THE COMPRESS-WITH-ANOTHER THRESHOLD OF THE SERIES $\mathcal{A}_k$: AN UPPER BOUND $4n/3$ FOR EVERY k, EXACT VALUES TO $n = 186$, AND A ONE-LETTER CORRECTION

KOREA SUPERINTELLIGENCE LABS

Abstract. Szykuła's recent survey of open problems on synchronising automata introduces the *compress-with-another* threshold of a state — the length of a shortest word sending that state and some other state to a common image — and records a conjecture, for a series $\mathcal{A}_k$ of binary automata on $n = 3k + 6$ states credited to a 2018 master's thesis of Dżyga: $\mathcal{A}_k$ is synchronising, and the shortest word compressing q_0 with another state is $(ab)^{k+2}a(ab)^{k+2}$, of length $\frac{4}{3}n$. The survey reports a computational check for $n \le 21$ only and says that even the synchronisation clause seems difficult to prove; it proves nothing about the series. We prove, for *every* $k \ge 1$ and with no computation, that the word $(ba)^{k+2}(ab)^{k+2}$, of length exactly $\frac{4}{3}n$, sends q_0 and q_2 to a common state, so the compress-with-another threshold of q_0 is at most $\frac{4}{3}n$ for every k. The same identity locates a one-letter slip in the conjecture: $(ab)^{k+2}a(ab)^{k+2} = a \cdot (ba)^{k+2}(ab)^{k+2}$, and a fixes q_0, so the exhibited word is the shorter word with a redundant leading letter and has length $\frac{4}{3}n+1$, one more than the value stated in the same sentence. Both halves of that sentence are true, of different quantities: we show that the merging distance of the *pair* $\{q_0, q_1\}$ is exactly $\frac{4}{3}n + 1$ and is attained by the exhibited word, while the threshold — the minimum over all partners — is exactly $\frac{4}{3}n$, both for $k = 1, \dots, 60$, that is $n = 9, \dots, 186$. The matching lower bounds rest on a ranking-function certificate whose three checked properties imply the bound however the certificate was produced. We also exhibit reset words showing $\mathcal{A}_k$ synchronising for $k \le 40$ ($n \le 126$), and determine the pairwise merging diameter of $\mathcal{A}_k$ for $k \le 16$: it is $(5k^2 + 32k + 47)/4$ for odd k and $(5k^2 + 34k + 44)/4$ for even k, asymptotically $\frac{5}{36}n^2$. Every theorem below is machine-checked in Lean 4. The general lower bound $\mathrm{cwa}(q_0) \ge \frac{4}{3}n$, and synchronisation for every k, remain open.

1. Introduction

The quantity. Let $\mathcal{A} = (Q, \Sigma, \delta)$ be a deterministic finite automaton with a totally defined transition function, extended to words by $\delta(q, \varepsilon) = q$ and $\delta(q, cw) = \delta(\delta(q, c), w)$. A word w is a *reset word* if $\delta(x, w) = \delta(y, w)$ for all $x, y \in Q$, and $\mathcal{A}$ is *synchronising* if it has one; the length of a shortest reset word is the reset threshold, and the Černý conjecture — that it is at most $(n-1)^2$ for an n-state synchronising automaton — is the organising problem of the area [6].

The survey [6] isolates a local relative of that quantity. We quote its definition verbatim ([6], §2.3):

> A word $w \in \Sigma^*$ *compresses a state with another* if $\delta(q, w) = \delta(p, w)$ for some $p \in Q \setminus \{q\}$. We call the length of the shortest such words the *compress-with-another* threshold.

So the quantity is attached to a state: writing

$$\mathrm{dist}(x, y) \;=\; \min\{\, |w| \;:\; w \in \Sigma^*,\ \delta(x, w) = \delta(y, w) \,\}$$

for the *merging distance* of a pair (with $\min \emptyset = \infty$), the compress-with-another threshold of q is

$$\mathrm{cwa}(q) \;=\; \min_{p \in Q \setminus \{q\}} \mathrm{dist}(q, p).$$

The survey adds that the problem "is very natural and closely related to the others, but so far largely overlooked and has not appeared in the published literature". Its Conjecture 6 states that in a synchronising strongly connected automaton with $n \ge 2$ states the compress-with-another

threshold of every state is $\mathcal{O}(n)$; strong connectivity is essential there, since without it a $\Theta(n^2)$ threshold is easy to arrange ([6], Proposition 7). For strongly connected automata, the survey reports, the threshold "rarely exceeds n", and "we know only one potential candidate for a larger value". That candidate is the series studied here.

The series, and what was conjectured. Section 2 reproduces the survey's definition of a series $\mathcal{A}_k$ of binary automata on $n = 3k + 6$ states, $k \geq 1$. The credit for it is one sentence carrying one reference — entry [9] of the survey's bibliography, which is Dżyga's master's thesis [2]: "There exists a series with a state whose compress-with-another threshold is likely to be $\frac{4}{3}n$ [9]. However, this fact seems to be very difficult to prove, and hence we leave it as a conjecture for this construction." About $\mathcal{A}_k$ the survey then states one conjecture, verbatim:

> **Conjecture 8.** For $k \geq 1$, the automaton $\mathcal{A}_k$ with $n = 3k + 6$ states is synchronizing, and the shortest word that compresses $\{q_0, q\}$ for some $q \in Q \setminus \{q_0\}$ is $(ab)^{k+2}a(ab)^{k+2}$ of length $\frac{4}{3}n$.

and, on the evidence for it:

> The series remains unproven, as proving its properties are difficult due to the lack of regularity in the behavior of this series. It seems to be difficult even to prove that these automata are synchronizing. The conjecture has been verified computationally for a small number of states $n \leq 21$.

Since $n = 3k + 6$, the window $n \leq 21$ is $k \leq 5$, and the smallest case the survey leaves open is $k = 6$, $n = 24$. The survey proves no statement about $\mathcal{A}_k$: apart from the assertion that the automaton is "obviously strongly connected", everything it says about the series is the conjecture, the remark that its properties are difficult to prove, and the report of that finite check.

Results. Write $u = ba$ and $v = ab$, and

$$w_k \;=\; u^{k+2}v^{k+2} \;=\; (ba)^{k+2}(ab)^{k+2}, \qquad |w_k| = 4k + 8 = \tfrac{4}{3}n.$$

- *An upper bound for every k* (Theorem 3.3). For every $k \geq 1$, $\delta_k(q_0, w_k) = \delta_k(q_2, w_k) = q_2$. Hence $\mathrm{cwa}(q_0) \leq \frac{4}{3}n$ for every $k \geq 1$. The proof is two inductions along alternating blocks and one case split on the parity of k; no computation enters it, and it settles the inequality on the whole series rather than on a finite window.
- *The conjecture's word is one letter too long* (Theorem 4.1). The free-monoid identity $(ab)^m a = a(ba)^m$ gives $(ab)^{k+2}a(ab)^{k+2} = a \cdot w_k$, and $\delta_k(q_0, a) = q_0$; so the word the conjecture exhibits is w_k with a redundant leading letter that q_0's a-self-loop absorbs. It does compress q_0 — with q_1 — but its length is $4k + 9 = \frac{4}{3}n + 1$, one more than the value $\frac{4}{3}n$ asserted for it in the same sentence.
- *The value is exact, well past the survey's window* (Theorem 5.2). $\mathrm{cwa}(q_0) = \frac{4}{3}n$ exactly, for $k = 1, \ldots, 60$, that is $n = 9, \ldots, 186$. The upper bound is Theorem 3.3 and needs no computation; the lower bound is one ranking-function certificate per k (Lemma 5.1).
- *The reading under which the conjecture names the right word* (Theorem 5.3). $\mathrm{dist}(q_0, q_1) = \frac{4}{3}n + 1$ exactly on the same range, and $(ab)^{k+2}a(ab)^{k+2}$ attains it. So both halves of the conjecture's sentence are true, of different quantities: $\frac{4}{3}n$ is the minimum over all partners, and the exhibited word is a shortest word for the single pair $\{q_0, q_1\}$, which costs one letter more.
- *Synchronisation* (Theorem 7.1). $\mathcal{A}_k$ is synchronising for $k = 1, \ldots, 40$, that is $n = 9, \ldots, 126$, by an explicitly constructed reset word. This is the conjecture's first clause, which the survey singles out as difficult on its own.
- *The pairwise merging diameter* (Theorem 8.1). The largest merging distance over all pairs of states of $\mathcal{A}_k$ is exactly $(5k^2 + 32k + 47)/4$ for odd k and $(5k^2 + 34k + 44)/4$ for even k, for $k = 1, \ldots, 16$; asymptotically $\frac{5}{36}n^2 \approx 0.139\,n^2$, against the elementary general bound $\binom{n}{2}$. The survey evaluates this quantity nowhere for the series.

Two further groups of statements support these and are of a different kind. Proposition 6.1 collects negative controls — among them that nothing of length $\frac{4}{3}n - 1$ compresses q_0, that the certificate machinery refuses the value $\frac{4}{3}n + 1$ instead of returning it, and that the survey's own hypothesis $k \geq 1$ is load-bearing, since at $k = 0$ the same formulas give threshold 7 rather than $4 \cdot 0 + 8 = 8$. Proposition 6.2 concerns the printed definition of $\mathcal{A}_k$: its case split is not total, at two arguments, and both readings supplied by the survey's own figure are shown to be forced, in the sense that either alternative destroys the survey's own value $\frac{4}{3}n$.

What is not claimed. We do not prove $\mathrm{cwa}(q_0) \geq \frac{4}{3}n$ for all k; that is the open half of the conjecture's value clause, and Section 9 says what makes it hard. We do not prove that $\mathcal{A}_k$ is synchronising for all k. We do not claim that w_k is the *unique* shortest compressing word, nor that q_2 and q_{2k+5} are the only partners attaining the minimum: both are true in every case we computed, and both are recorded below as computations outside the formal development (Remark 5.4). The closed form for the pairwise merging diameter is proved for $k \leq 16$ and conjectured beyond (Conjecture 8.2).

Provenance, and one caveat. Everything quoted above is from [6]; the arXiv source of v1 and the published version were both read, and they carry the same conjecture sentence and the same clause bounds in the definition of δ_k, so the discrepancy of Theorem 4.1 is not an artefact of one of the two. The two agree on the numbering as well — the conjecture is Conjecture 8, the eighth item of a single theorem counter, in the published text and in a compilation of the v1 source alike — and the bibliographic data of [6] (volume, pages, doi) were confirmed against the publisher's record and against Crossref.

The one caveat we cannot close concerns priority. The series is credited by [6] to Dżyga's 2018 master's thesis [2], which is not on arXiv; we tried to obtain it and could not. What we could establish is its provenance: the survey's author supervised the thesis and lists it publicly, as a 2018 master's thesis of Michalina Dżyga entitled *Synchronizing automata with extremal properties* (*Automaty synchronizowalne o ekstremalnych własnościach*), with no copy attached; and the single publication that the same list records as arising from it — [1], with the survey's author among the coauthors — was read and contains neither the series nor the compress-with-another quantity, so the part of the thesis at issue here has never been published. If the thesis nevertheless already contains the word $(ba)^{k+2}(ab)^{k+2}$ and the argument of Section 3, then Theorems 3.3 and 4.1 are re-derivations, and the one-letter discrepancy is a transcription slip between the thesis and the survey rather than a slip in the conjecture as conceived. The survey's own wording — "likely to be", "remains unproven", "seems to be very difficult to prove" — is evidence that the thesis contains no proof, but it is no evidence at all about which word the thesis prints. We state the results in the form that is correct either way: what is certainly new is that these facts now have proofs.

Beyond the survey, the compress-with-another quantity appears in no other source we could find: a full-text search of a large snapshot of arXiv sources for the phrases *compress-with-another* and *compresses a state with another* returns exactly one paper, [6] itself, and a second search for the literal string `(ab)^{k+2}` returns the same single paper. Two adjacent papers by the survey's author and coauthors, [4] and [3], were read and contain neither the series nor the quantity; their occurrences of "compress" concern compressing subsets, a different notion. A bibliographic database reports no works citing [6], which appeared in late August 2026, days before this was written.

2. Preliminaries

Automata and words. Throughout, $\Sigma = \{a, b\}$, words are read left to right, $|w|$ is the length of w and ε the empty word. For a pair of states x, y we use $\mathrm{dist}(x, y)$ and $\mathrm{cwa}(q)$ as defined in Section 1. Two standard facts are used and not reproved: a finite automaton is synchronising if and only if every pair of its states can be merged; and, by breadth-first search in the automaton of unordered pairs, a pair that can be merged at all can be merged by a word of length at most $\binom{n}{2}$, since each level of the search claims at least one new pair. The survey reports $n(n-1)/2$ as the *tight* upper

bound for synchronising a two-element subset ([6], §2.4), and it is attained: the Černý automaton on n states has a pair of merging distance exactly $\binom{n}{2}$, which we confirmed for $n \le 12$. Only the upper bound is used below.

The series $\mathcal{A}_k$. We reproduce the definition of [6] verbatim. Let $k \ge 1$, let $Q_{3k+6} = \{q_0, \dots, q_{3k+5}\}$, and define $\mathcal{A}_k = (Q_{3k+6}, \{a, b\}, \delta_k)$ by

$$\delta_k(q_i, a) = \begin{cases} q_0 & \text{if } i = 0, \\ q_{i+1} & \text{if } 1 \le i \le 2k+3,\ 2 \nmid i, \\ q_{i-1} & \text{if } 1 \le i \le 2k+3,\ 2 \mid i, \\ q_{i+1} & \text{if } 2k+4 \le i < 3k+5,\ 2 \mid i, \\ q_{i-1} & \text{if } 2k+7 \le i < 3k+5,\ 2 \nmid i, \\ q_3 & \text{if } i = 2k+5, \\ q_{3k+5} & \text{if } i = 3k+5,\ 2 \mid i, \end{cases} \tag{1}$$

$$\delta_k(q_i, b) = \begin{cases} q_{i+1} & \text{if } 1 \le i \le 2k+3,\ 2 \mid i, \\ q_{i-1} & \text{if } 1 \le i \le 2k+3,\ 2 \nmid i, \\ q_{i+1} & \text{if } 2k+5 \le i < 3k+5,\ 2 \nmid i, \\ q_{i-1} & \text{if } 2k+6 \le i < 3k+5,\ 2 \mid i, \\ q_2 & \text{if } i = 2k+4, \\ q_{3k+5} & \text{if } i = 3k+5,\ 2 \nmid i. \end{cases} \tag{2}$$

The survey adds that "the automaton $\mathcal{A}_k$ is obviously strongly connected".

Two gaps, and how they are filled. The case split just displayed is not total. Exactly two arguments match no clause:

(1) $\delta_k(q_0, b)$: every clause of (2) requires $i \ge 1$.
(2) $\delta_k(q_{3k+5}, x)$ for one of the two letters. At $i = 3k+5$ the printed rules cover a when $3k+5$ is even (that is, k odd) and b when $3k+5$ is odd (k even); the other letter matches no clause, because the two chain clauses that would reach it stop at $i < 3k+5$.

Figure 3 of [6], to which the definition explicitly refers ("the automaton $\mathcal{A}_k$ … illustrated in Figure 3"), supplies both values. It draws a b-arrow $q_0 \to q_1$ (together with the b-arrow $q_1 \to q_0$ and the a-self-loop at q_0), and it draws three arrows at q_{3k+5}, each labelled "a or b": one in from q_{3k+4}, one back out to q_{3k+4}, and a self-loop. We therefore read

$$\delta_k(q_0, b) = q_1, \qquad \delta_k(q_{3k+5}, x) = q_{3k+4} \ \text{ for the letter } x \text{ not covered above}, \tag{3}$$

the first of which is also what the first clause of (2) gives if its range $1 \le i \le 2k+3$ is read as $0 \le i \le 2k+3$. These two values are not assumptions of convenience: Proposition 6.2 shows that either alternative filling contradicts the survey's own value $\frac{4}{3}n$, so (3) is the only reading under which Conjecture 8 can be discussed at all.

The shape of $\mathcal{A}_k$. It is worth saying in words what (1)–(3) define, since every proof below walks in this picture. $\mathcal{A}_k$ is two chains glued at two states:

- an *upper chain* $q_0, \dots, q_{2k+3}$, on which b transposes the pairs (q_0, q_1), (q_2, q_3), …, (q_{2k+2}, q_{2k+3}), while a fixes q_0, transposes (q_1, q_2), (q_3, q_4), …, (q_{2k+1}, q_{2k+2}), and sends $q_{2k+3} \mapsto q_{2k+4}$;
- a *junction* q_{2k+4}, with $a\colon q_{2k+4} \mapsto q_{2k+5}$ and $b\colon q_{2k+4} \mapsto q_2$;
- a *lower chain* $q_{2k+5}, \dots, q_{3k+5}$, on which b transposes (q_{2k+5}, q_{2k+6}), (q_{2k+7}, q_{2k+8}), … and a transposes (q_{2k+6}, q_{2k+7}), (q_{2k+8}, q_{2k+9}), …, with $a\colon q_{2k+5} \mapsto q_3$ the exit back to the upper chain and, at the apex q_{3k+5}, one letter self-looping and the other returning to q_{3k+4}.

Both letters are therefore near-involutions, each with exactly one collision: a merges $\{q_4, q_{2k+5}\}$ and b merges $\{q_3, q_{2k+4}\}$, and no other pair of distinct states is merged by a single letter. Every compressing word must end by driving its pair onto one of those two pairs, which is the mechanism behind every lower bound in this paper.

3. The upper bound, for every k

Throughout this section $k \geq 1$ is arbitrary, $u = ba$, $v = ab$ and $w_k = u^{k+2}v^{k+2}$, so $|w_k| = 4k + 8$ and $3|w_k| = 4n$, i.e. $|w_k| = \frac{4}{3}n$.

Lemma 3.1 (block steps)**.** *For every $k \geq 1$:*

(1) $\delta_k(q_i, u) = q_{i+2}$ *for even i with $0 \leq i \leq 2k + 2$, and* $\delta_k(q_{2k+4}, u) = q_1$*;*
(2) $\delta_k(q_i, v) = q_{i+2}$ *for odd i with $1 \leq i \leq 2k + 1$, and* $\delta_k(q_{2k+3}, v) = q_2$*;*
(3) $\delta_k(q_i, v) = q_{i+2}$ *for even i with $2k + 4 \leq i \leq 3k + 3$, and* $\delta_k(q_i, v) = q_{i-2}$ *for odd i with $2k + 7 \leq i \leq 3k + 4$;*
(4) *at the apex: if k is even then* $\delta_k(q_{3k+4}, v) = q_{3k+5}$ *and* $\delta_k(q_{3k+5}, v) = q_{3k+3}$*; if k is odd then* $\delta_k(q_{3k+5}, v) = q_{3k+4}$*;*
(5) *at the exit:* $\delta_k(q_{2k+5}, v) = q_2$*.*

Proof. Each item is two applications of (1)–(3). For (1), if i is even and $i \leq 2k + 2$ then $\delta_k(q_i, b) = q_{i+1}$ and $\delta_k(q_{i+1}, a) = q_{i+2}$, both by the upper-chain clauses (for $i = 0$ the first step is (3)); and $\delta_k(q_{2k+4}, b) = q_2$, $\delta_k(q_2, a) = q_1$. For (2), if i is odd and $i \leq 2k + 1$ then $\delta_k(q_i, a) = q_{i+1}$ and $\delta_k(q_{i+1}, b) = q_{i+2}$; and $\delta_k(q_{2k+3}, a) = q_{2k+4}$, $\delta_k(q_{2k+4}, b) = q_2$. For (3), on the lower chain $\delta_k(q_i, a) = q_{i+1}$ and $\delta_k(q_{i+1}, b) = q_{i+2}$ for even i in the stated range, while $\delta_k(q_i, a) = q_{i-1}$ and $\delta_k(q_{i-1}, b) = q_{i-2}$ for odd i in its range. For (4), if k is even then $3k + 5$ is odd, so b self-loops at the apex and a is the letter supplied by (3): $\delta_k(q_{3k+4}, a) = q_{3k+5}$ (as $3k + 4$ is even) and $\delta_k(q_{3k+5}, b) = q_{3k+5}$, giving the first equality, while $\delta_k(q_{3k+5}, a) = q_{3k+4}$ and $\delta_k(q_{3k+4}, b) = q_{3k+3}$ give the second. If k is odd then $3k + 5$ is even, a self-loops at the apex and b is supplied by (3), so $\delta_k(q_{3k+5}, v) = \delta_k(q_{3k+5}, b) = q_{3k+4}$. For (5), $\delta_k(q_{2k+5}, a) = q_3$ and $\delta_k(q_3, b) = q_2$. □

Proposition 3.2. *For every $k \geq 1$,*

$$\delta_k(q_0, u^{k+2}) = q_{2k+4}, \quad \delta_k(q_{2k+4}, v^{k+2}) = q_2, \quad \delta_k(q_2, u^{k+2}) = q_1, \quad \delta_k(q_1, v^{k+2}) = q_2.$$

Proof. The first: by Lemma 3.1(1) the block u climbs the even states of the upper chain by two, so $q_0 \mapsto q_2 \mapsto \cdots \mapsto q_{2k+4}$ in $k + 2$ blocks, the last step starting at q_{2k+2}, which is within range.

The third: $k + 1$ climbs carry q_2 to q_{2k+4}, and one more block turns it, $\delta_k(q_{2k+4}, u) = q_1$. The fourth: by Lemma 3.1(2), $k+1$ blocks v carry q_1 up the odd states to q_{2k+3}, and one more gives q_2.

The second is the only place the parity of k enters. Suppose k is even, say $k = 2j + 2$ with $j \geq 0$. By Lemma 3.1(3), $j + 1$ blocks v carry q_{2k+4} up the even states of the lower chain to $q_{2k+4+2(j+1)} = q_{3k+4}$; by (4) one block reaches the apex q_{3k+5} and one more comes off it at q_{3k+3}; by (3) again, j blocks descend from q_{3k+3} to $q_{3k+3-2j} = q_{2k+5}$; and by (5) one final block exits to q_2. That is $(j + 1) + 1 + 1 + j + 1 = 2j + 4 = k + 2$ blocks. Suppose instead k is odd, $k = 2j + 1$. Then $j+1$ climbs carry q_{2k+4} to $q_{2k+4+2(j+1)} = q_{3k+5}$, the apex itself; by (4) one block leaves it at q_{3k+4}; j descents reach $q_{3k+4-2j} = q_{2k+5}$; and one exit block gives q_2. That is $(j+1)+1+j+1 = 2j+3 = k+2$ blocks. □

Theorem 3.3. *For every $k \geq 1$,*

$$\delta_k(q_0, w_k) \;=\; \delta_k(q_2, w_k) \;=\; q_2,$$

where $w_k = (ba)^{k+2}(ab)^{k+2}$ has length $4k + 8$; equivalently $3|w_k| = 4n$. Consequently w_k compresses q_0 with another state, and

$$\mathrm{cwa}(q_0) \;\leq\; \tfrac{4}{3}n \qquad \textit{for every } k \geq 1.$$

Proof. By Proposition 3.2, $\delta_k(q_0, u^{k+2}) = q_{2k+4}$ and then $\delta_k(q_{2k+4}, v^{k+2}) = q_2$; also $\delta_k(q_2, u^{k+2}) = q_1$ and then $\delta_k(q_1, v^{k+2}) = q_2$. The length is $2(k+2) + 2(k+2) = 4k+8$ and $3(4k+8) = 12k+24 = 4(3k+6) = 4n$. □

This is the first proved statement about the series: [6] asserts the value $\frac{4}{3}n$ only as a conjecture, supported by a check at $n \le 21$, and the word it exhibits is longer. Theorem 3.3 is a bound for the whole series and uses no computation whatever.

4. The conjecture's word

Theorem 4.1. *Let $s_k = (ab)^{k+2}a(ab)^{k+2}$ be the word exhibited in Conjecture 8. For every $k \ge 1$:*

(1) *$s_k = a \cdot w_k$ as words;*
(2) *$\delta_k(q_0, s_k) = \delta_k(q_1, s_k) = q_2$, so s_k compresses q_0 with q_1;*
(3) *$|s_k| = 4k+9$, that is $3|s_k| = 4n+3$, so $|s_k| = \frac{4}{3}n + 1$; while*
(4) *w_k compresses q_0 and $|w_k| + 1 = |s_k|$.*

Proof. (1) In the free monoid $(ab)^m a = a(ba)^m$ for every $m \ge 0$, by induction on m; with $m = k+2$ this gives $s_k = (ab)^{k+2}a \cdot (ab)^{k+2} = a(ba)^{k+2}(ab)^{k+2} = aw_k$. (2) $\delta_k(q_0, a) = q_0$ by the first clause of (1), so $\delta_k(q_0, s_k) = \delta_k(q_0, w_k) = q_2$ by Theorem 3.3; and $\delta_k(q_1, a) = q_2$, so $\delta_k(q_1, s_k) = \delta_k(q_2, w_k) = q_2$, again by Theorem 3.3. (3) $|s_k| = 2(k+2) + 1 + 2(k+2) = 4k+9$ and $3(4k+9) = 4(3k+6) + 3$. (4) is Theorem 3.3. □

The conjecture's sentence is thus internally inconsistent by exactly one letter: it names a word of length $\frac{4}{3}n + 1$ and calls its length $\frac{4}{3}n$. The next section resolves the inconsistency in favour of the value, and shows that the word is not wrong either — it is a shortest word for a different quantity. Two features of the situation are worth recording, because they explain how such a slip survives a computational check. First, the extra letter is invisible from q_0: the a-self-loop at q_0 absorbs it, so any check that looks only at "does the exhibited word compress q_0?" passes. Second, the extra letter does change the partner, from q_2 to q_1 — and the sentence in [6] preceding the definition of the series speaks unambiguously of a threshold ("a series with a state whose compress-with-another threshold is likely to be $\frac{4}{3}n$"), which is why we read the value clause as the author's intent and the word as the slip.

5. The exact threshold

The lower bounds all come from one device, stated here in the generality in which it is used.

Lemma 5.1 (ranking certificate)**.** *Let $\mathcal{A} = (Q, \Sigma, \delta)$ be a finite automaton and let $\varphi\colon Q \times Q \to \mathbb{N}$ satisfy*

(C1) *$\varphi(x, x) = 0$ for every $x \in Q$;*
(C2) *$\varphi(x, y) \le \varphi(\delta(x, c), \delta(y, c)) + 1$ for all $x, y \in Q$ and all $c \in \Sigma$.*

Then $\varphi(x, y) \le \varphi(\delta(x, w), \delta(y, w)) + |w|$ for every word w and all $x, y \in Q$. In particular, if $\delta(x, w) = \delta(y, w)$ then $|w| \ge \varphi(x, y)$; that is, φ is a lower bound on merging distance.

Proof. Induction on w. For $w = \varepsilon$ the inequality is an equality. For $w = cw'$, (C2) gives $\varphi(x, y) \le \varphi(\delta(x, c), \delta(y, c)) + 1$ and the inductive hypothesis applied at $\delta(x, c), \delta(y, c)$ gives $\varphi(\delta(x, c), \delta(y, c)) \le \varphi(\delta(x, w), \delta(y, w)) + |w'|$; add. For the last sentence, put $z = \delta(x, w) = \delta(y, w)$ and use (C1): $\varphi(x, y) \le \varphi(z, z) + |w| = |w|$. □

The point of Lemma 5.1 is that nothing about φ is trusted. Conditions (C1) and (C2) are finite checks over $Q \times Q$ and Σ; whatever procedure produced the table φ is irrelevant to the proof, and a defective procedure yields a table that fails a check rather than a false theorem. In practice we take for φ_k the table of true merging distances of $\mathcal{A}_k$, computed by a backward breadth-first search from the diagonal in the automaton of pairs; but that search is a *generator* and appears nowhere in the argument.

Theorem 5.2. *For every $k \in \{1, 2, \ldots, 60\}$, that is for $n = 9, 12, \ldots, 186$,*

$$\mathrm{cwa}(q_0) \;=\; 4k + 8 \;=\; \tfrac{4}{3}n,$$

attained by $w_k = (ba)^{k+2}(ab)^{k+2}$.

Proof sketch. The upper bound is Theorem 3.3 and involves no computation. For the lower bound, fix k and let φ_k be the table described above. Three properties are checked exhaustively: (C1) at all n states, (C2) at all n^2 ordered pairs and both letters, and the start condition $\varphi_k(q_0, q_p) \geq 4k + 8$ for each of the $n - 1$ states $q_p \neq q_0$. (A fourth, bookkeeping check confirms that δ_k maps Q into Q, which the induction of Lemma 5.1 needs.) By Lemma 5.1, any word w with $\delta_k(q_0, w) = \delta_k(q_p, w)$ for some $p \neq 0$ has $|w| \geq \varphi_k(q_0, q_p) \geq 4k + 8$.

This is the only part of the theorem that rests on exhaustive computation, and the search is finite and explicit: over the sixty automata it consists of 732,330 ordered pairs, on each of which two inequalities of type (C2) are evaluated (1,464,660 in total), together with 5,790 start inequalities and one diagonal and one range check at each of the 5,850 states. No search over words is performed anywhere; the certificate replaces it. □

Theorem 5.3. *For every $k \in \{1, 2, \ldots, 60\}$,*

$$\mathrm{dist}(q_0, q_1) \;=\; 4k + 9 \;=\; \tfrac{4}{3}n + 1,$$

and the word $s_k = (ab)^{k+2}a(ab)^{k+2}$ of Conjecture 8 attains it. In particular $\mathrm{dist}(q_0, q_1) = \mathrm{cwa}(q_0) + 1$ *on this range.*

Proof sketch. The upper bound is Theorem 4.1(2)–(3), which holds for every $k \geq 1$ with no computation. For the lower bound, use the same table φ_k and the same checks (C1), (C2) as in Theorem 5.2, with the start condition read at the single cell $\varphi_k(q_0, q_1) \geq 4k + 9$; Lemma 5.1 then bounds the length of any word merging that pair. The exhaustive part is the same finite check as before, plus one further comparison per k. □

Theorems 5.2 and 5.3 together give the precise diagnosis of Conjecture 8 promised in Section 1. The conjecture's sentence names two things, a word and a length, and each is correct for a different quantity: on the whole checked range the minimum over partners is $\frac{4}{3}n$, realised by $(ba)^{k+2}(ab)^{k+2}$ at the partner q_2, while the pair $\{q_0, q_1\}$ — the pair the exhibited word merges — costs $\frac{4}{3}n + 1$, realised by the exhibited word itself. A corrected statement of the conjecture would read: the compress-with-another threshold of q_0 in $\mathcal{A}_k$ is $\frac{4}{3}n$, and $(ba)^{k+2}(ab)^{k+2}$ is a shortest word attaining it.

Remark 5.4. The following were computed outside the formal development, in an unformalised reference implementation of δ_k and in a second implementation written independently of it; the two agree on every value reported here. They are recorded as computations only, and no statement of this paper depends on them. For $k = 1, \ldots, 100$ ($n \leq 306$) the automaton $\mathcal{A}_k$ is strongly connected and synchronising, $\mathrm{cwa}(q_0) = 4k + 8$, the minimum is attained at exactly two partners, q_2 and q_{2k+5}, and $\mathrm{dist}(q_0, q_1) = \mathrm{dist}(q_0, q_{2k+4}) = 4k + 9$. For $k = 1, \ldots, 6$ the shortest compressing word is moreover *unique* and equals w_k, and the shortest word merging $\{q_0, q_1\}$ is unique and equals s_k, the word of Conjecture 8. Uniqueness is not formalised anywhere; only the lengths are.

Remark 5.5. The survey attributes the difficulty of the series to "the lack of regularity in the behavior", and the profile $p \mapsto \mathrm{dist}(q_0, q_p)$ shows what is meant. At $k = 6$ ($n = 24$) it reads, for $p = 1, \ldots, 23$,

33, **32**, 50, 49, 69, 68, 90, 89, 103, 102, 78, 77, 55, 54, 34, 33, **32**, 50, 49, 69, 68, 90, 89.

It is not monotone in p: it rises to 103 at $p = 9$ before falling back, and the two states nearest q_0 in the pair automaton, q_2 and $q_{17} = q_{2k+5}$, are not the two nearest in $\mathcal{A}_k$ itself. (Computed in the same unformalised implementation; nothing depends on it.)

6. Controls

A finite verification is only as good as the things it would have refused. The following statements are proved by the same means as the theorems above and are recorded because they bound what those theorems show.

Proposition 6.1. (1) (Too small is refuted.) *For $k = 1, \dots, 60$, no word of length at most $4k + 7 = \frac{4}{3}n - 1$ compresses q_0 with another state.*

(2) (Too large is refuted.) *For $k = 1, \dots, 60$, the value $4k + 9$ is not the compress-with-another threshold of q_0.*

(3) (The certificate is falsifiable.) *For every $k = 1, \dots, 60$, the three checks of Theorem 5.2 asked for the lower bound $4k + 9$ instead of $4k + 8$ come out* false.

(4) (The hypothesis $k \geq 1$ is load-bearing.) *The formulas* (1)–(3) *at $k = 0$ define a 6-state automaton whose compress-with-another threshold of q_0 is 7, not $4 \cdot 0 + 8 = 8$; the word baaabab attains it, merging q_0 with q_5.*

(5) (Non-vacuity.) *The empty word compresses no state with another, and for every $k \geq 1$ some word does.*

Proof sketch. (1) is Lemma 5.1 with the certificate of Theorem 5.2; (2) follows from Theorem 3.3 alone, since w_k is shorter and compresses; (4) uses the same certificate machinery at $k = 0$ together with the exhibited witness, whose two trajectories are $q_0 \to q_1 \to q_2 \to q_1 \to q_2 \to q_3 \to q_4 \to q_2$ and $q_5 \to q_5 \to q_3 \to q_4 \to q_5 \to q_5 \to q_3 \to q_2$; (5) is immediate from the definitions and Theorem 3.3. Item (3) is the one that has to be checked and cannot be argued: it says that the same finite check, asked for a bound that is one too large, fails at every k, so that passing it is informative. □

Item (4) deserves a word. Conjecture 8 opens "For $k \geq 1$", and the restriction is not decoration: the same formulas make sense at $k = 0$ and the value there is 7, one below what $\frac{4}{3}n$ would predict. (The $k = 0$ threshold was also confirmed by brute force over all 2^7 words of length 7 in the unformalised reference implementation.)

Proposition 6.2. *Both readings in* (3) *are forced by the survey's own value $\frac{4}{3}n$:*

(1) *Under the alternative filling $\delta_k(q_0, b) = q_0$, the one-letter word b merges q_0 with q_1, so the threshold is 1 — for every $k \geq 1$.*

(2) *Under the alternative filling "q_{3k+5} is a sink under both letters", the word $(ba)^{k+2}$ followed by $k + 1$ alternating letters beginning with a, of length $3k + 5 < 4k + 8$, already merges q_0 with q_{2k+5} — for $k = 1, \dots, 8$.*

Proof sketch. (1) is immediate and holds for all k: $\delta_k(q_1, b) = q_0$ by the second clause of (2), so under that filling b itself merges the two states. (2) is a finite check: for each of the eight values of k the two trajectories of the displayed word of length $3k + 5$ are computed and compared. □

Neither alternative reproduces $\frac{4}{3}n$ — one gives 1 and the other at most $n - 1$ — so the figure's reading is the only one under which the survey's own statement can hold. In the unformalised reference implementation the sink filling in fact gives threshold exactly $3k + 5 = n - 1$ and destroys strong connectivity, which [6] asserts for $\mathcal{A}_k$; that computation is recorded here but is not needed, since the inequality above already excludes the reading.

7. Synchronisation

Theorem 7.1. *For every $k \in \{1, 2, \dots, 40\}$, that is for $n = 9, 12, \dots, 126$, the automaton $\mathcal{A}_k$ is synchronising.*

Proof sketch. For each such k a word is constructed by the standard greedy collapse — repeatedly pick a pair of the current image, append a word merging it, and replace the image by its image — where the pair to merge and the word merging it are chosen by following the table φ_k of Section 5 downhill. The constructed word is then *verified* to be a reset word by simulating it from all n states

and comparing the n results, which is 2,700 state simulations over the forty automata. As with the certificate, the guide is untrusted: a defective φ_k yields a word that fails the simulation check, not a false theorem. Nothing is stored — the words are recomputed as part of the verification — and no bound on their length is claimed. □

This settles the first clause of Conjecture 8 on a range about six times the survey's own in n. The survey calls this clause difficult on its own: "It seems to be difficult even to prove that these automata are synchronizing." It remains open for all k; see Section 9.

8. The pairwise merging diameter

Define the *pairwise merging diameter* of an automaton as $\max_{x\neq y} \operatorname{dist}(x, y)$, the worst-case cost of merging a single pair. For an n-state automaton in which every pair is mergeable it is at most $\binom{n}{2}$ (Section 2). The survey evaluates it nowhere for $\mathcal{A}_k$.

Theorem 8.1. *Put*

$$D(k) \;=\; \begin{cases} (5k^2+32k+47)/4 & \text{if } k \text{ is odd,} \\ (5k^2+34k+44)/4 & \text{if } k \text{ is even.} \end{cases}$$

For every $k \in \{1, 2, \ldots, 16\}$, every pair of states of $\mathcal{A}_k$ can be merged by a word of length at most $D(k)$, and some pair cannot be merged by any shorter word. Hence the pairwise merging diameter of $\mathcal{A}_k$ is exactly $D(k)$ for those k.

Proof sketch. Upper half: for each of the 18,936 ordered pairs occurring across the sixteen automata, a word is constructed by walking the φ_k-gradient from that pair and is then verified, by simulation, to merge the pair and to have length at most $D(k)$. Lower half: the certificate of Lemma 5.1 again, with (C1) and (C2) checked exhaustively and the start condition read at one extremal pair, exhibiting a pair x, y with $\varphi_k(x, y) \geq D(k)$. Since $D(k) > 0$, that pair is automatically a pair of distinct states. Both halves are exhaustive finite computations; the gradient walk and the table are untrusted generators, and only the simulations and the three checks enter the proof. □

Since $n = 3k + 6$, $D(k) = \frac{5}{36}n^2 + O(n) \approx 0.139\, n^2$, comfortably below $\binom{n}{2}$: at $k = 16$, $n = 54$, the values are 467 and 1431.

Conjecture 8.2. The pairwise merging diameter of $\mathcal{A}_k$ equals $D(k)$ for every $k \geq 1$.

The closed form was fitted on $k \leq 20$ and then checked against the unformalised reference implementation for $k = 1, \ldots, 30$, so thirty cells support it; Theorem 8.1 proves the first sixteen and the range $k = 17, \ldots, 30$ is a computation only, claimed here as evidence and not as a theorem. We note that the sequence of values $21, 33, 47, 65, 83, 107, 129, 159, 185, 221$ ($k = 1, \ldots, 10$) is not in the On-Line Encyclopedia of Integer Sequences (queried 3 September 2026: no matching sequence).

9. What remains open

Three questions are left, in decreasing order of interest.

The general lower bound. $\operatorname{cwa}(q_0) \geq \frac{4}{3}n$ for every k would, with Theorem 3.3, settle the value clause of Conjecture 8 outright. The structure described at the end of Section 2 makes the problem sharp: the only pairs merged by a single letter are $\{q_4, q_{2k+5}\}$ (by a) and $\{q_3, q_{2k+4}\}$ (by b), so the question is exactly how long a word must be to drive some pair (q_0, q_p) onto one of those two. The obstruction is the irregularity displayed in Remark 5.5: the profile $p \mapsto \operatorname{dist}(q_0, q_p)$ is not monotone, so no naive potential argument along the chains can work, and the certificates of Section 5 are per-k tables rather than a uniform formula. A closed form for the whole profile would presumably give the bound as a corollary.

Synchronisation for every k. Both letters are near-involutions with one collision each; a uniform family of reset words, rather than the per-k greedy words of Theorem 7.1, would settle the first clause of Conjecture 8.

The diameter. Conjecture 8.2, and with it the constant $\frac{5}{36}$ in the asymptotics.

10. Formal verification

Every lemma, proposition and theorem stated above is formalised and machine-checked in Lean 4 [5]; the exceptions, stated as such in the text, are Remarks 5.4 and 5.5, Conjecture 8.2, the two parenthetical sentences reporting computations from the unformalised reference implementation (the $k = 0$ brute force in Section 6 and the exact value under the sink filling), and the two standard facts quoted in Section 2. The development is five files — definitions, the general theory of Sections 3–4, the values of Section 5, the controls of Section 6, and the synchronisation and diameter results — and uses core Lean only, with no mathematical library: states are natural numbers, letters are booleans and words are lists of booleans, so that no coercion appears in any arithmetic side goal. There is no incomplete proof. The axiom set of every headline statement was printed. Theorems 3.3 and 4.1, Proposition 6.2(1) and all the block lemmas of Section 3 depend on propositional extensionality and quotient soundness alone — no compiled evaluation, no choice — which is the formal counterpart of their being proofs for all k rather than checks over a range. The finite-range statements (Theorems 5.2, 5.3, 7.1, 8.1 and Proposition 6.1(1), (3), (4), together with Proposition 6.2(2)) additionally depend on compiled evaluation, through exactly seven such axioms in the whole development, one for each named computation: the certificate check of Section 5, the falsifiability control of Proposition 6.1(3), the two computations behind Proposition 6.1(4), the sink check of Proposition 6.2(2), the reset-word check of Theorem 7.1 and the diameter check of Theorem 8.1. Items (2) and (5) of Proposition 6.1 use none of them; the first half of item (5) uses no axioms at all. Each of those seven runs a construction and then verifies its output, so that the constructions — the breadth-first search producing φ_k, the greedy collapse, the gradient walk — are never trusted. As a further guard against a mistranscribed automaton, the transition tables computed inside Lean were compared entry by entry, for $k = 1, \ldots, 13$ (702 transitions), against the independent reference implementation used for Remarks 5.4 and 5.5; they agree. The whole development checks in about 2.1 CPU-minutes. The repository is private; repository reference to be added at submission.

References

[1] M. Dżyga, R. Ferens, V. V. Gusev and M. Szykuła, *Attainable Values of Reset Thresholds*, in: 42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017), LIPIcs 83, Schloss Dagstuhl, 2017, Article 40, pp. 40:1–40:14; DOI 10.4230/LIPIcs.MFCS.2017.40. The one publication recorded as arising from [2]; read on 3 September 2026 and checked for the series and for the quantity of this paper, and it contains neither.

[2] M. Dżyga, *Synchronizing automata with extremal properties* (*Automaty synchronizowalne o ekstremalnych własnościach*), Master's thesis, University of Wrocław, 2018. Cited as the origin of the series $\mathcal{A}_k$ by [6], in whose bibliography the entry reads "M. Dżyga (2018)". The author's full name, Michalina Dżyga, the Polish title and the supervision by the author of [6] are from that author's public list of supervised diploma theses, consulted on 3 September 2026; the thesis itself is not on arXiv and we were unable to obtain a copy. See the caveat in Section 1.

[3] A. Kisielewicz, J. Kowalski and M. Szykuła, *Experiments with Synchronizing Automata*, in: Implementation and Application of Automata (CIAA 2016), Lecture Notes in Computer Science 9705, Springer, 2016, pp. 176–188; DOI 10.1007/978-3-319-40946-7_15; arXiv:1607.04025. Read on 3 September 2026 and checked for the series and the quantity of this paper; it contains neither.

[4] A. Kisielewicz and M. Szykuła, *Synchronizing Automata with Extremal Properties*, in: Mathematical Foundations of Computer Science 2015 (MFCS 2015), Lecture Notes in Computer Science, Springer, 2015, pp. 331–343; DOI 10.1007/978-3-662-48057-1_26; arXiv:1608.01268. Cited only because its title coincides with that of [2]; read on 3 September 2026 and checked for the series and the quantity of this paper, and it contains neither. Its occurrences of "compress" concern compressing subsets.

[5] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science, Springer, 2021, pp. 625–635; DOI 10.1007/978-3-030-79876-5_37.

[6] M. Szykuła, *Synchronizing Automata: Open Problems*, arXiv:2608.24245v1 (25 August 2026), cs.FL; Electronic Proceedings in Theoretical Computer Science **451** (2026), 33–47 (AFL 2026); DOI 10.4204/EPTCS.451.3 (volume, pages and doi confirmed against the publisher's volume record and against Crossref). All quotations in this paper are from the arXiv source of v1, retrieved on 3 September 2026, and were checked against the published version, which carries the same conjecture sentence and the same clause bounds in the definition of δ_k. In the published version and in a compilation of the v1 source alike, the conjecture we discuss is Conjecture 8 and the figure we read is Figure 3.