A uniform 7/48 cubic upper bound for reset thresholds

Sungbin Hong
Department of Electrical and Electronic Engineering
Faculty of Engineering, The University of Tokyo
Tokyo, Japan
hongsungbin2@g.ecc.u-tokyo.ac.jp

November 13, 2026

hexagon:2611.00506v1

Abstract

We prove that every complete synchronizing deterministic finite automaton A with n states satisfies $\mathrm{rt}(A) \le \frac{7}{48}n^3+o(n^3)$, uniformly over the automaton and its finite alphabet. The coefficient is the milestone proposed in Trahtman's 2011 half-rank approach. The argument fixes a short word whose kernel is stable over a dyadic length interval. Degree transport in the graphs of pairs mergeable by short words, combined with a simultaneous selection of times of small degree growth, gives a compression bound for every sufficiently large subset of states. The bound remains available after replacing the current word. We combine it with the cheap-gap construction and the Pin–Frankl rank upgrade: the cost of reaching half rank is paid by a saving in the subsequent rank reductions. The required scalar comparison has an elementary analytic proof over its entire parameter domain. An explicit two-parameter estimate proves uniformity and extends the result from strongly connected automata to arbitrary complete synchronizing automata. The full theorem, including the word constructions used as classical inputs, has also been formalized in Lean 4.

Keywords. Synchronizing automaton; reset threshold; Černý conjecture; degree transport; formal proof.
2020 Mathematics Subject Classification. 68Q70, 05C35.

1 Introduction

A reset word of a deterministic finite automaton sends all states to one state. The least length of such a word is its reset threshold. The Černý conjecture predicts the bound $(n-1)^2$ for every synchronizing automaton with n states. Our result concerns the cubic upper bound and does not establish the conjectured quadratic bound.

The classical Pin–Frankl estimate is $(n^3-n)/6$ [2, 3]. The specific milestone 7/48 comes from Trahtman's 2011 proposal [7], whose contribution and subsequent withdrawal are discussed below. Szykuła developed the avoiding-word approach into a proved improvement of the general cubic coefficient [4]. Shitov's subsequent estimate has coefficient 8257/49923, approximately 0.1654 [5]. For classical background and recent open problems we refer to [9, 10]. The quadratic benchmark is attained by Černý's classical family [1]. An earlier preprint of the present author studies degree transport and a different improvement of the cubic coefficient [11]. The proof below gives the required transport arguments directly and does not use the upper-bound theorem of that preprint.

Theorem 1.1. *Every complete synchronizing deterministic finite automaton A with n states over a finite input alphabet satisfies*

$$\mathrm{rt}(A) \le \frac{7}{48}n^3 + o(n^3).$$

The remainder is uniform over all such automata and input alphabets.

Equivalently, for every $\zeta > 0$ there is an integer $N(\zeta)$ such that, for every $n \ge N(\zeta)$ and every such A, $\mathrm{rt}(A) \le (7/48 + \zeta)n^3$. In particular, if $R(n)$ is the maximum reset threshold among all such n-state automata, then $\limsup_{n\to\infty} R(n)/n^3 \le 7/48$.

The order of quantifiers is part of the assertion: $N(\zeta)$ is independent of the alphabet and the transition function. All word lengths in this paper count letters of the original alphabet.

1.1 Trahtman's proposal and the origin of the milestone

Trahtman's paper *Modifying the upper bound on the length of minimal synchronizing word*, published at FCT 2011 [7], proposed the finite upper bound

$$\mathrm{rt}(\mathcal{A}) \le \frac{n(7n^2 + 6n - 16)}{48} \qquad \text{(the 2011 claim, subsequently withdrawn).} \tag{1.1}$$

Its structural idea was to reach an image of approximately half the states at quadratic cost and then apply the classical compression bound to the remaining image. In the published-version preprint, Lemma 4 claimed a word of length at most k^2 and corank at least k up to the halfway range; Theorem 2 combined this with the completion sum. This is the starting point for the numerical target of the present work.

The coefficient can be seen directly from the classical rank charges. At normalized corank a, their leading cost is $a^2n^2/2$ per rank. Therefore the completion after reaching half rank has leading cost

$$n^3 \int_{1/2}^{1} \frac{a^2}{2}\,\mathrm{d}a = \frac{7}{48}n^3, \qquad n^3 \int_{0}^{1/2} \frac{a^2}{2}\,\mathrm{d}a = \frac{1}{48}n^3. \tag{1.2}$$

A uniform quadratic construction of the first half would remove the second contribution and leave $7/48$ as the cubic coefficient. This explains both the significance of the proposed value and why the cost before half rank is decisive.

The claimed bound was not established by the 2011 proof. Its Lemma 3 asserted short avoidance of any prescribed state. Gonze, Jungers, and Trahtman [6] gave a four-state counterexample in which avoiding one state requires six letters, contradicting the asserted bound of n letters. The arXiv record [8] states that the paper was withdrawn because this key lemma was wrong and the paper's assertion had not been proved. We distinguish that unproved polynomial claim from the two-phase idea that continued to be useful.

That idea is explicitly credited to Trahtman in Shitov's introduction [5]; Szykuła's avoiding-word method [4] provides the intervening linear-algebraic ingredient. The singleton-fiber mechanism also links these constructions: omitting a singleton input of a current word can lower its rank after a suitable splice. In the stable-rank upgrade used here, avoidance is obtained under an explicit rank-stability condition, through the valid avoid-or-compress lemma. No assertion of uniformly short avoidance of a prescribed state is used. Theorem 1.1 establishes the uniform asymptotic $7/48$ coefficient; it does not assert the withdrawn finite polynomial (1.1) or the old quadratic half-rank lemma.

1.2 Contribution and proof strategy

The constant 7/48 is the leading cost of the usual Pin–Frankl completion after reaching half rank. The obstacle is the cost of reaching that rank. We pay this cost by making the completion cheaper, rather than requiring the entire first half to have subcubic cost. A short plateau word of corank C supplies a uniform first-compression cost near C^2, valid down to approximately half the number of its kernel blocks. Consequently it continues to save rank-reduction cost after the current word has already reached half rank.

For normalized parameters $c = C/n$ and $x = \rho/n$, where ρ marks the first long gap in minimum corank lengths, the leading prefix cost is at most

$$p(c,x) = \min\{c^2(1/2 - x),\ 2(x-c)_+ W(x)\}, \qquad W(x) = \begin{cases} 1/16, & x \le 1/4, \\ x(1/2-x), & x \ge 1/4. \end{cases} \tag{1.3}$$

Here $y_+ = \max\{y, 0\}$. The saving in the completion is

$$S(c) = \int_{1/2}^{(1+c)/2} (a^2/2 - c^2)_+ \, \mathrm{d}a. \tag{1.4}$$

The central scalar fact is $p(c,x) \le S(c)$ for $0 \le c, x \le 1/2$. Its proof in section 7 uses polynomial identities and interval restriction; it requires no sampling or numerical optimization.

The present argument uses short paths between two states, pays the rank loss of those paths, and applies avoid-or-compress with its explicit set and stability hypotheses. In particular, the positive leading prefix cost $p(c,x)$ is paid by the completion saving $S(c)$; it is not necessary to establish the quadratic half-rank claim discussed in section 1.1.

2 Words, compression, and classical upgrades

Let $\mathcal{A} = (Q, \Sigma, \delta)$ be a complete deterministic finite automaton, with $|Q| = n \ge 1$. Write $q \cdot w$ for the right action of a word $w \in \Sigma^*$, and $T \cdot w = \{q \cdot w : q \in T\}$ for a subset $T \subseteq Q$. Thus $q \cdot (uv) = (q \cdot u) \cdot v$. The rank of w is $\mathrm{rk}(w) = |Q \cdot w|$, and its corank is $n - \mathrm{rk}(w)$. Synchronization means that some word has rank one. A word compresses T if $|T \cdot w| < |T|$. For fixed n, there are only n^n distinct transition maps of a letter. Removing duplicate letters preserves shortest word lengths, so the maximum $R(n)$ over arbitrary finite alphabets is well defined.

For integers $t \ge 0$, let E_t be the simple graph on Q in which distinct p, q are adjacent if some word of length at most t merges them. The graphs E_t increase with t. Write $e_t = |E(E_t)|$, $\delta_t = \min_q \deg_{E_t}(q)$, and $\Delta_t = \max_q \deg_{E_t}(q)$. A subset is independent in E_t exactly when no word of length at most t compresses it.

We use the following standard constructions. They are also proved in the accompanying formalization, as described in section 9.

Lemma 2.1 (Pin–Frankl compression)**.** *If $\mathcal{A}$ is synchronizing and $T \subseteq Q$ has cardinality $s \ge 2$, some word of length at most $\binom{n-s+2}{2}$ compresses T. In particular, a word of corank $r \le n-2$ can be extended to a word of larger corank at extra cost at most*

$$\mathrm{PF}(r) = \binom{r+2}{2}.$$

Starting from a word of corank at least C, the remaining reset cost is at most

$$\sum_{r=C}^{n-2} \mathrm{PF}(r) = \binom{n+1}{3} - \binom{C+2}{3} \le \frac{n^3 - C^3}{6}. \tag{2.1}$$

The subset bound is the classical consequence of Frankl's set-pair theorem and Pin's shortest-word argument [2, 3]. The last identity telescopes. If a rank reduction skips coranks, omitting their nonnegative charges can only decrease the total cost.

Lemma 2.2 (Avoid-or-compress and stable-rank upgrade)**.** *Suppose $\varnothing \neq P \subsetneq T \subseteq Q$ and some word z satisfies $P \nsubseteq T \cdot z$. Then there is such an avoidance, or a compression of T, within $n - |P|$ letters. Consequently, if u has corank $r \geq 1$, $2(r+1) \leq n$, and a word v satisfies*

$$\mathrm{rk}(v) \leq \mathrm{rk}(w) \quad \text{whenever } |w| \leq |v| + 2r, \tag{2.2}$$

there is a word of corank at least $r+1$ and length at most $|u| + |v| + 2r$.

The first assertion is Szykuła's avoiding-word lemma; the second is the stable-rank construction in Shitov's argument, within the Trahtman–Szykuła half-rank approach discussed in section 1.1, [4, Lemma 2], [5, Theorem 2]. We recall the second deduction to specify the actual word involved. Let P be the inputs lying in singleton fibers of u. At most $2r$ inputs lie in nonsingleton fibers, so $|P| \geq n - 2r \geq 2$. Put $T = Q \cdot v$. If $P \nsubseteq T$, use the empty suffix. If $P \subsetneq T$, synchronization permits eventual avoidance of some element of P; avoid-or-compress and (2.2) give a word z, $|z| \leq n - |P| \leq 2r$, with $P \nsubseteq Q \cdot vz$. The compression alternative contradicts (2.2). If $P = T$, some letter must change P as a set, since otherwise every word preserves a set of at least two states. Such a letter already omits an element of P. In every case the actual word vzu misses the unique output associated with an omitted singleton input of u, and its image is contained in $Q \cdot u$. Its rank therefore decreases as required.

Lemma 2.3 (Accounting for replacement words)**.** *Suppose that for each actual word of corank $r < T$ under consideration, an actual word of larger corank and length at most $|u| + b_r$ exists, where $b_r \geq 0$. From a word of corank at least C one can reach corank at least T at additional cost at most $\sum_{r=C}^{T-1} b_r$. The new words need not extend the previous words.*

Proof. Induct on $\max\{T - \mathrm{corank}(u), 0\}$. Apply the stated upgrade if the target has not been reached. After a jump in corank, the remaining sum is at most the old tail with its first term removed, since all skipped terms are nonnegative. This gives the claimed bound. □

3 A fixed kernel and degree transport

In this and the next two sections the automaton is strongly connected and synchronizing. Put $h = n - 1$. Every ordered pair of states has a connecting word of length at most h: a shortest directed path has no repeated vertex. Define

$$D(t) = \max_{|w| \leq t} (n - \mathrm{rk}(w)).$$

The maximum is attained, D is nondecreasing, and $0 \leq D(t) \leq n - 1$.

Lemma 3.1 (Dyadic plateau)**.** *For every integer $J \geq 1$, there are $j < J$ and a word v such that, with $\ell = (2^{j+1} - 1)h$, $C = D(\ell)$, and $m = n - C$,*

$$|v| \leq \ell < 2^J n, \qquad \mathrm{rk}(v) = m, \qquad D(h) \leq C, \tag{3.1}$$

$$a_0 := D(2\ell + h) - C \leq \left\lfloor \frac{n-1}{J} \right\rfloor. \tag{3.2}$$

In particular, every word of length at most $2\ell + h$ has rank at least $m - a_0$.

Proof. Write $\ell_i = (2^{i+1} - 1)h$. Then $2\ell_i + h = \ell_{i+1}$, and the J increments $D(\ell_{i+1}) - D(\ell_i)$ telescope to at most $n - 1$. One increment is at most $\lfloor (n-1)/J \rfloor$. Choose an attaining word at its left endpoint. The other assertions follow from monotonicity and $j + 1 \le J$. □

Fix this word v for all subsequent queries. Its kernel partitions Q into m nonempty blocks, labeled by the states of $Q \cdot v$. For a state q define

$$K_t(q) = \#\{\text{kernel blocks wholly contained in } N_{E_t}(q)\}, \qquad K_t = \max_{q \in Q} K_t(q).$$

Let Γ_t be the graph on the m blocks in which two distinct blocks are joined if all cross pairs are edges of E_t. All these definitions use the same kernel. Hence K_t, the graphs Γ_t, and their minimum and maximum degrees are nondecreasing with time.

Lemma 3.2 (Exact transport)**.** *For every $t \ge 0$, putting $d = D(h)$ and $B = \ell + h$, we have*

$$\max \deg(\Gamma_t) \le K_t, \tag{3.3}$$

$$K_t \le \deg_{\Gamma_{t+B}}(A) + a_0 \qquad \text{for every kernel block } A, \tag{3.4}$$

$$\deg_{E_t}(q) \le d + \deg_{E_{t+h}}(p) \qquad \text{for every } p, q \in Q. \tag{3.5}$$

Proof. Every quotient neighbor of a block is wholly adjacent to any chosen state of that block. This proves (3.3).

For (3.4), choose x with $K_t(x) = K_t$ and a word w of length at most h sending the label $A \cdot v$ to x. The word vw sends the whole block A to x and is constant on each original kernel block. Its image meets at least $m - a_0$ original blocks: the number of such blocks is exactly $\mathrm{rk}(vwv)$, to which lemma 3.1 applies. Thus at most a_0 blocks are missed. For each wholly adjacent block at time t that is met by this image, choose a source block mapped into it by vw. Distinct met blocks have distinct source blocks. Such a source block is different from A, since the image of A is x and an open neighborhood excludes x. For each pair of states in A and this source block, concatenate vw with a merging witness of length at most t. This proves their complete join by time $t + B$. At least $K_t - a_0$ source blocks have been counted.

For (3.5), take a connector u, $|u| \le h$, from p to q. Its image omits at most d states. Every state in $N_{E_t}(q) \cap Q \cdot u$ has a preimage different from p. These preimages are neighbors of p in E_{t+h}: append the actual merging witness after u. Distinct image states give distinct preimages. At most d earlier neighbors are absent from the image, which proves the inequality. □

4 A simultaneous time of small growth

The following selection keeps the kernel fixed and works uniformly for every queried horizon. This matters when the horizon is the shortest compression length of a subsequently chosen subset.

Lemma 4.1. *Let $M \ge 1$, $B = \ell + h$, and $L = B + 3h$. If $T \ge B + ML$, there is an integer q such that*

$$B \le q \le T, \qquad T - q \le ML, \tag{4.1}$$

$$K_q \le \deg_{\Gamma_q}(A) + a_0 + \tau \quad \text{for every block } A, \tag{4.2}$$

$$\Delta_q \le \delta_q + d + \tau, \qquad \tau = \left\lfloor \frac{3n}{M} \right\rfloor. \tag{4.3}$$

Proof. Put $a = T - (B + ML)$ and $t_i = a + B + iL$ for $0 \le i < M$. At these times consider the nonnegative loads

$$(K_{t_i} - K_{t_i - B}) + (\delta_{t_i + h} - \delta_{t_i}) + (\Delta_{t_i} - \Delta_{t_i - 2h}).$$

For each of the three monotone sequences, the corresponding intervals have disjoint interiors, so their increments sum to at most n. All indices are nonnegative, since $B \ge 2h$. The total load over all i is at most $3n$. Thus one load is at most $\lfloor 3n/M \rfloor$. Take its time as q. It satisfies (4.1). Apply (3.4) at $q - B$ and use the small increment of K to obtain (4.2). Finally (3.5) gives $\Delta_q \le d + \delta_{q+h}$; the selected minimum-degree increment proves (4.3). □

The third increment is included in the formal selection lemma, which also controls $\Delta_q - \Delta_{q-2h}$. Only the first two consequences are needed in the cost argument below.

5 Compression of every large subset

Let s be the number of inputs belonging to singleton kernel blocks. If the nonsingleton blocks have sizes $b_i \ge 2$, then $\sum_i (b_i - 1) = C$ and $\sum_i b_i \le 2C$. Consequently

$$s \ge n - 2C. \tag{5.1}$$

Lemma 5.1 (Independent-set capacity)**.** *Suppose $q \ge |v|$, and an independent set T in E_q has size b. If all degrees in Γ_q lie between $K_q - \eta_0$ and K_q, then*

$$(2b - m)K_q \le b\eta_0. \tag{5.2}$$

Proof. Each kernel block is a clique in E_q, since v merges its states. Thus T occupies b distinct blocks. They are independent in Γ_q, by the definition of complete joins. The number of edges from those blocks to their complement is at least $b(K_q - \eta_0)$ and at most $(m - b)K_q$. Rearrangement proves (5.2). □

Lemma 5.2 (Pair geodesic)**.** *If T has at least two states and its shortest compression length is l, then $e_q \ge q$ for every integer $0 \le q \le l$.*

Proof. The value l is the minimum merging distance of a pair in T. Choose such a pair and a shortest merging word of length l. The pairs obtained after its first i letters, for $0 \le i < l$, have merging distances exactly $l - i$. A shorter continuation would contradict shortestness of the original word. They are distinct, and the q pairs with distances $1, \dots, q$ are edges of E_q. □

Proposition 5.3 (Uniform subset compression)**.** *Use the fixed plateau word of lemma 3.1, choose $M \ge 1$, and put*

$$\eta_0 = a_0 + \left\lfloor \frac{3n}{M} \right\rfloor, \quad \tau = \left\lfloor \frac{3n}{M} \right\rfloor, \quad \Omega = 1 + B + M(B + 3h).$$

For any $\epsilon, \delta > 0$ satisfying $C \le (1/2 - \delta)n$, every subset $T \subseteq Q$ with

$$|T| \ge 2, \qquad |T| \ge m/2 + \epsilon n$$

can be compressed by an original-alphabet word of length at most

$$C^2 + \frac{\eta_0 n}{8\epsilon\delta} + \frac{\tau n}{2} + \frac{\eta_0 n}{4\epsilon} + \Omega. \tag{5.3}$$

The word v is chosen before T and is common to all these subsets.

Proof. Let l be the shortest compression length of T, which exists by synchronization. If $l \le \Omega$, (5.3) is immediate. Otherwise apply lemma 4.1 to $T_0 = l - 1$. We obtain $q < l$, $q \ge B \ge |v|$, and $l \le q + \Omega$. The set T is independent in E_q. By lemma 5.1, writing $b = |T|$ and using $b \le n$,

$$2\epsilon n K_q \le (2b - m)K_q \le b\eta_0 \le n\eta_0, \qquad K_q \le \frac{\eta_0}{2\epsilon}.$$

If $b > m$, independence already contradicts the kernel clique property; thus the same reasoning covers every possible long-time case.

The sum of the ordinary E_q degrees of all singleton inputs is at most nK_q. Indeed, count incidences from each original vertex: its singleton neighbors are whole kernel blocks and number at most K_q. Since (5.1) gives $s \ge 2\delta n > 0$, some singleton input has degree at most nK_q/s. By (4.3),

$$\Delta_q \le d + \tau + \frac{nK_q}{s}.$$

At most $2C$ original states are nonsingletons. Summing their degrees by this maximum and singleton degrees by the incidence bound yields

$$\begin{aligned} e_q &\le \frac{nK_q}{2} + C(d + \tau) + \frac{CnK_q}{s} \\ &\le C^2 + \frac{\eta_0 n}{4\epsilon} + \frac{\tau n}{2} + \frac{\eta_0 n}{8\epsilon\delta}. \end{aligned}$$

Here $d \le C \le n/2$ and $s \ge 2\delta n$ were used. Finally lemma 5.2 gives $q \le e_q$, and $l \le q + \Omega$ proves the result. □

5.1 Fixing the auxiliary parameter first

For an integer $k \ge 1$, define

$$\begin{aligned} &J = k^2, && M = k, && \eta_k = k^{-2} + 3k^{-1}, \\ &\delta_k = \eta_k^{1/8}, && \epsilon_k = \eta_k^{1/4}, && \theta_k = \frac{\eta_k^{5/8}}{8} + \frac{3}{2k} + \frac{\eta_k^{3/4}}{4}, \\ &A_k = 2^{k^2}, && H_k = (k+1)2^{k^2} + 3k + 1. \end{aligned} \tag{5.4}$$

For this fixed k, lemma 3.1 supplies $|v| \le A_k n$. Moreover $\eta_0 \le \eta_k n$, $\tau \le 3n/k$, and $\Omega \le H_k n$. The last inequality follows by bounding $B \le A_k n$, $h \le n$, and $1 \le n$. Substitution in (5.3) gives the following form.

Corollary 5.4. *For each $k \ge 1$, every strongly connected synchronizing n-state automaton has an actual word v, $|v| \le A_k n$, such that, if its corank C satisfies $C \le (1/2 - \delta_k)n$, then every subset of at least two states and of size at least $(n - C)/2 + \epsilon_k n$ can be compressed at cost at most*

$$C^2 + \theta_k n^2 + H_k n. \tag{5.5}$$

The error θ_k tends to zero as $k \to \infty$, whereas H_k is fixed when $n \to \infty$. No parameter in this corollary depends on the alphabet or on the queried subset.

6 Reaching half rank with cheap gaps

Put $H = \lfloor n/2 \rfloor$. For $0 \le i \le n - 1$, let λ_i be the minimum length of a word of corank at least i. Set $\lambda_0 = 0$ and $g_i = \lambda_{i+1} - \lambda_i$ where defined. Choose the least $\rho < H$ with $g_\rho > n$ if it exists, and otherwise put $\rho = H$. Then

$$\lambda_\rho \le n\rho \le n^2. \tag{6.1}$$

For $n \geq 4$, $\rho \geq 1$: a synchronizing automaton with more than one state has a singular letter, so $\lambda_1 = 1$.

If $\rho = H$, (6.1) and lemma 2.1 already give a reset word of length at most $(7/48)n^3 + 2n^2$. In the rest of this section assume $1 \leq \rho < H$ and $g_\rho > n$.

Lemma 6.1 (Gap upgrade)**.** *For $\rho \leq r < H$, a word of corank r admits an upgrade of extra cost at most*

$$2r + \sum_{\substack{0 \leq j < \rho \\ i_j \leq r}} 2i_j, \qquad i_j = \lceil g_j/2 \rceil. \tag{6.2}$$

Proof. Choose the first index $\tau \leq \rho$ whose gap exceeds $2r$. It exists because $g_\rho > n > 2r$. A shortest word of length λ_τ has corank exactly τ and is rank-minimal among words of length at most $\lambda_\tau + 2r$. Apply lemma 2.2. All earlier gaps are at most $2r$, so

$$\lambda_\tau = \sum_{j<\tau} g_j \leq \sum_{\substack{j<\rho \\ i_j \leq r}} 2i_j.$$

This is also the gap construction underlying [5, Theorem 5]. □

Proposition 6.2 (Cheap-gap prefix)**.** *If an actual word v has corank C and length at most ℓ, there is an actual word of corank at least H and length at most*

$$2(\rho - C)_+ n^2 W(\rho/n) + n\ell + 3n^2. \tag{6.3}$$

Here $(\rho - C)_+ = \max\{\rho - C, 0\}$, with C and ρ integers.

Proof. Start with a word of length λ_ρ and corank at least ρ, and use lemmas 2.3 and 6.1. In the full nominal sum over $r = \rho, \ldots, H-1$, the weight $2i$ of a gap in bucket i occurs exactly

$$(H - \max\{\rho, i\})_+ \tag{6.4}$$

times. The actual construction may skip ranks, reducing this sum. Let $C_0 = \min\{C, \rho\}$. The first C_0 gaps have total $\lambda_{C_0} \leq \lambda_C \leq \ell$, and their rounded weights sum to at most $\ell + C_0$, because $2\lceil g/2 \rceil \leq g + 1$. Their total exposure is at most $n(\ell + C_0) \leq n\ell + n^2$.

For each of the at most $\rho - C_0$ remaining gaps, put $y = i/n$ and $x = \rho/n$. Its exposure is at most

$$2n^2 y \left[1/2 - \max\{x, y\}\right]_+ \leq 2n^2 W(x).$$

For $y \leq x$ this follows from $y(1/2 - x) \leq x(1/2 - x)$. For $y \geq x$, the parabola $y(1/2 - y)$ is maximized at $1/4$ if $x \leq 1/4$, and at x otherwise. Values $y \geq 1/2$ contribute zero. Finally $\sum_{r=\rho}^{H-1} 2r \leq n^2$, and the initial word costs at most n^2. Adding these bounds proves (6.3). The same argument handles $C \geq \rho$ through $C_0 = \rho$. □

Proposition 6.3 (Combined actual prefix)**.** *Suppose the same word v has corank C, length at most ℓ, and supplies a compression cap $C^2 + U$, $U \geq 0$, for every subset of size at least two and at least $(n - C)/2 + \epsilon n$. Put $c = C/n$, $x = \rho/n$, and $E = \lceil \epsilon n \rceil$. Then an actual word u of corank at least H exists with*

$$|u| \leq p(c, x)n^3 + \frac{nU}{2} + \frac{En^2}{2} + n\ell + 3n^2. \tag{6.5}$$

Proof. There are two constructions. The gap construction is (6.3), whose leading term is $2(x-c)_+W(x)n^3$. For the direct construction, start at corank ρ and use the cap whenever

$$r + E \le R, \qquad R = \left\lfloor \frac{n+C}{2} \right\rfloor.$$

This guarantees $n - r \ge (n-C)/2 + \epsilon n$. At the remaining ranks below H, use the PF upgrade. Since $R \ge H$, at most E such ranks occur. Each PF charge is at most $n^2/2$. Overpaying every nominal rank by C^2+U, and then adding the exceptional charges, bounds the direct word length by

$$n^2 + (H-\rho)(C^2+U) + En^2/2 \le c^2(1/2-x)n^3 + nU/2 + En^2/2 + n^2.$$

Select the cheaper of these two actual words. Enlarging the errors to the common nonnegative remainder in (6.5) proves the proposition. The images of the two candidate words have not been identified with one another. □

7 Paying the prefix from the completion saving

We first establish the scalar comparison from equations (1.3) and (1.4) independently of the automaton.

Lemma 7.1 (Scalar payment)**.** *For $0 \le c, x \le 1/2$, one has $p(c,x) \le S(c)$.*

Proof. If $x \le c$, then $p(c,x) = 0$ and the assertion follows from nonnegativity of the integral. Suppose $x > c$ and put $d = 1/2 - c$.

First let $0 \le c \le 1/3$. Since $W(x) \le 1/16$, a convex combination of the two bounds defining p gives

$$p(c,x) \le \min\{c^2(1/2-x), (x-c)/8\} \le \frac{c^2 d}{1+8c^2}. \tag{7.1}$$

Use weights $1/(1+8c^2)$ and $8c^2/(1+8c^2)$ in the last step. On the integration interval $a^2/2 \ge 1/8 \ge c^2$, so

$$S(c) = \frac{c(3+3c-23c^2)}{48}.$$

Subtracting the last expression in (7.1) yields

$$S(c) - \frac{c^2 d}{1+8c^2} = \frac{cF(c)}{48(1+8c^2)}, \qquad F(c) = 3 - 21c + 49c^2 + 24c^3 - 184c^4.$$

For $t = 3c \in [0,1]$, the exact identity

$$F(t/3) = 3(1-t)^4 + 5t(1-t)^3 + \frac{22}{9}t^2(1-t)^2 + \frac{25}{9}t^3(1-t) + \frac{5}{81}t^4$$

shows that $F(c) \ge 0$.

Now let $1/3 \le c \le 1/2$. Since $x > c > 1/4$ and $2x \le 1$,

$$p(c,x) \le 2x(x-c)(1/2-x) \le (x-c)(1/2-x) \le d^2/4.$$

Restrict the integral defining $S(c)$ to $[3c/2, (1+c)/2]$, an interval of length d contained in its original interval. On writing $a = 3c/2 + t$ we obtain

$$S(c) \ge \int_0^d \left(\frac{c^2}{8} + \frac{3c}{2}t + \frac{t^2}{2} \right) \mathrm{d}t = \frac{c^2 d}{8} + \frac{3cd^2}{4} + \frac{d^3}{6} \ge d^2/4.$$

The last inequality uses $3c \geq 1$. These arguments include $c = 0$ and $c = 1/2$, where no strictly positive common margin is needed. □

The formal scalar proof establishes the stronger inequality $p(c, x) + c/2000 \leq S(c)$. We use only its weaker consequence above; the shorter two-case proof is included to make the payment transparent.

Proposition 7.2 (Mixed completion)**.** *Under the cap assumptions of proposition 6.3, assume $0 \leq c = C/n \leq 1/2$. From any actual word u of corank at least H one can reach rank one with a word w satisfying*

$$|w| \leq |u| + \left(\frac{7}{48} - S(c)\right) n^3 + nU + \frac{En^2}{2} + 3n^2. \tag{7.2}$$

Proof. At nominal corank $r \leq n - 2$, use the cap or the PF upgrade, whichever is cheaper, when $r + E \leq R = \lfloor (n+C)/2 \rfloor$; otherwise use PF. All options produce actual words. The cap applies to the actual image of each current word, including a word obtained by replacement.

The PF charge obeys

$$\mathrm{PF}(r) \leq r^2/2 + 2n, \qquad \mathrm{PF}(r) \leq n^2/2.$$

For bulk ranks, the selected cost is at most $\min\{r^2/2, C^2\} + U + 2n$. Of the ranks with $r \leq R$, at most E fail the bulk condition; charge an extra $n^2/2$ for each. This count follows by injecting such an r into $\{0, \ldots, E-1\}$ via $r \mapsto R - r$.

Set $A = (1 + c)/2$ and define the nondecreasing function

$$f_c(a) = \begin{cases} \min\{a^2/2, c^2\}, & a \leq A, \\ a^2/2, & a > A. \end{cases}$$

It can have an upward jump at A. Its left Riemann sum is still bounded above by its integral. Since $H/n \geq 1/2 - 1/(2n)$ and $f_c \leq 1/2$ on $[0, 1]$,

$$n^2 \sum_{r=H}^{n-2} f_c(r/n) \leq n^3 \int_{1/2}^{1} f_c(a)\, \mathrm{d}a + n^2/4.$$

The integral equals $7/48 - S(c)$. The summed U terms cost at most nU, the PF rounding costs at most $2n^2$, and the exceptional ranks cost at most $En^2/2$. The remaining Riemann endpoint allowance fits in $3n^2$ together with PF rounding. Apply lemma 2.3 to obtain (7.2). □

Corollary 7.3 (Interior finite estimate)**.** *Under the same assumptions, $\mathcal{A}$ has a reset word of length at most*

$$\frac{7}{48} n^3 + \frac{3}{2} nU + En^2 + n\ell + 6n^2. \tag{7.3}$$

Proof. Use proposition 6.3 and then proposition 7.2 from its selected actual word. The leading remainder is $(p(c, x) - S(c))n^3 \leq 0$ by lemma 7.1. □

8 Uniform finite bounds and the general case

With the parameters in (5.4), define

$$\mathcal{E}_k = \frac{\delta_k}{8} + 2\theta_k + 2\epsilon_k, \qquad B_k = 4(k+1)2^{k^2} + 20k + 100. \tag{8.1}$$

Then $\mathcal{E}_k \to 0$ as $k \to \infty$, since $\eta_k \to 0$ and all powers in (5.4) are positive. The constants are chosen for bookkeeping and are not optimized.

Theorem 8.1 (Finite strongly connected bound)**.** *Let $k \geq 1$ satisfy $\delta_k < 1/2$. Every strongly connected complete synchronizing automaton with $n \geq 4$ states has a reset word of length at most*

$$\left(\frac{7}{48} + \mathcal{E}_k\right) n^3 + B_k n^2. \tag{8.2}$$

Proof. Choose v from corollary 5.4 and let C be its corank. If $C \geq (1/2 - \delta_k)n$, use v and complete by PF. Since $0 < \delta_k < 1/2$,

$$\frac{n^3 - C^3}{6} \leq \left(\frac{7}{48} + \frac{\delta_k}{8}\right) n^3.$$

For example, integrate the derivative $a^2/2 \leq 1/8$ over $[1/2 - \delta_k, 1/2]$; larger C only improves the bound. The initial word costs at most $A_k n \leq B_k n^2$, so (8.2) follows in this case.

Otherwise $C \leq (1/2 - \delta_k)n$ and the uniform cap applies. If $\rho = H$, the shortcut following (6.1) suffices. If $\rho < H$, use corollary 7.3 with

$$U = \theta_k n^2 + H_k n, \qquad \ell = |v| \leq A_k n, \qquad E = \lceil \epsilon_k n \rceil \leq \epsilon_k n + 1.$$

The resulting length is at most

$$\left(\frac{7}{48} + \frac{3}{2}\theta_k + \epsilon_k\right) n^3 + \left(\frac{3}{2}H_k + A_k + 7\right) n^2.$$

Finally

$$\tfrac{3}{2}\theta_k + \epsilon_k \leq \mathcal{E}_k, \qquad \tfrac{3}{2}H_k + A_k + 7 \leq B_k,$$

which proves the theorem. The shortcut error $2n^2$ also fits inside this quadratic allowance. □

Lemma 8.2 (Entry into a closed core)**.** *A complete synchronizing automaton has a unique closed strongly connected component K. If $|K| = s$, an actual word of length at most $(n-s)^2$ sends all states into K. The restricted automaton on K is complete and synchronizing over the same alphabet.*

Proof. The finite condensation graph has a closed component. Two different closed components cannot synchronize, so it is unique, and every state can reach it. A shortest path from a state outside K into K has length at most $n - s$. Apply its word to a chosen outside state in the current image. States already in K stay there; the number of outside image states decreases by at least one. Repeating at most $n - s$ times gives the entry word at the stated cost. Restriction preserves completeness and synchronization, since the set is closed and nonempty. □

Theorem 8.3 (Finite bound for all inputs)**.** *For the same k, every complete synchronizing n-state automaton with $n \geq 6$ satisfies*

$$\mathrm{rt}(\mathcal{A}) \leq \left(\frac{7}{48} + \mathcal{E}_k\right) n^3 + (B_k + 2)n^2. \tag{8.3}$$

Proof. First enter the component K of lemma 8.2, at cost at most n^2. If $s = |K| \geq 4$, apply theorem 8.1 inside K and append its reset word. Since $s \leq n$ and all coefficients are nonnegative, this gives (8.3). If $s \leq 3$, the PF bound, or the coarse bound s^3, resets the restriction in at most $27 \leq n^2$ letters. Together with entry this costs at most $2n^2$, which is also within (8.3). □

Proof of theorem 1.1. Given $\zeta > 0$, first choose $k \geq 1$ such that $\delta_k < 1/2$ and $\mathcal{E}_k < \zeta/2$. Then choose an integer

$$N \geq 6, \qquad N > \frac{2(B_k + 2)}{\zeta}.$$

For all $n \geq N$, theorem 8.3 implies the required estimate. Both choices depend only on ζ. In particular they precede the alphabet and transition function, which proves uniformity. □

For an explicit common remainder, one can set

$$e(n) = \inf_{\substack{k \geq 1 \\ \delta_k < 1/2}} \left(\mathcal{E}_k + \frac{B_k + 2}{n} \right).$$

The indexing set is nonempty, $e(n) \geq 0$, and $e(n) \to 0$ by the same two-stage choice. For $n \geq 6$, (8.3) for every admissible k gives $\mathrm{rt}(\mathcal{A}) \leq (7/48 + e(n))n^3$. Thus $e(n)n^3$ is one common $o(n^3)$ remainder.

9 Formal verification

The accompanying Lean 4 development establishes theorem 1.1 under the declaration

`CernyComplete.uniform_seven_forty_eight`.

The threshold N is chosen before the state count, alphabet size, and transition function. Transitions are total functions, synchronization means that an actual word sends the whole state set to one point, and word length is `List.length` over the original alphabet. Thus the formal statement has the same uniform scope and cost convention as the theorem in this paper. The PF and stable-rank constructions are proved within the development; their intermediate interface propositions are discharged before the final theorem. Its recorded axiom dependencies are the standard `propext`, `Classical.choice`, and `Quot.sound`, with no additional axiom or unfinished proof.

The accompanying artifact records clean compilation of 32 mathematical modules and one audit module, module-by-module kernel replay, and a successful fresh replay of the final import closure, using Lean 4.35.0 and pinned dependencies. Finite diagnostics supplement these checks; the scalar inequality and final theorem are proved over their full domains. Source correspondence and reproduction instructions are given in Appendices A and B. The complete formalization, certificate, execution history, and file hashes are preserved in a separate artifact. That artifact is not included in this Hexagon deposit; the deposited source contains the paper and its AI-use disclosure.

Acknowledgments and use of AI tools

OpenAI Codex and Anthropic Claude were used extensively in exploratory reasoning, proof development, programming, formalization, auditing, and manuscript preparation. The artifact distinguishes separate AI review contexts, finite computations, and actual Lean executions. Those review contexts are not external human peer review. At the time of this deposit, the human contributor claims no personal checking or understanding of any part of the proof. We acknowledge the authors of the classical results and the upstream formalization contributors identified in Appendix A. No AI system is listed as an author, and no claim of an external academic review or a priority adjudication is made by the verification certificate.

A Formal statement and source correspondence

The proposition of `CernyComplete.uniform_seven_forty_eight` can be written schematically as follows:

```
forall epsilon : Real, 0 < epsilon -> exists N : Nat,
  forall n a : Nat, N <= n ->
  forall step : Fin n -> Fin a -> Fin n,
    Synchronizing step ->
    ResetWithin step ((coefficient + epsilon) * (n : Real)^3)
```

Here `coefficient` is $7/48$ in $\mathbb{R}$; `Synchronizing` asserts an actual rank-one word, and `ResetWithin` asserts such a word with `List.length` at most the bound. Finite labelings identify this statement with theorem 1.1.

Mathematical component	Principal Lean modules
Actual words, graphs, quotient capacity	`CernyWordKernel`, `CernyKernel`
Whole-block and ordinary transport	`CernyWholeBlocks`, `CernyDegreeTransport`
Connectors, plateau, simultaneous times	`CernyExistence`, `CernyExistenceCalm`
All-subsets compression cap	`CernyCompression`, `CernyCompressionCount`, `CernyCompressionGlobal`
Classical constructions proved internally	`CernyPFVendor`, `CernyPF`, `CernyStableLinear`, `CernyStableVectors`, `CernyStableAvoid`, `CernyStableUpgrade`
Actual prefix and mixed completion	`CernyPrefixGaps`, `CernyGapAssembly`, `CernyFiniteMixed`, `CernyFiniteAssembly`
Scalar payment, errors, general inputs	`CernyScalar`, `CernyError`, `CernyReduction`, `CernyUniform`, `CernyComplete`

The interface propositions `PFUpgrade` and `StableRankUpgrade` are discharged by

`CernyPF.pfUpgrade_of_sync`
`CernyStable.stableRankUpgrade_of_sync`.

The PF component incorporates Baobing Zhang's (`baobingzhang`) Apache-2.0 Lean formalization [12], with source, attribution, and license preserved in the artifact. Its bridge identifies the vendor and original-word evaluators. The stable-rank component proves avoid-or-compress by the linear-space argument and constructs the splice vzu explicitly.

B Reproducing the formal verification

The complete formalization is preserved separately as `completed-audit.zip`, with its certificate and reproduction guide. It is not part of the files deposited on Hexagon. The artifact records compilation of all 33 project modules with `lean -trust=0`, module-by-module replay by `leanchecker`, and a final `leanchecker -verbose -fresh CernyComplete` replay. The fresh replay checks imported logical declarations in a new environment; it does not recompile every mathlib source. The certificate in `project/verification.json` records the pinned tool and dependency revisions, exact commands, and successful exit statuses. Detailed dependency audits and execution logs are preserved in the artifact.

For readers with a copy of that artifact, extract it and, on Linux x86_64, run

```
cd cerny-7-48-completed-audit-2026-11-13/project
bash bootstrap_and_verify.sh --fresh-target
```

The bootstrap requires Python 3, `curl`, `git`, GNU `tar` with zstd support, and network access to download fixed tool and dependency versions. If that environment is already prepared, run `python3 verify.py -fresh-target` instead. The large Lean and mathlib installations are not embedded in the source distribution.

These records belong to the supplied completed-audit artifact; manuscript preparation did not repeat the full kernel replay.

References

[1] J. Černý, *Poznámka k homogénnym experimentom s konečnými automatmi*, Matematicko-fyzikálny časopis **14**(3) (1964), 208–216. In Slovak. https://dml.cz/dmlcz/126647.

[2] P. Frankl, *An extremal problem for two families of sets*, European Journal of Combinatorics **3** (1982), 125–127. doi:10.1016/S0195-6698(82)80025-5.

[3] J.-E. Pin, *On two combinatorial problems arising from automata theory*, in Combinatorial Mathematics, North-Holland Mathematics Studies **75** (1983), 535–548. doi:10.1016/S0304-0208(08)73432-7.

[4] M. Szykuła, *Improving the upper bound on the length of the shortest reset words*, 35th Symposium on Theoretical Aspects of Computer Science (STACS 2018), LIPIcs **96**, 56:1–56:13. doi:10.4230/LIPIcs.STACS.2018.56. arXiv:1702.05455.

[5] Y. Shitov, *An improvement to a recent upper bound for synchronizing words of finite automata*, Journal of Automata, Languages and Combinatorics **24** (2019), 367–373. doi:10.25596/jalc-2019-367. arXiv:1901.06542.

[6] F. Gonze, R. M. Jungers, and A. N. Trahtman, *A note on a recent attempt to improve the Pin–Frankl bound*, Discrete Mathematics & Theoretical Computer Science **17**(1) (2015), 307–308. doi:10.46298/dmtcs.2101. arXiv:1412.0975.

[7] A. N. Trahtman, *Modifying the upper bound on the length of minimal synchronizing word*, in Fundamentals of Computation Theory (FCT 2011), Lecture Notes in Computer Science **6914** (2011), 173–180. doi:10.1007/978-3-642-22953-4_15. Published-version preprint: arXiv:1104.2409v4. The claimed upper bound was not established; see [6, 8].

[8] A. N. Trahtman, *Withdrawal notice for arXiv:1104.2409*, version 7, March 21, 2014; the first withdrawn version is version 5, December 10, 2012. https://arxiv.org/abs/1104.2409v7.

[9] M. V. Volkov, *Synchronizing automata and the Černý conjecture*, in Language and Automata Theory and Applications (LATA 2008), Lecture Notes in Computer Science **5196** (2008), 11–27. doi:10.1007/978-3-540-88282-4_4.

[10] M. Szykuła, *Synchronizing automata: Open problems*, Electronic Proceedings in Theoretical Computer Science **451** (2026), 33–47. doi:10.4204/EPTCS.451.3. arXiv:2608.24245.

[11] S. Hong, *Degree transport and a smaller cubic coefficient for reset thresholds*, preprint, Hexagon 2611.00172v1, November 2026. https://hexagonmath.org/2611.00172v1.

[12] B. Zhang, *Lean formalization of the Pin–Frankl bound and Kari's theorem*, software contribution, 2026. Source submission and preserved provenance: https://github.com/TheJustinSunPrize/awards/pull/605.