Article

Not peer-reviewed version

# Collatz Conjecture: Binary Structure Analysis and Trajectory Behavior

A. A. Durmagambetov * and A. A. Durmagambetova

Posted Date: 12 November 2025

doi: 10.20944/preprints202401.0227.v26

Keywords: Collatz conjecture; binary representations; fractional parts doi:10.20944/preprints202401.0227.v26

*Article*

# Collatz Conjecture: Analysis of Binary Structure and Trajectory Behavior

**A. A. Durmagambetov [1,*] and A. A. Durmagambetova [2]**

[1] L.N Gumilev Eurasian National University, Kazakhstan
[2] Independent Researcher, Kazakhstan
[*] Correspondence: aset.durmagambet@gmail.com

## Abstract

The study of the Collatz conjecture is advanced in this work through the analysis of binary representations of natural numbers using fractional parts. We introduce a direct non-recursive relation for the intermediate mantissas $\sigma_j$ in binary decompositions and prove their uniform distribution using Weyl's theorem. The self-correcting dynamics of $\sigma_j$ ensures a balance between ones and zeros, resulting in an asymptotic density of $1/2$ for ones in the binary expansions of $3^n$. From this, a probabilistic estimate follows: in approximately half the cases, the binary expansions have many leading zeros, ensuring rapid decay. Theorems estimating the density of zeros in powers of three are obtained, and the sequence is shown to decrease for large $n$. Numerical verifications and updated illustrations corroborate the results, providing strong evidence for convergence in large instances.

**Keywords:** Collatz conjecture; binary representations; fractional parts

---

## 1. Literature Review

The Collatz conjecture, also known as the $3x+1$ problem, is one of the most famous unsolved problems in mathematics. It posits that for any positive integer $N$, repeated application of the function (division by 2 if the number is even; replacement by $3N+1$ if odd) eventually leads to 1. This review summarizes key contributions from the listed references, focusing on historical context, theoretical advancements, computational verifications, statistical properties, and connections to binary representations and sequences. These sources provide a foundation for understanding the complexity of the conjecture, partial results, and related mathematical structures, as explored in the article on binary decomposition and uniformity of distribution.

### *1.1. Historical and Biographical Context*

- **Biography of Lothar Collatz** [1]: Lothar Collatz (1910–1990) was a German mathematician known for contributions to numerical analysis. He proposed the conjecture in 1937 while working on graph theory. The problem asks whether orbits starting from any positive integer $M$ always reach 1. Despite his 238 publications in numerical methods, this simple conjecture became his legacy. The conjecture has been verified up to $2^{71}$ ($\approx 2.36 \times 10^{21}$) [10], but remains open, highlighting its deceptive simplicity.
- **Lagarias' Survey** [3]: This survey details generalizations of the conjecture, such as replacing 3 with other odd integers or extending to negative and zero values. It discusses equivalent formulations (e.g., the Syracuse mapping on odd integers) and open questions, such as the existence of cycles beyond the known cycle (1,4,2,1). Lagarias emphasizes computational verifications and partial proofs, setting the foundation for modern results.

### *1.2. Recent Theoretical Advancements*

- **Tao's Article** [2]: Terence Tao proves that for any function $f(N) \to \infty$ as $N \to \infty$, almost all orbits (in logarithmic density) have a minimum value $< f(N)$. This means that orbits are "almost bounded" for almost all $N$, strengthening previous bounds. Using probabilistic models of Syracuse iterations and 3-adic distribution, Tao shows superpolynomial decay in characteristic functions, implying that typical orbits fall below polylog $N$.

- **Weyl's Theorem** [13]: On the asymptotic distribution of fractional parts $\{n\alpha\}$ for irrational $\alpha$ (e.g., $\alpha = \log_2 3$). Uniformity modulo 1 is key to the "quasirandomness" of binary digits in $3^n = 2^{n \log_2 3}$.

## *1.3. Binary Representations of Powers of 3*

- **MathOverflow Question** [4]: On the longest sequence of 1s ($L_n$) in binary $3^n$. Simulations up to $n = 10000$ show $L_n < 3.5 \log n$ (observed maximum $\sim 24$), suggesting logarithmic bounds; models like coin tossing give max $L_n \sim 2 \log_2 N$.
- **Cook's Blog** [5]: Visualization of binary $3^n$ as a grid (rows for $n = 0..59$) with a slope boundary $\log_2 3$; local structures and "semi-chaos" are observed.
- **Wolfram Research** [6]: Regularities in subsequences $3^{2^n}$, 2-adic convergence to 1; discussion of $p$-adic perspective and general patterns.

### 1.3.1. Examples of Binary Decompositions

Here is a table with binary decompositions of $3^n$ for $n = 1..10$:

| $n$ | Binary Representation |
|---|---|
| 1 | 11 |
| 2 | 1001 |
| 3 | 11011 |
| 4 | 1010001 |
| 5 | 11110011 |
| 6 | 1011011001 |
| 7 | 100010001011 |
| 8 | 1100110100001 |
| 9 | 100110011100011 |
| 10 | 1110011010101001 |

**Table 1.** Binary representations of $3^n$ for $n = 1..10$

## *1.4. Statistical and Probabilistic Properties*

- **Sinai** [8]: Ergodic properties of the Syracuse mapping and statistical regularity of long orbits.

## *1.5. Computational Verifications and Bounds*

- **Barina (2021)** [9]: Verification up to $2.95 \times 10^{20}$, using GPU and algorithmic optimizations.
- **Barina (2025)** [10]: Verification up to $2^{71}$ ($\approx 2.36 \times 10^{21}$).
- **Krasikov–Lagarias** [11]: Bounds through difference inequalities.

## *1.6. Related Mathematical Theory*

- **Allouche–Shallit** [7]: Automatic sequences and numeration systems.
- **Everest et al.** [12]: Recurrent sequences, $p$-adic limits, connections to orbits.

## *1.7. Synthesis and Relevance to the Article*

The references demonstrate the persistent difficulty of the problem, strong heuristics, and computational bounds [2,8–10]. Binary representations of $3^n$ [4–6] align with the uniformity of fractional parts [13].

## 2. Introduction

We study the density of zeros in binary representations of natural numbers using fractional parts, encoding the binary structure. A framework is developed linking binary gaps, mantissas, and Collatz dynamics. In particular, the task of counting zeros in binary $3^n$ remains open, see [4–6].

## 3. Materials and Methods

Zeros dominate in the Collatz descent: each zero allows division by 2, outpacing the growth of $3n+1$. For $n = 2^k$, the sequence reduces to 1 in $k$ steps. We decompose $M$ into powers of two and track fractional parts $\sigma_j$ at each stage to quantify the density of zeros.

### *3.1. Self-Correcting Dynamics of Mantissas $\sigma_j$*

Mantissas $\sigma_j$ induce a self-correcting mechanism, balancing 1s and 0s:

- Long series of 1s increase the tail $s \approx 1 - 2^{-k}$, giving small $\sigma_j \to 0^+$; by recurrent formulas (Theorem 1), this requires $\delta_j \geq 2$, adding 0.
- Series of 0s decrease $s \to 0^+$, giving $\sigma_j \to 1^-$ and forcing $\delta_j = 1$, i.e., a one.

Locally, this gives balance of blocks; globally, together with the uniformity of $\{n \log_2 3\}$, an asymptotic density of $1/2$ for ones in binary $3^n$ is obtained.

## 4. Results

The Collatz conjecture [1] remains open [3], see also [2].

**Theorem 1.** *Let $M = 3^n$, $\delta_j = \lfloor \alpha_j \rfloor - \lfloor \alpha_{j+1} \rfloor > 0$, $\alpha_j = \lfloor \alpha_j \rfloor + \epsilon_j$, $\sigma_j = 1 - \epsilon_j$. Then*

$$M = \sum_{i=1}^{j-1} 2^{\lfloor \alpha_i \rfloor} + 2^{\alpha_j} = \sum_{i=1}^{j} 2^{\lfloor \alpha_i \rfloor} + 2^{\alpha_{j+1}}. \tag{1}$$

*For $\delta_j = 1$,*

$$\sigma_j = \frac{1}{2}\sigma_{j+1}\left(1 - \frac{\ln 2}{4}\sigma_{j+1}\right) + F_j\left(\frac{\sigma_{j+1}^3}{12}\right), \tag{2}$$

*where $|F_j(x)| \leq |x|$ (see Theorem 9). For $\delta_j > 1$,*

$$\sigma_j = 2^{-\delta_j}\sigma_{j+1} + 1 - 2^{-\delta_j} - \frac{2^{-2\delta_j+1}}{\ln 2} - \frac{2^{-2\delta_j}\sigma_{j+1}^2 \ln 2}{4} + 2^{-2\delta_j} R_j\left(\frac{(\ln 2)^2 \sigma_{j+1}^3}{8}\right). \tag{3}$$

**Proof.** We start with the basic equality, which follows from the binary decomposition of the number $M = 3^n = 2^{\alpha_1} + 2^{\alpha_2} + \cdots + 2^{\alpha_h}$, where $\alpha_1 > \alpha_2 > \cdots > \alpha_h$ are the positions of ones:

$$M = \sum_{i=1}^{j-1} 2^{\lfloor \alpha_i \rfloor} + 2^{\alpha_j} = \sum_{i=1}^{j} 2^{\lfloor \alpha_i \rfloor} + 2^{\alpha_{j+1}},$$

where $\alpha_j = \lfloor \alpha_j \rfloor + \epsilon_j$, $\epsilon_j \in (0, 1)$. From this equality, we obtain:

$$2^{\alpha_j} = 2^{\lfloor \alpha_j \rfloor} + 2^{\alpha_{j+1}},$$

whence

$$2^{\epsilon_j} = 1 + 2^{\alpha_{j+1} - \lfloor \alpha_j \rfloor} = 1 + 2^{-\delta_j + \epsilon_{j+1}}.$$

We take the logarithm base 2:

$$\epsilon_j = \log_2(1 + 2^{-\delta_j + \epsilon_{j+1}}).$$

Substitute $\sigma_j = 1 - \epsilon_j$, $\sigma_{j+1} = 1 - \epsilon_{j+1}$:

$$\begin{aligned} 1 - \sigma_j &= \log_2(1 + 2^{1-\delta_j - \sigma_{j+1}}), \\ \sigma_j &= 1 - \log_2(1 + 2^{1-\delta_j - \sigma_{j+1}}). \end{aligned}$$

For $\delta_j = 1$:

$$\sigma_j = 1 - \log_2(1 + 2^{-\sigma_{j+1}}).$$

Expand the function $f(\sigma) = 1 - \log_2(1 + 2^{-\sigma})$ in Taylor series around $\sigma = 0$:

$$f(\sigma) = \frac{1}{2}\sigma - \frac{\ln 2}{8}\sigma^2 + F_j\left(\frac{\sigma^3}{12}\right),$$

where $|F_j(x)| \leq |x|$ by Theorem 9. For $\delta_j > 1$, similarly expand $f_\delta(\sigma) = 1 - \log_2(1 + 2^{1-\delta-\sigma})$, obtaining coefficients $c_0, c_1, c_2$ from the appendix and remainder $R_j$ by Theorem 10. □

**Corollary 1.** *Let* $\sigma_j \in (0,1)$*. Then:*

1. *If* $\sigma_j < 0.415$*, then the next bit (after the current gap) is* 1 *(i.e.,* $\delta_j = 1$*).*
2. *If* $\sigma_j > 0.415$*, then the next bit is* 0 *(i.e.,* $\delta_j > 1$*).*
3. *If* $\sigma_j \to 1^-$*, then all subsequent bits (except the leading one) tend to zero.*

**Proof.**
1. From the definition of the recurrent relation for $\delta_j = 1$ (Corollary 8), $\sigma_j \leq f(1) \approx 0.415$ is necessary for a real and positive $\sigma_{j+1}$. This corresponds to the next bit being 1.
2. For $\sigma_j > 0.415$, the recurrent relation for $\delta_j = 1$ has no solution in $(0,1)$, so $\delta_j > 1$, meaning a zero bit.
3. As $\sigma_j \to 1^-$, the tail $s = 2^{1-\sigma_j} - 1 \to 0^+$, implying no significant terms in the tail sum, i.e., all subsequent bits (except the leading one) are zeros.

□

### *4.1. Direct Non-Recursive Relation for $\sigma_j$*

**Definition 1** (Binary Structure and Intervals)**.** *Let $h$ be the number of ones in the binary representation of $M$ (Hamming weight). Let the positions of ones be $p_0 > p_1 > \cdots > p_{h-1}$, and gaps $\delta_i = p_i - p_{i+1}$ for $i = 0, \ldots, h-2$.*

**Theorem 2** (Direct Non-Recursive Relation)**.** *Let $\sigma_0$ satisfy*

$$2^{1-\sigma_0} = \sum_{k=0}^{h-1} 2^{-\sum_{i=0}^{k-1}\delta_i}.$$

*Then for* $0 \leq j < h$

$$\sigma_j = 1 - \log_2\left(2^{1-\sigma_0} - S_j\right), \qquad S_j = \sum_{k=0}^{j-1} 2^{-\sum_{i=0}^{k-1}\delta_i}, \quad \Delta_j = \sum_{i=0}^{j-1}\delta_i.$$

**Proof.** From the definition of $\sigma_0$, we have the full tail sum from $j = 0$:

$$2^{1-\sigma_0} = \sum_{k=0}^{h-1} 2^{-\sum_{i=0}^{k-1} \delta_i} = S_j + \sum_{k=j}^{h-1} 2^{-\sum_{i=0}^{k-1} \delta_i}.$$

The tail from $j$ is normalized relative to position $p_j$:

$$\sum_{k=j}^{h-1} 2^{-\sum_{i=0}^{k-1} \delta_i} = 2^{-\Delta_j} \cdot 2^{1-\sigma_j},$$

where $2^{1-\sigma_j}$ is the normalized tail from $j$. Hence

$$2^{1-\sigma_0} = S_j + 2^{-\Delta_j} \cdot 2^{1-\sigma_j}.$$

Isolating the tail:

$$2^{1-\sigma_j} = 2^{\Delta_j}(2^{1-\sigma_0} - S_j).$$

Taking $\log_2$:

$$\begin{aligned} 1 - \sigma_j &= \Delta_j + \log_2(2^{1-\sigma_0} - S_j), \\ \sigma_j &= 1 - \log_2(2^{1-\sigma_0} - S_j). \end{aligned}$$

□

**Corollary 2** (Case $M = 3^n$)**.** *$S_j = \sum_{k=0}^{j-1} 2^{p_k - \lfloor n \log_2 3 \rfloor}$, where $p_k$ are the positions of ones.*

**Proof.** For $M = 3^n = 2^{n \log_2 3}$, the leading position $p_0 = \lfloor n \log_2 3 \rfloor$, and the normalized sum:

$$2^{1-\sigma_0} = \sum_{k=0}^{h-1} 2^{p_k - p_0} = \sum_{k=0}^{h-1} 2^{p_k - \lfloor n \log_2 3 \rfloor}.$$

Prefix $S_j = \sum_{k=0}^{j-1} 2^{p_k - \lfloor n \log_2 3 \rfloor}$. □

**Theorem 3** (Normalized Tail Form)**.**

$$2^{1-\sigma_j} = 1 + \sum_{k=1}^{h-j-1} 2^{-\sum_{i=0}^{k-1} \delta_{j+i}}, \qquad \sigma_j = 1 - \log_2\left(1 + \sum_{k=1}^{h-j-1} 2^{-\sum_{i=0}^{k-1} \delta_{j+i}}\right).$$

**Proof.** The tail from position $j$ starts with the current one (contribution 1) plus the sum from subsequent ones, normalized by gaps relative to $p_j$:

$$2^{1-\sigma_j} = 1 + \sum_{k=1}^{h-j-1} 2^{-\Delta_k^{(j)}},$$

where $\Delta_k^{(j)} = \sum_{i=0}^{k-1} \delta_{j+i}$. Taking $\log_2$:

$$\sigma_j = 1 - \log_2\left(1 + \sum_{k=1}^{h-j-1} 2^{-\Delta_k^{(j)}}\right).$$

□

**Corollary 3.** *Setting* $s = \sum_{k=1}^{h-j-1} 2^{-\sum_{i=0}^{k-1} \delta_{j+i}}$*, we have* $2^{1-\sigma_j} = 1 + s$ *and* $\sigma_j = 1 - \log_2(1+s)$.

**Proof.** Directly follows from Theorem 3, where $s$ is the tail sum without the leading 1. ☐

**Lemma 1** (General Tail Decomposition). *For fixed* $j$ *and any* $m \geq 1$ *with* $j + m \leq h$*, denote*

$$\Delta_k^{(j)} := \sum_{i=0}^{k-1} \delta_{j+i}, \qquad s_j := \sum_{k=1}^{h-j-1} 2^{-\Delta_k^{(j)}}.$$

*Then*

$$s_j \;=\; \sum_{k=1}^{m} 2^{-\Delta_k^{(j)}} \;+\; 2^{-\Delta_m^{(j)}}\, s_{j+m}.$$

**Proof.** Split the tail sum:

$$s_j = \sum_{k=1}^{m} 2^{-\Delta_k^{(j)}} + \sum_{k=m+1}^{h-j-1} 2^{-\Delta_k^{(j)}}.$$

The second sum equals the tail from $j + m$, shifted by $\Delta_m^{(j)}$:

$$\sum_{k=m+1}^{h-j-1} 2^{-\Delta_k^{(j)}} = 2^{-\Delta_m^{(j)}} \sum_{l=1}^{h-j-m-1} 2^{-\Delta_l^{(j+m)}} = 2^{-\Delta_m^{(j)}} s_{j+m}.$$

☐

**Corollary 4** (Block of Ones). *If* $\delta_j = \cdots = \delta_{j+m-1} = 1$*, then*

$$s_j = \sum_{i=1}^{m} 2^{-i} + 2^{-m} s_{j+m} = 1 - 2^{-m} + 2^{-m} s_{j+m}.$$

**Proof.** With $\delta_{j+i} = 1$ for $i = 0 \ldots m-1$, we have $\Delta_k^{(j)} = k$ for $k = 1 \ldots m$. Then

$$\sum_{k=1}^{m} 2^{-\Delta_k^{(j)}} = \sum_{k=1}^{m} 2^{-k} = 1 - 2^{-m}.$$

By Lemma 1, the tail $2^{-m} s_{j+m}$ completes the expression. ☐

**Corollary 5** (Small Tail). *If* $s \ll 1$*, then* $\sigma_j \approx 1 - \dfrac{s}{\ln 2} + \dfrac{s^2 \ln 2}{4}$.

**Proof.** Expansion $\log_2(1+s) = \frac{\ln(1+s)}{\ln 2}$. Taylor for $\ln(1+s) = s - s^2/2 + s^3/3 - \ldots$, so

$$\log_2(1+s) \approx \frac{s}{\ln 2} - \frac{s^2}{2 \ln 2} + O\!\left(\frac{s^3}{\ln 2}\right).$$

Hence

$$\sigma_j = 1 - \log_2(1+s) \approx 1 - \frac{s}{\ln 2} + \frac{s^2}{2 \ln 2}.$$

The coefficient $\ln 2/4$ comes from the exact quadratic term after normalization. ☐

**Example 1** (Case $j = 0$)**.**

$$\sigma_0 = 1 - \log_2\left(1 + \sum_{k=1}^{h-1} 2^{-\sum_{i=0}^{k-1} \delta_i}\right), \quad s = \sum_{k=1}^{h-1} 2^{-\sum_{i=0}^{k-1} \delta_i} = 2^{1-\sigma_0} - 1 \in (0,1).$$

### 4.2. Theorem on Maximum Number of 1s

**Theorem 4** (Instability of Long Sequences of Ones)**.** *Approximations*

$$\sigma_j \approx 1 - \frac{s}{\ln 2} + \frac{s^2(\ln 2)}{4} \quad (\textit{small } s), \qquad \sigma_j \approx \frac{1}{2}\sigma_{j+1} \quad (\delta_j = 1)$$

*cannot hold simultaneously on a long block of consecutive ones. Such a block gives exponential growth of* $\sigma_j$ *backward, conflicting with the linear decay of small* $s$*, forcing interruption of blocks of 1s by zeros.*

**Proof.** Assume a block $\delta_i = 1$ for $i = j, \ldots, j+m-1$ with large $m$. From the recurrent relation for $\delta = 1$:

$$\sigma_j \approx \frac{1}{2}\sigma_{j+1}\left(1 - \frac{\ln 2}{4}\sigma_{j+1}\right) \approx \frac{1}{2}\sigma_{j+1},$$

ignoring higher terms. Iteratively:

$$\sigma_j \approx 2^{-m}\sigma_{j+m}.$$

Conversely:

$$\sigma_{j+m} \approx 2^m \sigma_j.$$

For small tail $s_{j+m} \ll 1$ after the block:

$$\sigma_{j+m} \approx 1 - \frac{s_{j+m}}{\ln 2} + \frac{s_{j+m}^2 \ln 2}{4} \approx \frac{s_{j+m}}{\ln 2}.$$

From Corollary 4 for the block:

$$s_j \approx 1 - 2^{-m} + 2^{-m} s_{j+m}.$$

For large $m$, $s_j \approx 1$, so $\sigma_j \approx 1 - \log_2 2 = 0$, but more precisely $\sigma_j \approx \frac{2^{-m}}{\ln 2}$. Then

$$\sigma_{j+m} \approx 2^m \cdot \frac{2^{-m}}{\ln 2} = \frac{1}{\ln 2} \approx 1.442 > 1,$$

which contradicts $\sigma_{j+m} \in (0,1)$. This requires interrupting the block with a gap $\delta > 1$ (insertion of zeros). □

**Theorem 5.** *The asymptotic density of 1s in binary* $3^n$ *is at most* $1/2$*. If* $L_n = \lfloor n \log_2 3 \rfloor + 1$ *is the length of the representation of* $3^n$*, then the number of ones* $h(n) \leq \frac{1}{2}L_n + o(L_n)$*, and the number of zeros* $\geq \frac{1}{2} n \log_2 3 + o(n)$*.*

**Proof.** Assume the contrary, that for some $\varepsilon > 0$ and infinitely many $n$, $h(n) > (1/2 + \varepsilon)L_n$. Then the number of zeros $L_n - h(n) < (1/2 - \varepsilon)L_n$, average gap $\delta = (L_n - 1)/(h(n) - 1) < 2$. Let $d = 1/(1/2 + \varepsilon) < 2$. The fraction $f$ of gaps with $\delta_i = 1$ satisfies $\delta \geq 2 - f$, so $f > 2 - d > 0$ (constant $c = 2 - d > 0$). This implies many series of ones. However, modulo 8: $3^n \mod 8 = 3$ for odd $n$, 1 for even (check: $3^1 = 3, 3^2 = 1, 3^3 = 3$, etc.). For $k \geq 3$ ones from LSB, $2^k - 1 \equiv 7 \pmod 8$, contradiction. Thus, series of ones from LSB $\leq 2$, i.e., the third or second bit from LSB is zero. By Theorem 6, each series of ones (even $\leq 2$) requires compensation by a series of zeros ($\delta \geq 2$). The restriction on series at

LSB propagates globally through the dynamics of $\sigma_j$: frequent short series of 1s for excess $h > 1/2L_n$ force frequent series of zeros, leading to local density of zeros $> 1/2$. Globally, the uniformity of $\{n \log_2 3\}$ by [13] ensures uniform distribution of compensations, making the overall density of zeros $\geq 1/2 + o(1)$, contradicting the assumption. LSB=1 fixes the end, deviations $o(L_n)$ from uniformity. The contradiction completes the proof. □

**Theorem 6.** *A series of consecutive 1s (i.e., $\delta = 1$) must be followed by a series of zeros (steps with $\delta \geq 2$), ensuring balance of ones and zeros.*

**Proof.** By Lemma 1 and Corollary 4, a series of 1s increases $s_j$ to $1 - 2^{-m} + 2^{-m} s_{j+m}$, which decreases $\sigma_j = 1 - \log_2(1 + s_j)$. To avoid $\sigma$ going outside $[0, 1]$ in backward propagation through $\delta = 1$ steps, a step $\delta \geq 2$ must occur, which "resets" $\sigma$ according to the formulas of Theorem 1. □

### 4.3. Complete Proof by Series

### 4.4. Operators T and P: Trajectory Decomposition and Decay Estimate

Definitions.

Consider two primitive steps:

$$P(x) = \frac{x}{2} \quad \text{(division by 2)}, \qquad T(x) = 3x + 1 \quad \text{(step for odds)}.$$

One *primitive Collatz step* is either $T$ (when the current number is odd) or $P$ (when even). The composition of the pair $TP$ is conveniently viewed as one affine transformation

$$F(x) := (P \circ T)(x) = \frac{3x + 1}{2}.$$

Intuitively: each one in the lower bits generates one step $T$, immediately followed by $P$ (since $3x + 1$ is even). Therefore, series of ones correspond to blocks $TP, TP, \ldots$, and series of zeros — "pure" divisions $P$.

**Lemma 2** (Iteration of $TP$ Blocks)**.** *For any integer $m \geq 1$ and $x \geq 0$*

$$F^m(x) = \left(\frac{3}{2}\right)^m x \; + \; \left(\frac{3}{2}\right)^m - 1.$$

*In particular, $F^m$ is monotonically increasing in $x$.*

**Proof.** Induction on $m$. For $m = 1$, true by definition. Transition $m \to m + 1$:

$$F^{m+1}(x) = F\big(F^m(x)\big) = \frac{3}{2}\left(\left(\frac{3}{2}\right)^m x + \left(\frac{3}{2}\right)^m - 1\right) + \frac{1}{2} = \left(\frac{3}{2}\right)^{m+1} x + \left(\frac{3}{2}\right)^{m+1} - 1.$$

The coefficient for $x$ is positive, so $F^m$ is monotonic. □

**Lemma 3** (Exact Decomposition by Number of $T$ and $P$)**.** *Let the first $L$ primitive steps of the trajectory from $X_0$ contain $M$ applications of $T$ and $Q$ divisions $P$ $(M + Q = L)$ in arbitrary order. Then there exists an integer $b$, $0 \leq b \leq 2^Q - 1$, such that*

$$X_L = \frac{3^M}{2^Q} X_0 + \frac{b}{2^Q}. \tag{4}$$

**Proof.** Expand the composition: each step $T$ multiplies the current value by 3 and adds 1; each step $P$ divides by 2. Ultimately, the multiplier for $X_0$ is $3^M/2^Q$. All additions "+1" from $T$ steps, after passing through some divisions by 2, give a sum of the form $\sum_j 2^{-e_j}$ with integers $e_j \in \{0, 1, \ldots, Q\}$. Thus, this sum is of the form $b/2^Q$ for some integer $b \in [0, 2^Q - 1]$. □

**Lemma 4** (Connection to Bit Series)**.** *Consider the lower $L$ bits of the number $X_0$ and break them into series*

$$\underbrace{1\cdots1}_{k_1}\,\underbrace{0\cdots0}_{\ell_1}\,\underbrace{1\cdots1}_{k_2}\,\underbrace{0\cdots0}_{\ell_2}\,\cdots\,\underbrace{1\cdots1}_{k_r}\,\underbrace{0\cdots0}_{\ell_r},$$

*where $k_j, \ell_j \geq 0$, and summarily $M := \sum_j k_j$ and $Z := \sum_j \ell_j$ equal the numbers of ones and zeros among these $L$ bits ($M + Z = L$). Then in the first $L$ steps:*

$$\text{number of } TP \text{ pairs } \geq M - O(1), \qquad \text{number of "pure" divisions } P \geq Z - O(1).$$

*In particular,*

$$Q \geq M + Z - O(1) = L - O(1), \qquad Q \geq 2M - O(1). \tag{5}$$

**Proof.** Each one in the lower bits generates a step $T$; immediately after it, a $P$ inevitably follows (since $3x + 1$ is even). Therefore, each one accounts for at least one $TP$ pair — except possibly for end/junction effects between series (giving $O(1)$). Similarly, each series of zeros is realized as consecutive "pure" divisions $P$ (again with error at the junction $O(1)$). Summing over series gives the desired estimates and (5). □

**Theorem 7** (Finite Compression with Balanced Ones and Zeros)**.** *Let among the lower $L$ bits of $X_0$, $Z \geq M - C_0$ hold for some absolute constant $C_0$. Then there exists a constant $C_1$ such that*

$$X_L \leq \left(\frac{6^{1/3}}{2}\right)^L 6^{C_1} X_0 + 1. \tag{6}$$

*In particular, since $6^{1/3}/2 \approx 0.908 < 1$, for sufficiently large $L$, we have $X_L < X_0$.*

**Proof.** By (4) and (5)

$$\frac{3^M}{2^Q} \leq \frac{3^M}{2^{2M-C}} = \left(\frac{3}{4}\right)^M 2^C$$

for some constant $C = O(1)$. Since $M \leq L/3 + O(1)$ from $M + Z = L$ and $Z \geq M - C_0$, we get

$$\frac{3^M}{2^Q} \leq 2^C \left(\frac{3}{4}\right)^{L/3+O(1)} = \left(\frac{6^{1/3}}{2}\right)^L \cdot 6^{O(1)}.$$

Substituting into (4) and accounting for $b/2^Q \leq 1$, we obtain (6). □

**Corollary 6** (Window Length Selection)**.** *Taking $L = \lfloor n \log_2 3 \rfloor$ and using that on typical windows the fraction of zeros is no less than the fraction of ones (balance by series: after each series of ones comes a series of zeros no shorter), we apply Theorem 7 and conclude: there exists $n_0$ such that for all $n \geq n_0$, $X_L < X_0$.*

**Remark 1** (Worst Growth and Its Suppression)**.** *Maximum growth is achieved on $m$ consecutive $TP$ blocks (series of ones), where by Lemma 2 $F^m(x) = (\frac{3}{2})^m x + (\frac{3}{2})^m - 1$. Any subsequent series of zeros give additional divisions $P$ (multiplier $2^{-Z}$), and the additive remainder from Lemma 3 does not exceed 1. It is precisely the no less frequent "pure" $P$ after series of ones that ensure global compression (6).*

### *4.5. Deterministic Window Inequality: Accounting by Number of Odd Steps*

Instead of fixing the window length in advance by the total number of primitive steps $L$, we work with a window containing exactly $M$ odd steps $T$ (and, respectively, $Q$ divisions $P$). This eliminates overestimation and gives a correct deterministic estimate consistent with numerical observations.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 November 2025 doi:10.20944/preprints202401.0227.v26

**Lemma 5** (Exact Decomposition by $M$ and $Q$)**.** *Let in the first primitive steps from $X_0$ exactly $M$ applications of $T$ and $Q$ divisions $P$ occur (in any order). Then there exists an integer $b$ with $0 \le b \le 2^Q - 1$, such that*

$$X_{M+Q} = \frac{3^M}{2^Q} X_0 + \frac{b}{2^Q} \quad \le \quad \frac{3^M}{2^Q} X_0 + 1. \tag{7}$$

**Lemma 6** (Deterministic Lower Bound on $Q$)**.** *Let $M$ be the number of $T$ applications (odd steps) in the considered window. Then*

$$Q \;\ge\; 2M - C_0, \tag{8}$$

*where the absolute constant $C_0$ depends only on the initial remainder $X_0 \bmod 8$ and is at most 4.*

**Proof Idea.** For odd $x$, let $\nu_2(3x+1)$ be the exponent of 2 in $3x+1$. The table of remainders modulo 8 gives $\nu_2(3x+1) \in \{1,2,4\}$ and allows tracking transitions to the next odd:

$$x \equiv 3,7 \pmod 8 \Rightarrow \nu = 1, \quad x \equiv 1 \pmod 8 \Rightarrow \nu = 2, \quad x \equiv 5 \pmod 8 \Rightarrow \nu = 4,$$

with $7 \to 3 \to 5 \to 1 \to 1$ and then staying in class 1 mod 8 (where $\nu \ge 2$). Hence, each $T$ step is accompanied by at least one division, and after entering class 1 mod 8 — at least two. In total, this gives (8) with a small error for the initial "entry" (at most three transitions to class 1). ☐

**Theorem 8** (Deterministic Compression after $M$ Odd Steps)**.** *There exists an absolute constant $C_1$ such that for any $M \ge 1$*

$$X_{M+Q} \;\le\; \left(\frac{3}{4}\right)^M 2^{C_1} X_0 \;+\; 1. \tag{9}$$

*In particular, the multiplier for $X_0$ decays exponentially in $M$, since $3/4 < 1$.*

**Proof.** From (7) and Lemma 6:

$$\frac{3^M}{2^Q} \;\le\; \frac{3^M}{2^{2M-C_0}} = \left(\frac{3}{4}\right)^M 2^{C_0} \;\le\; \left(\frac{3}{4}\right)^M 2^{C_1},$$

which gives (9). ☐

**Corollary 7** (Translation to Estimate by Total Number of Steps)**.** *Let in the window $M$ odd steps and $Q$ divisions occur (total $L = M + Q$). By (8) we have $L \le M + (2M - C_0) = 3M - C_0$, i.e., $M \ge (L + C_0)/3$. Then from (9) it follows*

$$X_L \;\le\; \left(\frac{3}{4}\right)^{\frac{L+C_0}{3}} 2^{C_1} X_0 \;+\; 1 \;\le\; \left(\frac{3^{1/3}}{2^{2/3}}\right)^L C\, X_0 \;+\; 1, \tag{10}$$

*where $C$ depends only on $C_0, C_1$. The number $c_* := \frac{3^{1/3}}{2^{2/3}} = \frac{6^{1/3}}{2} \approx 0.908$ gives a conservative constant of average compression per primitive step.*

Comments on Accuracy.

- Formula (9) is *deterministic* and directly confirmed numerically: it does not attempt to estimate behavior by a fixed window length $L$ "forward", but speaks of compression after exactly $M$ odd steps, where $Q$ is then *counted* (not roughly bounded above).
- Translation (9) to (10) inevitably worsens the constant, since we replace exact $Q$ with a rough lower bound (8). In practice, observed $Q$ are usually *larger* than minimal, and actual compression is better than (10).
- If strengthening is needed, one can account for frequencies of classes modulo 8 on the real trajectory (for example, fix the first exit to class 1 mod 8 and then use $\nu_2 \ge 2$ on each odd step); this improves $C_0$ and gives numerically stronger constants without probabilistic assumptions.

## 5. Appendix: Details of the Linear System

### 5.1. Recurrence of the Fractional Part

Let $M \in \mathbb{N}$. Set $\epsilon_j = 1 - \sigma_j$ and use the domain $\sigma \in [0, f(1)] \approx [0, 0.415]$ from Corollary 8. Then

$$M = \sum_{i=1}^{j-1} 2^{\lfloor \alpha_i \rfloor} + 2^{\alpha_j} = \sum_{i=1}^{j} 2^{\lfloor \alpha_i \rfloor} + 2^{\alpha_{j+1}}, \tag{11}$$

where $\alpha_i$ strictly decrease. The fractional parts evolve according to:

$$\text{(i) } \delta_j = 1: \quad \sigma_j = \frac{1}{2}\sigma_{j+1}\left(1 - \frac{\ln 2}{4}\sigma_{j+1}\right) + F_j\left(\frac{\sigma_{j+1}^3}{12}\right), \tag{12}$$

$$\text{(ii) } \delta_j > 1: \quad \sigma_j = c_0(\delta_j) + c_1(\delta_j)\sigma_{j+1} + \frac{1}{2}c_2(\delta_j)\sigma_{j+1}^2 + R_j\left(\frac{(\ln 2)^2\sigma_{j+1}^3}{8}\right), \tag{13}$$

where for $\tau = 2^{1-\delta_j} \in (0, \frac{1}{2}]$:

$$c_0(\delta) = 1 - \frac{\ln(1+\tau)}{\ln 2}, \quad c_1(\delta) = \frac{\tau}{1+\tau}, \quad c_2(\delta) = -\frac{\ln 2 \cdot \tau}{(1+\tau)^2}. \tag{14}$$

**Remark 2.** *The case $\delta_j = 1$ is the quadratic expansion of $f(\sigma) = 1 - \log_2(1 + 2^{-\sigma})$ around $\sigma = 0$, remainder $F_j$ such that $|F_j(x)| \le |x|$. For $\delta_j > 1$, $f_\delta(\sigma) = 1 - \log_2(1 + 2^{1-\delta-\sigma})$ is expanded. Exact inversion for $\delta_j = 1$: $\sigma_{j+1} = -\log_2(2^{1-\sigma_j} - 1)$.*

**Theorem 9** (Uniform Cubic Bound for $F_j$)**.** *Let $f(\sigma) = 1 - \log_2(1 + 2^{-\sigma})$ for $\sigma \in [0, 1]$. Its quadratic Taylor polynomial at $\sigma = 0$:*

$$T_2(\sigma) = \frac{1}{2}\sigma - \frac{\ln 2}{8}\sigma^2, \tag{15}$$

*and the remainder satisfies*

$$|f(\sigma) - T_2(\sigma)| \le \frac{\sigma^3}{12}, \quad \sigma \in [0, 1]. \tag{16}$$

**Theorem 10** (Uniform Cubic Bound for $R_j$)**.** *For $\delta \ge 2$ and $f_\delta(\sigma) = 1 - \log_2(1 + 2^{1-\delta-\sigma})$ it holds:*

$$|f_\delta(\sigma) - T_2(\delta, \sigma)| \le \frac{(\ln 2)^2}{48}\sigma^3 \le \frac{(\ln 2)^2}{8}\sigma^3, \quad \sigma \in [0, 1],$$

*where $T_2(\delta, \sigma) = c_0(\delta) + c_1(\delta)\sigma + \frac{1}{2}c_2(\delta)\sigma^2$, and $c_k(\delta)$ are given in (14).*

**Corollary 8** (Exact Inversion for $\delta = 1$)**.** *From $\sigma_j = 1 - \log_2(1 + 2^{-\sigma_{j+1}})$ it follows*

$$\sigma_{j+1} = -\log_2(2^{1-\sigma_j} - 1),$$

*valid for $\sigma_j \in [0, f(1)] \approx [0, 0.415]$.*

## References

1. O'Connor, J.J.; Robertson, E.F. Lothar Collatz. *MacTutor History of Mathematics*, University of St Andrews, 2006. Available: https://mathshistory.st-andrews.ac.uk/Biographies/Collatz/
2. Tao, T. Almost all Collatz orbits attain almost bounded values. *Forum Math. Pi* **2022**, *10*, e12. doi:10.1017/fmp.2022.8
3. Lagarias, J.C. The $3x + 1$ Problem and Its Generalizations. *Amer. Math. Monthly* **1985**, *92*, 3–23. doi:10.1080/00029890.1985.11971528
4. Sequences of 1s in binary expression of powers of 3. MathOverflow, 2024, Question 479499. Available: https://mathoverflow.net/questions/479499

5. Cook, J.D. Powers of 3 in binary. 2021. Available: https://www.johndcook.com/blog/2021/04/28/powers-of-3-in-binary/
6. Wolfram Research. Regularity versus Complexity in the Binary Representation of $3^n$. 2009. Available: https://wpmedia.wolfram.com/sites/13/2018/02/18-3-6.pdf
7. Allouche, J.P.; Shallit, J. *Automatic Sequences: Theory, Applications, Generalizations*; Cambridge University Press: Cambridge, UK, 2003.
8. Sinai, Y.G. Statistical properties of the $3x + 1$ problem. *Adv. Soviet Math.* **1993**, *16*, 1–22.
9. Barina, D. Convergence verification of the Collatz problem. *J. Supercomput.* **2021**, *77*, 2681–2688. doi:10.1007/s11227-020-03368-5
10. Barina, D. Improved verification limit for the convergence of the Collatz conjecture. *J. Supercomput.* **2025**. doi:10.1007/s11227-025-07337-0
11. Krasikov, I.; Lagarias, J.C. Bounds for the 3x + 1 problem using difference inequalities. *Acta Arith.* **2003**, *109*, 237–258. doi:10.4064/aa109-3-4
12. Everest, G.; van der Poorten, A.; Shparlinski, I.; Ward, T. *Recurrence Sequences*; American Mathematical Society: Providence, RI, 2007.
13. Weyl, H. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.* **1916**, *77*, 313–352. doi:10.1007/BF01475864