0}}\right), \tag{14}$$

using $2^{l(a_0)} \leq a_0$ (since $a_0 < 2^{l(a_0)+1}$). Generalizing:

$$a_{i+1} \leq 3a_i + \frac{a_i}{2^{d_0}} = a_i\left(3 + \frac{1}{2^{d_0}}\right), \tag{15}$$

yielding the closed-form upper bound:

$$a_i \leq a_0\left(3 + \frac{1}{2^{d_0}}\right)^i. \tag{16}$$

A lower bound follows from the dominant term in the closed form (12) with its sum $s = \sum$ being $s \in \mathbb{N}_0$:

$$a_i \geq 3^i a_0. \tag{17}$$

Applying $l(a_i) = \lfloor \log_2 a_i \rfloor$, we bound:

$$\lfloor \log_2 a_0 + i \log_2 3 \rfloor \leq l(a_i) \leq \lfloor \log_2 a_0 + i \log_2\left(3 + \frac{1}{2^{d_0}}\right)\rfloor. \tag{18}$$

Rewriting the upper bound:

$$l(a_i) \leq \lfloor \log_2 a_0 + i \log_2 3 + i \log_2\left(1 + \frac{1}{3 \cdot 2^{d_0}}\right)\rfloor. \tag{19}$$

The average growth rate of $l(a_i)$ per step lies between $\log_2 3 \approx 1.585$ and $\log_2(3 + 1/2^{d_0})$, which approaches $\log_2 3$ as d_0 increases. For $d_0 \geq 1$, the difference is small; computational results up to 2^{68} [Barina, 2020] imply no such $\mathbb{V}$ exists with $d_0 \leq 67$, and therefore:

$$\log_2\left(3 + \frac{1}{2^{67}}\right) - \log_2 3 < 3.26 \times 10^{-21}.$$

7 Lower Bound for non-trivial Cycle Length

Consider a non-trivial cycle of length k. Then there exists a minimal $m \in \mathbb{N}$ such that

$$a_k = a_0 \cdot 2^m \text{ and in general } a_{i \cdot k} = a_0 \cdot 2^{i \cdot m} \text{ for } i \in \mathbb{N}. \tag{20}$$

The value of m is actually $m = l(a_k)$, but in the following we keep m.

Utilizing the closed-form expression 12, we can rewrite the cycle condition as follows:

$$a_0 \cdot 2^m = a_k = 3^k a_0 + \sum_{i=0}^{k-1} 3^{k-1-i} \cdot 2^{t(a_i)} \tag{21}$$

$$a_0 \cdot \left(2^m - 3^k\right) = \sum_{i=0}^{k-1} 3^{k-1-i} \cdot 2^{t(a_i)} \tag{22}$$

To proceed, we establish an upper bound for $2^{t(a_i)}$. Recall that $t(a_i) = l(a_i) - d(a_i)$, where $d(a_i) \geq d_0$. Thus,

$$\begin{aligned} 2^{t(a_i)} &= 2^{l(a_i)-d(a_i)} && (23)\\ &\leq 2^{l(a_i)-d_0} && (24)\\ &\leq 2^{\lfloor \log_2 a_0 + i \log_2 3 + i \log_2\left(1+\frac{1}{3 \cdot 2^{d_0}}\right)\rfloor - d_0} && (25)\end{aligned}$$

Further, we can approximate by eliminating the terms inside the floor function:

$$2^{t(a_i)} \approx 2^{\lfloor i \log_2 3 \rfloor} \approx 3^i \tag{26}$$

Substituting this approximation in (21), we obtain

$$a_0 \cdot \left(2^m - 3^k\right) = \sum_{i=0}^{k-1} 3^{k-1-i} \cdot 2^{t(a_i)} \tag{27}$$

$$\leq \sum_{i=0}^{k-1} 3^{k-1-i} \cdot 3^i \tag{28}$$

$$= \sum_{i=0}^{k-1} 3^{k-1} \tag{29}$$

$$= k \cdot 3^{k-1} \tag{30}$$

Dividing by 3^k, we arrive at

$$a_0 \cdot \left(\frac{2^m}{3^k} - 1\right) \leq \frac{k}{3} \tag{31}$$

This yields the following lower bound for the cycle length:

$$k \geq 3a_0 \cdot \left(\frac{2^m}{3^k} - 1\right) \tag{32}$$

This relation is not straightforward, as the lower bound for k depends on k itself. Consequently, we need to test all fractions depending on k until left and right side are in sync.

The relation of k and m with $2^m > 3^k > 2^{m-1}$ showing the pairs k,m where the ratio $2^m/3^k$ reaches a new minimum towards 1 have been calculated and are listed in Table 1. The table shows the pairs k and m with the corresponding ratio $2^m/3^k$ and the logarithm of the ratio. The last column shows the logarithm of k.

Due to the computational results [Barina, 2020] we can assume:

$$a_0 > 2^{68} \text{ and thus } \log_2(3a_0) > 69.585 \tag{33}$$

To determine the minimal cycle length k, we solve for k such that $\log_2 k \geq \log_2(3a_0) + \log_2\left(2^m/3^k - 1\right)$, using $\log_2(3a_0) > 69.585$. From Table 1, consider $k = 6291$, with $\log_2 k = 12.6191$ and $\log_2(2^m/3^k - 1) = -10.6334$. This implies:

$$\log_2 k \geq 69.585 - 10.6334 = 58.9516,$$

but $\log_2 6291 \approx 12.6191 < 58.9516$, so $k = 6291$ is insufficient. Testing larger k, we find $k = 6586818670$, with $\log_2 k \approx 32.6169$ and $\log_2(2^m/3^k - 1) \approx -36.5235$, yielding:

$$\log_2 k \geq 69.585 - 36.5235 = 33.0615,$$

since $32.6169 < 33.0615$, this k is the smallest satisfying the bound. Thus, the minimal non-trivial cycle length is $k + m = 6586818670 + 10439860591 = 17.026.679.261$ including the trivial Collatz-operation (division by 2).

In [Eliahou, 1993] we find for Card Ω several values of our table related to m: 301994 for 2^{39}, 17087915 for 2^{48}, 102225496 for 2^{49} and 187363077 for 2^{52}. This makes sense due to the similar underlying interplay of 2^m and 3^k. But our investigation shows that the cycle length in [Eliahou, 1993] should include the trivial Collatz operation, leading to higher values for lower bound of non-trivial cycle lengths.

Our lower bound of 17.026.679.261 is confirmed by [Prost-Boucle, 2015] in its equation (12).

If m is chosen to be higher by e.g. one: $2^{m-1} > 3^k > 2^{m-2}$. Then the term $k \cdot 3^{k-1}$ need to increase to being a multiple of a_0 and in the magnitude of $a_0 \cdot 2^m$ to satisfy the equality. In this scenario $2^m > 2 \cdot 3^k$ and we get $k \cdot 3^{k-1} \approx \frac{k}{3} \cdot 2^m$. Consequently $k \approx 3a_0$, which would be much larger than the previous lower bound. Consequently, we ignore this case.

If the Collatz conjecture is (computational) proven for higher values of a_0, then we can derive directly the lower bound for non-trivial Collatz-cycle length from the table 2. If for example the Collatz conjecture is proven for $a_0 \leq 2^{72}$, then the lower bound for non-trivial cycle length is increasing to 186.265.759.595.

$\log_2\left(2^m/3^k-1\right)$	$\log_2 k$	$\log_2 a_0$	k	m	$k+m$
23.362	-26.2877	48.0647	10781274	17087915	27.869.189
25.9427	-26.7706	51.1283	64497107	102225496	166.722.603
26.8168	-27.5018	52.7336	118212940	187363077	305.576.017
27.3572	-29.058	54.8302	171928773	272500658	444.429.431
28.5666	-33.1373	60.1189	397573379	630138897	1.027.712.276
32.6169	-36.5235	67.5554	6586818670	10439860591	17.026.679.261
36.0684	-37.4008	71.8842	72057431991	114208327604	186.265.759.595
37.0009	-40.0174	75.4333	137528045312	217976794617	355.504.839.929
39.696	-40.2223	78.3333	890638885193	1411629234715	2.302.268.119.908
40.5801	-40.4608	79.4559	1643749725074	2605281674813	4.249.031.399.887
41.1243	-40.7473	80.2866	2396860564955	3798934114911	6.195.794.679.866
41.5185	-41.1052	81.0387	3149971404836	4992586555009	8.142.557.959.845
41.8278	-41.5821	81.8249	3903082244717	6186238995107	10.089.321.239.824

Table 2: Relation between a_0 and lower bound of non-trivial cycle length

8 Conclusion

This study introduces a novel reformulation of the Collatz conjecture by utilizing the binary structure of sequence terms, defined by the recurrence $a_{i+1} = 3a_i + 2^{t(a_i)}$. Through a closed-form expression for the sequence $\{a_i\}$, presented in Lemma 1, we enable a comprehensive analysis of bit-position dynamics, emphasizing the interaction between leading and trailing bit positions. Our approach provides a fresh perspective on the Collatz conjecture, shifting from numerical iterations to binary dynamics, thereby simplifying the examination of cycle conditions. In Section 7, we establish a lower bound of at least 17,026,679,261 steps for any non-trivial cycle, indicating that such cycles, if they exist, must be extraordinarily long and several orders of magnitude greater than the number of bits in the binary representation of the initial a_0. This result, supported by computational evidence [Barina, 2020] and consistent with prior studies [Prost-Boucle, 2015, Eliahou, 1993], highlights the stringent constraints on potential counterexamples.

The substantial lower bound strengthens the conjecture's resilience against non-trivial cycles, supporting convergence to the trivial cycle $1, 4, 2, 1$ for all starting values. Future research could explore the growth rate of $t(a_i)$ to further substantiate the Collatz conjecture's validity. As noted in [Kiemes, 2025], stochastic bit analysis and its impact on the growth rate of $t(a_i)$, combined with the significantly large cycle length, may contribute to progress toward a complete proof of the conjecture.

References

David Barina. Convergence verification of the collatz problem. *The Journal of Supercomputing*, 77(2):2001–2009, May 2020. doi: 10.1007/s11227-020-03368-x. URL `https://link.springer.com/article/10.1007/s11227-020-03368-x`. This paper discusses computational methods for verifying the Collatz conjecture, with verification up to 2^{68} reported elsewhere by Eric Roosendaal around the same period.

Shalom Eliahou. The 3x+1 problem: New lower bounds on nontrivial cycle lengths. *Discrete Mathematics*, 118: 45–56, 1993.

C. Hercher. Über die länge nicht-trivialer collatz-zyklen (ii). *Die Wurzel*, 7:146–154, 2018.

Jochen Kiemes. Collatz conjecture: The race to one, 2025. Preprint, available at `https://vixra.org/abs/2503.0161`, viXra:2503.0161.

Adrien Prost-Boucle. The 3x+1 problem: Existence of cycles under assumption of distance between odd values. 12 2015.

J. Simons and B. de Weger. Theoretical and computational bounds for m-cycles of the 3n + 1 problem, 2010. Preprint, available at `https://deweger.net/papers/[35a]SidW-3n+1-v1.44[2010].pdf`.

k	m	$\approx 2^m/3^k$	$\log_2\left(2^m/3^k - 1\right)$	$\log_2 k$
1	2	1.333333333333333333	-1.58496	0
3	5	1.185185185185185185	-2.43296	1.58496
5	8	1.053497942386831275	-4.22437	2.32193
17	27	1.039318248343855660	-4.66866	4.08746
29	46	1.025329407756841110	-5.30304	4.85798
41	65	1.011528851808608503	-6.43861	5.35755
94	149	1.009418849414340658	-6.73023	6.55459
147	233	1.007313248387464218	-7.09527	7.19967
200	317	1.005212039546930401	-7.58394	7.64386
253	401	1.003115213730841665	-8.32645	7.98299
306	485	1.001022761796411767	-9.93331	8.25739
971	1539	1.000979063991867884	-9.99631	9.92333
1636	2593	1.000935368094871156	-10.0622	10.676
2301	3647	1.000891674105338314	-10.1312	11.168
2966	4701	1.000847982023186090	-10.2037	11.5343
3631	5755	1.000804291848331222	-10.28	11.8262
4296	6809	1.000760603580690448	-10.3606	12.0688
4961	7863	1.000716917220180515	-10.4459	12.2764
5626	8917	1.000673232766718169	-10.5366	12.4579
6291	9971	1.000629550220220161	-10.6334	12.6191
6956	11025	1.000585869580603248	-10.7371	12.764
7621	12079	1.000542190847784186	-10.8489	12.8958
8286	13133	1.000498514021679740	-10.9701	13.0165
8951	14187	1.000454839102206673	-11.1024	13.1278
9616	15241	1.000411166089281757	-11.248	13.2312
10281	16295	1.000367494982821764	-11.41	13.3277
10946	17349	1.000323825782743471	-11.5925	13.4181
11611	18403	1.000280158488963658	-11.8015	13.5032
12276	19457	1.000236493101399109	-12.0459	13.5836
12941	20511	1.000192829619966613	-12.3404	13.6597
13606	21565	1.000149168044582960	-12.7108	13.732
14271	22619	1.000105508375164945	-13.2104	13.8008
14936	23673	1.000061850611629368	-13.9809	13.8665
15601	24727	1.000018194753893029	-15.7461	13.9294
47468	75235	1.000010929714251747	-16.4814	15.5347
79335	125743	1.000003664727390306	-18.0579	16.2757
190537	301994	1.000000064507504852	-23.886	17.5397
10781274	17087915	1.000000012206982855	-26.2877	23.362
64497107	102225496	1.000000008734393951	-26.7706	25.9427
118212940	187363077	1.000000005261805059	-27.5018	26.8168
171928773	272500658	1.000000001789216179	-29.058	27.3572
397573379	630138897	1.000000000105843488	-33.1373	28.5666
6586818670	10439860591	1.000000000010123125	-36.5235	32.6169
72057431991	114208327604	1.000000000005510890	-37.4008	36.0684
137528045312	217976794617	1.000000000000898654	-40.0174	37.0009
890638885193	1411629234715	1.000000000000779694	-40.2223	39.696
1643749725074	2605281674813	1.000000000000660733	-40.4608	40.5801
2396860564955	3798934114911	1.000000000000541772	-40.7473	41.1243
3149971404836	4992586555009	1.000000000000422811	-41.1052	41.5185
3903082244717	6186238995107	1.000000000000303850	-41.5821	41.8278

Table 1: List of Minima for the pairs k and m