REVISITING THE COLLATZ CONJECTURE: ANALYSING STRINGS, DISCOVERING BOUNDS, AND COMPUTING DISTRIBUTIONS IN BINARY COLLATZ ORBITS

Ashwat Prasanna[1]

1. TISB, Bangalore, India
ashwat.prasanna@gmail.com

December 2024

ABSTRACT

In 1937, Lothar Collatz defined the function $C : \mathbb{Z}^+ \to \mathbb{Z}^+$ as $C(x) = 3x + 1$ when x is odd and $C(x) = \frac{x}{2}$ when x is even. He conjectured that repeated computation of this function would eventually result in the value 1. Specifically, for every $x \in \mathbb{Z}^+$, $(C \circ C \circ \ldots \circ C)(x) = 1$, denoted $C^{(n)}(x) = 1$. As of 2025, this conjecture still remains unproven, after 88 years. In this research article, we employ a novel method using sequences and binary strings to offer a simplification of the conjecture and identify some novel patterns within Collatz orbits, which may potentially lead to a solution. First, we rigorously prove that the validity of the Collatz conjecture for any $x \in \mathbb{Z}^+$ is logically equivalent to the existence of positive integers y, k such that $R^{(y)}(x) = 2^k$, where $R(x)$ is three times x plus its binary least significant bit (LSB). The involvement of powers of two motivates an approach using binary representations of x. Specifically, we define a property, '`Changes`,' as the number of pairs of consecutive unequal digits in a binary string (when this goes to 1 or 0, the conjecture is true for x). For all $x \in \mathbb{Z}^+$, we show that $\texttt{Changes}(B(x)) \leq \texttt{Changes}(x)$, where $B(x)$ is x plus its binary LSB. We finally prove that, for any k-digit (binary) x, we have $\texttt{Changes}(x) > \left\lfloor \frac{2k-1}{3} \right\rfloor \implies \texttt{Changes}(C(x)) < \texttt{Changes}(x)$. We also prove that this is the best possible bound (effectively restricting the growth of `Changes`). We also elaborate on the distribution of $\texttt{Changes}(C(x))$ plotted against $\texttt{Changes}(x)$. These results may be used to further analyze the conjecture or analogues, improving the estimates of growth rates of ν_2 of values in R orbits; this is a step towards proving the complete conjecture.

Keywords: Collatz Conjecture · Binary Representation · Algorithm

MSC 2020 Classification: 37P99 (Primary), 11A99 (Secondary)

Contents

1 Introduction

Attributed to Lothar Collatz in 1937, the Collatz conjecture (also $3x+1$ problem, Ulam Conjecture, Syracuse Problem, etc.) defines the function $C : \mathbb{Z}^+ \to \mathbb{Z}^+$ as follows:

Let $C\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ be the function:

$$C(x) \coloneqq \begin{cases} 3x+1, & \text{if } x \equiv 1 \pmod 2 \\ \frac{x}{2}, & \text{if } x \equiv 0 \pmod 2 \end{cases}.$$

The conjecture states that for all $x \in \mathbb{Z}^+$, there exists some $n \in \mathbb{Z}^+$ such that $C^{(n)}(x) = 1$, where $C^{(n)}$ refers to the n-fold composition of C with itself.

For example, when $x = 3$, we have $C(x) = 10$, $C^{(2)}(x) = 5$, $C^{(3)}(x) = 16$, etc., with $C^{(7)}(x) = 1$.

Generally, such patterns of subsequent growth and decline when the Collatz function is applied tend to be unpredictable (earning them the name 'Hailstone sequences') and lending to their applications in encryption and pseudo-random number generators. Thus, analyzing patterns within Collatz orbits or Collatz analogs (which instead use $ax+b$, $a, b \in \mathbb{Z}^+$) is of importance.

Over the years, many different approaches have been utilized to investigate the Collatz conjecture. Some of the most related include Kaufman's [2] representation of Collatz orbits for $x \in \mathbb{Z}^+$ as binary strings, tracking the length of $C^{(k)}(x)$. Ren, Li, Xiao, and Bi also expressed the Collatz conjecture as an algorithm of $H(x) = \frac{x}{2}$ and $TPO(x) = 3x+1$, for more efficient computation [5]. There have also been other efforts to analyze the size of numbers within the Collatz orbit for some $x \in \mathbb{Z}^+$, determining the minimum $k \in \mathbb{Z}^+$ for $C^{(k)}(x) < x$ (see for reference [6]). Although they are still significant works, some of the limitations of prior literature include the 'unpredictability' of the Collatz orbits: namely, they cannot characterize a well-defined end-state or stopping time given a particular initial number, often due to the use of the conditional definition. Our work, which is an original research article, instead considers properties of binary strings and will delve into a novel analysis of the property `Changes` and the interesting results that arise when looking at Collatz orbits through this lens (which may eventually lead to a solution).

1.1 Notation

The following is an overview of the mathematical terminology used in the paper. New definitions introduced in this work are presented later in the main body.

(i) $\mathbb{Z}^+$ are the positive integers. $\mathbb{Z}_{\geq 0}$ denotes $\mathbb{Z}^+ \cup \{0\}$.

(ii) For any function f and any $k \in \mathbb{Z}^+$ we denote

$$f^{(k)} \equiv \overbrace{f \circ f \circ \cdots \circ f}^{k \text{ times}}.$$

(iii) We use $\{x\}$ for any real x to refer to the fractional part of x, with $\{x\} = x - \lfloor x \rfloor$.

(iv) The orbit of some $x \in \mathbb{Z}^+$ under a function f is the sequence $\mathrm{orbit}(x) = (x, f(x), f^{(2)}(x), \ldots)$.

(v) The 4-2-1 cycle refers to the only known cycle in the Collatz function: when $C^{(k)}(x) = 4$, $C^{(k+1)}(x) = 2$, $C^{(k+2)}(x) = 1$, $C^{(k+3)}(x) = 4$, and so on.

(vi) We say $\mathrm{OCC}(x)$ is TRUE or $\mathrm{OCC}(x)$ holds if and only if x satisfies the Original Collatz Conjecture (there exists $n \in \mathbb{Z}^+$ such that $C^{(n)}(x) = 1$).

1.2 Organization and Structure

The rest of the article is organized as follows. Section 2 delves into a detailed and rigorous proof using sequences to show the aforementioned simplification of the conjecture to the algorithm R. Section 3 then proposes analyzing R-orbits using `Changes`; following this, we prove a multitude of patterns in the behavior of `Changes`. Sections 4, 5, 6 are the conclusion, future work, and acknowledgements respectively. An appendix is provided, with proofs of additional results and the code used to verify our claims.

The specific contributions and claims made in Sections 2 and 3 are described in the following list:

(I) Algorithmic Representation of the Conjecture

(i) In order to allow the operations of multiplication and addition and division to be applied separately, we are motivated to express $C^{(n)}(x)$ as a product of the n^{th} terms of sequences $(\alpha_n(x))_{n\in\mathbb{Z}_{\geq 0}}$ and $(\beta_n(x))_{n\in\mathbb{Z}_{\geq 0}}$, with $C^{(n)}(x) = \alpha_n(x)\beta_n(x)$. Each sequence is defined to "perform" a different operation (see Definition 1 for full details).

(ii) Then, we can derive general rules for $\alpha_n(x)$ and $\beta_n(x)$. Hence, we reduce the conjecture to an equation and a repeated algorithm R involving two steps: multiplication by three ($A : \mathbb{Z}^+ \to \mathbb{Z}^+$) and then the addition of the largest dividing power of two ($B : \mathbb{Z}^+ \to \mathbb{Z}^+$). We formally prove that $OCC(x)$ is true if and only if the algorithm ever reaches a power of two (i.e. there exist positive integers (y, k) such that $R^{(y)}(x) = 2^k$).

(II) Binary and `Changes`

(i) Motivated by the involvement of powers of two, we investigate binary strings resulting from the previous algorithm (and relating to string growth rates as a result). Thus, we introduce $\texttt{Changes}(x)$ for $x \in \mathbb{Z}^+$ to explore how the conjecture may be satisfied, noting that $\texttt{Changes}(B(x)) \leq \texttt{Changes}(x)$ for all $x \in \mathbb{Z}^+$, where $B(x)$ is x plus its largest dividing power of two.

(ii) For all k-digit (binary) integers x, it is proven that if $\texttt{Changes}(x) > \left\lfloor \frac{2k-1}{3} \right\rfloor$, then we have $\texttt{Changes}(A(x)) < \texttt{Changes}(x)$.

(iii) We show other interesting patterns that arise when plotting $\texttt{Changes}(x)$ against $\texttt{Changes}(C(x))$.

2 Algorithmic Representation of The Conjecture

We begin by "separating" the division and multiplication followed by addition operations, by creating sequences whose n^{th}-term products give $C^{(n)}(x)$.

2.1 Constructing Sequences

To do this, we are motivated to construct sequences $(\alpha_n(x))_{n\in\mathbb{Z}_{\geq 0}}$ and $(\beta_n(x))_{n\in\mathbb{Z}_{\geq 0}}$ for all $x \in \mathbb{Z}^+$, such that $\alpha_n(x)\beta_n(x) = C^{(n)}(x)$. This allows a general rule to be identified, allowing for the Collatz Conjecture to be proven equivalent to simpler equations and an algorithm.

Definition 1. Define sequences $(\alpha_n)_{n\in\mathbb{Z}_{\geq 0}}$ and $(\beta_n)_{n\in\mathbb{Z}_{\geq 0}}$. For any $x \in \mathbb{Z}^+$, let $\alpha_0(x) = x$ and $\beta_0(x) = 1$.

$$\alpha_{n+1}(x) = \begin{cases} 3\alpha_n(x) + \frac{1}{\beta_n(x)}, & \text{if } \alpha_n(x)\beta_n(x) \equiv 1 \pmod 2 \\ \alpha_n(x), & \text{if } \alpha_n(x)\beta_n(x) \equiv 0 \pmod 2 \ . \end{cases}$$

$$\beta_{n+1}(x) = \begin{cases} \beta_n(x), & \text{if } \alpha_n(x)\beta_n(x) \equiv 1 \pmod 2 \\ \frac{\beta_n(x)}{2}, & \text{if } \alpha_n(x)\beta_n(x) \equiv 0 \pmod 2 \ . \end{cases}$$

By definition, it can be shown that $\alpha_{n+1}(x)\beta_{n+1}(x) = C(\alpha_n(x)\beta_n(x))$ for all $x, n \in \mathbb{Z}^+$. Therefore, $\alpha_n(x)\beta_n(x) = C^{(n)}(x)$ follows by a straightforward induction on n.

Note that $\beta_0 = 1$. It is only ever divided by 2. Thus, (β_n) is a sequence of reciprocals of non-decreasing powers of two. This motivates the below definition.

Definition 2. Let x be some positive integer. Define sequence $(k_i(x))_{i\in\mathbb{Z}_{\geq 0}}$ such that

$$k_n(x) = \log_2\left(\frac{1}{\beta_n(x)}\right), \text{ for all } n \in \mathbb{Z}_{\geq 0} \ .$$

By induction with base case $k_0 = 0$ and the recursive definition in Definition 1, it can be proven that $k_i(x) \in \mathbb{Z}_{\geq 0}$ for all integers $i \geq 0$ (because $\beta_n(x)$ is only divided by 2). Hence, $(k_i(x))$ is also a non-decreasing sequence.

Theorem 2 will derive a general rule for $\alpha_n(x)$ (using the recursion from Definition 1). The sequences $(j_i(x)), (a_i(x))$ and set $S(x)$ (all defined below) will be used to clarify the following proofs.

Definition 3. For all $x \in \mathbb{Z}^+$, define the set $S(x) \coloneqq \{i \mid i \in \mathbb{Z}^+,\ \alpha_i(x) \neq \alpha_{i+1}(x)\}$.

Remark 4. Observe that $S(x)$ must be infinite. If not, there exists a maximum value p, which implies $C^{(q+1)}(x) < C^{(q)}(x)$ for all $q > p$ (as for all $x \in \mathbb{Z}^+$, $C(x) \neq x$). This would be infinite descent in $\mathbb{Z}^+$, which is impossible.

Definition 5. Let x be any positive integer. By the well-ordering principle, $S(x)$ can be ordered in an increasing sequence. Let $(j_i(x))_{i\in\mathbb{Z}^+}$ be the elements of $S(x)$ in increasing order. Thus, the sequence $(j_i(x))_{i\in\mathbb{Z}^+}$ represents the indices at which the sequence $(\alpha_n(x))_{i\in\mathbb{Z}_{\geq 0}}$ changes.

Definition 6. Let x be any positive integer. Define the sequence $(a_i(x))_{i\in\mathbb{Z}^+}$ such that $a_i(x) \coloneqq k_{j_i(x)}(x)$ for all $i \in \mathbb{Z}^+$. As the sequences $(j_i(x))_{i\in\mathbb{Z}^+}$ and $(k_i(x))_{i\in\mathbb{Z}_{\geq 0}}$ are non-decreasing, $(a_i(x))_{i\in\mathbb{Z}^+}$ is non-decreasing.

Observe that, by definition, the sequence $(a_i(x))_{i\in\mathbb{Z}^+}$ gives the values of k_n when α_n increases. This will be used in following proofs.

Example: $x = 7$, see Table 1.

Table 1: Example values of α, β, a, j, k series for $x = 7$

n	$C^{(n)}(7)$	$\alpha_n(7)$	$\beta_n(7)$	$k_n(7)$	$j_n(7)$	$a_n(7)$
0	7	7	1	0		
1	22	22	1	0	0	0
2	11	22	$\frac{1}{2}$	1	2	1
3	34	68	$\frac{1}{2}$	1	4	2
4	17	68	$\frac{1}{4}$	2	7	4
5	52	208	$\frac{1}{4}$	2	11	7

Remark 7. Given some positive integer x, a relationship can be made between $a_i(x)$ and $j_i(x)$, with $a_i(x) = j_i(x) - i + 1$. This results from how we chose to define these series. As this will not be used in the rest of the argument, a proof of this can be found in Theorem 11 in the Appendix.

Theorem 1 will now prove that the sequence $(a_i(x))_{i\in\mathbb{Z}^+}$ is increasing. This will be used in the later proof of Theorem 5.

Theorem 1. *Let x be any positive integer. The sequence $(a_i(x))_{i\in\mathbb{Z}^+}$ is strictly increasing for all $i \in \mathbb{Z}^+$.*

Proof of Theorem 1. First, fix a value of x.

For clarity, we implicitly refer to $a_i(x)$ as a_i, $j_i(x)$ as j_i, $\alpha_n(x)$ as α_n, and $\beta_n(x)$ as β_n.

Note that the sequences (j_i) and (k_i) are both non-decreasing and $a_i = k_{j_i}$, so the sequence $(a_i)_{i\in\mathbb{Z}^+}$ is also non-decreasing ($a_i \leq a_{i+1}$ for all $i \in \mathbb{Z}^+$).

We prove the desired result by contradiction, supposing that there exists $i \in \mathbb{Z}^+$ such that $a_i = a_{i+1}$ (this assumption is sufficient since the sequence is already non-decreasing).

This implies $k_{j_i} = k_{j_{i+1}}$.

As $\{k_i\}$ is non-decreasing and $j_{(i+1)} > j_i$, we have $k_{j_i} = k_{(j_i)+1} = ... = k_{j_{(i+1)}}$.

This implies that $\beta_{j_i} = \beta_{j_i+1} = ... = \beta_{j_{i+1}}$ and so $\alpha_{j_i} < \alpha_{j_i+1} < ... < \alpha_{j_{i+1}}$, as they cannot change at the same time (Definition 1).

By the definition of j_i (as j_i represents the indices at which $\alpha_i(x) \neq \alpha_{i+1}(x)$), we have $\alpha_{j_{i+1}} < \alpha_{j_{i+1}+1}$, which implies $C^{(j_{i+1})}(x) \equiv 1 \pmod 2$.

From the definition of α_n, this implies $C^{(j_i)}(x) \equiv C^{(j_i+1)}(x) \equiv \cdots \equiv C^{(j_{i+1})}(x) \equiv 1 \pmod 2$.

But because $j_{i+1} > j_i$ and $j_i + 1 \leq j_{i+1}$, we have $C^{(j_i)}(x) \equiv C^{(j_i+1)}(x) \equiv 1 \pmod 2$. This is impossible (as $C(x)$ maps odd integers to even integers).

This is a contradiction and concludes the proof.

□

2.2 Expression as an Equation

We begin by proving that all $x \in \mathbb{Z}^+$ with $\mathrm{OCC}(x)$ being TRUE must also satisfy an equation:

Remark 8. Theorem 2 proves one direction of implication. The other direction (the converse) is proven later (by Theorem 5) with an analogous equation (2).

In the following theorem, we introduce the sequence $(m_n(x))_{i\in\mathbb{Z}^+}$ that will be used later in the algorithm we construct.

Theorem 2. *Let x be any positive integer. Then, there exists a non-decreasing sequence of positive integers $(m_n(x))_{i\in\mathbb{Z}^+}$ such that*

$$\alpha_n(x) = 3(\cdots 3(3(3x + 2^{a_1(x)}) + 2^{a_2(x)}) + 2^{a_3(x)})\cdots) + 2^{a_{m_n(x)}(x)}$$

for all $n \in \mathbb{Z}^+$ with $n > j_1(x)$.

Remark 9. Before we prove the theorem, it can be noted that for all $n \in \mathbb{Z}^+$ with $n \leq j_1(x)$, we have $\alpha_n(x) = n$, which trivially satisfies this statement. Henceforth, it will be implied that $n > j_1(x)$.

Proof of Theorem 2. This will be proven inductively.

Let us fix some $x \in \mathbb{Z}^+$. For clarity, we shall implicitly refer to $a_i(x)$ with a_i, $j_i(x)$ with j_i, and $m_n(x)$ with m_n.

Base case ($n = j_1 + 1$): For $n = j_1 + 1$, we have $\alpha_n(x) = 3\alpha_{n-1}(x) + 2^{k_{j_1}}$. By the recursive definition, $\alpha_{j_1} = x$. Hence, $\alpha_n(x) = 3x + 2^{a_1}$ (with $m_n = 1$).

Inductive step ($n \geq j_1 + 1, n \in \mathbb{Z}^+$): Assuming $\alpha_n(x) = 3(\cdots 3(3(3x + 2^{a_1}) + 2^{a_2}) + 2^{a_3})\cdots) + 2^{a_{m_n}}$, then $\alpha_{n+1}(x) = 3(\cdots 3(3(3x + 2^{a_1}) + 2^{a_2}) + 2^{a_3})\cdots) + 2^{a_{m_{(n+1)}}}$.

Proof of the inductive step. We will consider the possible values for $\alpha_{n+1}(x)$ (cases 1 and 2).

Case 1: $\alpha_{n+1}(x) = \alpha_n(x)$: this immediately satisfies the required statement, with $m_{n+1} = m_n$.

Case 2: $\alpha_{n+1}(x) \neq \alpha_n(x)$.

By definition, $\alpha_{n+1}(x) \neq \alpha_n(x)$ if and only if $n \in S(x)$.

For all $i \in \mathbb{Z}^+$, $\alpha_{(j_i+1)}(x) = \alpha_{(j_i+2)}(x) = \cdots = \alpha_{j_{(i+1)}}(x) \neq \alpha_{j_{(i+1)}+1}(x)$.

(*) As $j_i < j_{i+1}$ for all $i \in \mathbb{Z}^+$ and $S(x) \subseteq \mathbb{N}$ is infinite, there exists some smallest $j_i \geq n$. Denote this j_p. By the above, $\alpha_n(x) = \alpha_{j_p}(x)$.

As $\alpha_{n+1}(x) \neq \alpha_n(x)$, we have $\alpha_{n+1}(x) = 3\alpha_{j_p}(x) + \frac{1}{\beta_{j_p}(x)}$.

As $\frac{1}{\beta_{j_p}(x)} = 2^{k_{j_p}} = 2^{a_p}$, $\alpha_{n+1}(x) = 3\alpha_n(x) + 2^{a_p}$, which satisfies this equation.

Similarly to (*) (noting that j_i gives indices at which $\alpha_i(x)$ changes) we can observe that $2^{a_{m_n}} = 2^{a_{p-1}}$. As a_i strictly increases, $m_n = p - 1$ and $p = m_{n+1} = m_n + 1$.

Therefore, if $\alpha_n = 3(\cdots 3(3(3x + 2^{a_1}) + 2^{a_2}) + 2^{a_3})\cdots) + 2^{a_{m_n}}$, then $\alpha_{n+1} = 3(\cdots 3(3(3x + 2^{a_1}) + 2^{a_2}) + 2^{a_3})\cdots) + 2^{a_{m_{n+1}}}$.

This concludes the proof of the inductive step.

□

Observe that $m_{n+1} \in \{m_n, m_n + 1\}$ for all $n \in \mathbb{Z}^+$ and $m_i \geq 1$ for all valid $i \in \mathbb{Z}^+$ (with $i > j_1(x)$). Thus, $(m_n(x))_{i \in \mathbb{Z}^+}$ is a non-decreasing sequence of positive integers.

By the induction above, it can be seen that $\alpha_n = 3(\cdots 3(3(3x + 2^{a_1}) + 2^{a_2}) + 2^{a_3})\cdots) + 2^{a_{m_n}}$ for all $n \in \mathbb{Z}^+$.

This concludes the proof of Theorem 2.

□

The above definition of $(m_n(x))_{n \in \mathbb{Z}^+}$ is used in the following Corollaries.

Corollary 3. *Let x be a positive integer. If, for some $n \in \mathbb{Z}^+$, $C^{(n)}(x) = 1$ ($\mathrm{OCC}(x)$ is TRUE), then we have*

$$3(\cdots 3(3(3x + 2^{a_1(x)}) + 2^{a_2(x)}) + 2^{a_3(x)})\cdots) + 2^{a_{m_n(x)}(x)} = 2^{k_n(x)} \tag{1}$$

Proof. If $C^{(n)}(x) = 1$, then $\alpha_n(x)\beta_n(x) = 1$.

By Theorem 2, $\alpha_n(x) = 3(\cdots 3(3(3x + 2^{a_1(x)}) + 2^{a_2(x)}) + 2^{a_3(x)})\cdots) + 2^{a_{m_n(x)}(x)} = \frac{1}{\beta_n(x)} = 2^{k_n(x)}$.

This immediately concludes the proof.

□

Remark 10. We can also show that, from our choice of definitions, $k_n(x) = n - m_n(x)$ for all $x, n \in \mathbb{Z}^+$. This proof is not directly relevant to the paper. Hence, it is provided in Appendix 13.

Corollary 4. *Let x be some positive integer. For all $k \in \mathbb{Z}^+$, there exists some positive integer n for which $m_n(x) = k$.*

Proof. Let us fix some x.

As $m_{n+1}(x) \in \{m_n(x), m_n(x)+1\}$, it suffices to show there exists no maximum value in the sequence $(m_n(x))$. This is proven by contradiction.

Suppose there existed such a value, denoted $m_p(x)$.

For all positive integers $n > p$, this implies $\alpha_{n+1}(x) = \alpha_n(x)$ and therefore $C^{(n+1)}(x) = \frac{C^{(n)}(x)}{2}$.

As $C^{(n+1)}(x)$ is finite and $C^{(k)} \in \mathbb{Z}^+$ for all $k \in \mathbb{Z}^+$, this is impossible (a contradiction).

This concludes the proof.

□

Remark 11. Theorem 2 implied Equation (1) from $\mathrm{OCC}(x)$. The converse is proven in Theorem 5. The proof works by showing that the satisfaction of Equation (2), analogous to (1), is only possible with complete adherence to the 'rules' of $C(n)$. This proves that (2) being true and $\mathrm{OCC}(x)$ are actually logically equivalent (the crux of the simplification).

Theorem 5. *Let x be a positive integer. If, for some $k \in \mathbb{Z}^+$ and some increasing sequence $b_1, b_2, \cdots, b_m \in \mathbb{Z}^+$, we have*

$$3(\cdots 3(3(3x + 2^{b_1}) + 2^{b_2}) + 2^{b_3})\cdots) + 2^{b_m} = 2^k \tag{2}$$

then $\mathrm{OCC}(x)$ is TRUE.

Remark 12. Note that equation (2) is a converse of (1), as this equation is always satisfied when $\mathrm{OCC}(x)$ is TRUE, as $a_i(x)$ is the required value for all b_i.

Proof of Theorem 5. Let us fix an x.

Outline: This statement is proved by first showing that sequences $(\omega_n)_{n\in\mathbb{Z}_{\geq 0}}$ and $(\tau_n)_{n\in\mathbb{Z}_{\geq 0}}$ (defined below) can always be constructed such that there exists some $p \in \mathbb{Z}^+$ with $\omega_p \tau_p = 1$. It is then proven that $\omega_n = \alpha_n(x)$ and $\tau_n = \beta_n(x)$, allowing us to prove that $\mathrm{OCC}(x)$ is TRUE.

For this x and sequence (b_i), let $(\omega_n)_{n\in\mathbb{Z}_{\geq 0}}$ be a finite sequence of positive integers.

$$\omega_0 = x$$

$$\omega_{n+1}(x) = \begin{cases} 3\omega_n(x) + 2^{b_i}, & \text{if } n = b_i + i - 1 \text{ for any } i \in \mathbb{Z}^+ \\ \omega_n(x), & \text{otherwise} \end{cases}$$

For this x, let $(\tau_n)_{n\in\mathbb{Z}_{\geq 0}}$ be a sequence of rational numbers.

$$\tau_0 = 1$$

$$\tau_{n+1} = \begin{cases} \tau_n, & \text{if } n = b_i + i - 1 \text{ for any } i \in \mathbb{Z}^+ \\ \frac{\tau_n}{2}, & \text{otherwise} \end{cases}$$

Claim 5.1. *By this definition, we have*

$$\omega_{b_m+m} = 3(\cdots 3(3(3x + 2^{b_1}) + 2^{b_2}) + 2^{b_3})\cdots) + 2^{b_m}$$

Proof of Claim 5.1. As $b_i < b_{i+1}$ for all integers $1 \leq i < m$, there are exactly m distinct values of n such that $\omega_{n+1} \neq \omega_n$, by the recursive definition of ω_n.

Therefore, this expansion is a direct consequence of the recursive definition of (ω_n).

This concludes the proof of Claim 5.1.

□

Define $p \in \mathbb{Z}^+$ as the least positive integer for which $\tau_p = \frac{1}{2^k}$ (with k given in the Theorem statement). This is always well defined, as shown below.

Claim 5.2. *There always exists some $p \in \mathbb{Z}^+$ such that $\tau_p = \frac{1}{2^k}$, with $p > b_m + m$.*

Proof of Claim 5.2. By the definition of (τ_n), it can be seen that $\tau_{b_m+m} = \frac{1}{2^{b_m}}$.

From the theorem hypothesis, as $3(\cdots 3(3(3x + 2^{b_1}) + 2^{b_2}) + 2^{b_3})\cdots) + 2^{b_m} = 2^k$, we must have $2^{b_m} < 2^k$.

Note that b_m is the maximum of (b_i). As $\tau_{n+1} = \frac{\tau_n}{2}$ for all integers $n > b_m + m$ (a geometric sequence with ratio $\frac{1}{2}$) and both $b_m, k \in \mathbb{Z}^+$. As $\tau_{b_m+m} \geq \frac{1}{2^{b_m}}$ there exists at least one positive integer $p > b_m + m$ such that $\tau_p = \frac{1}{2^k}$.

Hence, there is some least p.

This concludes the proof of Claim 5.2.

□

Note that $\omega_{n+1} = \omega_n = 2^k$ for all $n \geq b_m + m$ (as b_m is the maximum term of (b_i)). Also note that $\omega_{b_m+m} = 3(\cdots 3(3(3x + 2^{b_1}) + 2^{b_2}) + 2^{b_3})\cdots) + 2^{b_m} = 2^k$, by the hypothesis of Theorem 5. Thus, by Claim 5.2, there exists a $p > b_m + m$ for which $\omega_p \tau_p = 1$.

The following claims will prove that $\omega_n = \alpha_n(x)$ and $\tau_n = \beta_n(x)$ for all $n \leq p \in \mathbb{Z}^+$.

Claim 5.3. *$\omega_n \tau_n \in \mathbb{Z}^+$ for all $n \in \mathbb{Z}^+$ with $n \leq p$.*

Proof of Claim 5.3. Given $\omega_0 = x \in \mathbb{Z}^+$ and $\omega_{n+1} \in \{3\omega_n + 2^{b_n}, \omega_n\}$, it is always true that $\omega_n \in \mathbb{Z}^+$, for all $n \in \mathbb{Z}^+$.

The proof will proceed by contradiction.

Suppose there exists some $q \leq p \in \mathbb{Z}^+$ such that $\omega_q \tau_q \notin \mathbb{Z}^+$.

Then there exist some $c \in \mathbb{Z}_{\geq 0}$ and $d \in \mathbb{Z}^+$ such that $\{\omega_q \tau_q\} = \{\frac{3^c}{2^d}\}$.

Thus, for any $n \in \mathbb{Z}^+$ with $n \geq q$, there exist some $c \in \mathbb{Z}_{\geq 0}$ and $d \in \mathbb{Z}^+$ such that $\{\omega_n \tau_n\} = \{\frac{3^c}{2^d}\}$.

Thus, for all $n \geq q \in \mathbb{Z}^+$, $\omega_n \tau_n \notin \mathbb{Z}^+$.

By the hypothesis of Theorem 5 and Claim 5.2, it must be true that $\omega_p \tau_p = 1$. However, as $q \leq p$, this is a contradiction.

This concludes the proof of Claim 5.3.

□

Note that by definition, if and only if $\tau_{n+1} = \tau_n$, then $\omega_{n+1} \neq \omega_n$. This will be used in the following proof.

Claim 5.4. *For all $n \in \mathbb{Z}^+$ with $n < p$, if $\omega_n \tau_n \equiv 1 \pmod 2$, then $\tau_{n+1} = \tau_n$.*

Proof of Claim 5.4. This will be proved by contradiction.

For some $n < p \in \mathbb{Z}^+$ where $\omega_n \tau_n \equiv 1 \pmod 2$, suppose $\tau_{n+1} = \frac{\tau_n}{2}$.

As $\tau_{n+1} \neq \tau_n$, by the recursive definition, $\omega_{n+1} = \omega_n$. Given $\omega_n \tau_n \equiv 1 \pmod 2$, $\{\omega_{n+1} \tau_{n+1}\} = \frac{1}{2}$

As $n + 1 \leq p$, this leads to a contradiction by Claim 5.3.

This concludes the proof of Claim 5.4.

□

Claim 5.5. *For any $n < p \in \mathbb{Z}^+$, if $\omega_n \tau_n \equiv 0 \pmod 2$, then $\omega_{n+1} = \omega_n$.*

Proof of Claim 5.5. This can also be proved through contradiction.

For some $n < p \in \mathbb{Z}^+$ where $\omega_n \tau_n \equiv 0 \pmod 2$, suppose $\omega_{n+1} \neq \omega_n$.

By the recursive definition of (ω_n) and (τ_n), note that $\omega_{n+1} \tau_{n+1} \equiv 1 \pmod 2$

By Claim 5.4, $\omega_{n+1} \neq \omega_{n+2}$.

Thus, $\omega_n < \omega_{n+1} < \omega_{n+2}$.

This implies there exists some $i \in \mathbb{Z}^+$ such that $b_i = b_{i+1}$, since $b_i + i - 1 + 1 = b_{i+1} + (i + 1) - 1$.

This is a contradiction as the sequence (b_i) is required to be strictly increasing.

This concludes the proof of Claim 5.5.

□

From the above, we see that Claims 5.3, 5.4, and 5.5 prove that $\omega_n = \alpha_n(x)$ and $\tau_n = \beta_n(x)$ for all $n \leq p \in \mathbb{Z}^+$.

By Claim 5.2, there exists some $p \in \mathbb{Z}^+$ such that $\omega_p \tau_p = 1$, this implies $\alpha_p(x)\beta_p(x) = 1$ and $C^{(p)}(x) = 1$. Thus, $\mathrm{OCC}(x)$ is TRUE.

This concludes the proof of Theorem 5.

□

Therefore, by Theorems 2 and 5, the satisfaction of (2) for some $x \in \mathbb{Z}^+$ and any $n \in \mathbb{Z}^+$ is logically equivalent to $\mathrm{OCC}(x)$.

2.3 Analysing powers of 2

Definition 13. For all $x \in \mathbb{Z}^+$, let $\Gamma(x) \coloneqq 2^k$, where k is the largest integer for which $2^k | x$.

Noting the structure of the equation, we are motivated to define the following functions A, B, and R:

Definition 14.

(i) Define $A\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ such that $A(x) = 3x$

(ii) Define $B\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ such that $B(x) = x + \Gamma(x)$

(iii) Finally, define the composition $R\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ such that $R(x) = B(A(x))$

Remark 15. This definition of B is analogous to defining $B(x) = x + 2^{a_1}(x)$, as $a_1(x) = \log_2(\Gamma(x))$. The proof of this can be found in Theorem 12 in the Appendix. This result will be used in a subsequent proof.

Remark 16. It must be noted that as B is applied repeatedly, the exponent of the added power of two will increase. Specifically, $\Gamma(B^{(n+1)}(x)) > \Gamma(B^{(n)}(x))$ for all $x, n \in \mathbb{Z}^+$. This follows from the definition of Γ (and is easier to see in binary).

Theorem 6. *Let x be a positive integer for which $\mathrm{OCC}(x)$ is TRUE. For all positive integers $n > j_1(x)$, we have $\alpha_n(x) = R^{(m_n(x))}(x)$.*

Remark 17. Much like Theorem 2, this is trivially true for $n \leq j_1(x)$ as $R^{(0)}(x) = \alpha_n(x) = x$.

The proof of Theorem 6 will first require Lemma 7 to be proven.

Lemma 7. *Let x, n be any positive integers for which $\mathrm{OCC}(x)$ is TRUE. Suppose $\alpha_{n+1}(x) \neq \alpha_n(x)$. Then, $2^{a_{m_n(x)}(x)} = \Gamma(\alpha_n(x))$.*

Proof. This will be proved by contradiction.

For clarity, we shall refer to $a_i(x)$ as a_i, $k_i(x)$ as k_i, $j_i(x)$ as j_i, $\alpha_n(x)$ as α_n, and $\beta_n(x)$ as β_n.

First, observe that since $\mathrm{OCC}(x)$ is TRUE there exists some smallest $p \in \mathbb{Z}^+$ with $p > n$ such that $\alpha_p\beta_p = 1$ (there exist infinitely many suitable values for p, by the 4-2-1 loop).

Let us fix some positive integer x.

Given $\mathrm{OCC}(x)$ is TRUE, by Theorem 2, α_p is an integer power of 2.

Claim 7.1. $2^{a_{m_n}} \not> \Gamma(\alpha_n)$.

Proof. Assume $2^{a_{m_n}} > \Gamma(\alpha_n)$

Hence, $2\Gamma(\alpha_n) | 2^{a_{m_n}}$ but $2\Gamma(\alpha_n) \nmid \alpha_n$. Therefore, $\alpha_{n+1} \equiv \Gamma(\alpha_n) \pmod{2\Gamma(\alpha_n)})$.

Note that $2^{a_{i+1}} > 2^{a_i}$ for all $i \in \mathbb{Z}^+$ (by Theorem 1). Therefore, for all $q \in \mathbb{Z}^+$ with $q > n$, $2^{a_{m_q}} \equiv 0 \pmod{2^{a_{m_n}+1}})$.

Thus, note that $\alpha_q \equiv \Gamma(\alpha_n) \pmod{2\Gamma(\alpha_n)})$, for all integers $q > n$.

As $p > n$, we have $\alpha_p \equiv \Gamma(\alpha_n) \pmod{2\Gamma(\alpha_n)})$.

Note that α_p and $\Gamma(\alpha_n)$ are integer powers of 2. Hence, as $\Gamma(\alpha_n) < \alpha_p$, $\alpha_p \equiv 0 \pmod{2\Gamma(\alpha_n)})$. However, this contradicts $\alpha_p \equiv \Gamma(\alpha_n) \pmod{2\Gamma(\alpha_n)})$.

This concludes the proof of Claim 7.1.

□

Claim 7.2. $2^{a_{m_n}} \not< \Gamma(\alpha_n)$.

Proof. Like the above claim, assume (for contradiction) that $2^{a_{m_n}} < \Gamma(\alpha_n)$.

As $2^{a_{m_n}} < \Gamma(\alpha_n)$ and $\Gamma(\alpha_n) | \alpha_n$ it must be true that $\alpha_n \equiv 0 \pmod{2^{a_{m_n}+1}}$ and $\alpha_{n+1} \equiv 2^{a_{m_n}} \pmod{2^{a_{m_n}+1}}$.

Note that $2^{a_{i+1}} > 2^{a_i}$ for all $i \in \mathbb{Z}^+$ (by Theorem 1). Therefore, for all $q \in \mathbb{Z}^+$ with $q > n$, $2^{a_{m_q}} \equiv 0 \pmod{2^{a_{m_n}+1}}$.

As $3x \equiv x \pmod 2$, we have $\alpha_q \equiv 2^{a_{m_n}} \pmod{2^{a_{m_n}+1}}$ for all integers $q > n$.

Since $p > n$, we have $\alpha_p \equiv 2^{a_{m_n}} \pmod{2^{a_{m_n}+1}}$.

Note that α_p and $2^{a_{m_n}+1}$ are integer powers of 2. As $2^{a_{m_n}+1} < \alpha_p$, this implies that $\alpha_p \equiv 0 \pmod{2^{a_{m_n}+1}}$. However, this contradicts $\alpha_p \equiv 2^{a_{m_n}} \pmod{2^{a_{m_n}+1}}$.

This concludes the proof of Claim 7.2.

□

By Claims 7.1 and 7.2, $2^{a_{m_n(x)}(x)} = \Gamma(\alpha_n(x))$. This concludes the proof of Lemma 7.

□

Now, we prove Theorem 6, using Lemma 7.

Proof of Theorem 6. Fix some positive integer x. This statement will be proven inductively.

For clarity, we will implicitly refer to $\alpha_n(x)$ with α_n, $j_i(x)$ with j_i, and $m_n(x)$ with m_n.

Base Case ($n = j_1 + 1$)*:* $\alpha_{j_1+1} = 3x + 2^{a_1}$. We have $2^{a_1} = \Gamma(x)$ (by Theorem 12 in Appendix A) and $\Gamma(x) = \Gamma(3x)$. Hence, $\alpha_{j_1+1} = R^{(m_{(j_1+1)})}$. This immediately gives the required result.

Inductive step: If $\alpha_n = R^{(m_n)}$ for some $n \geq j_1 + 1$, then $\alpha_{n+1} = R^{(m_{n+1})}$.

Proof of the inductive step. Observe, from the proof of Theorem 2, that $m_{n+1} \in \{m_n + 1, m_n\}$.

Case 1 ($m_{n+1} = m_n$)*:* This is immediately true for x, as $\alpha_{n+1} = \alpha_n$ (by the general rule).

Case 2 ($m_{n+1} = m_n + 1$)*:*
Observe that $\alpha_{n+1} = 3\alpha_n + 2^{a_{m_n}}$, by the general rule.

As $a_{m_n} \in \mathbb{Z}^+$ and is well-defined for $n > j_1$, there exists some positive integer r such that $\alpha_{n+1} = 3\alpha_n + 2^r$.

Note that $\mathrm{OCC}(x)$ is TRUE. Hence, by Lemma 7, we see that $2^r = \Gamma(x)$.

Observing that $\Gamma(x) = \Gamma(3x)$ for all $x \in \mathbb{Z}^+$, $\alpha_{n+1}(x) = 3\alpha_n(x) + \Gamma(3\alpha_n(x)) = R^{(m_{n+1})}(x)$ (by the inductive hypothesis).

This concludes the proof of the inductive step.

□

From the induction, it can be seen that $\alpha_n(x) = R^{(m_n(x))}(x)$ for all $x \in \mathbb{Z}^+$ for which $\mathrm{OCC}(x)$ is TRUE.

This concludes the proof of Theorem 6.

□

Note that Theorem 6 (in conjunction with Theorem 2) shows that $OCC(x)$ is TRUE implies the existence of some positive integer pair (n, k) such that $R^{(n)}(x) = 2^k$ (for all $x \in \mathbb{Z}^+$). The converse is also true by the use of Theorem 5, with $\Gamma(\alpha_i(x))$ used as the values for each b_i. This is valid, observing that $\Gamma(R(x)) > \Gamma(3x)$ for all $x \in \mathbb{Z}^+$.

In a similar manner, we also have the converse of Theorem 6 being true.

Therefore, the following are equivalent for any arbitrary positive integer x:

1. $\mathrm{OCC}(x)$.

2. Existence of an increasing sequence $b_1, b_2, ...b_m$ of positive integers and some $k \in \mathbb{Z}^+$ such that $3(\cdots 3(3(3x + 2^{b_1}) + 2^{b_2}) + 2^{b_3})\cdots) + 2^{b_m} = 2^k$.

3. Existence of a pair of positive integers (n, k) such that $R^{(n)}(x) = 2^k$.

Remark 18. By the nature of the 4-2-1 loop, there are actually infinite pairs of positive integers (n, k) for which $R^{(n)}(x) = 2^k$ (for all $x \in \mathbb{Z}^+$ such that $\mathrm{OCC}(x)$ is TRUE).

To summarize this section, R involves repeated alternating multiplication by 3 and addition of the largest dividing integer power of 2. By Theorem 6 and Corollary 4, for any positive integer x, if and only if $\mathrm{OCC}(x)$ is TRUE, there exists some pair $(n, k) \in \mathbb{Z}^+ \times \mathbb{Z}^+$ such that $R^{(n)}(x) = 2^k$.

3 Binary and `Changes`

In order to better represent and analyze the addition of powers of 2, we are motivated to use a binary (base-2) representation and investigate the strings of 1s and 0s.

For any positive integer x, the Collatz Conjecture is satisfied for $x \in \mathbb{Z}^+$ if and only if there exists an ordered pair $(n, k) \in \mathbb{Z}^+ \times \mathbb{Z}^+$ such that $R^{(n)}(x) = 2^k$.

A binary number will also be represented as a string of 1s and 0s, beginning with a 1 (leading 0s truncated).

3.1 Analysing numbers with `Changes`

We introduce $\texttt{Changes}(x)$ of a positive integer x to analyze binary strings in an easy-to-compute manner that can easily represent the desired output of the algorithm (if $OCC(x)$ is TRUE, then `Changes` eventually goes to 0 or 1). More specific properties we observe with `Changes` will also be discussed.

Definition 19 (Changes)**.** Let $d_k, d_{k-1}, d_2, d_1, \cdots, d_0$ be the digits of x in base-2 (in order of decreasing value). $d_k = 1$, as leading 0s are omitted.

Define $(\gamma_n)_{n\in\mathbb{Z}_{\geq 0}}$ such that:

$$\gamma_n = \begin{cases} 0, & \text{if } d_n = d_{n+1} \\ 1, & \text{if } d_n \neq d_{n+1} \end{cases}$$

Then, $\texttt{Changes}(x) = \sum_{i=0}^{k-1} \gamma_i$

Corollary 8. *For $x \in \mathbb{Z}^+$, $\texttt{Changes}(x) \leq 1$ if and only if $B(x) = 2^k$ for some $k \in \mathbb{Z}^+$.*

Proof. If and only if x is of the form 111....1000....0, 111...111, or 100...000 is it true that $B(x) = 2^k$. We have $x = 2^k$ or $x = 2^k - 2^m$ for some nonnegative integer m with $m < k$. Hence, $\texttt{Changes}(x) = 0, 1$. □

Remark 20. The motivation behind the use of Changes is that the property works well with B and Γ (as elaborated below) and is relatively easy to compute. Further note that powers of 2 greater than 1 will always have 1 change, the desired output of the earlier algorithm. When R is applied, the growth rates of the trailing string of 0s can be considered. This average growth rate when applying B, for instance, can potentially be approximated using Changes (if we can show that this is greater than the growth rate of the main string, this proves the conjecture).

Remark 21. We note that $\texttt{Changes}(R(x)) = \texttt{Changes}(C(x))$ or $\texttt{Changes}(R(x)) = \texttt{Changes}(C(x)) + 1$ for any $x \in \mathbb{Z}^+$, because $R(x) = C(x)2^k$ for some $k \in \mathbb{Z}_{\geq 0}$.

Theorem 9. *For all* $x \in \mathbb{Z}^+$, $\texttt{Changes}(B(x)) \leq \texttt{Changes}(x)$.

For example, $x = 10001110$ has $\texttt{Changes}(x) = 3$ and $\texttt{Changes}(B(x)) = 3$.

For $x = 10111$, we have $\texttt{Changes}(x) = 2$ and $\texttt{Changes}(B(x)) = 1$.

Proof of Theorem 9. Consider the unique binary representation of x, which can be constructed for any $x \in \mathbb{Z}^+$:

$$x = \sum_{i=0}^{s} b_i 2^i \text{ with all } b_i \in \{0, 1\}, i \in \mathbb{Z}_{\geq 0}$$

Similarly, construct y with $y = B(x) = x + \Gamma(x)$:

$$y = \sum_{i=0}^{t} c_i 2^i \text{ with all } c_i \in \{0, 1\}, i \in \mathbb{Z}_{\geq 0}$$

Let $k = \log_2 \Gamma(x)$. By definition of $\Gamma(x)$, $k \in \mathbb{Z}_{\geq 0}$

It is apparent that $b_k = 1$. As $2^k | x$, for all $j \in \mathbb{Z}^+$ with $j < k$, we have $b_j = 0$.

Therefore, as $y = x + 2^k$, we also have $c_j = 0$ for all $j \in \mathbb{Z}^+$ with $j \leq k$.

Let d be the largest positive integer such that $b_d = b_{d-1} = \cdots = b_{k+1} = b_k = 1$ and $b_{d+1} = 0$.

The value d (possibly equal to k) is well-defined, as we have $b_k = 1$ and a finite number of digits.

Thus, $c_{d+1} = 1, c_d, ..., c_0 = 0$, and $c_{d+2} = a_{d+2}$. Also, $c_i = b_i$ for all $j \geq d + 2, j \in \mathbb{Z}^+$.

As $k \neq 0$:

$$\texttt{Changes}(x) - \texttt{Changes}(B(x)) = \begin{cases} 1, & \text{if } c_{d+2} = 1 \\ 0, & \text{if } c_{d+2} = 0 \end{cases}$$

Thus, $\texttt{Changes}(B(x)) \leq \texttt{Changes}(x)$, which concludes the proof of Theorem 9.

□

Remark 22. The process explored indicates the behaviour of the collapse of substrings of 1s to a single digit (or a bit shift to the left).

Remark 23. Consider some $x \in \mathbb{Z}^+$ and apply R. Note that in $R(x)$, we have that the string of trailing zeroes increases in length by at least one compared to x (in other words, $\Gamma(R(x)) > \Gamma(x)$). However, this increase is often more than one, as substrings of 1s collapse to single digits. `Changes` can potentially provide some further insight as to this number of digits on average (i.e. lower the changes, more the digits on average, and vice versa). In fact, if this growth rate is larger than around $\log_2(3)$ on average (which is an approximation of the growth rate of the main string), the conjecture may be proven true (possible extension of this work). Later, more patterns regarding `Changes` in $R(x)$ and $A(x)$ will be discussed and proven.

Thus $(B(x))$ can either reduce or maintain the 'changes' present in the binary representation. Now, we attempt to analyze the behavior of `Changes` when A is applied. Since $R = B \circ A$, this helps us evaluate $\texttt{Changes}(R^{(n)}(x))$ for all $x, n \in \mathbb{Z}^+$.

3.2 Upper Limit to Growth of `Changes`

It can be shown that there is an upper limit to increases in changes of $A(x)$, based on the number of digits, for all $x \in \mathbb{Z}^+$. First, we define the following:

Definition 24. For every $x \in \mathbb{Z}^+$, we use $d(x)$ to refer to the number of digits of x in binary. In other words, $2^{d(x)} \leq x < 2^{d(x)+1}$.

Remark 25. Some efforts have been made to study the relationship between the values of $d(x)$ and the Collatz Conjecture [2]. They arrive at other interesting results that can possibly be extended using `Changes`.

Now, since the application of B must reduce or maintain changes, we can attempt to determine how $\texttt{Changes}(A(x))$ relates to $\texttt{Changes}(x)$, so we can therefore ultimately consider R, which is $B \circ A$.

Definition 26. For all $k \in \mathbb{Z}^+$, h_k is defined to be the least $h \in \mathbb{Z}^+$ such that, for all x with $d(x) = k$:

$$\text{If } \texttt{Changes}(x) > h, \text{ then } \texttt{Changes}(A(x)) < \texttt{Changes}(x).$$

Noting that $h_k \leq k - 1$ (which is a suitable value of h) for all $k \in \mathbb{Z}^+$, h_k is always well-defined.

Now, we can prove some stronger results with $\texttt{Changes}$ and h_k. The following theorem was first verified computationally for all positive integers $x \leq 2^{32}$ (details in the Appendix), then proven rigorously, as shown below.

Theorem 10. *For all $k \in \mathbb{Z}^+$, $h_k = \lfloor \frac{2k-1}{3} \rfloor$.*

The proof of the theorem will rely on the following definitions.

Definition 27. For some $x \in \mathbb{Z}^+$ (with binary representation $a_k a_{k-1} \dots a_0$), refer to some $i \in \mathbb{Z}^+$ as a *change-point* if $a_i \neq a_{i+1}$. Hence, the number of change points for x is $\texttt{Changes}(x)$.

Definition 28. Refer to i as a *mutual change-point* of x, y if it is a change-point for both x and y.

Remark 29. This proof will proceed by considering $A(x) = 3x$ as $2x + x$ (in binary, this is a bit shift to the right). Therefore, consider the number of instances where changes must overlap (by Pigeonhole). By showing that $a_i = a_{i+1}$ wherever this occurs (where a_i are the digits), we show that $\texttt{Changes}(A(x)) < \texttt{Changes}(x)$.

Proof of Theorem 10. Fix some $x \in \mathbb{Z}^+$, with binary expansion $a_k a_{k-1} \dots a_0$ and $\texttt{Changes}(x) > \lfloor \frac{2k-1}{3} \rfloor$.

Consider $x + 2x$, and the set $P = \{i | i < k, i \text{ is a mutual change-point of } x \text{ and } 2x\}$.

Claim 10.1. *For this x, if $i \in P$, i is not a change-point of $3x$.*

Proof of Claim 10.1. Let the binary expansion of $2x$ be $b_{k+1} b_k \dots b_1 b_0$, with $b_{i+1} = a_i$ and $b_0 = 0$.

Thus, let the binary representation of $3x$ be $c_k c_{k-1} \dots c_0$.

Consider some $i \in P$.

Note that if i is a change point in x, then $i + 1$ is a change point in $2x$.

We have $(a_i, a_{i+1}, b_i, b_{i+1}) = (0, 1, 1, 0), (1, 0, 0, 1)$.

With $c_0, c_1, \dots$ being the digits of $3x$, either $c_i = c_{i+1} = 1$ (no carry) or $c_i = c_{i+1} = 0$ (carry).

Therefore, if $i \in P$, i is not a change point of $3x$.

This concludes the proof of Claim 10.1.

□

Let us assume $\texttt{Changes}(x) > \lfloor \frac{2k-1}{3} \rfloor$. By the pigeonhole principle, $|P| \geq 2(\lfloor \frac{2k-1}{3} \rfloor + 1) - k$ (these values of i are mutual change points).

Now, we can count change points of $3x$, applying Claim 10.1. Noting that $3x$ has $k+1$ digits, we have $\texttt{Changes}(3x) \leq (k+1) - (2(\lfloor \frac{2k-1}{3} \rfloor + 1) - k) \leq \lfloor \frac{2k-1}{3} \rfloor$.

Therefore, $\texttt{Changes}(3x) \leq \lfloor \frac{2k-1}{3} \rfloor < \texttt{Changes}(x)$, when $\texttt{Changes}(x) > \lfloor \frac{2k-1}{3} \rfloor$.

Therefore, $\lfloor \frac{2k-1}{3} \rfloor$ is a possible value for h, so $h_k \leq \lfloor \frac{2k-1}{3} \rfloor$.

Claim 10.2. *For every k, there exists at least one $m \in \mathbb{Z}^+$ such that we have $\texttt{Changes}(x) = \lfloor \frac{2k-1}{3} \rfloor$ as well as $\texttt{Changes}(A(x)) = \texttt{Changes}(x)$.*

Proof of Claim 10.2. We claim that the following values of x satisfy the claim, depending on the value of k.

i If $k \equiv 0 \pmod 3$, let $x = \sum_{i=1}^{i=k/3} 2^{3i-1}$. Thus $x = 100100...100_2$.

ii If $k \equiv 1 \pmod 3$, let $x = \sum_{i=0}^{i=\lfloor k/3 \rfloor} 2^{3i}$. Thus $x = 100100...1001_2$.

iii If $k \equiv 2 \pmod 3$, let $x = \sum_{i=0}^{i=\lfloor k/3 \rfloor} 2^{3i+1}$. Thus $x = 100100...10010_2$

It can then be noted that for each case, $\texttt{Changes}(x) = \lfloor \frac{2k-1}{3} \rfloor$, directly from the definition.

For each of the cases, respectively, we can determine $3x$:

i If $k \equiv 0 \pmod 3$, we have $3x = \sum_{i=1}^{i=k/3} 2^{3i} + 2^{3i-1} = 110110...1100_2$.

ii If $k \equiv 1 \pmod 3$, we have $3x = \sum_{i=0}^{i=\lfloor k/3 \rfloor} 2^{3i+1} + 2^{3i} = 110110...11011_2$.

iii If $k \equiv 2 \pmod 3$, we have $3x = \sum_{i=0}^{i=\lfloor k/3 \rfloor} 2^{3i+2} + 2^{3i+1} = 110110...110110_2$.

Thus, $\texttt{Changes}(3x) = \texttt{Changes}(x)$, as required. This concludes the proof of Claim 10.2 □

From Claim 10.2, it must also be true that $h_k \geq \lfloor \frac{2k-1}{3} \rfloor$.

Using the previous statement, equality must be true, with $h_k = \lfloor \frac{2k-1}{3} \rfloor$.

This concludes the proof of Theorem 10.

□

Remark 30. Thus, for k-digit (in binary) values of x with more than h_k changes, we have $\texttt{Changes}(A(x)) < \texttt{Changes}(x)$ (limiting the growth of changes as $A(x)$ is applied).

Remark 31. This (alongside the fact that $\texttt{Changes}(B(x)) \leq \texttt{Changes}(x)$) suggests that growth in `Changes` is possibly bounded as R is applied. Specifically, it is not uncommon for an increase in the `Changes` on application of R to be immediately followed by a decrease in the `Changes` and vice versa, in part for this reason (Theorem 10).

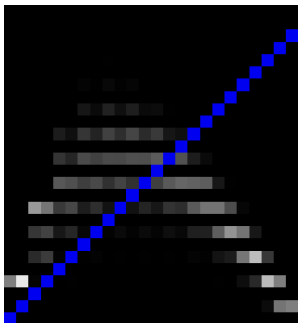
Analyzing this behavior may explain how all numbers appear to attain $\texttt{Changes}(R^{(n)}(x)) \leq 1$ (proving this would reduce the Collatz conjecture to a much more trivial case, only for x of the form $2^k - 1$). Thus, `Changes` could present a major step in proving the Collatz conjecture.

3.3 Extension: Distribution of `Changes`

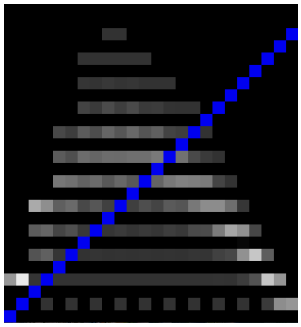
Figure 1 plots $\texttt{Changes}(x)$ (x-coordinate) to $\texttt{Changes}(R(x))$ (y-coordinate) for all k-digit binary x (in this Figure, $k = 24$). The line corresponding to $\texttt{Changes}(x) = \texttt{Changes}(R(x))$ is also shown here in blue. Pixel brightness is given by relative frequency of each configuration of x-values and y-values (normalized by number of occurrences per column) for Figure (a). In Figure (b), the brightness of the pixels is similarly defined (also for $k = 24$) although all nonzero values are assigned a base brightness of RGB (50, 50, 50) (for better visibility).

Remark 32. Notably, this distribution appears nearly uniform for arbitrarily large values of k up to 24 (the highest tested in this article). We conjecture that this trend continues, and the distribution appears identical as k tends to infinity.

Remark 33. Note that this is also a similar pattern as to that for $\texttt{Changes}(x)$ against $\texttt{Changes}(C(x))$, because either $\texttt{Changes}(C(x)) = \texttt{Changes}(R(x))$ or $\texttt{Changes}(C(x)) = \texttt{Changes}(R(x)) - 1$.



(a) No Added Visibility



(b) Added Visibility (50, 50, 50) RGB

Figure 1: Distribution of $\texttt{Changes}(R(x))$ (y-axis) vs. $\texttt{Changes}(x)$ (x-axis) for $k = 24$

3.4 Extension: Generalizing `Changes`

Base-2 provides us with a number of perspectives when we consider `Changes`, since it allows us to analyze the growth rate of the trailing substring of zeros. But other bases, such as 4 and 8, are insightful as well.

Base-4 for instance, pairs up digits in the binary representation: Any substring of `101010` multiplied by three will be mapped to `101010`, which does not affect the changes in Base-4. Moreover, multiplication by three in base-4 is equivalent to a bitwise shift followed by subtraction, and powers of two have predictable patterns in base-4, all of which are factors making this representation desirable.

Definition 34 (k-Changes)**.** Let $d_k, d_{k-1}, d_2, d_1, \cdots, d_0$ be the digits of x in base-k (in order of decreasing value). $d_k \neq 0$, as leading 0s are omitted.

$$\text{Define } \texttt{Changes}(x) = \sum_{i=0}^{k-1} |d_i - d_{i+1}|$$

From the definition, the desired end state of the repeated R algorithm has 4-`changes` ≤ 2.

Remark 35. We see similar interesting properties when considering the 4-`changes` of a number. This is because using base-4 has the same effect as grouping together adjacent bits in binary. For example, a perfectly alternating substring of the form `101010`... in base-2 will have 4-`changes` of 0, as will a substring of the form `111111`.... This is, of course, a very specific *example*, but the behavior also possibly extends to more general properties, an area of future exploration.

4 Conclusion

In conclusion, we define an alternative representation of the Collatz conjecture and prove its equivalence to the OCC. For all $x \in \mathbb{Z}^+ < 2^{32}$, it was found that $A(x)$ will always have fewer changes than x if $\texttt{Changes}(x)$ exceeds $\left\lfloor \frac{2k-1}{3} \right\rfloor$. This effectively bounds the potential growth in `Changes`. Moreover, the use of `Changes` provides a property that relates well with the growth of the trailing substring of zeroes (or ν_2) of values in R-orbits. Since this growth rate's average value for some x can determine whether OCC is true for that x, this novel and useful method of `Changes` is a significant step towards a solution of the Collatz conjecture, highlighting a new approach and potentially leading to a full proof in the future (this may also help identify patterns in Collatz-based encryption schemes).

The main contributions of the paper are the following:

(i) Proposing a representation of the initial Collatz Conjecture as an algorithm that can be applied efficiently in binary and involves elementary operations, where the desired final value has exactly 1 change (when $OCC(x)$ is TRUE).

(ii) Defining the function $\texttt{Changes}\colon \mathbb{Z}^+ \to \mathbb{Z}_{\geq 0}$ to analyze values in the representation of the Collatz Conjecture (above), which loosely relates to the growth rate of the trailing string of zeroes.

(iii) Proving growth in `Changes` to be restricted with an upper bound, with all k-digit binary integers x with $\texttt{Changes}(x) > h_k$ always decreasing in changes after applying $A(x)$ (for all integers $< 2^{32}$ in $\mathbb{Z}^+$).

(iv) Proving a mathematical relationship for h_k (verified computationally for $k \leq 32$).

(v) Constructing plots of the distribution of $\texttt{Changes}(R(x))$ against $\texttt{Changes}(x)$ for all k-digit binary $x \in \mathbb{Z}^+$. These distributions were generated for $2 \leq k \leq 24$.

Applications: The exploration of the Collatz Conjecture and its orbits has applications in various other fields, such as Blockchain [5], Pseudo-random Number generators [7], Digital Watermarking [3], and Encryption [1, 4]. This specific research allows us to track new patterns (e.g. behaviour of `Changes` under A and B) and create alternative encryption schemes based on the Collatz conjecture and similar dynamical systems.

Limitations: Though we were able to formally prove the bound in the growth of `Changes` using h_k, we weren't able to mathematically show all of the observed patterns in the distribution of changes. By computing this for larger values of k, we can verify that these patterns do indeed exist, in the distribution, and thus attempt to prove them. Secondly, we have observed (not proven yet) that `Changes` always decreases as the R algorithm is applied, but the time taken for this to happen hasn't been considered. Lastly, the development of `Changes` generally opens up a field of study and allows a multitude of patterns to be studied: this is only a minor subset of all the patterns that exist when considering `Changes` under the Collatz process.

Many of these identified areas of exploration are outlined in the 'Future Work' section.

5 Future Work

Some of the most pertinent questions are listed below. These are novel and interesting directions for further research on the conjecture and `Changes` theory:

1. When R is repeatedly applied to some values, what can be said about the length of the growing substring of trailing 0s (or ν_2 of these values)? If this average growth rate exceeds the growth rate of the string, this would nearly prove the conjecture. In particular, note that the growth rate of this string is dependent on `Changes`. Specifically, when the changes are smaller (when there are large uniform strings consisting of only 1s or 0s), we often see a higher growth rate (this is because each operation of $B(x)$ either skips past a string of 0s or collapses a long string of 1s into a single 1). For instance, $\Gamma(\mathtt{B}(\mathtt{10011111000})) > \Gamma(\mathtt{B}(\mathtt{100110110}))$.
2. Can we consider a 'stronger' version of changes in the form of a tuple, giving the lengths of each substring consisting of only one type of digit (for example `100100111` would have $(1, 2, 1, 2, 3)$).
3. Can we prove that the distribution in Section 3.3 converges at higher k? What would this mean?
4. Can we work backwards? This would involve using the same algorithm of $B(A(x))$, but using the inverse relations (B^{-1}, A^{-1}) to prove that any number can be attained with x initially being of the form 2^y?
5. How does this extend for analogs of the Collatz function (perhaps of the form $ax + 1$ instead of $3x + 1$)? When this method is generalized to other bases, do the same properties still hold?
6. As detailed in Remark 35, what similar patterns emerge when plotting the 4-`changes`? Do they help discriminate effectively between *perfectly alternating* and *perfectly uniform* substrings, offering a better way bound the growth of $\nu_2(R^{(n)}(x))$ in the Collatz conjecture?

References

[1] Renza Diego, Sebastián Mendoza, and Miguel Ballesteros. High-uncertainty audio signal encryption based on the collatz conjecture. *Journal of Information Security and Applications*, 46:62–69, 2019.

[2] Richard Kaufman. A Reduced Forward Collatz Algorithm: How Binary Strings Change Their Length Under 3x+1. *arXiv (Cornell University)*, 1 2023.

[3] Haoyu Ma, Chunfu Jia, Shijia Li, Wantong Zheng, and Dinghao Wu. Xmark: Dynamic software watermarking using collatz conjecture. *IEEE Transactions on Information Forensics and Security*, 14(11):2859–2874, 2019.

[4] Haoyu Ma, Shijia Li, Debin Gao, and Chunfu Jia. Secure repackage-proofing framework for android apps using collatz conjecture. *IEEE Transactions on Dependable and Secure Computing*, 19(5):3271–3285, 2022.

[5] Wei Ren, Simin Li, Ruiyang Xiao, and Wei Bi. Collatz conjecture for $2^{100000} - 1$ is true - algorithms for verifying extremely large numbers. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 411–416, 2018.

[6] Michael R. Schwob, Peter Jau-Shyong Shiue, and Rama Venkat. Novel theorems and algorithms relating to the collatz conjecture. *International Journal of Mathematics and Mathematical Sciences*, 2021:1–10, 2021.

[7] David S. Xu and Dan E. Tamir. Pseudo-random number generators based on the collatz conjecture. *International Journal of Information Technology*, 11(3):453–459, 2019.

A Appendix

This section details some other theorems that were observed and proven throughout the research. These proofs are more mechanical and similar to those presented earlier (and thus would confuse the structure of the paper), but are provided here as they are required to preserve rigor.

Theorem 11. *Let x be some positive integer. For all $i \in \mathbb{Z}^+$, $a_i(x) = j_i(x) - i + 1$.*

Proof. For clarity, we shall refer to $a_i(x)$ as a_i, $k_i(x)$ as k_i and $j_i(x)$ as j_i.

Fix some $i \in \mathbb{Z}^+$. Observe that $a_i = k_{j_i}$.

For all $0 \leq i < j_i$, $k_{i+1} = k_i + 1$ if and only if $\beta_i \neq \beta_{i+1}$ and $\alpha_i = \alpha_{i+1}$.

Let $P = \{i | i \in \mathbb{Z}^+, 0 \leq i < j_i, k_{i+1} \neq k_i\}$.

Observe that $|P| = j_i - i + 1$, by the definition of $j_i(x)$.

As it is always true that $k_{i+1} \in \{k_i, k_i + 1\}$ (it always increments by 1), it is immediately true that $a_i = k_{j_i} = j_i - i + 1$.

This concludes the proof. □

Theorem 12. *Let x be any positive integer. Then, we have $a_1(x) = \log_2(\Gamma(x))$.*

Proof. For clarity, we shall refer to $a_i(x)$ as a_i, $k_i(x)$ as k_i and $j_i(x)$ as j_i.

Observe that for all non-negative integers $i < j_1$, $\alpha_i(x)\beta_i(x) \equiv 0 \pmod 2$, with $\alpha_{j_1}(x)\beta_{j_1}(x) \equiv 1 \pmod 2$.

By definition, for all non-negative integers $i < j_1$, $\alpha_i(x) = \alpha_{i+1}(x) = x$. Therefore, $\beta_{i+1}(x) = \frac{1}{2}\beta_i(x)$.

Hence, it can be seen that $\frac{1}{\beta_{j_1}(x)} = \Gamma(x)$.

Therefore, $k_{j_1} = a_1 = \log_2(\Gamma(x))$.

This concludes the proof. □

Corollary 13. *Let x be any positive integer. Then we have $k_n(x) = n - m_n(x)$.*

Proof. This can be observed by counting the number of distinct integers $0 \leq i < n$ for which $\beta_i(x) \neq \beta_{i+1}(x)$ (number of divisions by 2).

By Definition 3, observe that $\beta_i(x) \neq \beta_{i+1}(x)$ for all $i \notin S(x)$.

Let $P = \{i | i \in \mathbb{Z}^+, 0 \leq i < n, i \notin S(x)\}$.

By Theorem 2, note that there are exactly $m_n(x)$ positive integers i with $0 \leq i < n$ for which $\alpha_i(x) \neq \alpha_{i+1}(x)$ (in other words, for which $i \in S(x)$).

Therefore, $|P| = n - m_n(x)$, counting the divisions needed to reach $\beta_n(x)$, which is $\log_2 \frac{1}{\beta_n(x)} = k_n$.

Hence, $\beta_n(x) = \frac{1}{2^{n-m_n(x)}}$ and hence, we have $k_n(x) = n - m_n(x)$.

This concludes the proof.

□

B Code

A version of the following program was run on an AMD EPYC with 96 vCPUs and 192 GiB memory, iterating over each value of $k \in \mathbb{Z}^+$, $1 < k \leq 32$, returning h_k.

The complete code can be found at:
`https://github.com/AshwatPrasanna/Simulation/tree/main`

B.1 Output

The output of the code is an ordered list of h_k (in the previously elaborated format).

```
[1, 1, 2, 3, 3, 4, 5, 5, 6, 7, 7, 8, 9, 9, 10, 11, 11, 12, 13, 13, 14, 15, 15,
    16, 17, 17, 18, 19, 19, 20, 21]
```