Complexity of suffix-free regular languages

Janusz A. Brzozowski[a], Marek Szykuła[b]

[a] *David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1*

[b] *Institute of Computer Science, University of Wrocław, Joliot-Curie 15, PL-50-383 Wrocław, Poland*

Abstract

We study various complexity properties of suffix-free regular languages. A sequence $(L_k, L_{k+1}, \dots)$ of regular languages in some class, where n is the quotient complexity of L_n, is *most complex* if its languages L_n meet the complexity upper bounds for all basic measures. It is known that there exist such most complex sequences in several classes of regular languages. In contrast to this, we prove that there does not exist a most complex sequence in the class of suffix-free regular languages. However, we do exhibit two such sequences that together meet upper bounds for all basic measures.

Keywords: most complex, regular language, state complexity, suffix-free, syntactic complexity, transition semigroup

1. Introduction

We study complexity properties of suffix-free regular languages. A much shorter preliminary version of these results appeared in [19] without any proofs.

Motivation The *state complexity* $\kappa(L)$ of a regular language L over an alphabet Σ is the number of states in a minimal deterministic finite automaton (DFA) with input alphabet Σ recognizing L. The state complexity of a regularity preserving unary operation $\circ$ on regular languages is the maximal value of $\kappa(L_n^\circ)$ as a function of n, where L_n varies over all regular languages L_n with state complexity n. Similarly, the state complexity of a regularity preserving binary operation $\circ$ on regular languages is the maximal value of $\kappa(K_m \circ L_n)$ as a function of m and n, where K_m and L_n vary over all regular languages of state complexities m and n, respectively. Of special

Email addresses: `brzozo@uwaterloo.ca` (Janusz A. Brzozowski), `msz@cs.uni.wroc.pl` (Marek Szykuła)

interest are the state complexities of common operations on regular languages. The state complexities of union, product (concatenation), (Kleene) star and reversal were studied by Maslov [33] in 1970, but this work was not well known for many years. The paper by Yu, Zhuang, and Salomaa [37] inspired considerable interest in these problems and much research has been done on this topic in the past 22 years. The state complexity of an operation gives a worst-case lower bound on the time and space complexities of the operation. For this reason it has been studied extensively; see [5, 28, 36] for additional references.

Consider, for example, the product of two regular languages. To find its state complexity we need to establish an upper bound for it, and find two languages of state complexities m and n, respectively that meet this bound. It is known that $(m-2)2^n + 2^{n-1}$ is a tight upper bound on product. The languages that meet this bound are called *witnesses*. In general, different witnesses have been used for the two arguments and for different operations. However, Brzozowski [6] has shown that one witness and its slightly modified version called a *dialect* is sufficient to meet the bounds on all binary Boolean operations, product, star, and reversal. The DFA of this witness is shown in Figure 1. Let the language recognized by this DFA be $L_n(a,b,c)$, and let $L_n(b,a,c)$ be the language of the DFA obtained from that of Figure 1 by interchanging the roles of a and b. Then $L_m(a,b,c)$ and $L_n(b,a,c)$ meet the bound mn for all binary boolean operations, $L_m(a,b,c)$ and $L_n(a,b,c)$ meet the bound $(m-1)2^n + 2^{n-1}$ for product, and $L_n(a,b,c)$ meets the bounds $2^{n-1} + 2^{n-2}$ for star and 2^n for reversal.

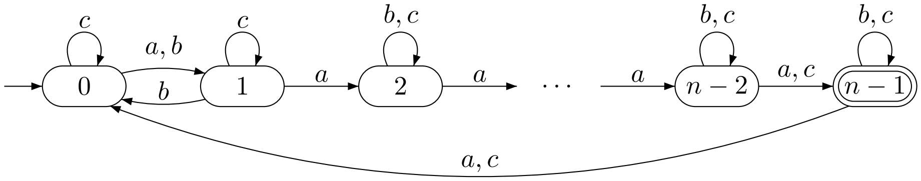

Figure 1: Minimal DFA of a most complex regular language.

Although state complexity is a useful measure, it has some deficiencies. The language $L_n(a,b,c)$ and the language $\{a,b,c\}^{n-2}$ both have state complexity n, but the second language is intuitively much simpler. Secondly, all the bounds for common operations – except reversal, which misses the bound by one state – are also met by star-free languages [13], where the class of star-free languages is the smallest class containing the finite languages and closed under Boolean operations and product, but not star. For these reasons Brzozowski [6] suggested additional complexity measures: the maximal size of the syntactic semigroup of the language and the state complexities of atoms (discussed later). It turns out that the language defined in Figure 1 also meets these bounds. For these reasons this language has been called a *most complex regular language*. The added measure of the size of the syntactic semigroup distinguishes well between $L_n(a,b,c)$ and $\{a,b,c\}^{n-1}$ since the semigroup of the first language has n^n elements, while the second has $n-1$ elements.

In previous work on state complexity it was always assumed that the two arguments in product and binary Boolean operations are restricted to be over the same alphabet. But Brzozowski pointed out in 2016 [7] that operations on languages over different alphabets, *unrestricted operations*, have higher tight upper bounds. It was shown that witnesses very similar to those of Figure 1 also meet these bounds [7, 16], and thus are most complex for both restricted and unrestricted operations. For suffix-free languages the restricted and unrestricted complexities are the same.

Most complex languages are useful for testing the efficiency of systems. Hence to check the maximal size of the objects that a system can handle, we can use most complex languages. It is certainly simpler to have just one or two worst-case examples.

The question now arises whether most complex languages also exist in subclasses of regular languages. A natural source of subclasses is obtained from the notion of convexity. Convex languages were introduced by Thierrin [35] and studied later in [1]. They can be defined with respect to arbitrary binary relations on an alphabet Σ^*, but the relations "is a prefix of", "is a suffix of" and "is a factor of" turned out to be of considerable interest, where if $w = xyz \in \Sigma^*$, then x is a *prefix* of w, y is a *factor*, and z is a *suffix*. A language L is *prefix-convex* if, whenever $w = xyz$ and w and x are in L, then so is xy. *Factor-convex* and *suffix-convex* languages are defined in a similar way. Brzozowski began studying the complexity properties of these languages in [4].

The class of prefix-convex languages has four natural subclasses: *right ideals* (languages L satisfying $L = L\Sigma^*$), *prefix-closed* languages (where w in L implies that every prefix of w is in L), *prefix-free* languages (where w in L implies that no prefix of w other than w is in L), and *prefix-proper* languages that are prefix-convex but do not belong to any one of the three special subclasses. Similarly, there are four subclasses of suffix-free languages: *left ideals* (languages L satisfying $L = \Sigma^*L$), *suffix-closed* languages, *suffix-free* languages, and *suffix-proper* languages. Finally, there are four subclasses of factor-free languages: *two-sided ideals* (languages L satisfying $L = \Sigma^*L\Sigma^*$), *factor-closed* languages, *factor-free* languages, and *factor-proper* languages. Ideals appear in pattern matching [24]: if we are looking for all the words in a given text that begin with words in a pattern language L, then we are dealing with the right ideal $L\Sigma^*$, and similar statements apply to left and two-sided ideals. Prefix-closed (respectively, suffix-closed, factor closed) languages are complements of right ideals (respectively, left ideals, two-sided ideals). Prefix-free (respectively, suffix-free, factor-free) languages, other than the language consisting of the empty word, are codes [3], and have many applications, particularly in cryptography, data compression and error correction.

Most complex left, right and two-sided ideals for restricted operations were found in [10], and unrestricted operations were added in [16]. Most complex prefix-closed, and prefix-free languages were exhibited in [17] and proper prefix-convex languages in [15]. Most complex suffix-closed languages were found in [14]. In contrast to these results, the first example of a subclass of the class of regular languages that does not have a most complex language is that of suffix-free languages. This result was reported in [20], and is the subject of the present paper. It is also known (unpublished result) that most complex proper suffix-convex languages do not exist.

Quotient Complexity A basic complexity measure of a regular language L over an alphabet Σ is the number n of distinct left quotients of L, where a *(left) quotient* of L by a word $w \in \Sigma^*$ is $w^{-1}L = \{x \mid wx \in L\}$. We denote the set of quotients of L by $K = \{K_0, \dots, K_{n-1}\}$, where $K_0 = L = \varepsilon^{-1}L$ by convention. Each quotient K_i can be represented also as $w_i^{-1}L$, where $w_i \in \Sigma^*$ is such that $w_i^{-1}L = K_i$. The number of quotients of L is its *quotient complexity* [5]. Since the set of quotients of L is the same as the number of states in a minimal DFA recognizing L, the quotient complexity of L is the same as its state complexity; however, quotient complexity suggests language-theoretic methods, whereas state complexity deals with automata. The quotient complexities of unary operation and binary operations are defined analogously to state complexities.

To establish the state/quotient complexity of a regularity preserving unary operation $\circ$ we need a sequence $(L_n, n \geqslant k) = (L_k, L_{k+1}, \dots)$, called a *stream*, of witness languages that meet this bound; here k is usually some small integer because the bound may not apply for $n < k$. In all the cases that have been studied the languages in a stream are defined in the same way, differing only in the parameter n. For example, the languages of the DFAs of Figure 1 define the stream $(L_n(a, b, c) \mid n \geqslant 3)$.

To establish the state/quotient complexity of a regularity preserving binary operation $\circ$ on regular languages, for each pair (m, n), $m \geqslant h, n \geqslant k$, we need to find two languages $K_{m,n}$ and $L_{m,n}$ meeting this bound; here $K_{m,n}$ and $L_{m,n}$ are of complexities m and n, respectively. The notation $K_{m,n}$ and $L_{m,n}$ implies that $K_{m,n}$ depends on n, and $L_{m,n}$ depends on m. If $K_{m,n}$ together with $L_{m,n}$ meet the required bound, it may occur that $K_{m,n}$ with $L_{m',n}$ may *not* meet that bound. Indeed, two such examples have been found by Han and Salomaa [25]: the union and intersection of finite languages require such witnesses. Fortunately, however, in all other cases studied in the literature, it suffices to use witness streams $(K_m, m \geqslant h)$ and $(L_n, n \geqslant k)$, where K_m is independent of n and L_n is independent of m. This is the case for suffix-free languages. The state/quotient complexity of suffix-free languages using various witnesses was examined in [23, 26, 30].

We also extend the notions of *maximal complexity*, *stream*, and *witness* to DFAs.

Syntactic Complexity A second measure of complexity of a regular language is its syntactic complexity. Let Σ^+ be the set of non-empty words of Σ^*. The *syntactic semigroup* of L is the set of equivalence classes of the Myhill congruence $\approx_L$ on Σ^+ defined by $x \approx_L y$ if and only if $uxv \in L \Leftrightarrow uyv \in L$ for all $u, v \in \Sigma^*$. The syntactic semigroup of L is isomorphic to the *transition semigroup* of a minimal DFA $\mathcal{D}$ recognizing L [34], which is the semigroup of transformations of the state set of $\mathcal{D}$ induced by non-empty words. The *syntactic complexity* of L is the cardinality of its syntactic/transition semigroup.

Holzer and König [27], and independently Krawetz, Lawrence and Shallit [31] studied the syntactic complexity in the classes of unary and binary regular languages. This problem was also solved for the classes of right ideals [16, 22], left ideals [16, 18, 22], two-sided ideals [16, 18, 22], prefix-free languages [12], and suffix-free languages [12, 19].

Complexities of Atoms A possible third measure of complexity of a regular language L is the number and quotient complexities, which we call simply *complexities*, of certain languages, called atoms, uniquely defined by L. Atoms arise from an equivalence on Σ^* which is a left congruence refined by the Myhill congruence, where two words x and y are equivalent if $ux \in L$ if and only if $uy \in L$ for all $u \in \Sigma^*$ [29]. Thus x and y are equivalent if $x \in u^{-1}L \Leftrightarrow y \in u^{-1}L$. An equivalence class of this relation is called an *atom* [21] of L. It follows that an atom is a non-empty intersection of complemented and uncomplemented quotients of L. The quotients of a language are unions of its atoms.

Terminology and Notation A *deterministic finite automaton (DFA)* is defined as a quintuple $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite non-empty set of *states*, Σ is a finite non-empty *alphabet*, $\delta \colon Q \times \Sigma \to Q$ is the *transition function*, $q_0 \in Q$ is the *initial* state, and $F \subseteq Q$ is the set of *final* states. We extend δ to a function $\delta \colon Q \times \Sigma^* \to Q$ as usual. A DFA $\mathcal{D}$ *accepts* a word $w \in \Sigma^*$ if $\delta(q_0, w) \in F$. The language accepted by $\mathcal{D}$ is denoted by $L(\mathcal{D})$. If q is a state of $\mathcal{D}$, then the language L^q of q is the language accepted by the DFA $(Q, \Sigma, \delta, q, F)$. A state is *empty* if its language is empty. Two states p and q of $\mathcal{D}$ are *equivalent* if $L^p = L^q$. A state q is *reachable* if there exists $w \in \Sigma^*$ such that $\delta(q_0, w) = q$. A DFA is *minimal* if all of its states are reachable and no two states are equivalent. Usually DFAs are used to establish upper bounds on the quotient complexity of operations and also as witnesses that meet these bounds.

A *nondeterministic finite automaton (NFA)* is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, I, F)$, where Q, Σ and F are defined as in a DFA, $\delta \colon Q \times \Sigma \to 2^Q$ is the *transition function*, and $I \subseteq Q$ is the *set of initial states*. An *ε-NFA* is an NFA in which transitions under the empty word ε are also permitted.

The *quotient DFA* of a regular language L with n quotients is defined by $\mathcal{D} = (K, \Sigma, \delta_{\mathcal{D}}, K_0, F_{\mathcal{D}})$, where $\delta_{\mathcal{D}}(K_i, w) = K_j$ if and only if $w^{-1}K_i = K_j$, and $F_{\mathcal{D}} = \{K_i \mid \varepsilon \in K_i\}$. To simplify the notation, without loss of generality we use the set $Q_n = \{0, \dots, n-1\}$ of subscripts of quotients as the set of states of $\mathcal{D}$; then $\mathcal{D}$ is denoted by $\mathcal{D} = (Q_n, \Sigma, \delta, 0, F)$, where $\delta(p, w) = q$ if $\delta_{\mathcal{D}}(K_p, w) = K_q$, and F is the set of subscripts of quotients in $F_{\mathcal{D}}$. The quotient DFA of L is unique and it is isomorphic to each complete minimal DFA of L.

A *transformation* of Q_n is a mapping $t \colon Q_n \to Q_n$. The *image* of $q \in Q_n$ under t is denoted by qt. The *range* of t is $\operatorname{rng}(t) = \{q \in Q_n \mid pt = q \text{ for some } p \in Q_n\}$. In any DFA, each letter $a \in \Sigma$ induces a transformation δ_a of the set Q_n defined by $q\delta_a = \delta(q, a)$. By a slight abuse of notation we use the letter a to denote the transformation it induces; thus we write qa instead of $q\delta_a$. We also extend the notation to sets of states: if $P \subseteq Q_n$, then $Pa = \{pa \mid p \in P\}$. If s, t are transformations of Q, their composition is denoted by $s{*}t$ and defined by $q(s{*}t) = (qs)t$; the $*$ is usually omitted. Then also for a word $w = a_1 \cdots a_k$, δ_w denotes the transformation $\delta_{a_1} \cdots \delta_{a_k}$ induced by w. Let $\mathcal{T}_{Q_n}$ be the set of all n^n transformations of Q_n; then $\mathcal{T}_{Q_n}$ is a monoid under composition.

For $k \geqslant 2$, a transformation (permutation) t of a set $P = \{q_0, q_1, \dots, q_{k-1}\} \subseteq Q$ is a *k-cycle* if $q_0t = q_1, q_1t = q_2, \dots, q_{k-2}t = q_{k-1}, q_{k-1}t = q_0$. This k-cycle is denoted by $(q_0, q_1, \dots, q_{k-1})$. A 2-cycle (q_0, q_1) is called a *transposition*. A transformation that changes only one state p to a state $q \neq p$ is denoted by $(p \to q)$. A transformation mapping a subset P of Q to a single state and acting as the identity on $Q \setminus P$ is denoted by $(P \to q)$. We also denote by $[q_0, \dots, q_{n-1}]$ the transformation that maps $p \in \{0, \dots, n-1\}$ to q_p.

We now define dialects of languages and DFAs following [10]. Let $\Sigma = \{a_1, \dots, a_k\}$ be an alphabet; we assume that its elements are ordered as shown. Let π be a *partial permutation* of Σ, that is, a partial function $\pi \colon \Sigma \to \Gamma$ where $\Gamma \subseteq \Sigma$, for which there exists $\Delta \subseteq \Sigma$ such that π is bijective when restricted to Δ and undefined on $\Sigma \setminus \Delta$. We denote undefined values of π by the symbol "$-$".

If L is a language over Σ, we denote it by $L(a_1, \dots, a_k)$ to stress its dependence on Σ. If π is a partial permutation, let s_π be the language substitution defined as follows: for $a \in \Sigma$, $a \mapsto \{\pi(a)\}$ when $\pi(a)$ is defined, and $a \mapsto \emptyset$ when $\pi(a)$ is not defined. For example, if $\Sigma = \{a, b, c\}$, $L(a, b, c) = \{a, b, c\}^*\{ab, acc\}$, and $\pi(a) = c$, $\pi(b) = -$, and $\pi(c) = b$, then $s_\pi(L) = \{b, c\}^*\{cbb\}$. In other words, the letter c plays the role of a, and b plays the role of c. A *permutational dialect* of $L(a_1, \dots, a_k)$ is a language of the form $s_\pi(L(a_1, \dots, a_k))$, where π is a partial permutation of Σ; this dialect is denoted by $L(\pi(a_1), \dots, \pi(a_k))$. If the order on Σ is understood, we use $L(\Sigma)$ for $L(a_1, \dots, a_k)$ and $L(\pi(\Sigma))$ for $L(\pi(a_1), \dots, \pi(a_k))$. Undefined values appearing at the end of the alphabet are omitted. For example, if $\Sigma = \{a, b, c, d\}$ then we write $L(a, b)$ instead of $L(a, b, -, -)$.

Let $\Sigma = \{a_1, \dots, a_k\}$, and let $\mathcal{D} = (Q, \Sigma, \delta, q_1, F)$ be a DFA; we denote it by $\mathcal{D}(a_1, \dots, a_k)$ to stress its dependence on Σ. If π is a partial permutation, then the *permutational dialect*

$$\mathcal{D}(\pi(a_1), \dots, \pi(a_k))$$

of $\mathcal{D}(a_1, \dots, a_k)$ is obtained by changing the alphabet of $\mathcal{D}$ from Σ to $\pi(\Sigma)$, and modifying δ so that in the modified DFA $\pi(a_i)$ induces the transformation induced by a_i in the original DFA; thus $\pi(a_i)$ plays the role of a_i. One verifies that if the language $L(a_1, \dots, a_k)$ is accepted by DFA $\mathcal{D}(a_1, \dots, a_k)$, then $L(\pi(a_1), \dots, \pi(a_k))$ is accepted by $\mathcal{D}(\pi(a_1), \dots, \pi(a_k))$.

In the sequel we refer to permutational dialects simply as *dialects*.

Contributions

1. We prove that a most complex stream of suffix-free languages does not exist. This is in contrast with the existence of streams of most complex regular languages, right, left, and two-sided ideals, prefix-free and proper prefix-convex languages.

2. We exhibit a single ternary witness that meets the bounds for star, product, and Boolean operations.

3. We exhibit a single quinary witness that meets the bounds for Boolean operations, reversal, number of atoms, syntactic complexity, and quotient complexities of atoms.

4. We show that when $m, n \geqslant 6$ and $m - 2$ and $n - 2$ are relatively prime, there are binary witnesses that meet the bound $(m - 1)2^{n-2} + 1$ for product.

5. We prove that any witness DFA for star and any second witness DFA for product must have transition semigroups that are subsemigroups of the suffix-free semigroup of transformations $\mathbf{T}^{\leqslant 5}(n)$ which has maximal cardinality for $2 \leqslant n \leqslant 5$; that the witness DFAs for reversal, syntactic complexity and quotient complexities of atoms must have transition semigroups that are subsemigroups of the suffix-free semigroup of transformations $\mathbf{T}^{\geqslant 6}(n)$ which has maximal cardinality for $n = 2, 3$ and $n \geqslant 6$; and that the witness DFAs for Boolean operations can have transition semigroups that are subsemigroups of $\mathbf{T}^{\leqslant 5} \cap \mathbf{T}^{\geqslant 6}$.

2. Suffix-Free Transformations

In this section we discuss some properties of suffix-free languages with emphasis on their syntactic semigroups as represented by the transition semigroups of their quotient DFAs. We assume that our basic set is always $Q_n = \{0, \dots, n-1\}$.

2.1. Suffix-Free Languages

Let $\mathcal{D}_n = (Q_n, \Sigma, \delta, 0, F)$ be the quotient DFA of a suffix-free language L, and let T_n be its transition semigroup. For any transformation t of Q_n, the sequence $(0, 0t, 0t^2, \dots)$ is called the *0-path* of t. Since Q_n is finite, there exist i, j such that $0, 0t, \dots, 0t^i, 0t^{i+1}, \dots, 0t^{j-1}$ are distinct but $0t^j = 0t^i$. The integer $j - i$ is the *period* of t and if $j - i = 1$, t is *initially aperiodic.* The following properties of suffix-free languages are known [12, 26]:

Lemma 1. *If L is a regular suffix-free language, then*

1. *There exists $w \in \Sigma^*$ such that $w^{-1}L = \emptyset$; hence $\mathcal{D}_n$ has an empty state, which is state $n-1$ by convention.*
2. *For $w, x \in \Sigma^+$, if $w^{-1}L \neq \emptyset$, then $w^{-1}L \neq (xw)^{-1}L$.*
3. *If $L \neq \emptyset$ and $w^{-1}L = L$, then $w = \varepsilon$.*
4. *For any $t \in T_n$, the 0-path of t in $\mathcal{D}_n$ is aperiodic and ends in $n-1$.*

Property 3 is known as the *non-returning* property [26] and also as *unique reachability* [11].

An (unordered) pair $\{p, q\}$ of distinct states in $Q_n \setminus \{0, n-1\}$ is *colliding* (or p *collides* with q) in T_n if there is a transformation $t \in T_n$ such that $0t = p$ and $rt = q$ for some $r \in Q_n \setminus \{0, n-1\}$. A pair of states is *focused* by a transformation u of Q_n if u maps both states of the pair to a single state $r \notin \{0, n-1\}$. We then say that $\{p, q\}$ is *focused to state r*. If L is a suffix-free language, then from Lemma 1 (2) it follows that if $\{p, q\}$ is colliding in T_n, there is no transformation $t' \in T_n$ that focuses $\{p, q\}$. So colliding states can be mapped to a single state by a transformation in T_n only if that state is the empty state $n-1$.

Following [12], for $n \geqslant 2$, we let

$$\begin{aligned}\mathbf{B}(n) = \{t \in \mathcal{T}_Q \mid\ & 0 \notin \operatorname{rng}(t),\ (n-1)t = n-1, \text{ and for all } j \geqslant 1, \\ & 0t^j = n-1 \text{ or } 0t^j \neq qt^j,\ \forall q \text{ such that } 0 < q < n-1\}.\end{aligned}$$

Example 1. *We have $\mathbf{B}(2) = \{[1,1]\}$ and $\mathbf{B}(3) = \{[1,2,2],[2,1,2],[2,2,2]\}$. For $n = 4$, there are 17 transformations satisfying $0t \neq qt$. However, if $t = [1,2,2,3]$ or $t = [2,1,1,3]$, then $0t^2 = 1t^2$, which violates the third condition for $\mathbf{B}$; hence $\mathbf{B}(4)$ has 15 elements. The cardinality of $\mathbf{B}(5)$ is 115.* ■

Proposition 1 ([12]). *If L is a regular language with $\mathcal{D}_n = (Q_n, \Sigma, \delta, 0, F)$ as its minimal DFA and syntactic semigroup T_L, then the following hold:*

1. *If L is suffix-free, then T_L is a subset of $\mathbf{B}(n)$.*
2. *If L has the empty quotient, only one final quotient, and $T_L \subseteq \mathbf{B}(n)$, then L is suffix-free.*

Since the transition semigroup of a minimal DFA of a suffix-free language must be a subsemigroup of $\mathbf{B}(n)$, the cardinality of $\mathbf{B}(n)$ is an upper bound on the syntactic complexity of suffix-free regular languages with quotient complexity n. This upper bound, however, cannot be reached since $\mathbf{B}$ is not a semigroup for $n \geqslant 4$: We have $s = [1, 2, n-1, \dots, n-1]$ and $t = [n-1, 2, 2, \dots, 2, n-1]$ in $\mathbf{B}(n)$, but $st = [2, 2, n-1, \dots, n-1]$ is not in $\mathbf{B}(n)$.

2.2. Semigroups $\mathbf{T}^{\leqslant 5}(n)$ with Maximal Cardinality when $2 \leqslant n \leqslant 5$

For $n \geqslant 2$, let

$$\mathbf{T}^{\leqslant 5}(n) = \{t \in \mathbf{B}(n) \mid \text{ for all } p, q \in Q_n \text{ where } p \neq q, \text{ either } pt = qt = n-1 \text{ or } pt \neq qt\}.$$

Proposition 2. *For $n \geqslant 4$, the semigroup $\mathbf{T}^{\leqslant 5}(n)$ is generated by the following set $\mathbf{H}^{\leqslant 5}(n)$ of transformations of Q:*

- a: $(0 \to n-1)(1, \dots, n-2)$,
- b: $(0 \to n-1)(1, 2)$,
- *for* $1 \leqslant p \leqslant n-2$, c_p: $(p \to n-1)(0 \to p)$.

For $n = 4$, a and b coincide, and so $\mathbf{H}^{\leqslant 5}(4) = \{a, c_1, c_2\}$. Also, $\mathbf{H}^{\leqslant 5}(3) = \{a, c_1\} = \{[2, 1, 2], [1, 2, 2]\}$ and $\mathbf{H}^{\leqslant 5}(2) = \{c_1\} = \{[1, 1]\}$.

Proof. It was proved in [12] that for $n \geqslant 4$, the semigroup $\mathbf{T}^{\leqslant 5}(n)$ is generated by the following set $\mathbf{G}^{\leqslant 5}(n)$ of n transformations of Q_n: a and b as above, and c'_p for $1 \leqslant p \leqslant n-2$, defined by $qc'_p = q+1$ for $q = 0, \dots, p-1$, $pc'_p = n-1$, and $qc'_p = q$ for $q = p+1 \dots, n-1$. Since $\mathbf{H}^{\leqslant 5}(n) \subseteq \mathbf{T}^{\leqslant 5}$, the semigroup generated by $\mathbf{H}^{\leqslant 5}(n)$ is a subsemigroup of $\mathbf{T}^{\leqslant 5}(n)$. So it is sufficient to show that every transformation in $\mathbf{G}^{\leqslant 5}(n)$ can be generated by $\mathbf{H}^{\leqslant 5}(n)$. Transformation c'_p changes $\{0, 1, \dots, p-1\}$ to $\{1, 2, \dots, p\}$, p is mapped to $n-1$, and $\{p+1, p+2, \dots, n-2)$ is mapped to itself. The image of $Q_n \setminus \{p, n-1\}$ is thus $\{1, 2, \dots, p, p+1, \dots n-2\}$. The image of $Q_n \setminus \{p, n-1\}$ under c_p is $\{p, 1, 2, \dots, p-1, p+1, \dots, n-2\}$. However, since transformations a

and b restricted to $Q_n \setminus \{0, n-1\}$ generate all permutations of $Q_n \setminus \{0, n-1\}$, $\{p, 1, 2, \dots, p-1, p+1, \dots, n-2\}$ can be transformed to $\{1, 2, \dots, p, p+1, \dots n-2\}$. Hence the claim holds. □

From now on we use the transformations of Proposition 2 for $\mathbf{T}^{\leqslant 5}(n)$. A DFA using these transformations is illustrated in Figure 2.

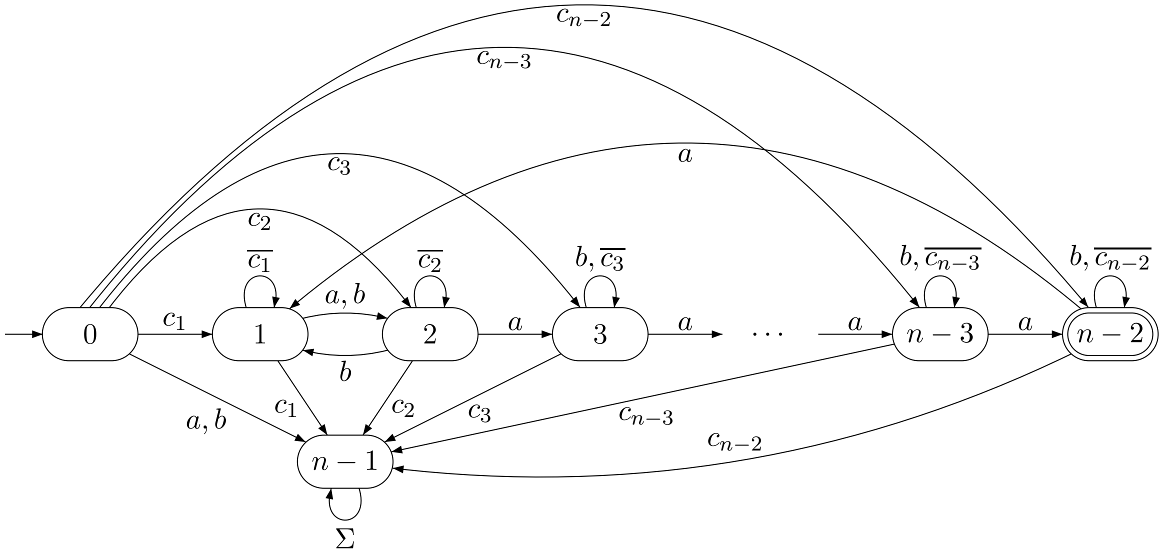

Figure 2: DFA with $\mathbf{T}^{\leqslant 5}(n)$ as its transformation semigroup; $\overline{c_p} = \{c_1, \dots, c_{n-2}\} \setminus \{c_p\}$.

Proposition 3. *For $n \geqslant 2$, $\mathbf{T}^{\leqslant 5}(n)$ is the unique maximal semigroup of a suffix-free language in which all possible pairs of states are colliding.*

Proof. For each pair $p, q \in Q \setminus \{0, n-1\}$, $p \neq q$, there is a transformation $c_p \in \mathbf{T}^{\leqslant 5}(n)$ with $0c_p = p$ and $qc_p = q$. Thus all pairs are colliding. If all pairs are colliding, then for each $p, q \in Q \setminus \{n-1\}$, there is no transformation t with $pt = qt \neq n-1$, for this would violate suffix-freeness. By definition, $\mathbf{T}^{\leqslant 5}(n)$ has all other transformations that are possible for a suffix-free language, and hence is unique. □

Proposition 4. *For $n \geqslant 5$, the number n of generators of $\mathbf{T}^{\leqslant 5}(n)$ cannot be reduced.*

Proof. If a generator t maps 0 to $p \in \{1, \dots, n-2\}$, then it must also map a state $q \in \{1, \dots, n-2\}$ to $n-1$, since the 0-path of t is aperiodic and ends in $n-1$. Thus t cannot generate a permutation of $\{1, \dots, n-2\}$. Since $\mathbf{T}^{\leqslant 5}(n)$ has all transformations that map 0 to $n-1$, permute $\{1,\dots,n-2\}$ and fix $n-1$, we need two generators, say a and b, with $0a = 0b = n-1$ to induce all the permutations.

For $p \in \{1,\dots,n-2\}$, consider the transformation c_p of Proposition 2. Note that every transformation that maps 0 to a state in $\{1,\dots,n-2\}$ must also map a state in $\{1,\dots,n-2\}$ to $n-1$. Since c_p maps $\{0,1,\dots,n-2\} \setminus \{p\}$ onto $\{1,\dots,n-2\}$, it must be generated by $c'_p u$, where c'_p is a transformation mapping 0 to some state $q \in \{1,\dots,n-2\}$ and p to $n-1$, and where u permutes $\{1,\dots,n-2\}$. Since each c_p requires a different c'_p, one for each $p \in \{1,\dots,n-2\}$, we need at least $n-2$ generators c'_p. □

Example 2. $\mathbf{T}^{\leqslant 5}(4)$ *has 13 elements. All transitions of* $\mathbf{B}(4)$ *are present except* $[3,1,1,3]$ *and* $[3,2,2,3]$*. The semigroup* $\mathbf{T}^{\leqslant 5}(5)$ *has 73 elements.* ■

Semigroups $\mathbf{T}^{\leqslant 5}(n)$ are suffix-free semigroups that have maximal cardinality when $2 \leqslant n \leqslant 5$ [12].

2.3. Semigroups $\mathbf{T}^{\geqslant 6}(n)$ *with Maximal Cardinality when* $n = 2, 3$ *and* $n \geqslant 6$

For $n \geqslant 2$, let

$$\mathbf{T}^{\geqslant 6}(n) = \{t \in \mathbf{B}(n) \mid 0t = n-1, \text{ or } qt = n-1 \text{ for all } q \text{ such that } 1 \leqslant q \leqslant n-2\}.$$

Proposition 5 ([20])**.** *For* $n \geqslant 4$, $\mathbf{T}^{\geqslant 6}(n)$ *is a semigroup contained in* $\mathbf{B}(n)$*, its cardinality is* $(n-1)^{n-2} + (n-2)$*, and it is generated by the set* $\mathbf{G}^{\geqslant 6}(n)$ *of the following transformations:*

- a: $(0 \to n-1)(1,\dots,n-2)$;
- b: $(0 \to n-1)(1,2)$;
- c: $(0 \to n-1)(n-2 \to 1)$;
- d: $(\{0,1\} \to n-1)$;
- e: $(Q \setminus \{0\} \to n-1)(0 \to 1)$.

For $n = 4$*,* a *and* b *coincide, and so* $\mathbf{G}^{\geqslant 6}(4) = \{a,c,d,e\}$*. Also* $\mathbf{G}^{\geqslant 6}(3) = \{a,e\} = \{[2,1,2],[1,2,2]\}$ *and* $\mathbf{G}^{\geqslant 6}(2) = \{e\} = \{[1,1]\}$.

Example 3. $\mathbf{T}^{\geqslant 6}(4)$ *has 11 elements. All transitions of* $\mathbf{B}(4)$ *are present except* $[1,2,3,3]$*,* $[1,3,2,3]$*,* $[2,1,3,3]$ *and* $[2,3,1,3]$*. Semigroup* $\mathbf{T}^{\geqslant 6}(5)$ *has size 67.* ■

A DFA using the transformations of Proposition 5 is shown in Figure 3.

Semigroups $\mathbf{T}^{\geqslant 6}(n)$ are suffix-free semigroups that have maximal cardinality when $n \geqslant 6$ [20].

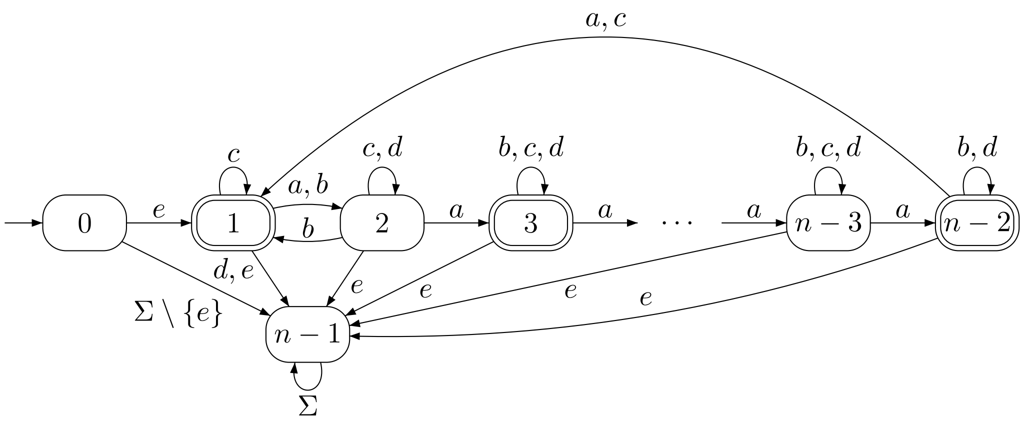

Figure 3: DFA with $\mathbf{T}^{\geqslant 6}(n)$ as its transformation semigroup for the case when n is odd.

3. Witnesses with Transition Semigroups in $\mathbf{T}^{\leqslant 5}(n)$

In this section we consider DFA witnesses whose transition semigroups are subsemigroups of $\mathbf{T}^{\leqslant 5}(n)$. We show that there is one witness that satisfies the bounds for star, product and Boolean operations.

Definition 1. *For $n \geqslant 6$, we define the DFA $\mathcal{D}_n = (Q_n, \Sigma, \delta, 0, \{1\})$, where $Q_n = \{0, \dots, n-1\}$, $\Sigma = \{a, b, c\}$, and δ is defined by the transformations*

- a: $(0 \to n-1)(1, 2, 3)(4, \dots, n-2)$,
- b: $(2 \to n-1)(1 \to 2)(0 \to 1)(3, 4)$,
- c: $(0 \to n-1)(1, \dots, n-2)$.

Theorem 1 (Star, Product, Boolean Operations)**.** *Let $\mathcal{D}_n(a, b, c)$ be the DFA of Definition 1, and let the language it accepts be $L_n(a, b, c)$. For $n \geqslant 6$, L_n and its permutational dialects meet the bounds for star, product and Boolean operations as follows:*

1. $L_n^*(a, b, -)$ *meets the bound* $2^{n-2} + 1$. [Cmorik and Jirásková [23]]
2. $L_m(a, b, c) \cdot L_n(b, c, a)$ *meets the bound* $(m-1)2^{n-2} + 1$.
3. $L_m(a, b, -)$ *and* $L_n(-, b, a)$ *meet the bounds* $mn - (m + n - 2)$ *for union and symmetric difference,* $mn - 2(m+n-3)$ *for intersection, and* $mn - (m+2n-4)$ *for difference.*

The claim about the star operation was proved in [23]. We add a result about the transition semigroup of the star witness and prove the remaining two claims in this section.

3.1. Star

In 2009 Han and Salomaa [26] showed that the language of a DFA over a four-letter alphabet meets the bound $2^{n-2}+1$ for the star operation for $n \geqslant 4$. The transition semigroup of this DFA is a subsemigroup of $\mathbf{T}^{\leqslant 5}(n)$. In 2012 Cmorik and Jirásková [23] showed that for $n \geqslant 6$ a binary alphabet $\{a,b\}$ suffices. The transition semigroup of this DFA is again a subsemigroup of $\mathbf{T}^{\leqslant 5}(n)$. We prove that these are special cases of the following general result:

Theorem 2. *For $n \geqslant 4$, the transition semigroup of a minimal DFA $\mathcal{D}(Q_n, \Sigma, \delta, 0, F)$ of a suffix-free language L that meets the bound $2^{n-2}+1$ for the star operation is a subsemigroup of $\mathbf{T}^{\leqslant 5}(n)$ and is not a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$.*

Proof. To show that the transition semigroup of $\mathcal{D}$ is a subsemigroup of $\mathbf{T}^{\leqslant 5}(n)$, by Proposition 3 it suffices to show that every pair of states is colliding.

We construct an NFA $\mathcal{N}$ for L_n^* by making 0 a final state in $\mathcal{D}$ – this is possible since 0 is uniquely reachable – and adding an empty-word transition from every final state to 0. We then determinize $\mathcal{N}$ using the subset construction to get a DFA $\mathcal{D}^*$ for L_n^*. The states of $\mathcal{D}^*$ are sets of states of $\mathcal{D}$.

Consider a subset $S \subseteq Q_n$. We can assume that $n-1 \in S$, since S and $S \cup \{n-1\}$ cannot be distinguished. If $S \neq \{0, n-1\}$ and $S \neq \{n-1\}$, then S can be reached only if $0 \in S$ and $S \cap F \neq \emptyset$, or $0 \notin S$ and $S \cap F = \emptyset$, because by the construction for star in $\mathcal{N}$ there is an ε-transition from every final state to the initial state 0 and no transformation fixes 0. Thus, to meet the bound $2^{n-2}+1$, for each possible subset of $\{1, \dots, n-2\}$ there must be a reachable subset S containing that subset.

Suppose that $p, q \in \{1, \dots, n-2\}$ are not colliding, that is, there is no transformation t with $0t = p$ and $q \in \operatorname{rng}(t)$. Consider S that contains both p and q. Since the ε-transitions from final states to initial state 0 are the only sources of nondeterminism, S must be reached from a subset S' containing 0 and $q' \in \{1, \dots, n-2\}$ by a transformation t with $0t = p$ and $q't = q$ (or $0t = q$ and $q't = p$), which contradicts that p, q are not colliding.

Since transformations causing colliding pairs are not in $\mathbf{T}^{\geqslant 6}(n)$, the transition semigroup of $\mathcal{D}$ cannot be a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$. □

3.2. Product

To avoid confusing the states of the two DFAs in a product, we label the states of the first DFA differently. Let $\mathcal{D}'_m = \mathcal{D}'_m(a,b,c) = (Q'_m, \Sigma, \delta', 0', \{1'\})$, where $Q'_m = \{0', \dots, (m-1)'\}$, and $\delta'(q', x) = p'$ if $\delta(q, x) = p$, and let $\mathcal{D}_n = \mathcal{D}_n(b,c,a)$. We use the standard construction of the ε-NFA $\mathcal{N}$ for the product: the final state

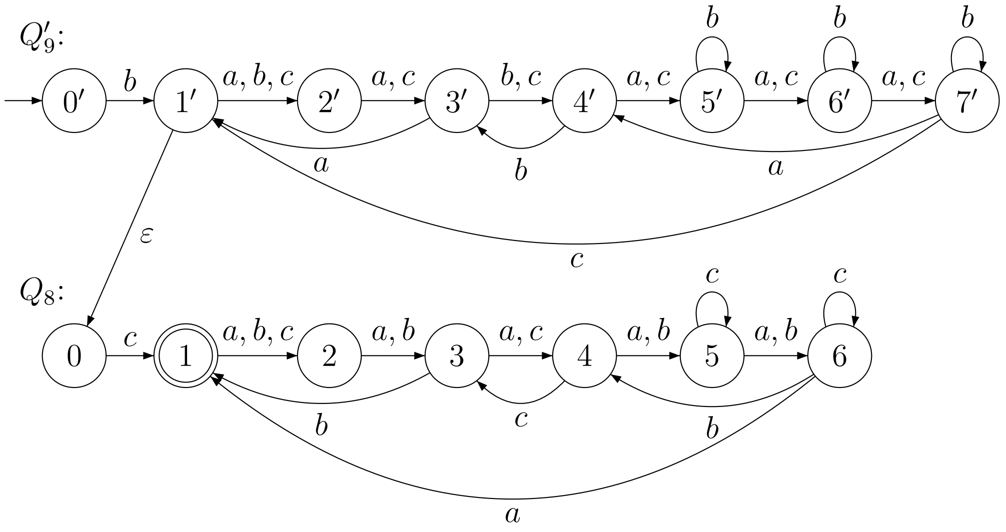

Figure 4: The NFA $\mathcal{N}$ for product $L'_9(a, b, c) \cdot L_8(b, c, a)$. The empty states $8'$ and 7 and the transitions to them are omitted.

of $\mathcal{D}'_m$ becomes non-final, and an ε-transition is added from that state to the initial state of $\mathcal{D}_n$. This is illustrated in Figure 4 for $m = 9, n = 8$.

We use the subset construction to determinize $\mathcal{N}$ to get a DFA $\mathcal{P}$ for the product. The reachable states of $\mathcal{P}$ are subsets of $Q'_m \cup Q_n$ and have one of three forms: $\{0'\}$, $\{1', 0\} \cup S$ or $\{p'\} \cup S$, where $p' = 2', \dots, (m-1)'$ and $S \subseteq \{1, \dots, n-1\}$.

Note that for each $x \in \Sigma$ every state $q \in Q_n \setminus \{0, n-1\}$ has a unique predecessor state $p \in Q_n \setminus \{n-1\}$ such that $px = q$. For $w \in \Sigma^*$, the w-predecessor of $S \subseteq Q_n \setminus \{0, n-1\}$ is denoted by Sw^{-1}.

Lemma 2. *For each $n \geqslant 6$ and each $q \in Q_n$ there exists a word $w_q \in c\{a, b\}^*$ such that $1'w_q = 3'$, $0w_q = q$, and each state of $Q_n \setminus \{0, q, n-1\}$ has a unique w_q-predecessor in $Q_n \setminus \{0, n-1\}$. In fact, the following w_q satisfy these requirements:*

$$w_q = \begin{cases} cab^2, & \text{if } q = 1; \\ ca, & \text{if } q = 2; \\ cab^4, & \text{if } q = 3; \\ cab^2a^3b^{q-4}, & \text{if } q \geqslant 4 \text{ and } q \text{ is even}; \\ ca^4b^{q-5}, & \text{if } q \geqslant 5 \text{ and } q \text{ is odd}. \end{cases} \tag{1}$$

Proof. It is easily verified that in each case $1'w_q = 3'$ and $0w_q = q$. Note that a and b induce permutations on $Q_n \setminus \{0, n-1\}$ and c is a one-to-one mapping from $Q_n \setminus \{0, n-1\}$ to $Q_n \setminus \{0, 1\}$. Thus every state in $Q_n \setminus \{0, n-1\}$ that is mapped by w_q to $Q_n \setminus \{0, n-1\}$ has a w_q-predecessor in $Q_n \setminus \{0, n-1\}$, and state q has 0 as its w_q-predecessor. □

Theorem 3 (Product: Ternary Case)**.** *For $m, n \geqslant 6$, the product $L'_m(a, b, c) \cdot L_n(b, c, a)$ meets the bound $(m-1)2^{n-2} + 1$.*

Proof. Let P consist of the following states $\{0'\}$, 2^{n-2} sets of the form $\{1', 0, n-1\} \cup S$ and $(m-2)2^{n-2}$ sets of the form $\{p', n-1\} \cup S$, where $S \subseteq \{1, \dots, n-2\}$, and $p' = 2', \dots, (m-1)'$ – a total of $(m-1)2^{n-2} + 1$ sets. We shall prove that all these states of $\mathcal{P}$ are reachable and pairwise distinguishable. This together with the known upper bound will prove that the witnesses indeed meet the bound for product.

Consider the distinguishability of two states in P. If we apply any word ending in c to any subset of $Q_n \setminus \{0\}$, the resulting set does not contain the final state 1. Suppose that one of the states in our pair is $\{0'\}$; this is the only state accepting bc. Next, if one of the states has the form $\{1', 0, n-1\} \cup S$ and the other is $\{p', n-1\} \cup R$, then the former state accepts c, whereas the latter does not. If $p < q$, then $\{p', n-1\} \cup R$ is distinguished from $\{q', n-1\} \cup S$ by c^{m-q}. This leaves the case where the state in $Q'_m \setminus \{0'\}$ is the same in both sets in our pair. If the sets in Q_n are R and S, $R \neq S$, and $q \in R \oplus S$, then a^{m-1-q} distinguishes these states.

Now we turn to reachability. Since state $0'$ is initial in $\mathcal{N}$, $\{0'\}$ is reachable. First we show that the sets in P are reachable if $S = \emptyset$. The set $\{1', 0, n-1\}$ is reached by ba^3, $\{2', n-1\}$ by ba, $\{p', n-1\}$ by bac^{p-2} for $p = 3, \dots, (m-2)$, and $\{(m-1)', n-1\}$ by b^3.

Now suppose all sets of the form $\{1', 0, n-1\} \cup S$ and $\{p', n-1\} \cup S$, $p = 2, \dots, m-1$, with $S \subseteq \{1, \dots, n-2\}$ and $|S| = k$, are reachable. We show that if $k < n-2$, then every set with S of size $k+1$ can be reached. In each case we assume that $q \notin S$.

1. Sets with $3'$. We add q to S by applying w_q to $(\{1', 0, n-1\} \cup Sw_q^{-1})$. By Lemma 2, every state except q in $Q_n \setminus \{0, n-1\}$ has a unique w_q-predecessor in $Q_n \setminus \{0, n-1\}$. Hence, assuming we have $\{3', n-1\} \cup Sw_q^{-1}$ we can add q since
$$(\{1', 0, n-1\} \cup Sw_q^{-1})w_q = \{3', n-1\} \cup S \cup \{q\}.$$
2. Sets with $4'$:
   (a) $(\{3', n-1\} \cup Sb^{-1} \cup \{3\})b = \{4', n-1\} \cup S \cup \{1\}$,
   (b) $(\{3', n-1\} \cup Sb^{-1} \cup \{1\})b = \{4', n-1\} \cup S \cup \{2\}$,

(c) $(\{3', n-1\} \cup Sb^{-1} \cup \{2\})b = \{4', n-1\} \cup S \cup \{3\}$,

(d) $(\{3', n-1\} \cup Sb^{-1} \cup \{m-2\})b = \{4', n-1\} \cup S \cup \{4\}$,

(e) $(\{3', n-1\} \cup Sb^{-1} \cup \{q\})b = \{4', n-1\} \cup S \cup \{q+1\}$, for $q = 4, \dots, n-2$.

3. Sets with p', $p = 5, \dots, m-2$:

$$(\{(p-1)', n-1\} \cup S(b^2ab)^{-1} \cup \{3\})b^2ab = \{p', n-1\} \cup S \cup \{1\},$$

$(\{(p-1)', n-1\} \cup Sa^{-1} \cup \{q-1\})a = \{p', n-1\} \cup S \cup \{q\}$ for $q = 2, \dots, n-2$.

4. Sets with $(m-1)'$: If $1 \notin S$, then S has a c-predecessor in $Q_n \setminus \{0, n-1\}$ and $(\{1', 0, n-1\} \cup Sc^{-1})c = \{2', n-1\} \cup S \cup \{1\}$. Thus we have

$$(\{2', n-1\} \cup S(bab)^{-1} \cup \{1\})bab = \{(m-1)', n-1\} \cup S \cup \{1\},$$
$$(\{2', n-1\} \cup Sb^{-1} \cup \{1\})b = \{(m-1)', n-1\} \cup S \cup \{2\},$$
$$(\{(m-1)', n-1\} \cup Sa^{-1} \cup \{q-1\})a = \{(m-1)', n-1\} \cup S \cup \{q\},$$

for $q = 3, \dots, n-2$.

5. Sets with $1'$:

$$(\{3', n-1\} \cup Sa^{-1} \cup \{n-2\})a = \{1', 0, n-1\} \cup S \cup \{1\},$$
$$(\{3', n-1\} \cup Sa^{-1} \cup \{1\})a = \{1', 0, n-1\} \cup S \cup \{2\},$$
$$(\{2', n-1\} \cup S(a^2)^{-1} \cup \{1\})a^2 = \{1', 0, n-1\} \cup S \cup \{3\},$$
$$(\{1', 0, n-1\} \cup S(a^3)^{-1} \cup \{q-3\})a^3 = \{1', 0, n-1\} \cup S \cup \{q\},$$

for $q = 4, \dots, n-2$.

6. Sets with $2'$:

$$(\{1', 0, n-1\} \cup Sc^{-1})c = \{2', n-1\} \cup S \cup \{1\}, \text{ as in Case 4},$$
$$(\{1', 0, n-1\} \cup Sa^{-1} \cup \{q-1\})a = \{2', n-1\} \cup S \cup \{q\},$$

for $q = 2, \dots, n-2$.

Thus the induction step goes through and all required sets are reachable. □

Cmorik and Jirásková [23, Theorem 5] also found binary witnesses that meet the bound $(m-1)2^{n-2}$ in the case where $m-2$ and $n-2$ are relatively prime. It remained unknown whether the bound $(m-1)2^{n-2}+1$ is reachable with a binary alphabet. We show in the Appendix that a slight modification of the first witness of [23] meets the upper bound exactly.

Theorem 4. *Suppose $m, n \geqslant 4$ and $L'_m L_n$ meets the bound $2^{n-2}+1$. Then the transition semigroup T_n of a minimal DFA $\mathcal{D}_n$ of L_n is a subsemigroup of $\mathbf{T}^{\leqslant 5}(n)$ and is not a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$.*

Proof. Let $\mathcal{D}'_m = (Q'_m, \Sigma, \delta', 0', F')$ with $Q'_m = \{0', \dots, (m-1)'\}$, and let $\mathcal{D}_n = (Q_n, \Sigma, \delta, 0, F)$ with $Q_n = \{0, \dots, n-1\}$. We construct the NFA for $L'_m L_n$ as usual.

From [26, Lemma 9] we know that the set of reachable and distinguishable subsets of $Q'_m \cup Q_n$ can be represented by: $\{p'\} \cup X$ for each $p \in \{1, \dots, m-1\}$ and $X \subseteq \{1, \dots, n-2\}$, or by $\{0'\}$. A reachable subset S contains 0 if and only if $S \cap F' \neq \emptyset$. Also S is not distinguishable from $S \cup \{(m-1)'\}$ and $S \cup \{n-1\}$. To reach the bound $(m-1)2^{n-2}+1$, all these subsets or their equivalents must be reachable.

Suppose that $p, q \in \{1, \dots, n-2\}$ are not colliding in Q_n. Consider S that contains both p and q. Then S is reached from some S' by a transformation t, where S' contains 0 and a state $r \in \{1, \dots, n-2\}$ such that $0t = p$ and $rt = q$ (or $0t = q$ and $rt = p$). But since p and q are not colliding, there is no such transformation in T_n. Thus all pairs in $Q_n \setminus \{0, n-1\}$ must be colliding and T_n is a subsemigroup of $\mathbf{T}^{\leqslant 5}$. □

3.3. Boolean Operations

Let $\mathcal{D} = \mathcal{D}(a,b,c)$ be the DFA of Definition 1, and let $L = L(a,b,c)$ be the language of this DFA. Consider the partial permutation $\pi'(a) = a$, $\pi'(b) = b$, and $\pi'(c) = -$. The dialect associated with π' is $\mathcal{D}'(a,b,-)$, which is $\mathcal{D}(a,b,c)$ restricted to the alphabet $\{a,b\}$. This is our first witness for Boolean operations, and we prime its states to distinguish them from those of the second witness defined below. This DFA is illustrated in Figure 5 for $m = 9$.

Now take the partial permutation π, where $\pi(a) = -$, $\pi(b) = b$, and $\pi(c) = a$; here a plays the role of c. Thus $\mathcal{D}(-,b,a)$ is the DFA in which $a\colon (0 \to n-1)(1, \dots, n-2)$ and $b\colon (2 \to n-1)(1 \to 2)(0 \to 1)(3,4)$, as illustrated in Figure 5 for $n = 8$. This DFA is our second witness for Boolean operations. The language of $\mathcal{D}(-,b,a)$ is the dialect $L(-,b,a)$ of L.

Theorem 5. *For $m, n \geqslant 6$, $L'_m(a,b,-)$ and $L_n(-,b,a)$ meet the bounds $mn-(m+n-2)$ for union and symmetric difference, $mn-2(m+n-3)$ for intersection, and $mn-(m+2n-4)$ for difference.*

Proof. As usual, we construct the direct product $\mathcal{P}$ of $\mathcal{D}'_m$ and $\mathcal{D}_n$. The initial state of $\mathcal{P}$ is $(0',0)$ which reaches only $(1',1)$ and can never be re-entered. Thus $0'$ and 0 do not appear in any reachable state of $\mathcal{P}$ other than $(0',0)$. We show that the remaining

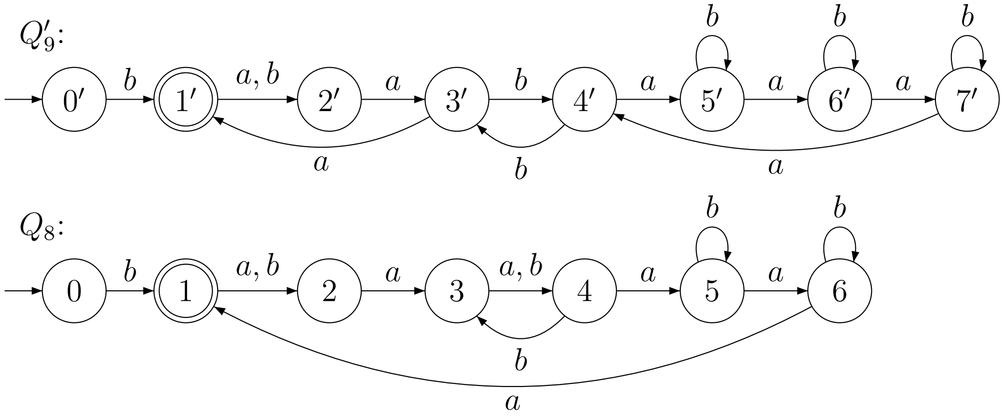

Figure 5: Witnesses $\mathcal{D}_9'(a, b, -)$ and $\mathcal{D}_8(-, b, a)$ for Boolean operations. The empty states $8'$ and 7 and the transitions to them are omitted.

$(m-1)(n-1)$ states are all reachable, for a total of $1+(m-1)(n-1) = mn-(m+n-2)$ states. Let m define the row and n the column in the direct product.

As a preliminary step we show how $((m-2)', n-2)$ can be reached. First we reach $(4', 4)$ by ba^2b. If $m = 6$, then $(4', 4)a^{q-4} = (4', q)$ for $q = 5, \dots, (n-2)$. Otherwise, if $m > n$, then $(0', 0)ba^3(ba)^2 = (5', 4)$, $(5', 4)(ba)^{m-n-1} = ((m-n+4)', 4)$, and $(m-n+4, 4)a^{n-6} = ((m-2)', n-2)$. If $m = n$, then $(4', 4)a^{n-6} = ((m-2)', n-2)$. Let $m < n$. Observe that if $(n-m) \equiv p \pmod 3$ for $0 \leqslant p \leqslant 2$, then $((3-p)', 4)a^{n-m}ba^{m-6} = ((m-2)', n-2)$. Hence we need to reach $((3-p)', 4)$. We consider the following three cases:

1. $(n-m) \equiv 0 \pmod 3$.
   We reach $(3', 4)$ by b^2a^2ba, and then apply $a^{n-m}ba^{m-6}$.
2. $(n-m) \equiv 2 \pmod 3$.
   We reach $(1', 4)$ by b^2a^2, and then apply $a^{n-m}ba^{m-6}$.
3. $(n-m) \equiv 1 \pmod 3$.
   If $m \leqslant n-4$ then $(1', 4)a^3ba^{n-m-3}ba^{m-6} = ((m-2)', n-2)$. Otherwise $m = n-1$. If $n < 12$, then we have five cases, where $(0', 0)ba^2 = (3', 3)$:
   (a) $m = 6$, $n = 7$: $(4', 4)a^2ba^2 = (2', 4)$.
   (b) $m = 7$, $n = 8$: $(3', 3)(ab)^2a^4ba^2 = (2', 4)$.
   (c) $m = 8$, $n = 9$: $(3', 3)a^8 = (2', 4)$.
   (d) $m = 9$, $n = 10$: $(3', 3)aba^9 = (2', 4)$.
   (e) $m = 10$, $n = 11$: $(3', 3)aba^4ba^3baba^6ba^2 = (2', 4)$.
   Otherwise we reach $(2', 4)$ as follows: $(4', 4)a^{m-6} = (4', n-2)$, $(4', n-2)a^{m-6} = (4', n-7)$, $(4', n-7)baba^2ba^3 = (2', 1)$, $(2', 1)a^3 = (2', 4)$.

Now, having reached $((m-2)', n-2)$, we can reach all the remaining pairs: We have $((m-2)', n-2)(ab)^2b = ((m-1)', 3)$, and from $((m-1)', 3)$ all pairs $((m-1)', q)$ for $1 \leqslant q \leqslant n-1$ can be reached, since $Q_n \setminus \{0, n-1\}$ is a strongly connected component in $\mathcal{D}_n$. Similarly, $((m-2)', n-2)ab^2 = (4', n-1)$, and all pairs $(p', n-1)$ for $1 \leqslant p \leqslant m-1$ can be reached. For the remaining pairs we proceed as follows:

1. Column 3:
   (a) $((m-2)', n-2)aba = (1', 3)$,
   (b) $(0,0)ba^3b = (2', 3)$,
   (c) $(0,0)ba^2 = (3', 3)$,
   (d) $(2', 3)ab = (4', 3)$,
   (e) $((p-1)', 3)ab = (p', 3)$ for $p = 5, \dots, m-2$.
2. Column 4:
   (a) $(3', 3)a = (1', 4)$,
   (b) $(3', 3)b = (4', 4)$,
   (c) $((p-1)', 3)a = (p', 4)$ for $p = 2, 3, 5, \dots, m-2$.
3. Column q for $q = 5, \dots, n-2$:
   (a) $(3', q-1)a = (1', q)$,
   (b) $(3', q)b = (4', q)$,
   (c) $((p-1)', q-1)a = (p', q)$ for $p = 2, 3, 5, \dots, m-2$.
4. Column 1:
   (a) $(0', 0)b = (1', 1)$,
   (b) $((m-2)', n-2)a = (4', 1)$.
   (c) $((p-1)', n-2)a = (p', 1)$ for $p = 2, 3, 5, \dots, m-2$.
5. Column 2:
   (a) $(3', 1)a = (1', 2)$,
   (b) $(0', 0)a^2 = (2', 2)$,
   (c) $(3', 1)b = (4', 2)$,
   (d) $((p-1)', 1)a = (p', 2)$ for $p = 3, 5, \dots, m-2$.

As we discussed in Subsection 3.2, for each $x \in \Sigma$ every state $q \in Q_n \setminus \{0, n-1\}$ has a unique predecessor state $p \in Q_n \setminus \{n-1\}$. It follows that if $qw = p \in Q_n \setminus \{0, n-1\}$, for some state q and word w, then $rw \neq p$ for $r \neq q$. The same facts apply to Q'_m. We shall need the following two claims:

Claim 1. From any pair (q', p) with $1 \leqslant q \leqslant m-2$, $1 \leqslant p \leqslant n-2$ we can reach $(1', 1)$.

First we find a word w such that $q'w \in \{1', 2', 3'\}$ and $pw = 1$. Note that it is sufficient to find u such that $q'u \in \{1', 2', 3'\}$ and $pu \in Q_n \setminus \{0, n-1\}$, because then

$w = ua^{n-1-(pu)}$ does the job. So if $q' \in \{1', 2', 3'\}$, then we are done. Otherwise use a^{m-1-q}, which maps q' to $4'$. If $pa^{m-1-q} \neq 2$, then we use b to map $4'$ to $3'$ and keep the state of $\mathcal{D}_n$ in $Q_n \setminus \{0, n-1\}$, and again we are done. Suppose $pa^{m-1-q} = 2$. If $m = 6$ then $4'ab = 3'$ and $pa^{m-1-q}ab \neq 2$, and we are done. Otherwise $4'a^{m-5} = 4'$, and also $4'aba^{m-6} = 4'$. If $2a^{m-5} \neq 2$ then we can apply an additional b and be done. If $2a^{m-5} = 2$, then $2aba^{m-6} = 3$, and again by applying an additional b we are done.

Now let $q' \in \{2', 3'\}$. We show that there is a word w such that $q'w = 1$ and $1w = 1$. If $q'a^{n-2} = 1'$ then we are done, and if $q'a^{n-2} \in \{2', 3'\} \setminus \{q'\}$, then $q'(a^{n-2})^2 = 1$ and we have reached $(1', 1)$. So assume $q'a^{n-2} = q'$; thus $(n-2) \equiv 0$ mod 3, and so $n \geqslant 8$.

If $q' = 3'$, then $(3', 1)a^4ba^{n-1-4} = (r', 1)$ with $r' \neq 3'$, so we are done by $(a^4ba^{n-1-4})^2$. If $q' = 2'$, then $(2', 1)a^5ba^{n-1-5} = (r', 1)$ with $r' \neq 2'$, so we are done by $a^5b(a^{n-1-5})^2$.

Claim 2. For any $q' \in Q'_m \setminus \{0', (m-1)'\}$ and $p \in Q_n \setminus \{0, (n-1)\}$ there is a word such that $q'w \in Q'_m \setminus \{0', (m-1)'\}$ and $pw = n-1$. It follows that (q', p) can be mapped to $(r', n-1)$, for any state $r' \in Q'_m \setminus \{0'\}$.

If $q' \in \{1', 2', 3'\}$ then let $w_1 = a^2b$, ab, or b, respectively, and ε otherwise; so w_1 maps q' to $\{4', \ldots, (m-2)'\}$. Then let $w_2 = a^{n-(pw_1)}$; so w_2 maps pw_1 to 2, and qw_1 is still in $\{4', \ldots, (m-2)'\}$. Then $w = w_1w_2b$ satisfies the claim. Since $Q'_m \setminus \{0', (m-1)'\}$ is strongly connected, we can map $(q'w, n-1)$ to any $r' \in Q'_m \setminus \{0'\}$.

Now we consider the four cases of the operations. In all cases pair $(0', 0)$ is distinguished as the only non-empty state which does not accept any word starting with a.

1. Union and symmetric difference.
   Consider (q'_1, p_1) and (q'_2, p_2) with $q'_1 \neq q'_2$, where $1 \leqslant q_1, q_2 \leqslant m-1$ and $1 \leqslant p_1, p_2 \leqslant n-1$. Without loss of generality, $q'_1 \neq (m-1)'$. The same arguments here work for both operations, since we are not mapping the pairs to $(1', 1)$. By Claim 2 we can map (q'_1, p_1) to $(1, n-1)$ by some word w. Then $q'_2w \neq 1$. If also $p_2 \neq 1$, then (q'_2, p_2) is not final and w distinguishes our pairs. Otherwise we apply Claim 2 once more for $(1, p_2)$, obtaining a word u such that $(1, n-1)u = (1, n-1)$, and $(q'_2, 1)u = (q'_2u, n-1)$. Since $1u = u$, we have $q'_2u \neq 1$, and so $w'u$ distinguishes our pairs.
   Assume now $q'_1 = q'_2$ and $p_1 \neq p_2$, where $p_1 \neq n-1$. Let w be a word mapping p_1 to 1; then $p_2w \neq 1$. If $q'_1w \neq 1'$, $(q'_2, p_2)w = (q'_1, p_2)w$ is not final so w distinguishes our pairs. Otherwise wa^3b^2 maps q'_1 to $(m-1)'$, and p_1 to 4. Thus $wa^3b^2a^{n-6}$ maps p_1 to 1, and since p_2 is mapped elsewhere and $q'_1 = q'_2$ to $(m-1)'$, our pairs are distinguished.
2. Intersection.

Consider (q'_1, p_1) and (q'_2, p_2) with $q'_1 \neq q'_2$ or $p_1 \neq p_2$, where $1 \leqslant q_1, q_2 \leqslant m-2$ and $1 \leqslant p_1, p_2 \leqslant n-2$. By Claim 1 we can map (q'_1, p_1) to the final state $(1', 1)$ by some word w. Then either $q'_2 w \neq 1'$ or $p_2 w \neq 1$, (q'_2, p_2) is not final and our pairs are distinguished. Together with $(0', 0)$ and $((m-1)', n-1)$ these give $2 + (m-2)(n-2) = mn - 2(m+n-3)$ distinguished pairs.

3. Difference.
Consider (q'_1, p_1) and (q'_2, p_2) with $q'_1 \neq q'_2$, where $1 \leqslant q_1, q_2 \leqslant m-2$ and $1 \leqslant p_1, p_2 \leqslant n-1$. This follows in exactly the same way as the corresponding case of union and symmetric difference. Assume now $q'_1 = q'_2$ and $p_1 \neq p_2$. Without loss of generality, $p_1 \neq n-1$. By Claim 1 we can map (q'_1, p_1) to the non- final state $(1', 1)$ by some word w. Then $(q'_2, p_2)w = (1', p_2 w)$, and since $p_2 w \neq p_1 w = 1$, $(1', p_2 w)$ a final state. Thus w distinguishes our pairs.
Together with $(0', 0)$ and $((m-1)', n-1)$ these give $2 + (m-2)(n-1) = mn - (m + 2n - 4)$ distinguished pairs.

□

4. Witnesses with Semigroups in $\mathbf{T}^{\geqslant 6}(n)$

We now turn to the operations which cannot have witnesses with transition semigroups in $\mathbf{T}^{\leqslant 5}$.

Definition 2. *For $n \geqslant 4$, we define the DFA $\mathcal{D}_n(a, b, c, d, e) = (Q_n, \Sigma, \delta, 0, F)$, where $Q_n = \{0, \dots, n-1\}$, $\Sigma = \{a, b, c, d, e\}$, δ is defined by the transformations*

- a: $(0 \to n-1)(1, \dots, n-2)$,
- b: $(0 \to n-1)(1, 2)$,
- c: $(0 \to n-1)(n-2 \to 1)$,
- d: $(\{0, 1\} \to n-1)$,
- e: $(Q \setminus \{0\} \to n-1)(0 \to 1)$,

and $F = \{q \in Q_n \setminus \{0, n-1\} \mid q \text{ is odd}\}$. For $n = 4$, a and b coincide, and we can use $\Sigma = \{b, c, d, e\}$. The structure of $\mathcal{D}_5(a, b, c, d, e)$ is illustrated in Figure 3.

Our main result in this section is the following theorem:

Theorem 6 (Boolean Operations, Reversal, Number and Complexity of Atoms, Syntactic Complexity)**.** *Let $\mathcal{D}_n(a,b,c,d,e)$ be the DFA of Definition 2, and let the language it accepts be $L_n(a,b,c,d,e)$. Then $L_n(a,b,c,d,e)$ meets the bounds for Boolean operations, reversal, number and quotient complexity of atoms, and syntactic complexity as follows:*

1. *For $n, m \geqslant 4$, $L_m(a,b,-,d,e)$ and $L_n(b,a,-,d,e)$ meet the bounds $mn-(m+n-2)$ for union and symmetric difference, $mn-2(m+n-3)$ for intersection, and $mn-(m+2n-4)$ for difference.*
2. *For $n \geqslant 4$, $L_n(a,-,c,-,e)$ meets the bound $2^{n-2}+1$ for reversal and number of atoms.*
3. *For $n \geqslant 6$, $L_m(a,b,c,d,e)$ meets the bound $(n-1)^{n-2}+n-2$ for syntactic complexity, and the bounds on the quotient complexities of atoms.*

The claim about syntactic complexity is known from [20]. It was shown in [21] that the number of atoms of a regular language L is equal to the quotient complexity of L^R. In the next subsections we prove the claim about Boolean operations, reversal, and atom complexity. First we state some properties of $\mathcal{D}_n$.

Proposition 6. *For $n \geqslant 4$ the DFA of Definition 2 is minimal, accepts a suffix-free language, and its transition semigroup T_n has cardinality $(n-1)^{n-2}+n-2$. In particular, T_n contains (a) all $(n-1)^{n-2}$ transformations that send 0 and $n-1$ to $n-1$ and map $Q_n \setminus \{0, n-1\}$ to $Q_n \setminus \{0\}$, and (b) all $n-2$ transformations that send 0 to a state in $Q_n \setminus \{0, n-1\}$ and map all the other states to $n-1$. Also, T_n is generated by $\{a,b,c,d,e\}$ and cannot be generated by a smaller set of transformations.*

Proof. To prove minimality, note first that only the initial state 0 accepts e, and only state $n-1$ accepts nothing. Suppose $p, q \in Q_n \setminus \{0, n-1\}$, $p < q$, and p and q have the same parity; then $pa^{n-2-p}c = 1$, which is a final state, but $qa^{n-2-q}c$ is an even state, which is non-final. If p and q have different parities, then one is final while the other is not. Hence $\mathcal{D}_n$ is minimal.

The language L_n is suffix-free, because every word in L_n has the form ex with $x \in \{a,b,c,d\}$. The claims about cardinality of T_n, transformations, and generators were proved in [12, 20]. □

4.1. Boolean Operations

We now show that witness DFAs for Boolean operations may have transition semigroups in $\mathbf{T}^{\geqslant 6}$. Thus Boolean operations have witnesses with transition semigroups that are subsemigroups of either $\mathbf{T}^{\leqslant 5}$ or $\mathbf{T}^{\geqslant 6}$. This is sharp contrast with all the other complexity measures: any witness for star and any second witness for product can only be associated with $\mathbf{T}^{\leqslant 5}$, while the size of the syntactic semigroup, reversal, and complexities of atoms must have witnesses associated with $\mathbf{T}^{\geqslant 6}$.

Theorem 7. *For $n, m \geqslant 4$, $L_m(a, b, -, d, e)$ and $L_n(b, a, -, d, e)$ meet the bounds $mn - (m + n - 2)$ for union and symmetric difference, $mn - 2(m + n - 3)$ for intersection, and $mn - (m + 2n - 4)$ for difference.*

Proof. Our two DFAs $\mathcal{D}_m(a, b, -, d, e)$ and $\mathcal{D}_n(b, a, -, d, e)$ are illustrated in Figure 6; consider their direct product. Let $H = \{((m - 1)', q) \mid 1 \leqslant q \leqslant n - 2\}$, $V = \{(p', n - 1) \mid 1 \leqslant p \leqslant m - 2\}$, and $M = \{(p', q) \mid 1 \leqslant p \leqslant m - 2, 1 \leqslant q \leqslant n - 2\}$.

Let $\mathcal{A}_n(a, b) = (Q_n \setminus \{0, n - 1\}, \{a, b\}, \delta, 1, F)$ be the DFA obtained from the DFA $\mathcal{D}_n(a, b, -, d, e)$ by restricting the alphabet to $\{a, b\}$ and the set of states to $\{1, \dots, n - 2\}$. Since DFAs $\mathcal{A}'_m(a, b)$ and $\mathcal{A}_n(b, a)$ have ordered pairs $\{\delta'_a, \delta'_b\}$ and $\{\delta_a, \delta_b\}$ of transformations that generate the symmetric groups of degrees m and n and are not conjugate, the result from [2, Theorem 1] applies, except in our case where $m = 4$ and $n = 4$ (we add two states to the m and n in [2]). We have verified this case by computation. Therefore we know that all states in M are reachable from state $(1', 1)$ and all pairs of such states are distinguishable.

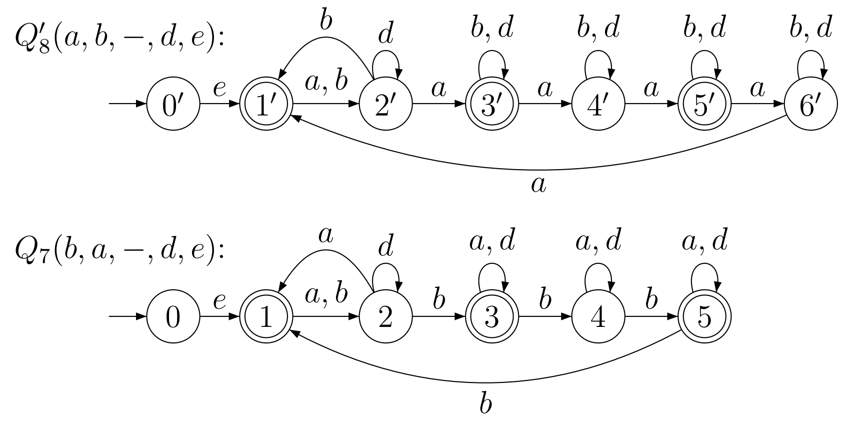

Figure 6: The DFAs $\mathcal{D}'_8(a, b, -, d, e)$ and $\mathcal{D}_7(b, a, -, d, e)$ for Boolean operations. The empty states $8'$ and 7 and the transitions to them are omitted.

We show that $(0', 0)$, $((m-1)', n-1)$ and all states in H, V and M are reachable. State $(0', 0)$ is the initial state, $(1', 1)$ is reached from $(0', 0)$ by e and $((m-1)', n-1)$ by a. By Theorem 1 of [2] all the states in M, and in particular $(1', 2)$ and $(2', 1)$, are reachable from $(1', 1)$. From state $(1', 2)$ we reach $((m - 1)', 2)$ by d and from there, state $((m - 1)', q)$ by a word in b^* for $1 \leqslant q \leqslant n - 2$. Symmetrically, from $(2', 1)$ we reach $(2', n-1)$, and from there, state $(p', n-1)$ for $1 \leqslant p \leqslant m-2$ by a word in a^*. Hence all $(m-1)(n-1)+1 = mn-(m+n-2)$ states are reachable.

State $(0', 0)$ is the only state accepting a word beginning with e. State $((m-1)', n-1)$ is the only empty state.

For union we consider six possibilities for the distinguishability of two states:

1. States in H. If n is odd, and $p', r' \in Q'_n \setminus \{0', (n-1)'\}$ with $p < r$, where p' and r' are both final or both non-final, then r' accepts a^{n-1-r} whereas p' rejects it. If n is even, r' accepts $a^{n-r}b$ whereas p' rejects it. Hence if $((m-1)', p)$ and $((m-1)', r)$ are in H, they are distinguishable.
2. States in V. The argument is symmetric to that in H.
3. H and V. Consider $(p', n-1) \in V$ and $((m-1)', q) \in H$. Note that $(1', n-1)$ accepts a^2b whereas $((m-1)', 1)$ rejects it. For every (p', q) there exists a permutation t induced by a word w in $\{a, b\}^*$ such that $(1', 1)t = (p', q)$. Then $(p', q)t^{-1} = (1', 1)$. Hence $(p', n-1)$ and $((m-1)', q)$ are mapped by t^{-1} to $(1', n-1)$ and $((m-1)', 1)$, which are distinguishable.
4. H and M. If $(p', r) \in M$, then there exists a permutation t induced by a word w in $\{a, b\}^*$ such that $(p', r)t = (1', 2)$. The same t maps $((m-1)', q) \in H$ to some $((m-1)', s)$ with $1 \leqslant s \leqslant n-2$. Then $(1', 2)d = ((m-1)', 2)$. If $s \neq 1$, then $sd \in \{2, \dots, n-2)$, and we are done, since states $((m-1)', 2)$ and $((m-1)', s)$ are in H and hence are distinguishable. Otherwise, $((m-1)', s)d = ((m-1)', n-1)$ and again we are done.
5. V and M. The argument is symmetric to that for H and M.
6. States in M. Apply Theorem 1 of [2].

For symmetric difference, we have as many states as for union. The arguments for distinguishability are exactly the same as for union.

For intersection, all states in $H \cup V \cup \{((m-1)', n-1)\}$ are empty, and all states in M are distinguishable by [2, Theorem 1]. Hence there are $(m-2)(n-2)+2 = mn-2(m+n-3)$ states altogether.

For difference, all states in $H \cup \{((m-1)', n-1)\}$ are empty, so this leaves $(m-2)(n-1)+2 = mn-(m+2n-4)$ states. The states in V are all distinct by the argument used for union. Also $(p', q) \in M$ is distinguishable from $(r', n-1) \in V$, by the argument used for union. □

Since $\delta_d = \delta_{c_1}\delta_{c_1}$ and $\delta_e = \delta_{c_1}\delta_{c_2}\cdots\delta_{c_{n-1}}$, where the c_i are from Proposition 2, the semigroup of $\mathcal{D}_n(a, b, -, d, e)$ is in $\mathbf{T}^{\leqslant 5}(n) \cap \mathbf{T}^{\geqslant 6}(n)$. In fact, one can verify that the semigroup of $\mathcal{D}_n(a, b, -, d, e)$ is $\mathbf{T}^{\leqslant 5}(n) \cap \mathbf{T}^{\geqslant 6}(n)$.

4.2. Reversal

Han and Salomaa [26] showed that to meet the bound for reversal one can use the binary DFA of Leiss [32] and add a third input to get a suffix-free DFA. Cmorik and Jirásková [23] showed that a binary alphabet will not suffice. We show a different ternary witness below, and prove that any witness must have its transition semigroup in $\mathbf{T}^{\geqslant 6}$.

Theorem 8 (Semigroup of Reversal Witness)**.** *For $n \geqslant 4$, the transition semigroup of a minimal DFA $\mathcal{D}_n(Q_n, \Sigma, \delta, 0, F)$ of a suffix-free language L_n that meets the bound $2^{n-2}+1$ for reversal is a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$.*

Proof. It suffices to show that no pair of states is colliding.

We construct an NFA $\mathcal{N}$ for L_n^R by reversing the transitions of $\mathcal{D}$, and interchanging the sets of final and initial states. We then determinize $\mathcal{N}$ using the subset construction to get a DFA $\mathcal{R}$ for L_n^R. The states of $\mathcal{R}$ are sets of states of $\mathcal{D}$.

Consider a reachable subset $S \subseteq Q_n$. We can assume that $n-1 \notin S$, since $n-1$ is not reachable from a start state in $\mathcal{N}$. Also, if $0 \in S$ then $S = \{0\}$, as otherwise in $\mathcal{D}$ some transformation would map both 0 and $q \in \{1, \dots, n-1\}$ to a final state so the language would not be suffix-free [26, Lemma 6]. Hence, every subset $S \subseteq \{1, \dots, n-2\}$ must be reachable.

Suppose that there are two distinct states $p, q \in \{1, \dots, n-2\}$ such that the pair $\{p, q\}$ is colliding. This means that $0\delta_w = p$ and $q'\delta_w = q$ for some word w and $q' \in \{1, \dots, n-2\}$. Since the subset $\{p, q\}$ is reachable in $T_{\mathcal{D}^R}$, there is some $(\delta_u)^{-1}$ such that $F(\delta_u)^{-1} = \{p, q\}$. So $p\delta_u \in F$ and $q(\delta_u) \in F$. But then $wu \in L_n$, and also $vwu \in L_u$, where v is such that $0\delta_v = q'$, which contradicts that L_n is suffix-free. □

Theorem 9 (Reversal Complexity)**.** *If $n \geqslant 4$, then $L_n(a, -, c, -, e)$ of Definition 2 meets the bound $2^{n-2}+1$ for reversal.*

Proof. As before, we construct NFA $\mathcal{N}$ for L^R – state $n-1$ is now not reachable – and determinize $\mathcal{N}$ to get the DFA $\mathcal{R}$. We shall show that we can reach $\{0\}$ and all subsets of $Q_n \setminus \{0, n-1\}$.

The initial state of $\mathcal{N}$ consists of all the odd states in $Q_n \setminus \{0, n-1\}$. By applying e we reach the set $\{0\}$. By applying e again, we reach $\emptyset$. To reach $Q_n \setminus \{0, n-1\}$ from F we apply $c(a^2c)^{\lceil n/2 \rceil - 1}$.

Consider a non-empty proper subset $S \subsetneq Q_n \setminus \{0, n-1\}$. Since a acts cyclically on $Q_n \setminus \{0, n-1\}$, it is sufficient to prove reachability for S that contains $n-2$ if n is odd, and does not contain $n-2$ if n is even; thus S agrees with F on $n-2$. We start from $F = \{1, 3, \dots, n-4, n-2\}$ and transform it inductively. Assume that for $i \geqslant 0$, we have reached F_i which agrees with S on states $1, \dots, i$ and with F on states

$i+1, \dots, n-2$. Of course, $F_0 = F$. If S and F_i agree on $i+1$, then we already have $F_{i+1} = F_i$. Otherwise $i+1 \in F_i$ if and only if $i+1 \notin S$. We have $i+1 < n-2$, since we know that S agrees with F on $n-2$. Then $i+1 \in F_i$ if and only if $i+2 \notin F_i$. We apply $a^{i+1}ca^{n-3-i}$ to F_i, which yields the subset F_{i+1} differing from F_i only in state $i+1$; thus $i+1 \in F_{i+1}$ if and only if $i+1 \in S$. The induction follows, and so $S = F_{n-2}$ is reachable.

Now we prove that these $2^{n-2}+1$ states (sets of states of $\mathcal{D}$) are pairwise distinguishable. Set $\{0\}$ is the only final state. Consider two sets $R, S \subseteq Q_n \setminus \{0, n-1\}$ and suppose that $q \in R \oplus S$. Then $a^{q-1}e$ is accepted from the set that contains q but not from other set. Therefore there are no equivalent states. □

Although $L_n(a, -, c, -, e)$ meets the bound for the number of atoms, it does not meet the bounds on the quotient complexity of atoms.

4.3. Complexity of Atoms in Suffix-Free Languages

Let $Q_n = \{0, \dots, n-1\}$ and let L be a non-empty regular language with quotient set $K = \{K_0, \dots, K_{n-1}\}$. Let $\mathcal{D} = (Q_n, \Sigma, \delta, 0, F)$ be the minimal DFA of L in which the language of state q is K_q.

Denote the complement of a language L by $\overline{L} = \Sigma^* \setminus L$. Each subset S of Q_n defines an *atomic intersection* $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$, where $\overline{S} = Q_n \setminus S$. An *atom* of L is a non-empty atomic intersection. Since atoms are pairwise disjoint, every atom A has a unique atomic intersection associated with it, and this atomic intersection has a unique subset S of K associated with it. This set S is called the *basis* of A.

Let $A_S = \bigcap_{i \in S} K_i \cap \bigcap_{i \in \overline{S}} \overline{K_i}$ be an atom. For any $w \in \Sigma^*$ we have

$$w^{-1}A_S = \bigcap_{i \in S} w^{-1}K_i \cap \bigcap_{i \in \overline{S}} \overline{w^{-1}K_i}.$$

Since a quotient of a quotient of L is also a quotient of L, $w^{-1}A_S$ has the form:

$$w^{-1}A_S = \bigcap_{i \in X} K_i \cap \bigcap_{i \in Y} \overline{K_i},$$

where $|X| \leqslant |S|$ and $|Y| \leqslant n - |S|$, $X, Y \subseteq Q_n$.

Proposition 7. *Suppose L is a suffix-free language with $n \geqslant 4$ quotients. Then L has at most $2^{n-2}+1$ atoms. Moreover, the complexity $\kappa(A_S)$ of atom A_S satisfies*

$$\kappa(A_S)\begin{cases}\leqslant 2^{n-2}+1, & \textit{if } S=\emptyset;\\ = n, & \textit{if } S=\{0\};\\ \leqslant 1+\sum_{x=1}^{|S|}\sum_{y=0}^{n-2-|S|}\binom{n-2}{x}\binom{n-2-x}{y}, & \emptyset\neq S\subseteq\{1,\dots,n-2\}.\end{cases}\tag{2}$$

Proof. If $n-1 \in S$, then A_S is not an atom. Also, $K_0 \cap K_i = \emptyset$ for $i = 1, \dots, n-1$, since L is suffix-free. Hence $0 \in S$ implies $S = \{0\}$. It follows that there are at most $2^{n-2}+1$ atoms.

For atom complexity, consider the following cases:

1. $S = \emptyset$. Here $\overline{S} = Q_n$, $A_\emptyset = \bigcap_{i\in Q_n} \overline{K_i}$ and for $w \in \Sigma^+$, $w^{-1}A_S = \bigcap_{i\in Y} \overline{K_i}$, where $0 \notin Y$, since $w^{-1}L = L$ only if $w = \varepsilon$, and $n-1$ is always in Y since $w^{-1}K_{n-1} = K_{n-1}$ because $K_{n-1} = \emptyset$. Thus we have the initial quotient $A_\emptyset$ with $0 \in \overline{S}$ and at most 2^{n-2} choices for the other quotients of A_S. Hence the complexity of $A_\emptyset$ is at most $2^{n-2}+1$.
2. $S = \{0\}$. Since the language is suffix-free, we have that $K_0 \cap K_q = \emptyset$ for all $q \in Q \setminus \{0\}$. Thus $A_0 = K_0 \cap \overline{K_1} \cap \ldots \cap \overline{K_{n-1}} = K_0$, and $K_0 = L_n$ has n quotients.
3. Since $n-1$ always appears in $\overline{S}$ and the $S = \{0\}$ case is done, there remain the cases where $\emptyset \neq S \subseteq Q_n \setminus \{0, n-1\}$. Suppose $(X, Y \cup \{n-1\})$, where $n-1 \notin Y$, represents a non-empty quotient of A_S by a non-empty word w. Since 0 appears only initially, Y cannot contain 0. Then X must have at least one and at most $|S|$ elements from $\{1, \dots, n-2\}$, and $Y \subseteq \{1, \dots, n-2\} \setminus X$. If $0\delta_w = n-1$, then all the states that are mapped to Y by δ_w are from $\{1, \dots, n-2\} \setminus S$, so in Y there must be from 0 to $n-2-|S|$ states. Hence, together with the empty quotient, we have the formula from Equation 2. Suppose that $0\delta_w \neq n-1$. The 0-path in δ_w is aperiodic and ends in $n-1$, so there exists a state $p \in \{1, \dots, n-2\}$ such that $p\delta_w = n-1$. If $p \in S$, then $n-1 \in X$, and so (X, Y) represents the empty quotient. If $p \in \{1, \dots, n-2\} \setminus S$, then again Y contains at most $n-2-|S|$ states.

□

Following Iván [29] we define a DFA for each atom:

Definition 3. *Suppose $\mathcal{D} = (Q_n, \Sigma, \delta, 0, F)$ is a minimal DFA and let $S \subseteq Q_n$. Define the DFA $\mathcal{D}_S = (Q_S, \Sigma, \Delta, (S, \overline{S}), F_S)$, where*

- $Q_S = \{(X, Y) \mid X, Y \subseteq Q_n, X \cap Y = \emptyset\} \cup \{\bot\}$.

- *For all $a \in \Sigma$, $(X,Y)a = (Xa, Ya)$ if $Xa \cap Ya = \emptyset$, and $(X,Y)a = \perp$ otherwise; and $\perp a = \perp$.*
- $F_S = \{(X,Y) \mid X \subseteq F, Y \subseteq \overline{F}\}$.

DFA $\mathcal{D}_S$ recognizes the atomic intersection A_S of L. If $\mathcal{D}_S$ recognizes a non-empty language, then A_S is an atom.

Theorem 10. *For $n \geqslant 4$, the language $L_n(\mathcal{D}_n(a,b,c,d,e))$ of Definition 2 meets the bounds of Proposition 7 for the atoms.*

Proof. There are three cases to consider:

1. $S = \emptyset$. The initial state of $\mathcal{D}_S$ is $(\emptyset, Q_n)$. Since $(\emptyset, Q_n)ed = (\emptyset, \{n-1\})$, which is a final state, $A_\emptyset$ is an atom. For any non-empty word w, we have a quotient of $A_\emptyset$ represented by $(\emptyset, Y)$, where $0 \notin Y$, $n-1 \in Y$, and $Y \setminus \{n-1\}$ is a subset of $Q_n \setminus \{0, n-1\}$. By Proposition 6 the transition semigroup of $\mathcal{D}$ contains all transformations that send 0 and $n-1$ to $n-1$ and map $Q_n \setminus \{0, n-1\}$ into $Q_n \setminus \{0\}$; hence all 2^{n-2} quotients of $A_\emptyset$ by non-empty words are reachable. Together with $A_\emptyset$ we have $2^{n-2}+1$ quotients of $A_\emptyset$.
   Since $(\emptyset, Q_n)e = (\emptyset, \{1, n-1\})$, which is non-final, and $(\emptyset, Y)e = (\emptyset, \{n-1\})$, which is final, $(\emptyset, Y)$ is distinguished from the initial state. We claim also that all pairs $(\emptyset, Y)$ and $(\emptyset, Y')$, $Y \neq Y'$, are distinguishable. Without loss of generality, suppose that $q \in Y \setminus Y'$. If we map q to 1 and $(Y \cup Y') \setminus \{1\}$ to $n-1$ by a word w, then $(\emptyset, Y)w = (\emptyset, \{1, n-1\})$, which is non-final, and $(\emptyset, Y')w = (\emptyset, \{n-1\})$, which is final. Hence the bound $2^{n-2}+1$ is met.
2. $S = \{0\}$. The initial state of $\mathcal{D}_S$ is $(\{0\}, Q_n \setminus \{0\})$. We have $(\{0\}, Q_n \setminus \{0\})x = \perp$ for all $x \in \Sigma \setminus \{e\}$, and $(\{0\}, Q_n \setminus \{0\})e = (\{1\}, \{n-1\})$. From $(\{1\}, \{n-1\})$ we can reach any state $(\{q\}, \{n-1\})$ with $q \in Q_n \setminus \{0, n-1\}$. Thus we can reach n states altogether: the initial state, $n-2$ states $(\{q\}, \{n-1\})$ with $q \in Q_n \setminus \{0, n-1\}$, and $\perp$.
   Since $(\{0\}, Q_n \setminus \{0\})e = (\{1\}, \{n-1\})$, which is final, the initial state is non-empty and so differs from $\perp$. For $q \in Q_n \setminus \{0, n-1\}$, consider the word w that maps q to 1 and $Q_n \setminus \{0, q\}$ to $\{n-1\}$. Then $(\{q\}, \{n-1\})w = (\{1\}, \{n-1\})$, which is final, while $(\{0\}, Q_n \setminus \{0\})w = (\{n-1\}, Y)$ (since w does not contain e), which is non-final since Y contains 1. Hence $(\{q\}, \{n-1\})$ is distinguishable from the initial state and $\perp$. Finally, for $p, q \in Q_n \setminus \{0, n-1\}$, $(\{p\}, \{n-1\})$ is distinguishable from $(\{q\}, \{n-1\})$ by the mapping that sends p to 1 and q to $n-1$. Hence the bound n is met.

3. $\emptyset \neq S \subseteq Q_n \setminus \{0, n-1\}$. The initial state here is $(S, \overline{S})$, where $\emptyset \neq S \subseteq \{1, \dots, n-2\}$. Since we can map S to $\{1\}$ and $\overline{S}$ to $\{n-1\}$, each such intersection is an atom. If $M = \{1, \dots, n-2\}$, the atom has the form $(S, \overline{S}) = (S, (M \setminus S) \cup \{0, n-1\})$. By applying e, we get $\perp$. To get a state of A_S different from $\perp$, we can map S to any non-empty subset X of M of cardinality $|X|$, where $1 \leqslant |X| \leqslant |S|$; we can map $M \setminus S$ to any subset Y of $(M \cup \{n-1\}) \setminus X$ of cardinality $|Y|$, where $0 \leqslant |Y| \leqslant |M| - |X|$; and we can map 0 and $n-1$ to $n-1$. Each such intersection $(X, Y \cup \{n-1\})$ represents a non-empty quotient of A_S by a non-empty word because we can map X to $\{1\}$ and Y to $\{n-1\}$. Hence we have the formula in Equation 2 for the number of reachable states of $\mathcal{D}_S$.
   Now consider two states $(X, Y \cup \{n-1\})$ and $(X', Y' \cup \{n-1\})$, where one of these two states could be the initial one. If $X \neq X'$, assume that $q \in X' \setminus X$. Let w map X to $\{1\}$ and $Q_n \setminus X$ to $n-1$. Then $(X, Y \cup \{n-1\})w = (\{1\}, \{n-1\})$, while the first component of $(X', Y' \cup \{n-1\})w$ contains $n-1$, meaning that that state is $\perp$; so these states are distinguishable. A symmetric argument applies if $q \in X \setminus X'$.
   This leaves the case where $X = X'$. Then X is disjoint from both Y and Y'. If $q \in Y' \setminus Y$, let w map X to $\{1\}$, $Y \cup \{0\}$ to $\{n-1\}$, and $Y' \setminus Y$ to $\{1\}$. Then $(X, Y \cup \{n-1\})w = (\{1\}, \{n-1\})$, while $(X, Y' \cup \{n-1\})w = (\{1\}, \{1, n-1\}) = \perp$, and we have distinguishablity. A symmetric argument applies if $q \in Y \setminus Y'$.

□

Theorem 11. *The transition semigroup T_n of a minimal DFA $\mathcal{D}_n(Q_n, \Sigma, \delta, 0, F)$ with $n \geqslant 4$ of a suffix-free language L_n that meets the bounds for atom complexities from Proposition 7 is a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$ and is not a subsemigroup of $\mathbf{T}^{\leqslant 5}(n)$.*

Proof. It is sufficient to show that every pair of states $p, q \in \{1, \dots, n-2\}$ is focused by some transformation from T_n.

Let $p, q \in \{1, \dots, n-2\}$ be two distinct states. Consider the atom $A_{\{p,q\}}$. From the proof of Proposition 7, to meet the bound for the quotient complexity of $A_{\{p,q\}}$ every pair $(X, Y \cup \{n-1\})$ must represent a quotient of the atom by a non-empty word w, where $X, Y \subseteq \{1, \dots, n-2\}$ are disjoint, $1 \leqslant |X| \leqslant 2$, and $|Y| \leqslant n-2-|X|$. In particular, when $|X| = 1$ we know that the transformation of w maps both p and q to the state from X, and so p and q are focused. □

Tables 1 and 2 show the quotient complexities of atoms for small n.

Table 1: Suffix-free atom complexity. The entries from left to right are suffix-free language, left ideal, regular language. The $*$ stands for "not applicable".

n	1	2	3	4	5	$\cdots$
$\lvert S\rvert = 0$	$*/\mathbf{1}/\mathbf{1}$	$*/\mathbf{2}/\mathbf{3}$	$*/4/7$	$5/8/15$	$9/16/31$	$\cdots$
$\lvert S\rvert = 1$	$*/\mathbf{1}/\mathbf{1}$	$*/\mathbf{2}/\mathbf{3}$	$*/\mathbf{5}/\mathbf{10}$	$\mathbf{5}/13/29$	$13/33/76$	$\cdots$
$\lvert S\rvert = 2$		$*/\mathbf{2}/\mathbf{3}$	$*/4/\mathbf{10}$	$4/\mathbf{16}/\mathbf{43}$	$\mathbf{16}/\mathbf{53}/\mathbf{141}$	$\cdots$
$\lvert S\rvert = 3$			$*/3/7$	$*/8/29$	$8/43/\mathbf{141}$	$\cdots$
$\lvert S\rvert = 4$				$*/4/15$	$*/16/76$	$\cdots$
$\lvert S\rvert = 5$					$*/5/31$	$\cdots$
max	$*/1/1$	$*/2/3$	$*/5/10$	$5/16/43$	$16/53/141$	$\cdots$
ratio	$-$	$*/2.00/3.00$	$*/2.50/3.33$	$*/3.20/4.30$	$3.20/3.31/3.28$	$\cdots$

5. Conclusions

It may appear that the semigroup $\mathbf{T}^{\leqslant 5}(n)$ should not be of great importance, since it exceeds $\mathbf{T}^{\geqslant 6}(n)$ only for $n = 4$ and $n = 5$, and therefore should not matter when n is large. However, our results show that this is not the case. We conclude with our result about the non-existence of single universal suffix-free witness.

Theorem 12. *For $n \geqslant 4$, there does not exist a most complex suffix-free language.*

Proof. If there exists a most complex DFA $\mathcal{D}(Q_n, \Sigma, \delta, 0, F)$, then in particular it would have to meet the bound for reversal and star. From Theorem 8 the transition semigroup of $\mathcal{D}$ must be a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$, but from Theorem 2 we know that it cannot be a subsemigroup of $\mathbf{T}^{\geqslant 6}(n)$. Consequently no most complex suffix-free language exists for $n \geqslant 4$. □

The first four studies of most complex languages were done for the classes of regular languages [6], right ideals [8, 9, 16], left ideals [9, 10, 16], and two-sided ideals [9, 10, 16]. In those cases there exists a single witness over a minimal alphabet which, together with its permutational dialects, covers all the complexity measures. In the case of suffix-free languages such a witness does not exist. Our study is an example of a general problem: Given a class of regular languages, find the smallest set of witnesses over minimal alphabets that together cover all the measures. The witness of Definition 1 is conjectured to be over a minimal alphabet, unless the bound for product can be met by binary DFAs for every $n, m > c$, for some c; this is an

Table 2: Suffix-free atom complexity continued.

n	6	7	8	9
$\|S\| = 0$	17/32/63	33/64/127	65/128/255	129/256/511
$\|S\| = 1$	33/81/187	81/193/442	193/449/1,017	449/1,025/2,296
$\|S\| = 2$	**53**/156/406	156/427/1,086	427/1,114/2,773	1,114/2,809/6,859
$\|S\| = 3$	43/**166/501**	**166/542/1,548**	**542**/1,611/4,425	1,611/4,517/12,043
$\|S\| = 4$	16/106/406	106/462/**1,548**	462/**1,646/5,083**	**1,646/5,245/15,361**
$\|S\| = 5$	$*$/32/187	32/249/1,086	249/1,205/4,425	1,205/4,643/**15,361**
$\|S\| = 6$	$*$/6/63	$*$/64/442	64/568/2,773	568/3,019/12,043
$\|S\| = 7$		$*$/7/127	$*$/128/1,017	128/1,271/6,859
$\|S\| = 8$			$*$/8/255	$*$/256/2,296
$\|S\| = 9$				$*$/9/511
max	53/166/501	166/542/1,548	542/1,646/5,083	1,646/5,245/15,361
ratio	3.31/3.13/3.55	3.13/3.27/3.09	3.27/3.04/3.28	3.04/3.19/3.02

open problem. The witness of Definition 2 is over a minimal alphabet, since five letters are required to meet the bound for syntactic complexity.

Acknowledgments
This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant No. OGP000087, and by the National Science Centre, Poland under grant number 2014/15/B/ST6/00615.
We are grateful to Sylvie Davies for careful proofreading, and for anonymous reviewers for their insightful comments.

References

[1] T. Ang, J.A. Brzozowski, Languages convex with respect to binary relations, and their closure properties, Acta Cybernet. 19 (2009) 445–464.

[2] J. Bell, J.A. Brzozowski, N. Moreira, R. Reis, Symmetric groups and quotient complexity of boolean operations, in: J. Esparza, et al. (Eds.), ICALP 2014, volume 8573 of *LNCS*, Springer, 2014, pp. 1–12.

[3] J. Berstel, D. Perrin, C. Reutenauer, Codes and Automata, Cambridge University Press, 2009.

[4] J.A. Brzozowski, Complexity in convex languages, in: A. Dediu, H. Fernau, C. Martin-Vide (Eds.), LATA 2010, volume 6031 of *LNCS*, Springer, 2010, pp. 1–15.

[5] J.A. Brzozowski, Quotient complexity of regular languages, J. Autom. Lang. Comb. 15 (2010) 71–89.

[6] J.A. Brzozowski, In search of the most complex regular languages, Int. J. Found. Comput. Sc. 24 (2013) 691–708.

[7] J.A. Brzozowski, Unrestricted state complexity of binary operations on regular languages, in: F.M. C. Câmpeanu, J. Shallit (Eds.), DCFS 2016, volume 9777 of *LNCS*, Springer Berlin / Heidelberg, 2016, pp. 60–72. For a revised version see `http://arxiv.org/abs/1602.01387`.

[8] J.A. Brzozowski, G. Davies, Most complex regular right ideals, in: H. Jürgensen, et al. (Eds.), DCFS, volume 8614 of *LNCS*, Springer, 2014, pp. 90–101.

[9] J.A. Brzozowski, S. Davies, Quotient complexities of atoms in regular ideal languages, Acta Cybernet. 22 (2015) 293–311.

[10] J.A. Brzozowski, S. Davies, B.Y.V. Liu, Most complex regular ideal languages, Discrete Math. Theor. Comput. Sci. (2016). To appear, also at `http://arxiv.org/abs/1511.00157`.

[11] J.A. Brzozowski, G. Jirásková, B. Li, J. Smith, Quotient complexity of bifix-, factor-, and subword-free regular languages, Acta Cybernet. 21 (2014) 507–527.

[12] J.A. Brzozowski, B. Li, Y. Ye, Syntactic complexity of prefix-, suffix-, bifix-, and factor-free regular languages, Theoret. Comput. Sci. 449 (2012) 37–53.

[13] J.A. Brzozowski, B. Liu, Quotient complexity of star-free languages, Int. J. Found. Comput. Sci. 23 (2012) 1261–1276.

[14] J.A. Brzozowski, C. Sinnamon, Complexity of left-ideal, suffix-closed, and suffix-free regular languages, 2016. `http://arxiv.org/abs/1610.00728`.

[15] J.A. Brzozowski, C. Sinnamon, Complexity of prefix-convex regular languages, 2016. `http://arxiv.org/abs/1605.06697`.

[16] J.A. Brzozowski, C. Sinnamon, Unrestricted state complexity of binary operations on regular and ideal languages, 2016. `http://arxiv.org/abs/1602.01387`.

[17] J.A. Brzozowski, C. Sinnamon, Complexity of right-ideal, prefix-closed, and prefix-free regular languages, Acta Cybernet. (2017). To appear. Also at `http://arxiv.org/abs/1605.06697`.

[18] J.A. Brzozowski, M. Szykuła, Upper bounds on syntactic complexity of left and two-sided ideals, in: A.M. Shur, M.V. Volkov (Eds.), DLT 2014, volume 8633 of *LNCS*, Springer, 2014, pp. 13–24.

[19] J.A. Brzozowski, M. Szykuła, Complexity of suffix-free regular languages, in: A. Kosowski, I. Walukiewicz (Eds.), FCT 2015, volume 9210 of *LNCS*, Springer, 2015, pp. 146–159.

[20] J.A. Brzozowski, M. Szykuła, Upper bound for syntactic complexity of suffix-free languages, in: A. Okhotin, J. Shallit (Eds.), DCFS 2015, volume 9118 of *LNCS*, Springer, 2015, pp. 33–45. Full paper at `http://arxiv.org/abs/1412.2281`.

[21] J.A. Brzozowski, H. Tamm, Theory of átomata, Theoret. Comput. Sci. 539 (2014) 13–27.

[22] J.A. Brzozowski, Y. Ye, Syntactic complexity of ideal and closed languages, in: G. Mauri, A. Leporati (Eds.), DLT, volume 6795 of *LNCS*, Springer, 2011, pp. 117–128.

[23] R. Cmorik, G. Jirásková, Basic operations on binary suffix-free languages, in: Z. Kotásek, et al. (Eds.), MEMICS, pp. 94–102.

[24] M. Crochemore, C. Hancart, Automata for pattern matching, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, volume 2, Springer, 1997, pp. 399–462.

[25] Y.S. Han, K. Salomaa, State complexity of union and intersection of finite languages, Int. J. Found. Comput. Sci. 19 (2008) 581–595.

[26] Y.S. Han, K. Salomaa, State complexity of basic operations on suffix-free regular languages, Theoret. Comput. Sci. 410 (2009) 2537–2548.

[27] M. Holzer, B. König, On deterministic finite automata and syntactic monoid size, Theoret. Comput. Sci. 327 (2004) 319–347.

[28] M. Holzer, M. Kutrib, Descriptional and computational complexity of finite automata—a survey, Information and Computation 209 (2011) 456 – 470.

[29] S. Iván, Complexity of atoms, combinatorially, Inform. Process. Lett. 116 (2016) 356–360.

[30] G. Jirásková, P. Olejár, State complexity of union and intersection of binary suffix-free languages, in: H. Bordihn, et al. (Eds.), NCMA, Austrian Computer Society, 2009, pp. 151–166.

[31] B. Krawetz, J. Lawrence, J. Shallit, State complexity and the monoid of transformations of a finite set, in: M. Domaratzki, A. Okhotin, K. Salomaa, S. Yu (Eds.), Proceedings of the Implementation and Application of Automata, (CIAA), volume 3317 of *LNCS*, Springer, 2005, pp. 213–224.

[32] E. Leiss, Succinct representation of regular languages by boolean automata, Theoret. Comput. Sci. 13 (2009) 323–330.

[33] A.N. Maslov, Estimates of the number of states of finite automata, Dokl. Akad. Nauk SSSR 194 (1970) 1266–1268 (Russian). English translation: Soviet Math. Dokl. **11** (1970), 1373–1375.

[34] J.E. Pin, Syntactic semigroups, in: Handbook of Formal Languages, vol. 1: Word, Language, Grammar, Springer, New York, NY, USA, 1997, pp. 679–746.

[35] G. Thierrin, Convex languages, in: M. Nivat (Ed.), Automata, Languages and Programming, North-Holland, 1973, pp. 481–492.

[36] S. Yu, State complexity of regular languages, J. Autom. Lang. Comb. 6 (2001) 221–234.

[37] S. Yu, Q. Zhuang, K. Salomaa, The state complexities of some basic operations on regular languages, Theoret. Comput. Sci. 125 (1994) 315–328.

6. Appendix

Product of Suffix-Free Languages: Binary Case

For $m \geqslant 6, n \geqslant 3$, let the first DFA be that of [23], except that the set of final states is changed to $\{2', 4'\}$; thus let $\Sigma = \{a, b\}$, $\mathcal{D}'_m(a,b) = (Q'_m, \Sigma, \delta', 0', \{2', 4'\})$, and let $\mathcal{D}_n(a,b) = (Q_n, \Sigma, \delta, 0, \{1\})$, with the transitions shown in Figure 7. Let $L'_m(a,b)$ and $L_n(a,b)$ be the corresponding languages.

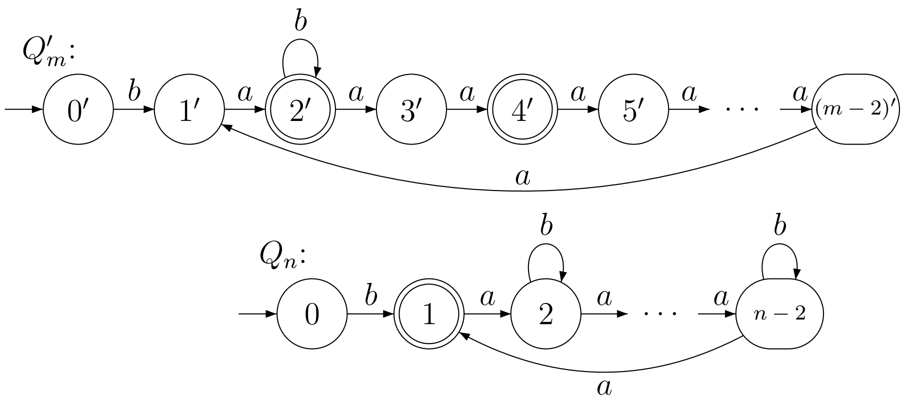

Figure 7: The witness DFA $\mathcal{D}'_m$ and $\mathcal{D}_n$ for product. The transitions to the empty states $(m-1)'$ and $n-1$ are omitted.

Theorem 13 (Product: Binary Case)**.** *For $m, n \geqslant 6$, $L'_m(a,b)$ is suffix-free and $L'_m(a,b) \cdot L_n(a,b)$ meets the bound $(m-1)2^{n-2}+1$ when $m-2$ and $n-2$ are relatively prime.*

Proof. First we need to show that $\mathcal{D}'_m(a,b)$ is minimal and $L'_m(a,b)$ is suffix-free. For minimality, it is easy to verify that every pair of states is distinguished by a word of the form a^ib. Suppose that $L'_m(a,b)$ is not suffix-free; then there are some words u, v such that $v \in L'_m(a,b)$ and $uv \in L'_m(a,b)$. Since there is no transformation mapping $0'$ to itself, this means that v maps both $0'$ and $q' = 0'u \in \{1', \dots, (m-2)'\}$ to a final state. Clearly $v = bw$ for some w and so q' must be $2'$. Then w maps $1'$ and $2'$ to a final state. Since these two states cannot be focused to any state other than $(m-1)'$, and b sends every state from $\{1', \dots, (m-2)'\}$ except $2'$ to $(m-1)'$, we have $w = a^i$. But a^i preserves the distance between the states in the cycle, and so $1'$ and $2'$ cannot be mapped simultaneously to $2'$ and $4'$. Minimality and suffix-freeness for $\mathcal{D}_n(a,b)$ has been shown in [23].

The proof for meeting the bound $(m-1)2^{n-2}+1$ is similar to that of [23, Theorem 5]. We construct the NFA for $L'_m(a,b)L_n(a,b)$ as usual. We show that $(m-1)2^{n-2}+1$ subsets of the states are reachable and distinguishable. Since a subset S without $n-1$ cannot be distinguished from $S \cup \{n-1\}$, we consider only S as the representative for both S and $S \cup \{n-1\}$. Similarly, a reachable subset S contains 0 if and only if it contains either $2'$ or $4'$; we use only $S \setminus \{0\}$ as the representative for these subsets.

Let $X \subseteq \{1, \dots, n-2\}$. If $X = \emptyset$, then clearly each of the $m-1$ subsets $\{p'\}$ for $p \in \{1, \dots, m-2\}$ is reachable by ba^{p-1}. Now let $X \neq \emptyset$. To show that each of the $(m-2)(2^{n-2}-1)$ subsets $S = \{p'\} \cup X$ for $p \in \{1, \dots, m-2\}$ is reachable, we follow the proof of [23, Theorem 5] exactly. In this proof note that $4' \in F$ makes no difference, since b maps $4'$ to $(m-1)'$, and so if applied to a subset containing $4'$, it results in a set with $(m-1)'$.

For the $2^{n-2}-1$ subsets $S = \{(m-1)'\} \cup X$, suppose first $1 \notin X$; then S is reachable from $\{1'\} \cup X$ by b. If $1 \in X$, then S is reachable from $\{4'\} \cup X$ by b.

Finally, we show that all these $(m-1)2^{n-2}$ sets together with the initial state $\{0'\}$ are distinguishable. Set $\{0'\}$ is distinguished from every other subset by bab. Consider $\{p'\} \cup X$ and $\{q'\} \cup Y$ for distinct $p, q \in \{1, \dots, m-2\}$, and some $X, Y \subseteq \{1, \dots, n-2\}$. Then $w = a^{m-p}b$ maps p' to $2'$, and q' to $(m-1)'$. Thus $(\{p'\} \cup X)w$ contains the final state 1, and $(\{q'\} \cup X)w$ does not. Consider $\{p'\} \cup X$ and $\{q'\} \cup Y$ with $X \neq Y$; then X and Y differ in some $r \in \{1, \dots, n-2\}$. Thus a^{n-r-1} distinguishes these subsets. □