A Regular-Language and Tree Representation of Odd Collatz Dynamics

Jon Seymour
jon@wildducktheories.com
Sydney, Australia

June 2026

Abstract

We study the odd Collatz map $T(n) = (3n+1)/2^{v_2(3n+1)}$ by classifying each step according to the 2-adic valuation $v = v_2(3n+1)$: an OE step ($v = 1$, $n \equiv 3$ or $7 \pmod 8$), an OEE step ($v = 2$, $n \equiv 1 \pmod 8$), or an OEEE$^+$ step ($v \geq 3$, $n \equiv 5 \pmod 8$). We show that every Collatz orbit is partitioned into *Steiner circuits*, each matching the regular expression (7*3)?(1 | 5) over the mod-8 residues $\{1, 3, 5, 7\}$, with full orbits matching ((7*3)?(1 | 5))*. We further introduce the 5 (mod 8) *overlay tree* for a forward Collatz path from an initial node a to a final node b: its nodes are a, b, all 5 (mod 8) nodes on the path, and for each such node the 0 (mod 3) node m such that the forward OEEE-free path from m reaches n. We illustrate the construction on the path $27 \to 1$, which is a concatenation of several OEEE-free paths joined at 5 (mod 8) nodes. For any two odd integers a, b connected by a Collatz path with $o \geq 1$ steps, we prove $2^e < 3^o \implies a < b$; for OEEE-free pairs we prove the biconditional $2^e \geq 3^o \iff a \geq b$ conditional on Conjecture 3 (OE negative parity), via an inductive argument on the ordering target function $\Delta = (3^o - 2^e)a + K$; the OEE cases are proved unconditionally and the conjecture precisely identifies the remaining gap. All 289 failures of the ordering implication found across an adversarial sample of 240,000 pairs occurred on paths containing at least one interior node $\equiv 5 \pmod 8$, with none on the 57,381 OEEE-free pairs examined, confirming that the 5 (mod 8) condition is the essential mechanism by which the ordering can fail. We prove that the only all-OEE cycle is $1 \to 1$ and conjecture this extends to all OEEE-free cycles; hence any other Collatz cycle must be an OEEE-path. We further conjecture that no Collatz cycle contains a 5 (mod 8) node; conditional on this and the unique-cycle conjecture, there are no Collatz cycles other than $1 \to 1$.

1 The Collatz Map and Step Classification

We work with odd positive integers under the Collatz map. For an odd integer n, define

$$T(n) \;=\; \frac{3n+1}{2^{v_2(3n+1)}},$$

where $v_2(m)$ denotes the 2-adic valuation of m (the largest power of 2 dividing m). By construction $T(n)$ is always odd.

Definition 1 (OE, OEE, OEEE$^+$, and OEE$^+$ steps)**.** *Let n be an odd positive integer and set $v = v_2(3n+1)$.*

- *An **OE step** occurs when $v = 1$, i.e. $3n+1 \equiv 2 \pmod 4$, equivalently $n \equiv 3 \pmod 4$, i.e. $n \equiv 3$ or $7 \pmod 8$. Interior OE steps (those not immediately preceding an OEE$^+$ step) have $n \equiv 7 \pmod 8$; the unique final OE step of each OE-run, which immediately precedes the OEE$^+$ termination, has $n \equiv 3 \pmod 8$.*

- *An **OEE step** occurs when $v = 2$, i.e. $3n+1 \equiv 4 \pmod 8$, equivalently $n \equiv 1 \pmod 8$.*

- *An **OEEE$^+$ step** occurs when $v \geq 3$, i.e. $3n+1 \equiv 0 \pmod 8$, equivalently $n \equiv 5 \pmod 8$ (when $v = 3$) or higher powers of 2 divide $3n+1$.*

- *An **OEE$^+$ step** is either an OEE or an OEEE$^+$ step, i.e. $v \geq 2$; it corresponds to residue 1 or 5 $\pmod 8$.*

Definition 2 (Steiner circuit)**.** *A **Steiner circuit** is a maximal run of Collatz steps of the form*

$$(OE)^*\,(OEE^+),$$

that is, a (possibly empty) sequence of OE steps followed by a single terminating OEE$^+$ step (which is either OEE or OEEE$^+$).

Definition 3 (OEEE-free pair, OEEE-free path, OEEE-path)**.** *Let a and b be odd positive integers with b reachable from a under repeated application of T. Write $T^{(k)}(a)$ for the k-th iterate of T starting at a, with $T^{(0)}(a) = a$, and let $o \geq 1$ be the number of steps from a to b, so that $T^{(o)}(a) = b$. We call (a, b) an **OEEE-free pair** if*

$$\{\, T^{(k)}(a) \;\mid\; 0 \leq k < o \,\} \;\cap\; \{\, n \in \mathbb{Z}_{>0} \mid n \equiv 5 \pmod 8 \,\} \;=\; \emptyset,$$

*i.e. no iterate strictly before b is $\equiv 5 \pmod 8$; the endpoint b itself may or may not be $\equiv 5 \pmod 8$. The forward path from a to b is then called an **OEEE-free path**; otherwise it is an **OEEE-path**.*

We use **path** exclusively for forward traversals of the Collatz map (from a node toward 1) and **walk** for traversals in the reverse direction (from a node toward its predecessors, i.e. from root toward leaves in the Collatz tree).

Definition 4 (OEEE-free walk)**.** *An **OEEE-free walk** is a sequence of nodes $n_0, n_1, n_2, \ldots$ governed by the following state machine on $n \bmod 3$:*

$n \bmod 3$	*action*	*next node*
0	***halt***	–
1	*multiply-shift*	$(4n-1)/3$
2	*multiply-shift*	$(2n-1)/3$

Each multiply-shift step corresponds to $k \leq 2$ halvings in the forward Collatz map ($k = 1$ for $n \equiv 2 \pmod 3$; $k = 2$ for $n \equiv 1 \pmod 3$), so every step has $v \leq 2$ and the walk is OEEE-free by construction. The walk halts at the first node $\equiv 0 \pmod 3$; it is postulated to always halt (Postulate 1). If a walk starts at b and halts at a, then there exists a forward Collatz path from a to b.

A note on exponential versus logarithmic form

Throughout this paper we state ordering conditions in the exponential form $2^e \geq 3^o$ rather than the equivalent logarithmic form $e \geq o \log_2 3$. The two are equivalent:

Lemma 1.1 (Exponential–logarithmic equivalence)**.** *For non-negative integers e and o,*

$$2^e \;\geq\; 3^o \iff e \;\geq\; o \log_2 3.$$

Proof. Since $x \mapsto 2^x$ is strictly increasing and $2^{o \log_2 3} = 3^o$, we have $2^e \geq 3^o \iff 2^e \geq 2^{o \log_2 3} \iff e \geq o \log_2 3$. □

The logarithmic form $e \geq o \log_2 3$ has intuitive appeal as a comparison of step-count ratios against $\log_2 3 \approx 1.585$, and we retain it in Section 5.3 for motivational purposes. However, $\log_2 3$ is transcendental (a consequence of the Gel'fond–Schneider theorem), so any inequality involving it carries an implicit appeal to real analysis. The exponential form $2^e \geq 3^o$ involves only integers and is decidable by direct comparison. This preference for exact integer arithmetic over analytic approximation is in the spirit of Erdős's proof of Bertrand's postulate [1], which replaced analytic estimates with explicit binomial coefficient bounds. All theorems and conjectures in this paper are therefore stated in the exponential form, even though the two forms are logically interchangeable by Lemma 1.1.

2 Mod 8 Encoding and the Regular Language

2.1 Why each step type carries a fixed residue mod 8

We track residues of successive iterates modulo 8.

Proposition 1 (Mod-8 residues of OE/OEE/OEEE$^+$ steps)**.** *The step type of each odd iterate n is determined by $n \bmod 8$:*

- $n \equiv 7 \pmod 8$*:* $3n+1 \equiv 6 \pmod{16}$*, so* $v_2(3n+1) = 1$ *(**OE step**).*
- $n \equiv 3 \pmod 8$*:* $3n+1 \equiv 10 \pmod{16}$*, so* $v_2(3n+1) = 1$ *(**OE step**). This is the* final *OE step of an OE-run, since* $T(n) = (3n+1)/2 \equiv 5 \pmod 8$ *is itself an OEEE$^+$ node.*
- $n \equiv 1 \pmod 8$*:* $3n+1 \equiv 4 \pmod{16}$*, so* $v_2(3n+1) = 2$ *(**OEE step**).*
- $n \equiv 5 \pmod 8$*:* $3n+1 \equiv 0 \pmod 8$*, so* $v_2(3n+1) \geq 3$ *(**OEEE$^+$ step**); the exact value of* $v_2(3n+1)$ *and of* $T(n) \bmod 8$ *depend on higher bits of n.*

Note that $T(n) \bmod 8$ *is* not *determined by* $n \bmod 8$ *alone for OE and OEE steps; the encoding uses only the* input *residue* $n \bmod 8$*.*

The key point is that the *input* residue $n \bmod 8$ determines the symbol we assign in the encoding, independently of what happens after the even divisions.

For the purpose of symbolic encoding we record the *residue class of the input node n* within each step:

Residue $n \bmod 8$	Step type	Symbol
7	OE (interior of run)	7
3	OE (final step before OEE$^+$ termination)	3
1	OEE (exactly two even divisions)	1
5	OEEE$^+$ (three or more even divisions)	5

2.2 The regular expression

The correspondence above implies a definite structure:

- A maximal **OE-run** starts with zero or more residue-7 nodes and ends at a residue-3 node; it is encoded as 7*3.
- The symbol **3** marks the *final OE step before the OEE$^+$ termination* of the current OE-run, not the end of the Steiner circuit itself.

- Termination by **OEE** (exactly two even divisions) occurs at a node congruent to 1 (mod 8), encoded as 1.
- Termination by **OEEE$^+$** (three or more even divisions) occurs at a node congruent to 5 (mod 8), encoded as 5.
- Together, OEE and OEEE$^+$ constitute an **OEE$^+$** termination (two or more even divisions), encoded as 1 | 5.

Thus a single Steiner circuit—zero or more OE-runs followed by a terminating OEE$^+$ step (OEE or OEEE$^+$)—is encoded by the regular expression:

$$\boxed{(7^*3)?(1 \mid 5)}.$$

An entire odd Collatz orbit, decomposed into successive Steiner circuits, is encoded by:

$$\boxed{\big((7^*3)?(1 \mid 5)\big)^*}.$$

This is a regular language over the alphabet $\{1,3,5,7\}$, giving a finite-state symbolic grammar for the mod-8 projection of odd Collatz dynamics.

3 Example: the odd path $27 \to 1$

Table 1 shows the full odd Collatz path from 27 to 1 under the map T. The residue n mod 8 serves directly as the encoding symbol (Proposition 1), so no separate symbol column is needed.

The 41-symbol string over $\{1,3,5,7\}$, with Steiner circuits marked by vertical bars, is:

$$\underbrace{3\,1}_{1} \mid \underbrace{7\,7\,7\,3\,1}_{2} \mid \underbrace{1}_{3} \mid \underbrace{3\,1}_{4} \mid \underbrace{7\,3\,1}_{5} \mid \underbrace{7\,7\,3\,1}_{6} \mid \underbrace{5}_{7} \mid \underbrace{7\,3\,1}_{8} \mid \underbrace{3\,1}_{9}$$
$$\mid \underbrace{7\,7\,7\,7\,3\,5}_{10} \mid \underbrace{7\,7\,3\,5}_{11} \mid \underbrace{1}_{12} \mid \underbrace{1}_{13} \mid \underbrace{5}_{14} \mid \underbrace{5}_{15} \mid \underbrace{7\,3\,5}_{16} \mid \underbrace{5}_{17}$$

Each group is one Steiner circuit matching $(7^*3)?(1 \mid 5)$. The path comprises 17 Steiner circuits.

4 The 5 (mod 8) Overlay Tree

4.1 Definition

Let a be an odd positive integer (the **initial node**) and b an odd positive integer (the **final node**, typically $b = 1$). The 5 (mod 8) **overlay tree** with

Step	n	$n$mod8	v	Type	Step	n	$n$mod8	v	Type
0	27	3	1	OE	21	283	3	1	OE
1	41	1	2	OEE	22	425	1	2	OEE
2	31	7	1	OE	23	319	7	1	OE
3	47	7	1	OE	24	479	7	1	OE
4	71	7	1	OE	25	719	7	1	OE
5	107	3	1	OE	26	1079	7	1	OE
6	161	1	2	OEE	27	1619	3	1	OE
7	121	1	2	OEE	28	2429	5	3	OEEE$^+$
8	91	3	1	OE	29	911	7	1	OE
9	137	1	2	OEE	30	1367	7	1	OE
10	103	7	1	OE	31	2051	3	1	OE
11	155	3	1	OE	32	3077	5	4	OEEE$^+$
12	233	1	2	OEE	33	577	1	2	OEE
13	175	7	1	OE	34	433	1	2	OEE
14	263	7	1	OE	35	325	5	4	OEEE$^+$
15	395	3	1	OE	36	61	5	3	OEEE$^+$
16	593	1	2	OEE	37	23	7	1	OE
17	445	5	3	OEEE$^+$	38	35	3	1	OE
18	167	7	1	OE	39	53	5	5	OEEE$^+$
19	251	3	1	OE	40	5	5	4	OEEE$^+$
20	377	1	2	OEE	41	1	1	—	terminal

Table 1: Odd Collatz path $27 \to 1$: 41 steps, 42 nodes. The symbol for each step equals n mod 8.

initial node a and final node b is defined if and only if $b = T^k(a)$ for some $k \geq 0$, i.e. b lies on the forward Collatz path from a.

Nodes. The tree always contains:

- the initial node a,
- the final node b,
- every node $\equiv 5 \pmod 8$ on the forward path from a to b,
- for each such $5 \pmod 8$ node n: the $0 \pmod 3$ node m reached by the OEEE-free walk starting from n (Definition 4), if it exists (assumed by Postulate 1).

Edges. All edges are directed in the forward Collatz direction (from a toward b). Each node type carries a specific outgoing edge:

- The **initial node** a points to the first $5 \pmod 8$ node on the forward path from a to b, or directly to b if no $5 \pmod 8$ node lies between a and b.
- Each **5** (mod **8**) **node** points to the next $5 \pmod 8$ node on the forward path toward b, or directly to b if no further $5 \pmod 8$ node exists between it and b.
- Each **0** (mod **3**) **node** m points to the first $5 \pmod 8$ node on the forward Collatz path from m. Such a node m is the terminus of the OEEE-free walk starting from the corresponding $5 \pmod 8$ node n.

Postulate 1. *The OEEE-free walk (Definition 4) always halts, i.e. for every node* $n \equiv 5 \pmod 8$*, the walk starting from* n *eventually reaches a node* $\equiv 0 \pmod 3$*. Equivalently, for every node* $n \equiv 5 \pmod 8$*, there exists a node* $m \equiv 0 \pmod 3$ *such that the forward Collatz path from* m *reaches* n *as an OEEE-free path.*

This is supported by computation but is not proved in general.

4.2 Example: $a = 27$, $b = 1$

Figure 1 shows the $5 \pmod 8$ overlay tree with initial node $a = 27$ and final node $b = 1$. The forward path from 27 to 1 passes through seven nodes $\equiv 5 \pmod 8$, namely $445, 2429, 3077, 325, 61, 53, 5$ (in forward order, a to b). Each of these points to the next $5 \pmod 8$ node in the forward direction, with the last one (5) pointing to $b = 1$. All seven have $0 \pmod 3$ termini, consistent with Postulate 1. The initial node $a = 27$ points to the first $5 \pmod 8$ node 445. For each $5 \pmod 8$ node n, the $0 \pmod 3$ node m whose forward OEEE-free path reaches n has path length ranging from 1 step ($n = 5$, $n = 61$) to 17 steps ($n = 445$, where $m = 27 = a$).

5 The Path Identity and the $3^o/2^e$ Growth Factor

5.1 Setup

Let a and b be odd positive integers such that b is reachable from a on the forward Collatz path and no node strictly between a and b is $\equiv 5 \pmod 8$; b itself may or may not be $\equiv 5 \pmod 8$. (The case $b = a$ is the null path with $e = o = 0$.) Define:

- e = the total number of even halvings on the forward path from a to b, not counting b itself, and

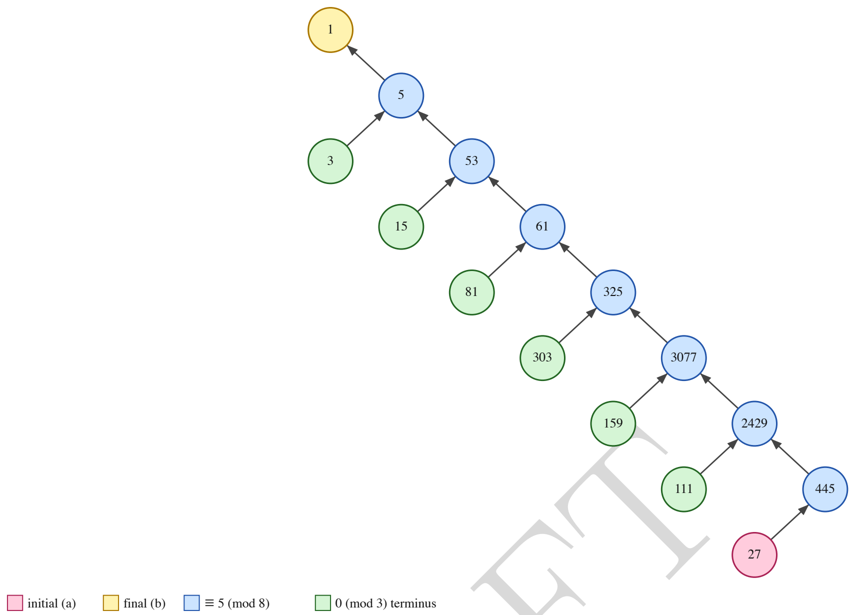

Figure 1: $5 \pmod 8$ overlay tree with initial node $a = 27$ (pink) and final node $b = 1$ (yellow). Blue nodes are $\equiv 5 \pmod 8$; green nodes are $0 \pmod 3$ termini. All arrows point in the forward Collatz direction. Each $5 \pmod 8$ node points to the next $5 \pmod 8$ node on the forward path, or to b if none exists; $a = 27$ points to the first $5 \pmod 8$ node 445. Each $0 \pmod 3$ node points to the first $5 \pmod 8$ node on its forward path. By definition, no OEEE-free path visits a $5 \pmod 8$ node.

- $o =$ the number of odd Collatz steps on the forward path from a to b, not counting b itself.

Example. Take $a = 75$. The forward path is

$$75 \to 226 \to 113 \to 340 \to 170 \to 85 = b,$$

giving $o = 2$ odd steps (at $75, 113$) and $e = 3$ even halvings (at $226, 340, 170$), with $b = 85$.

5.2 The Path Identity

Each odd step multiplies by 3 and adds 1; each even step divides by 2. Tracking the accumulated powers along the forward path from a to b yields the **path identity**:

$$2^e \cdot b \;=\; 3^o \cdot a \;+\; K,$$

where $K \geq 0$ is an integer that collects the carry terms from each $3n+1$ step, weighted by the remaining powers of 2. Dividing through by $2^e a$:

$$\frac{b}{a} \;=\; \frac{3^o}{2^e} + \frac{K}{2^e \cdot a}.$$

Since $K \geq 0$, the growth factor b/a always exceeds the bare ratio $3^o/2^e$; the correction $K/(2^e a)$ is non-negative.

5.3 The Growth Factor $b/a \approx 3^o/2^e$

For large a and b the correction term $K/(2^e a)$ is small, giving

$$\frac{b}{a} \;\approx\; \frac{3^o}{2^e}.$$

This can seem surprising: why should the growth factor of a Collatz path mirror the ratio of powers of 3 and 2 raised to the counts of odd and even steps? The path identity shows it is in fact an exact algebraic consequence, with the correction vanishing as $a \to \infty$. Since $K \geq 0$ the growth factor always satisfies $b/a \geq 3^o/2^e$; the correction is always non-negative.

Taking logarithms base 2:

$$\log_2(b/a) \;\approx\; o \log_2 3 - e.$$

Thus $b > a$ when $3^o > 2^e$ (a surplus of odd steps relative to even halvings, so the path grows) and $b < a$ when $3^o < 2^e$ (a deficit of odd steps, so the path contracts). This makes structural sense: a surplus of odd steps inflates the iterate relative to where it started.

5.4 Sharpness on OEEE-free Paths

Let a and b be odd positive integers connected by an OEEE-free path. Since e and o count operations (steps), not nodes, T is never applied to the final node b, so b contributes nothing to e or o. Every step source on an OEEE-free path is not $\equiv 5 \pmod 8$, so every step has $v \leq 2$, giving $e \leq 2o$.

Proposition 2 (OEEE-free carry bounds)**.** *Let (a, b) be an OEEE-free pair with $o \geq 1$ steps. Then:*

$$3^o - 2^o \;\leq\; K \;\leq\; 4^o - 3^o.$$

Both bounds are tight: the lower bound is achieved when every step is an OE step ($v = 1$); the upper bound when every step is an OEE step ($v = 2$).

Proof. We use the suffix recurrence: $R_o = 0$ and $R_m = 2^{v_m} R_{m+1} + 3^{o-1-m}$ for $0 \leq m < o$, where $v_m \in \{1, 2\}$ since the path is OEEE-free. Then $K = R_0$.

Base case ($o = 1$): The single step has $v_0 \in \{1, 2\}$ and $K = R_0 = 2^{v_0} \cdot 0 + 3^0 = 1$. Since $3^1 - 2^1 = 1 = 4^1 - 3^1$, the bounds hold with equality.

Inductive step: Suppose the bounds hold for a path of length o, i.e. $3^o - 2^o \leq K' \leq 4^o - 3^o$. Prepending one step with valuation $v \in \{1, 2\}$ gives a path of length $o + 1$ with carry

$$K \;=\; 2^v K' + 3^o.$$

For the lower bound, $v \geq 1$, so

$$K \;\geq\; 2(3^o - 2^o) + 3^o \;=\; 3 \cdot 3^o - 2^{o+1} \;=\; 3^{o+1} - 2^{o+1}.$$

For the upper bound, $v \leq 2$, so

$$K \;\leq\; 4(4^o - 3^o) + 3^o \;=\; 4^{o+1} - 3 \cdot 3^o \;=\; 4^{o+1} - 3^{o+1}.$$

Both inequalities are equalities when every step has $v = 1$ (all-OE) or $v = 2$ (all-OEE) respectively, confirming tightness. □

Hence on any OEEE-free path:

$$3^o - 2^o \;\leq\; K \;\leq\; 4^o - 3^o \;<\; 4^o.$$

Theorem 1 (Collatz path ordering, positive parity state)**.** *Let a and b be odd positive integers connected by a Collatz path with $o \geq 1$, and let e, o, K be as above. Then:*

$$2^e < 3^o \quad\Longrightarrow\quad a < b.$$

Proof. From the growth-factor form $b/a = 3^o/2^e + K/(2^e a)$ with $K \geq 0$: if $3^o > 2^e$ then $b/a \geq 3^o/2^e > 1$, so $b > a$.

The approximation bound $|b/a - 3^o/2^e| < (4/3)^o/a$ follows from the OEEE-free bound $K \leq 4^o - 3^o < 4^o$ of Proposition 2. □

6 Ordering on OEEE-free Paths

6.1 The unified ordering target function

For an OEEE-free path $a \to b$ with parameters (o, e, K), the path identity (Section 5.3) gives $2^e b = 3^o a + K$. Subtracting $2^e a$ from both sides isolates the spatial displacement:

$$2^e(b - a) \;=\; (3^o - 2^e)\, a + K.$$

Since $2^e > 0$, the sign of $b - a$ is determined entirely by the right-hand side. We define the **ordering target function**

$$\Delta \;=\; (3^o - 2^e)\, a + K,$$

so that $\operatorname{sgn}(b - a) = \operatorname{sgn}(\Delta)$. The two binary states used to track alignment are:

$$s_e \;=\; \operatorname{sgn}(\Delta), \qquad s_p \;=\; \operatorname{sgn}(3^o - 2^e),$$

where $s_e = +$ (resp. $-$) means $b > a$ (resp. $b \le a$) and $s_p = +$ (resp. $-$) means $3^o > 2^e$ (resp. $3^o \le 2^e$).

Theorem 2 states that $s_e = s_p$ on every OEEE-free path. Each step of the path is either an **OE step** ($v = 1$, $b \equiv 3$ or $7 \pmod 8$, $b_1 = (3b+1)/2$) or an **OEE step** ($v = 2$, $b \equiv 1 \pmod 8$, $b_1 = (3b+1)/4$), with carry update $K_1 = 3K + 2^e$ in both cases. We say an OE step is **diverging** when $s_p = +$ and **converging** otherwise; an OEE step is **diverging** when $s_p = -$ and **converging** otherwise.

Lemma 6.1 (Forward sign alignment)**.** *For any OEEE-free path with $o \ge 1$: if $3^o > 2^e$ then $b > a$.*

Proof. Since 3^o and 2^e are integers, $3^o > 2^e$ implies $3^o - 2^e \ge 1$, so $\Delta = (3^o - 2^e)a + K \ge a + K \ge a \ge 1 > 0$. ☐

6.2 Step dynamics and the Δ recurrence

Appending a step with valuation $v \in \{1, 2\}$ to a path with parameters (o, e, K) gives $(o + 1, e + v, K_1)$ with $K_1 = 3K + 2^e$. The updated target function satisfies:

$$\begin{aligned}
\Delta_1 &= (3^{o+1} - 2^{e+v})\, a + K_1 \\
&= 3\big[(3^o - 2^e)\, a + K\big] + 2^e(3a - 2^v a + 1) \\
&= 3\Delta + 2^e\,(3 - 2^v)\, a + 2^e.
\end{aligned}$$

Lemma 6.2 (Step modifiers)**.** *Under a single step extension:*

(OE) $\Delta_1 = 3\Delta + 2^e(a + 1)$.

(OEE) $\Delta_1 = 3\Delta - 2^e(a - 1)$.

Proof. Substitute $v = 1$ (giving $3 - 2^v = 1$) and $v = 2$ (giving $3 - 2^v = -1$) into the recurrence above. ☐

An OE step drives Δ upward (by $2^e(a + 1) > 0$); an OEE step drives it downward (by $-2^e(a - 1) \le -2^{e+1} < 0$ since $a \ge 3$).

6.3 Negative parity invariant and main theorem

A key simplification follows from substituting $K = 2^e b - 3^o a$ (the path identity) directly into the Δ recurrence. For any step:

$$K_1 = 3K + 2^e = 3(2^e b - 3^o a) + 2^e = 2^e(3b+1) - 3^{o+1}a,$$

so

$$\begin{aligned}\Delta_1 &= (3^{o+1} - 2^{e+v})a + K_1 \\ &= (3^{o+1} - 2^{e+v})a + 2^e(3b+1) - 3^{o+1}a \\ &= 2^e(3b+1) - 2^{e+v}a \\ &= 2^e\big(3b+1-2^v a\big).\end{aligned}$$

Since $2^e > 0$, the sign of Δ_1 is determined entirely by $3b+1-2^v a$:

$$\Delta_1 \;\le\; 0 \quad\Longleftrightarrow\quad 3b+1 \;\le\; 2^v a. \tag{†}$$

For an OEE step ($v = 2$): $\Delta_1 \le 0 \iff 3b+1 \le 4a$, i.e. $b \le (4a-1)/3$; for an OE step ($v = 1$): $\Delta_1 \le 0 \iff 3b+1 \le 2a$, i.e. $b \le (2a-1)/3$.

We first prove the two OEE cases, then identify precisely what remains open.

Lemma 6.3 (OEE negative parity)**.** *Let (a, b) be an OEEE-free pair with $o \ge 1$. If the step appended is an OEE step ($v = 2$) and $3^{o+1} \le 2^{e+2}$ (new state in $s_p = -$), then $\Delta_1 < 0$, i.e. $b_1 < a$.*

Proof. By (†), $\Delta_1 < 0$ iff $3b+1 < 4a$.

Subcase A — prior state in $s_p = -$ ($3^o \le 2^e$): The inductive hypothesis $\Delta \le 0$ gives $b \le a$, so $3b+1 \le 3a+1$. Since $a \ge 3$ (odd, $o \ge 1$), $3a+1 \le 4a-2 < 4a$. Hence $3b+1 < 4a$ strictly.

Subcase B — prior state in $s_p = +$ ($3^o > 2^e$): We need $3b+1 < 4a$. Substituting $a = (2^e b - K)/3^o$ and using $K \ge 0$:

$$4a \;=\; \frac{4(2^e b - K)}{3^o} \;\ge\; \frac{4\cdot 2^e b - 4K}{3^o}.$$

The condition $3b+1 < 4a$ is equivalent (after rearranging with $a = (2^e b - K)/3^o$) to $(3b+1)\cdot 3^o < 4(2^e b - K)$, i.e. $b(4\cdot 2^e - 3^{o+1}) > 4K + 3^o$. Since $3^{o+1} \le 2^{e+2}$ we have $4\cdot 2^e - 3^{o+1} = 2^{e+2} - 3^{o+1} \ge 0$. We prove

$$b\,(2^{e+2} - 3^{o+1}) \;>\; 4K + 3^o \tag{$*$}$$

by induction on the number of OEE steps. *Base* ($o = 0$, $K = 0$): ($*$) gives $b > 1$, which holds for all odd $b \geq 3$. *Step*: after an OEE step $(e, o, K, b) \mapsto (e+2, o+1, 3K+2^e, (3b+1)/4)$, using $2^{e+4} - 3^{o+2} = 4(2^{e+2} - 3^{o+1}) + 3^{o+1}$:

$$\begin{aligned}(3b+1)(2^{e+4} - 3^{o+2}) &= 4(3b+1)(2^{e+2} - 3^{o+1}) + (3b+1) \cdot 3^{o+1} \\ &> 4\big(3(4K + 3^o) + 2^{e+2} - 3^{o+1}\big) + (3b+1) \cdot 3^{o+1} \\ &= 48K + 2^{e+4} + (3b+1) \cdot 3^{o+1}.\end{aligned}$$

The target is $16(3K + 2^e) + 4 \cdot 3^{o+1} = 48K + 2^{e+4} + 4 \cdot 3^{o+1}$, so it suffices that $3b + 1 > 4$, i.e. $b \geq 3$. Hence ($*$) holds and $\Delta_1 < 0$.

Both subcases give $\Delta_1 < 0$ whenever the negative parity state is entered or maintained via an OEE step. □

Remark 1. *Subcase A formalises directly in Lean 4 as a two-line structural induction on* `OEF`*. Subcase B is proved above by an auxiliary induction* ($*$) *on the number of OEE steps; this induction is on a* different *measure from the structural* `OEF` *induction, and no clean Lean 4 proof has yet been found. All numerical evidence confirms the subcase is true, and the paper proof is mathematically complete, but the Lean formalisation of Lemma 6.3 subcase B remains open.*

Remark 2 (Steiner circuit structure and the crossing step)**.** *A **Steiner circuit** is a maximal run from one* $b \equiv 1 \pmod 8$ *node to the next, consisting of* $k \geq 0$ *OE steps followed by exactly one OEE step. The negative parity state can only be* entered *— as opposed to sustained — via the terminal OEE step of a circuit. We call this terminal step a **crossing step** when it changes* s_p *from + to −, i.e. when* $3^o > 2^e$ *(before) and* $3^{o+1} \leq 2^{e+2}$ *(after).*

At a crossing step the exponents satisfy the tight sandwich

$$2^e \ < \ 3^o \ \leq \ \tfrac{4}{3} \cdot 2^e,$$

equivalently $3^o/2^e \in (1, 4/3]$*. This forces* $4 \cdot 2^e - 3^{o+1}$ *to be a small non-negative integer (it lies in* $[0, 2^e)$*), and the condition* ($*$) *becomes a tight numerical inequality between b, K, o, and e at that single step.*

Numerical search ($a \leq 5{,}000$*) finds zero violations of* $3b+1 < 4a$ *among all crossing steps. The tightest observed margin is* $4a - (3b+1) = 16$ *at* $(a, b, o, e, K) = (107, 137, 4, 6, 101)$*, and the ratio* $3^o/2^e$ *at crossing steps lies in* $(1.068, 1.266]$*, always strictly below the critical value* $4/3$*.*

The proposed proof strategy for subcase B is therefore to establish condition ($*$) *via an induction restricted to the OE-steps* within a single circuit*,* *starting from the* $1 \pmod 8$ *circuit-entry node, rather than a global induction over all OEE steps. Since each circuit contains exactly one OEE step, this circuit-level induction collapses the multi-step auxiliary argument to a single local verification at the crossing step.*

We can now state precisely what is known and what remains open.

Conjecture 1 (OE negative parity)**.** *Let* (a, b) *be a genuine OEEE-free pair in the negative parity state* ($3^o \leq 2^e$) *with* $b \not\equiv 5 \pmod 8$ *and* $b \equiv 3$ *or* $7 \pmod 8$ *(so the next step is an OE step) and* $3^{o+1} \leq 2^{e+1}$ *(the new state is also* $s_p = -$*). Then* $3b + 1 \leq 2a$*, equivalently* $b \leq (2a - 1)/3$*.*

Why this is not yet proved. By (†) with $v = 1$, the condition $\Delta_1 \leq 0$ reduces *exactly* to $3b + 1 \leq 2a$. We know $b \leq a$ (from the negative parity state), but $b \leq a$ does not imply $3b + 1 \leq 2a$: the latter requires $b \leq (2a - 1)/3$, which is strictly stronger than $b < a$ for $a \geq 5$ (e.g. $a = 7$, $b = 5$ satisfies $b < a$ but $3 \cdot 5 + 1 = 16 > 14 = 2 \cdot 7$). The condition $3b + 1 \leq 2a$ is a genuine constraint on the orbit — it asserts that when the path is in the negative parity state and takes an OE step that stays in $s_p = -$, the node b is bounded by $(2a - 1)/3$, not merely by a. This cannot be deduced from the carry bounds $0 \leq K \leq 4^o - 3^o$ alone: a phantom predecessor with $b = a - 2$ and $K = 2^e(a - 2) - 3^o a$ satisfies the carry bounds and $b < a$, yet violates $3b + 1 \leq 2a$ whenever $a \geq 5$. What is needed is a property of the actual orbit structure of OEEE-free paths — one that rules out nodes b with $(2a - 1)/3 < b < a$ in the negative parity state.

Theorem 2 (OEEE-free path ordering, conditional)**.** *Let* (a, b) *be an OEEE-free pair with* $o \geq 1$*. Then:*

$$2^e < 3^o \implies a < b.$$

Conditional on Conjecture 1, the full biconditional holds:

$$2^e \geq 3^o \iff a \geq b.$$

Proof. The forward implication $2^e < 3^o \implies a < b$ is Theorem 1, unconditional.

For the reverse direction $2^e \geq 3^o \implies a \geq b$, assuming Conjecture 1: the negative parity state $3^o \leq 2^e$ can only be entered via an OEE step, for which Lemma 6.3 gives $\Delta < 0$ strictly. Conjecture 1 then handles all subsequent OE steps sustaining the negative parity state. Hence $\Delta \leq 0$ throughout, giving $b \leq a$. □

The OEEE-free condition is essential: it restricts steps to $v \in \{1, 2\}$. For an OEEE$^+$ step ($v \geq 3$) the modifier becomes $2^e(3-2^v)a+2^e \leq 2^e(1-5a) < 0$ even when $s_p = +$, allowing K to escape the parity constraint — as the $a = 165$, $b = 167$ example illustrates.

See Appendix A for worked examples illustrating Theorems 1 and 2 and a non-example showing the necessity of the OEEE-free condition.

7 Cycles and the $5 \pmod 8$ Constraint

Theorem 3 (All-OEE cycle)**.** *The only OEEE-free cycle in which every step is an OEE step ($e = 2o$) is $a = b = 1$.*

Proof. Suppose $a = b$ with $e = 2o$. The path identity gives $K = (4^o - 3^o)a$. Since $K \leq 4^o - 3^o$, we get $a \leq 1$, hence $a = 1$. The known $1 \to 1$ cycle ($o = 1$, $e = 2$) confirms $a = b = 1$ is valid. □

Conjecture 2 (Unique OEEE-free cycle)**.** *The only cycle on an OEEE-free path is $a = b = 1$.*

The all-OEE case is proved by Theorem 3. The mixed case ($o < e < 2o$) requires ruling out cycles with $3^o < 2^e < 4^o$, which does not follow from the carry bounds alone and sits at Collatz-conjecture-level difficulty.

Corollary 1. *Assuming Conjecture 2, if a non-trivial Collatz cycle exists (other than $1 \to 1$), then it must be an OEEE-path.*

Proof. By Conjecture 2, any OEEE-free cycle is $a = b = 1$ (the all-OEE case is established unconditionally by Theorem 3; the mixed case is the conjecture). Hence any other cycle is not OEEE-free, i.e. it is an OEEE-path. □

Conjecture 3. *No non-trivial Collatz cycle contains a $5 \pmod 8$ node.*

This conjecture is unproven. It is offered here not as an established result but as a suggested avenue for future research: a proof of this conjecture together with a proof of Conjecture 2 would establish that there are no non-trivial Collatz cycles (Theorem 4).

Theorem 4 (No non-trivial cycles, conditional)**.** *Assuming Conjecture 2 and Conjecture 3, there are no Collatz cycles other than $1 \to 1$.*

Proof. By Corollary 1 (which assumes Conjecture 2), any non-trivial cycle must be an OEEE-path and hence contains an interior $5 \pmod 8$ node. But Conjecture 3 asserts no such cycle exists. The conclusion follows. □

8 Conclusion

The odd Collatz dynamics admit two complementary representations:

- A **proved regular language structure** over $\{1, 3, 5, 7\}$:

$$(7^*3)?(1 \mid 5) \quad \text{and} \quad \big((7^*3)?(1 \mid 5)\big)^*,$$

  encoding Steiner circuits and full orbits. Here 7 encodes interior OE steps, 3 encodes the final OE step of a run (immediately before the OEE$^+$ termination), 1 encodes an OEE termination (residue 1 mod 8, exactly two even divisions), and 5 encodes an OEEE$^+$ termination (residue 5 mod 8, three or more even divisions); together $1 \mid 5$ encodes any OEE$^+$ termination (two or more even divisions). This encoding is a theorem following directly from mod-8 arithmetic of $3n + 1$.

- A 5 (mod 8) **overlay tree** whose nodes are the initial node a, the final node b, all 5 (mod 8) nodes on the forward path, and for each such 5 (mod 8) node n, the 0 (mod 3) node m whose forward OEEE-free path reaches n, if it exists. All edges point in the forward Collatz direction: each 5 (mod 8) node points to the next 5 (mod 8) node or to b; each 0 (mod 3) node m points to the first 5 (mod 8) node on the forward path from m; a points to the first 5 (mod 8) node or directly to b. By definition, no OEEE-free path visits a 5 (mod 8) node. Whether every 5 (mod 8) node has such a 0 (mod 3) node remains open (Postulate 1).

- A **Collatz path ordering** (Theorems 1 and 2): for any two odd integers a, b connected by a Collatz path with $o \geq 1$, we prove $2^e < 3^o \implies a < b$; for OEEE-free pairs the biconditional $2^e \geq 3^o \iff a \geq b$ is established conditional on Conjecture 1 (Theorem 2), via the ordering target function $\Delta = (3^o - 2^e)a + K$; the OEE cases are proved, and the open case is OE steps sustaining the negative parity state. All 289 failures of the ordering implication found in adversarial sampling occurred on paths with at least one interior 5 (mod 8) node, none on OEEE-free paths (Table 2).

- A **cycle constraint**: the only all-OEE cycle is $1 \to 1$ (Theorem 3); we conjecture this extends to all OEEE-free cycles (Conjecture 2), so any other Collatz cycle must be an OEEE-path. We conjecture that no Collatz cycle contains a 5 (mod 8) node (Conjecture 3); conditional on this conjecture and Conjecture 2, there are no Collatz cycles other than $1 \to 1$ (Theorem 4).

Together, these results reveal a tight algebraic structure underlying odd Collatz dynamics. The residue-5 (mod 8) class plays a distinguished role throughout: alongside 1 (mod 8) it marks the OEE^+ termination of each Steiner circuit, but uniquely it is the sole source of OEEE^+ steps ($v \geq 3$), it determines the branching structure of the overlay tree, and its absence from OEEE-free path interiors is the key condition under which the ordering theorem holds. The empirical evidence (Table 2) further suggests that interior 5 (mod 8) nodes are not merely sufficient but necessary for ordering failures, pointing to the OEEE-free condition as the load-bearing structural constraint for the ordering theorem.

Open problems. One postulate and three conjectures remain open:

1. **Walk termination** (Postulate 1): the OEEE-free walk always halts, i.e. for every node $n \equiv 5 \pmod 8$ the walk eventually reaches a node $\equiv 0 \pmod 3$.

2. **OE negative parity** (Conjecture 1): when an OE step sustains the negative parity state ($3^o \leq 2^e$ and $3^{o+1} \leq 2^{e+1}$), the target function satisfies $\Delta_1 \leq 0$. The OEE subcase A of Lemma 6.3 is proved and Lean-formalised; OEE subcase B (entering $s_p = -$ from $s_p = +$) is proved in the paper via auxiliary induction ($*$) but the Lean formalisation remains open (see the remark after Lemma 6.3). The OE step is the sole remaining gap in Theorem 2. See §6 for a precise account of why the carry bounds alone are insufficient.

3. **Unique OEEE-free cycle** (Conjecture 2): the only cycle on an OEEE-free path is $a = b = 1$. The all-OEE case is proved (Theorem 3); the mixed case sits at Collatz-conjecture-level difficulty.

4. **No** 5 (mod 8) **node in any cycle** (Conjecture 3): no Collatz cycle contains a node $\equiv 5 \pmod 8$. Conditional on this conjecture and Conjecture 2, there are no Collatz cycles other than $1 \to 1$ (Theorem 4).

References

[1] P. Erdős, *Beweis eines Satzes von Tschebyschef*, Acta Litt. Sci. Szeged **5** (1932), 194–198.

A Empirical Evidence

Confirming example: $a = 7$, $b = 13$

The forward path $7 \to 22 \to 11 \to 34 \to 17 \to 52 \to 26 \to 13$ has $o = 3$ odd steps (at $7, 11, 17$) and $e = 4$ even halvings (at $22, 34, 52, 26$), not counting the terminus 13. The interior odd nodes are $7, 11, 17$, none $\equiv 5 \pmod 8$, so $(7, 13)$ is an OEEE-free pair (Definition 3). Then $2^e = 16 < 27 = 3^o$, confirming $2^e < 3^o$ and $a = 7 < 13 = b$, consistent with Theorem 1.

OEEE-path consistent with Theorem 1: $a = 55$, $b = 71$

The forward path is

$$55 \to 166 \to 83 \to 250 \to 125 \to 376 \to 188 \to 94 \to 47 \to 142 \to 71,$$

with $o = 4$ odd steps (at $55, 83, 125, 47$) and $e = 6$ even halvings, not counting the terminus 71. The interior node $125 \equiv 5 \pmod 8$, so $(55, 71)$ is *not* an OEEE-free pair. Nevertheless, $2^6 = 64 < 81 = 3^4$, so $2^e < 3^o$, and Theorem 1 applies unconditionally to give $a = 55 < 71 = b$ — as observed. The OEEE-free hypothesis is irrelevant here: Theorem 1 holds for all Collatz paths.

Confirming example: $a = 107$, $b = 91$

The forward path is

$$107 \to 322 \to 161 \to 484 \to 242 \to 121 \to 364 \to 182 \to 91,$$

with $o = 3$ odd steps (at $107, 161, 121$) and $e = 5$ even halvings (at $322, 484, 242, 364, 182$), not counting the terminus 91. The interior odd nodes are $107 \equiv 3 \pmod 8$, $161 \equiv 1 \pmod 8$, and $121 \equiv 1 \pmod 8$, none $\equiv 5 \pmod 8$, so $(107, 91)$ is an OEEE-free pair. Here $2^e = 32 > 27 = 3^o$, so $2^e \geq 3^o$, and indeed $a = 107 > 91 = b$, consistent with Theorem 2.

Non-example: $a = 165$, $b = 167$

The forward path is

$$165 \to 31 \to 47 \to 71 \to 107 \to 161 \to 121 \to 91 \to 137 \to 103 \to 155 \to 233 \to 175 \to 263 \to 395 \to 59$$

with $o = 17$ odd steps and $e = 27$ even halvings.

The OEEE-free condition fails. The interior nodes include $165 \equiv 5 \pmod 8$ and $445 \equiv 5 \pmod 8$, so $(165, 167)$ is *not* an OEEE-free pair and Theorem 2 does not apply.

The exponential comparison. $2^{27} = 134{,}217{,}728 > 129{,}140{,}163 = 3^{17}$, so $2^e \geq 3^o$ holds — but only just: the excess is $2^e - 3^o = 5{,}077{,}565$, a margin of less than 4% of 3^o.

How close are b/a and $3^o/2^e$? From the path identity:

$$\frac{b}{a} = \frac{3^o}{2^e} + \frac{K}{2^e \cdot a} = \frac{129{,}140{,}163}{134{,}217{,}728} + \frac{1{,}106{,}233{,}681}{134{,}217{,}728 \times 165}.$$

Numerically, $3^o/2^e \approx 0.9622$ while $b/a = 167/165 \approx 1.0121$, a difference of $K/(2^e a) \approx 0.0499$. The correction term is large enough to push b/a above 1 even though $3^o/2^e < 1$: the path carries enough extra weight in $K = 1{,}106{,}233{,}681$ to overcome the deficit $2^e - 3^o = 5{,}077{,}565$. This is precisely the mechanism that Theorem 2 asserts cannot occur on an OEEE-free path.

The admissibility coupling. The path identity forces $b \equiv K \cdot (2^e)^{-1} \pmod{3^o}$. Here $K \cdot (2^{27})^{-1} \bmod 3^{17} = 167 = b$, confirming the coupling holds — but since $b = 167 \ll 3^{17} = 129{,}140{,}163$, the residue class places no useful lower bound on b. The OEEE-free condition is what would otherwise constrain K and force $a \geq b$.

Adversarial sampling

To provide independent empirical support for Theorem 2, 2,000 random odd starting points were drawn uniformly from $[1, 10^{12}]$. For each trajectory the 16 nodes closest in absolute magnitude to the trajectory median were selected. This is the most adversarial choice: near the median, a and b are closest in size, so $b/a \approx 1$ and the margin by which $b \leq a$ might fail is smallest. All ancestor-descendant pairs within each such window were examined (240,000 pairs in total, of which 57,381 were OEEE-free).

What the table shows. Table 2 records, for every pair where the ordering implication $2^e \geq 3^o \implies a \geq b$ fails, how many interior nodes $\equiv 5 \pmod 8$ the path contains. The striking feature is the top row: among the 57,381 OEEE-free pairs, *not a single failure* was observed. Every one of the 289 failures found across all 240,000 pairs occurs on a path containing at least one interior $5 \pmod 8$ node, i.e. an OEEE-path. Moreover the failure count grows with the number of such nodes, suggesting that each interior

5 (mod 8) node contributes independently to inflating the offset K beyond the level the ordering can absorb.

Interior nodes $\equiv 5 \pmod 8$	Failures	% of failures
0 (OEEE-free paths)	**0**	**0.0%**
1	12	4.2%
2	5	1.7%
3	133	46.0%
4	77	26.6%
5	38	13.1%
6	21	7.3%
7	3	1.0%
Total	**289**	**100.0%**

Table 2: Failures of $2^e \geq 3^o \implies a \geq b$ classified by the number of interior nodes $\equiv 5 \pmod 8$ on the path, across 240,000 adversarially-selected ancestor-descendant pairs drawn from 2,000 random Collatz trajectories with starting points in $[1, 10^{12}]$. Pairs are concentrated near trajectory medians where $b/a \approx 1$, the region most likely to produce violations. Of 57,381 OEEE-free pairs examined, zero violations were found. All 289 failures occur on OEEE-paths.

What the table shows. The data confirm Theorem 2: among 57,381 OEEE-free pairs, zero violations of $2^e \geq 3^o \implies a \geq b$ were found. The table also illuminates the necessity of the OEEE-free condition: all 289 failures involve at least one interior 5 (mod 8) node. Notably, many OEEE-paths satisfy $a \geq b$ even when $2^e \geq 3^o$; the interior 5 (mod 8) condition is necessary but not sufficient for failure.

The proof of Theorem 2 is organised via the ordering target function $\Delta = (3^o - 2^e)a + K$ (Section 6), whose sign equals $\text{sgn}(b - a)$. The negative parity state can only be entered via OEE steps (Lemma 6.3), both subcases of which give $\Delta < 0$ strictly and are fully proved. OE steps sustaining the negative parity state are the subject of Conjecture 1, on which Theorem 2 is conditional.