Operations on Unambiguous Finite Automata

Jozef Jirásek, Jr., Galina Jirásková [1,⋆], and Juraj Šebej [2,⋆⋆]

[1] Mathematical Institute, Slovak Academy of Sciences
Grešákova 6, 040 01 Košice, Slovakia
jirasekjozef@gmail.com, jiraskov@saske.sk
[2] Institute of Computer Science, Faculty of Science, P. J. Šafárik University,
Jesenná 5, 040 01 Košice, Slovakia
juraj.sebej@gmail.com

Abstract. A nondeterministic finite automaton is unambiguous if it has at most one accepting computation on every input string. We investigate the complexity of basic regular operations on languages represented by unambiguous finite automata. We get tight upper bounds for intersection (mn), left and right quotients ($2^n - 1$), positive closure ($3/4 \cdot 2^n - 1$), star ($3/4 \cdot 2^n$), shuffle ($2^{mn} - 1$), and concatenation ($3/4 \cdot 2^{m+n} - 1$). To prove tightness, we use a binary alphabet for intersection and left and right quotients, a ternary alphabet for star and positive closure, a five-letter alphabet for shuffle, and a seven-letter alphabet for concatenation. We also get some partial results for union and complementation.

1 Introduction

A nondeterministic machine is unambiguous if it has at most one accepting computation on every input string. Ambiguity was studied intensively mainly in connection with context-free languages and it is well known that the classes of ambiguous, unambiguous, and deterministic context-free languages are all different. Ambiguity in finite automata was first considered by Schmidt [22] in his unpublished thesis, where he obtained a lower bound $2^{\Omega(\sqrt{(n)})}$ on the conversion of unambiguous finite automata into deterministic finite automata, as well as for the conversion of nondeterministic finite automata into unambiguous finite automata. He also developed an interesting lower bound method for the size of unambiguous automata based on the rank of certain matrices.

Stearns and Hunt [24] provided polynomial algorithms for the equivalence and containment problems for unambiguous finite automata (UFAs), and they extended them to ambiguity bounded by a fixed integer k. Chan and Ibarra [5] provided a polynomial space algorithm to decide, given a nondeterministic finite automaton (NFA), whether it is finitely ambiguous. They also showed that it is PSPACE-complete to decide, given an NFA M and an integer k, whether M is k-ambiguous.

⋆ Research supported by VEGA grant 2/0084/15.
⋆⋆ Research supported by VEGA grant 1/0142/15.

Ibarra and Ravikumar [13] defined the ambiguity function $a_M(n)\colon \mathbb{N} \to \mathbb{N}$ of a an NFA M such that $a_M(n)$ is the maximum number of distinct accepting computations of M on any string of length n, and they proved that the exponential ambiguity problem is decidable for NFAs. Weber and Seidl [25] showed that if an n-state NFA is finitely ambiguous, then it is at most $5^{n/2}n^n$-ambiguous. Allauzen et al. [1] considered ε-NFAs, and they proved that, given a trim ε-cycle free NFA A, it is decidable in time that is cubic in the number of transitions of A, whether A is finitely, polynomially, or exponentially ambiguous.

Ravikumar and Ibarra [21] considered the relationship between different types of ambiguity of NFAs to the succinctness of their representations, and they provided a complete picture for unary and bounded languages. Exponentially and polynomially ambiguous NFAs were separated by Leung [16] by providing, for every n, an exponentially ambiguous n-state NFA such that every equivalent polynomially ambiguous NFA requires $2^n - 1$ states.

The UFA-to-DFA tradeoff was improved to the optimal bound 2^n by Leung [17]. He described, for every n, a binary n-state UFA with a unique initial state whose equivalent DFA requires 2^n states. Similar binary example with multiple initial states was given by Leiss [15], and a ternary one was presented already by Lupanov [18]; note that the reverse of Lupanov's ternary witness for NFA-to-DFA conversion is deterministic. Leung [17] elaborated the Schmidt's lower bound method for the number of states in a UFA. He considered, for a language L, a matrix whose rows are indexed by strings x_i and columns by strings y_i, and the entry in row x_i and column y_j is 1 if $x_i y_j \in L$ and it is 0 otherwise. He showed that the rank of such a matrix provides a lower bound on the number of states in any UFA for L. Using this method, he was able to describe for every n an n-state finitely ambiguous NFA, whose equivalent UFA requires $2^n - 1$ states.

A lower bound method was further elaborated by Hromkovič et al. [12]. They used communication complexity to show that so called exact cover of all 1's with monochromatic sub-matrices in a communication matrix of a language provides a lower bound on the size of any UFA for the language. This allowed them to simplify proofs presented in [22,24]. Using communication complexity methods, Hromkovič and Schnitger [11] showed a separation of finitely and polynomially ambiguous NFAs, and even proved a hierarchy for polynomial ambiguity.

A survey paper on unambiguity in automata theory was presented by Colcombet [6], where he considered word automata, tropical automata, infinite tree automata, and register automata. He showed that the notion of unambiguity is not well understood so far, and that some challenging problems, including complementation of UFAs, remain open.

Unary unambiguous automata were examined by Okhotin [20], who proved that the tight upper bound for UFA-to-DFA conversion in the unary case is given by a function in $\mathrm{e}^{\Theta(\sqrt[3]{n(\ln n)^2})}$, while the trade-off for NFA-to-UFA conversion is $\mathrm{e}^{\sqrt{n \ln n}(1+o(1))}$. He also considered the operations of star, concatenation, and complementation on unary UFA languages, and obtained the tight upper bound $(n-1)^2+1$ for star, an upper bound mn for concatenation which is tight if m, n are relatively prime, and a lower bound $n^{2-\epsilon}$ for complementation.

In this paper, we continue this research and study the complexity of basic regular operations on languages represented by unambiguous finite automata. First, we restate the lower bound method from [17,22]. Using the notions of reachable and so called co-reachable states in an NFA N, we assign a matrix M_N to the NFA N in such a way that the rank of M_N provides a lower bound on the number of states in any UFA for the language $L(N)$. We use this to get all our lower bounds. To get upper bounds, we first construct an NFA for the language resulting from an operation, and then we apply the (incomplete) subset construction to this NFA to get an incomplete DFA, so also UFA, for the resulting language.

As a result, we get tight upper bounds for intersection (mn), left and right quotients ($2^n - 1$), positive closure ($3/4 \cdot 2^n - 1$), star ($3/4 \cdot 2^n$), shuffle ($2^{mn} - 1$), and concatenation ($3/4 \cdot 2^{m+n} - 1$). To prove tightness, we always use small fixed alphabets. Since the reverse of an ambiguous finite automata is unambiguous, the complexity of a language and its reverse is the same. Finally, we get some partial results for complementation and union.

2 Preliminaries

We assume that the reader is familiar with basic notions in formal languages and automata theory. For details and all the unexplained notions, the reader may refer to [10,23,26].

Let Σ be a finite alphabet of symbols. Then Σ^* denotes the set of strings over the alphabet Σ including the empty string ε. The length of a string w is denoted by $|w|$, and the number of occurrences of a symbol a in a string w by $\#_a(w)$. A language is any subset of Σ^*. For a finite set X, the cardinality of X is denoted by $|X|$, and its power-set by 2^X.

A *nondeterministic finite automaton* (NFA) is a 5-tuple $N = (Q, \Sigma, \Delta, I, F)$, where Q is a finite nonempty set of states, Σ is a finite nonempty input alphabet, $\Delta \subseteq Q \times \Sigma \times Q$ is the transition relation, $I \subseteq Q$ is the set of initial states, and $F \subseteq Q$ is the set of final states. Each element (p, a, q) of Δ is called a *transition* of N. A *computation* of N on an input string $a_1 \cdots a_n$ is a seguence of transitions $(q_0, a_1, q_1)(q_1, a_2, q_2) \cdots (q_{n-1}, a_n, q_n) \in \Delta^*$. The computation is *accepting* if $q_0 \in I$ and $q_n \in F$; in such a case we say that the string $a_1 \cdots a_n$ is accepted by N. The *language accepted by* the NFA N is the set of strings $L(N) = \{w \in \Sigma^* \mid w \text{ is accepted by } N\}$.

An NFA $N = (Q, \Sigma, \Delta, I, F)$ is *unambiguous* (UFA) if it has at most one accepting computation on every input string, and it is *deterministic* (DFA) if $|I| = 1$ and for each state p in Q and each symbol a in Σ, there is at most one state q in Q such that (p, a, q) is a transition of N. Let us emphasize that we allow NFAs to have multiple initial states, and DFAs to be incomplete.

The transition relation Δ may be viewed as a function from $Q \times \Sigma$ to 2^Q, which can be extended to the domain $2^Q \times \Sigma^*$ in the natural way. We denote this function by $\cdot$. Using this notation we get $L(N) = \{w \in \Sigma^* \mid I \cdot w \cap F \neq \emptyset\}$.

Every NFA $N = (Q, \Sigma, \cdot, I, F)$ can be converted to an equivalent incomplete DFA $N' = (2^Q \setminus \{\emptyset\}, \Sigma, \cdot', I, F')$, where $F' = \{R \in 2^Q \setminus \{\emptyset\} \mid R \cap F \neq \emptyset\}$, and for each R in $2^Q \setminus \{\emptyset\}$ and each a in Σ, the partial transition function $\cdot'$ is defined as follows: $R \cdot' a = R \cdot a$ if $R \cdot a \neq \emptyset$ and $R \cdot' a$ is undefined otherwise. We call the DFA N' the *incomplete subset automaton* of NFA N. Since every incomplete DFA is a UFA, we get the following observation.

Proposition 1. *If a language L is accepted by an n-state* NFA*, then L is accepted by a* UFA *of at most $2^n - 1$ states.* $\square$

The *reverse* w^R of a string w is defined by $\varepsilon^R = \varepsilon$ and $(va)^R = av^R$ where $a \in \Sigma$ and $v \in \Sigma^*$. The *reverse of a language* L is the language L^R defined by $L^R = \{w^R \mid w \in L\}$. The *reverse of an automaton* $N = (Q, \Sigma, \cdot, I, F)$ is the NFA N^R obtained from N by swapping the role of initial and final states and by reversing all the transitions. Formally, we have $N^R = (Q, \Sigma, \cdot^R, F, I)$, where $q \cdot^R a = \{p \in Q \mid q \in p \cdot a\}$ for each state q in Q and each symbol a in Σ. The NFA N^R accepts the reverse of the language $L(N)$.

Let $N = (Q, \Sigma, \cdot, I, F)$ be an NFA. We say that a set S is *reachable* in N if there is a string w in Σ^* such that $S = I \cdot w$. Next, we say that a set T is co-reachable in N if T is reachable in N^R. In what follows we are interesting in *non-empty* reachable and co-reachable sets, and we use the following notation:

$$\mathcal{R} = \{S \subseteq Q \mid S \text{ is reachable in } N \text{ and } S \neq \emptyset\}, \tag{1}$$

$$\mathcal{C} = \{T \subseteq Q \mid S \text{ is co-reachable in } N \text{ and } T \neq \emptyset\}. \tag{2}$$

The next observation uses the notions of reachable and co-reachable sets in an NFA to get a characterization of unambiguous automata.

Proposition 2. *Let $\mathcal{R}$ and $\mathcal{C}$ be the families of non-empty reachable and co-reachable sets in an* NFA *N. Then N is unambiguous if and only if $|S \cap T| \leq 1$ for each S in $\mathcal{R}$ and each T in $\mathcal{C}$.*

Proof. (If) Assume that $|S \cap T| \leq 1$ for each S in $\mathcal{R}$ and each T in $\mathcal{C}$. Suppose for a contradiction, that there is a string w, on which N has two distinct accepting computations. Let a be the first symbol in w such that after reading a, the two corresponding states in these computations are distinct; denote the two states by p and q, respectively. Then $w = uav$ for some strings u and v. Let $S = I \cdot ua$ and $T = F \cdot^R v^R$. Then $S \in \mathcal{R}$, $T \in \mathcal{C}$, and $S \cap T \supseteq \{p, q\}$, a contradiction.

(Only if) For a contradiction, suppose that N is unambiguous and there are two distinct states p, q in $S \cap T$. Let S be reachable by u and T be co-reachable by v. Then there are two distinct accepting computations on uv^R: $s_1 \xrightarrow{u} p \xrightarrow{v^R} f_1$ and $s_2 \xrightarrow{u} q \xrightarrow{v^R} f_2$ for some states s_1, s_2 in I and f_1, f_2 in F, a contradiction. $\square$

If N^R is deterministic, then each co-reachable set in N is of size one, and we get the following result.

Corollary 3. *Let N be an* NFA*. If N^R is deterministic, then N is unambiguous.* $\square$

Recall that the *state complexity* of a regular language L, $\mathrm{sc}(L)$, is the smallest number of states in any complete DFA accepting the language L. The state complexity of a regular operation is the maximal state complexity of languages resulting from the operation, considered as the function of state complexities of the arguments. The *nondeterministic state complexity* of languages and operations is defined analogously using NFA representation of languages. We define the *unambiguous state complexity* of a regular language L, $\mathrm{usc}(L)$, as the smallest number of states in any UFA for L.

To prove that a DFA is minimal, we only need to show that all its states are reachable from the initial state, and that no two distinct states are equivalent. To prove minimality of NFAs, a fooling set lower bound method may be used [2,8]. To prove a lower bound for the size of a UFA, a method based on ranks of certain matrices was developed by Schmidt [22, Theorem 3.9], Leung [17, Theorem 2] and Hromkovič et al. [12]. We use it in the following statement.

Proposition 4 ([12,17,22]). *Let L be accepted by an NFA N. Let $\mathcal{R}$ and $\mathcal{C}$ be the families of non-empty reachable and co-reachable sets in N, respectively. Let M_N be the matrix in which the rows are indexed by sets in $\mathcal{R}$, the columns are indexed by sets in $\mathcal{C}$, and in the entry (S,T), we have 0/1 if S and T are/ are not disjoint. Then* $\mathrm{usc}(L) \ge \mathrm{rank}(M_N)$.

Proof. Let A be a minimal UFA accepting L. Consider a matrix M'_A, in which rows are indexed by the states of A and columns are indexed by strings generating the co-reachable sets in $\mathcal{C}$. The entry (q,w) is 1 if w^R is accepted by A from q, and it is 0 otherwise. Then every row of M_N is a sum of the rows of M'_A corresponding to the states in S: Notice that since A is a UFA, for every column there is at most one such row that contains a 1. Thus every row of M_N is a linear combination of rows in M'_A, and therefore $\mathrm{rank}(M_N) \le \mathrm{rank}(M'_A) \le \mathrm{usc}(L)$. □

Throughout our paper, we use the following observation from [16] and its corollary stated in the proposition below.

Lemma 5 ([16, Lemma 3]). *Let $|Q| = n$ and M_n be a $2^n - 1 \times 2^n - 1$ matrix over the field with characteristic 2 with rows and columns indexed by a non-empty subsets of Q such that $M_n(S,T) = 1$ if $S \cap T \neq \emptyset$ and $M_n(S,T) = 0$ otherwise. Then the rank of M_n is $2^n - 1$.* □

Proposition 6. *Let L be accepted by an* NFA *N. Let $\mathcal{R}$ be the family of all non-empty reachable sets in N. If each non-empty set is co-reachable in* NFA *N, then* $\mathrm{usc}(L) \ge |\mathcal{R}|$.

Proof. Consider the matrix M_N given by Proposition 4. Notice that M_N contains $|\mathcal{R}|$ rows of the matrix M_n given in Lemma 5. By Lemma 5, the rank of M_n is $2^n - 1$, so the rows of M_n are linearly independent. Therefore all the rows of M_N must be linearly independent, and we have $\mathrm{rank}(M_N) = |\mathcal{R}|$. Hence $\mathrm{usc}(L) \ge \mathrm{rank}(M_N) = |\mathcal{R}|$ by Proposition 4. □

3 Operations on Unambiguous Finite Automata

We start with the reversal and intersection operations. Then we continue with left and right quotients. Notice that if an NFA N is unambiguous then also N^R is unambiguous. Hence we get the following result.

Theorem 7 (Reversal). *Let L be a regular language. Then* $\text{usc}(L^R) = \text{usc}(L)$.

Proof. If L is accepted by an unambiguous NFA N, then L^R is accepted by N^R which is unambiguous as well, and has the same number of states as N. Hence $\text{usc}(L^R) \le \text{usc}(L)$. On the other hand, the language L^R cannot be accepted by any smaller UFA because otherwise the language $L = (L^R)^R$ would be accepted by a smaller UFA as well. □

Theorem 8 (Intersection). *Let K and L be languages over Σ with* $\text{usc}(K) = m$ *and* $\text{usc}(L) = n$. *Then* $\text{usc}(K \cap L) \le mn$, *and the bound is tight if* $|\Sigma| \ge 2$.

Proof. To get the upper bound, let K and L be accepted m-state and n-state UFAs $A = (Q_A, \Sigma, \cdot_A, I_A, F_A)$ and $B = (Q_B, \Sigma, \cdot_B, I_B, F_B)$, respectively. Construct the product automaton $N = (Q_A \times Q_B, \Sigma, \cdot, I_A \times I_B, F_A \times F_B)$, where $(p, q) \cdot a = p \cdot_A a \times q \cdot_B a$. Then N is an mn-state UFA for $K \cap L$.

To prove tightness, consider languages $K = \{w \in \{a, b\}^* \mid \#_a(w) = m - 1\}$, and $L = \{w \in \{a, b\}^* \mid \#_b(w) = n - 1\}$, which are accepted by UFAs $A = (\{0, 1, \ldots, m-1\}, \{a, b\}, \cdot_A, \{0\}, \{m-1\})$, where $i \cdot_A a = i + 1$ if $i \le m - 2$ and $i \cdot_A b = i$ for each state i, and $B = (\{0, 1, \ldots, n-1\}, \{a, b\}, \cdot_B, \{0\}, \{n-1\})$, where $j \cdot_B a = j$ for each j and $j \cdot_B b = j + 1$ if $j \le n - 2$, respectively. Construct the product automaton N as described above, and notice that each singleton set $\{(i, j)\}$ is reachable and co-reachable in N. Hence M_N is an $mn \times mn$ identity matrix, and the theorem follows by Proposition 4. □

The *left quotient* of a language L by a string w is $w \backslash L = \{x \mid wx \in L\}$, and the left quotient of a language L by a language K is the language $K \backslash L = \bigcup_{w \in K} w \backslash L$. The state complexity of the left quotient operation is $2^n - 1$ [27], and its nondeterministic state complexity is $n+1$ [14]. In both cases, the witness languages are defined over a binary alphabet. Our next result shows that the tight upper bound for UFAs is $2^n - 1$. To prove tightness we use a binary alphabet.

Theorem 9 (Left Quotient). *Let $K, L \subseteq \Sigma^*$,* $\text{usc}(K) = m$, *and* $\text{usc}(L) = n$. *Then* $\text{usc}(K \setminus L) \le 2^n - 1$, *and the bound is tight if* $|\Sigma| \ge 2$.

Proof. To get an upper bound, let A be an n-state UFA for L. Construct an n-state NFA N for $K \setminus L$ from A by making initial all states of A that are reachable from the initial set by some string in K. By Proposition 1, $\text{usc}(K \setminus L) \le 2^n - 1$.

For tightness, let $K = \{a^k \mid k \ge m - 1\}$ and L be the language accepted by the n-state DFA $A = (\{0, 1, \ldots, n-1\}, \{a, b\}, \{0\}, \{0, 1, \ldots, n-1\})$ shown in Fig. 1. Notice that each state of A is reachable by some string in K. Construct an n-state NFA N for $K \setminus L$ from A by making all the states initial. Hence the initial set of N is $\{0, 1, \ldots, n-1\}$. Next, we can shift every reachable subset right

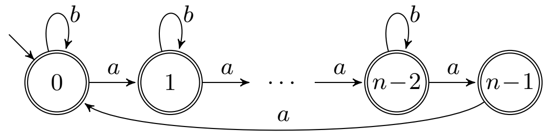

Fig. 1. The UFA of a language L with $\mathrm{usc}(K \setminus L) = 2^n - 1$, where $K = a^{\geq m-1}$.

by one (modulo n) by reading a, and we can remove the state n from any subset containing state n by reading b. Therefore each non-empty set is reachable in N.

To construct N^R, we only need to reverse the transitions on a in N. The initial subset of N^R is $\{0, 1, \ldots, n-1\}$, and we can again shift any subset and remove one state as before. It follows that every non-empty set is reachable in N^R, that is, co-reachable in NFA N. By Proposition 6, we have $\mathrm{usc}(K \setminus L) \geq 2^n - 1$. □

The *right quotient* of a language L by a string w is $L/w = \{x \mid x\,w \in L\}$, and the right quotient of a language L by a language K is $L/K = \bigcup_{w \in K} L/w$. If a language L is accepted by an n-state DFA or NFA A, then the language L/K is accepted by an automaton that is exactly the same as A, except for the set of final states that consists of all states of A, from which some string in K is accepted by A [27]. Thus $\mathrm{sc}(L/K) \leq n$ and $\mathrm{nsc}(L/K) \leq n$. The tightness of the first upper bound has been shown using binary languages in [27]. The second upper bound is met by unary languages $a^{\geq m-1}$ and $a^{\leq n-1}$. Our next aim is to show that the tight upper bound for unambiguous finite automata is $2^n - 1$, with witnesses defined over a binary alphabet.

Theorem 10 (Right Quotient). *Let* $K, L \subseteq \Sigma^*$, $\mathrm{usc}(K) = m$, *and* $\mathrm{usc}(L) = n$. *Then* $\mathrm{usc}(L \,/\, K) \leq 2^n - 1$, *and the bound is tight if* $|\Sigma| \geq 2$.

Proof. To get an upper bound, let A be an n-state UFA for L. Construct an n-state NFA for $L \,/\, K$ as described above. By Proposition 1, $\mathrm{usc}(L \,/\, K) \leq 2^n - 1$.

To prove tightness, let $K = \{a^k \mid k \geq m-1\}$ and L be the language accepted by the n-state NFA $A = (\{0, 1, \ldots, n-1\}, \{a, b\}, \{0, 1, \ldots, n-1\}, \{n-1\})$ shown in Fig. 2. Since the automaton A^R is deterministic, the NFA A is unambiguous by Corollary 3. Since a string in K is accepted by A from each state of A, we construct an NFA N for $L \,/\, K$ from A by making all the states of A final. Notice that we obtain the same NFA as in the proof of the previous lemma, thus by the same arguments $\mathrm{usc}(L \,/\, K) \geq 2^n - 1$. □

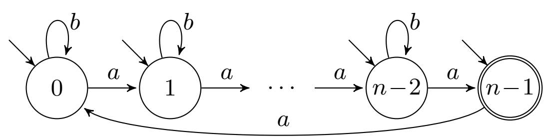

Fig. 2. The UFA of a language L with $\mathrm{usc}(L \,/\, K) = 2^n - 1$, where $K = a^{\geq m-1}$.

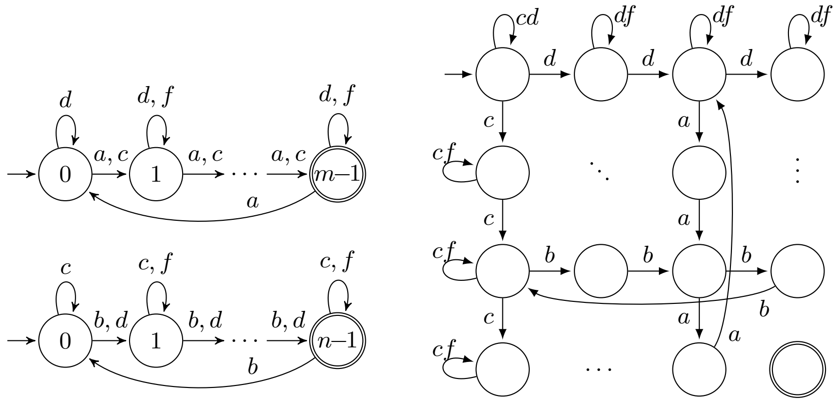

Fig. 3. Witness UFAs for shuffle meeting the upper bound $2^{mn} - 1$ (left) and a sketch of the resulting NFA N.

Now let us continue with the shuffle and concatenation operations. The *shuffle* of two strings u and v over an alphabet Σ is defined as the set of strings $u \sqcup\!\sqcup\, v = \{u_1 v_1 \cdots u_k v_k \mid u = u_1 \cdots u_k, v = v_1 \cdots v_k, u_1, \ldots, u_k, v_1, \ldots, v_k \in \Sigma^*\}$. The shuffle of languages K and L over Σ is defined by $K \sqcup\!\sqcup\, L = \bigcup_{u \in K, v \in L} u \sqcup\!\sqcup\, v$. The state complexity of the shuffle operation on languages represented by incomplete deterministic automata was studied by Câmpeanu et al. [3]. They proved that $2^{mn} - 1$ is a tight upper bound for that case. Here we show that the same upper bound is tight also for UFAs, and to prove tightness, we use almost the same languages as in [3, Theorem 1]. Up to our best knowledge, the problem is still open for complete deterministic automata.

Theorem 11 (Shuffle). *Let* $K, L \subseteq \Sigma^*$, $\operatorname{usc}(K) = m$, *and* $\operatorname{usc}(L) = n$. *Then* $\operatorname{usc}(L \sqcup\!\sqcup\, K) \leq 2^{mn} - 1$, *and the bound is tight if* $|\Sigma| \geq 5$.

Proof. Let $A = (Q_A, \Sigma, \cdot_A, I_A, F_A)$ and $B = (Q_B, \Sigma, \cdot_B, I_B, F_B)$ be m- and n-state UFAs for K and L respectively. Then $K \sqcup\!\sqcup\, L$ is accepted by an mn-state NFA $N = (Q_A \times Q_B, \Sigma, \cdot, I_A \times I_B, F_A \times F_B)$, where for each state (p, q) in $Q_A \times Q_B$ and each symbol a in Σ, we have $(p, q) \cdot a = (p \cdot_A a \times \{q\}) \cup (\{p\} \times q \cdot_B a)$. Hence $\operatorname{usc}(K \sqcup\!\sqcup\, L) \leq 2^{mn} - 1$ by Proposition 1.

To prove tightness, let $\Sigma = \{a, b, c, d, f\}$. Let K and L be the regular languages accepted by DFAs $A = (\{0, 1, \ldots, m-1\}, \Sigma, \cdot_A, \{0\}, \{m-1\})$ and $B = (\{0, 1, \ldots, n-1\}, \Sigma, \cdot_B, \{0\}, \{n-1\})$ shown in Fig. 3 (left); notice that these DFAs are the same as in [3, Theorem 1] up to the position of final states. Construct an NFA N for $K \sqcup\!\sqcup\, L$ as described above. Fig. 3 (right) shows a sketch of the resulting NFA. It is shown in [3] that each non-empty set is reachable in N: The initial set $\{(0, 0)\}$ goes to the "full" set $\{0, 1, \ldots, m-1\} \times \{0, 1, \ldots, n-1\}$ by $c^m d^n$, and for each subset S with $(i, j) \in S$, we have $S \cdot a^{m-i} b^{n-j} f a^i b^j = S \setminus \{(i, j)\}$. Next, in N^R we have $\{(m-1, n-1)\} \cdot^R c^m d^n = \{0, 1, \ldots, m-1\} \times \{0, 1, \ldots, n-1\}$, and $S \cdot^R a^i b^j f a^{m-i} b^{n-j} = S \setminus \{(i, j)\}$ for each subset S with $(i, j) \in S$. It follows that each non-empty set is co-reachable in N, so $\operatorname{usc}(L) \geq 2^{mn} - 1$. $\square$

The *concatenation* of languages K and L is $KL = \{uv \mid u \in K \text{ and } v \in L\}$. The state complexity of concatenation is $m2^n - 2^{n-1}$, and its nondeterministic state complexity is $m + n$. In both cases, the witnesses are defined over a binary alphabet [9,14,19,27]. In the next theorem we get a tight upper bound for concatenation on UFAs. To prove tightness, we use a seven-letter alphabet.

Theorem 12 (Concatenation). *Let $K, L \subseteq \Sigma^*$, $\text{usc}(K) = m$, and $\text{usc}(L) = n$, where $m, n \geq 2$. Then $\text{usc}(KL) \leq 3/4 \cdot 2^{m+n} - 1$, and the bound is tight if $|\Sigma| \geq 6$.*

Proof. Let $A = (Q_A, \Sigma, \cdot_A, I_A, F_A)$ and $B = (Q_B, \Sigma, \cdot_B, I_B, F_B)$ be UFAs for languages K and L, respectively. Let $|Q_A| = m$, $|F_A| = k$, $|Q_B| = n$, $|I_B| = \ell$. Construct an NFA $N = (Q_A \cup Q_B, \Sigma, \cdot, I, F_B)$ for KL, where for each q in $Q_A \cup Q_B$ and each a in Σ,

$$q \cdot a = \begin{cases} q \cdot_A a, & \text{if } q \in Q_A \text{ and } q \cdot_A a \cap F_A = \emptyset; \\ q \cdot_A a \cup I_B, & \text{if } q \in Q_A \text{ and } q \cdot_A a \cap F_A \neq \emptyset; \\ q \cdot_B a, & \text{if } q \in Q_B, \end{cases}$$

and

$$I = \begin{cases} I_A, & \text{if } I_A \cap F_A = \emptyset; \\ I_A \cup I_B, & \text{otherwise.} \end{cases}$$

Notice that if a set S is reachable in the NFA N and $S \cap F_A \neq \emptyset$, then $I_B \subseteq S$. It follows that the number of reachable sets is $2^{m-k}2^n + (2^m - 2^{m-k})2^{n-\ell}$, which is maximal if $\ell = 1$. In such a case, this number equals $(2^m + 2^{m-k})2^{n-1}$, which is maximal if $k = 1$. After excluding the empty set, we get the upper bound.

For tightness, let K and L be languages over $\{a, b, c, d, \alpha, \beta, \gamma\}$ accepted by automata A and B shown in Fig. 4, where $Q_A = \{q_0, q_1, \ldots, q_{m-1}\}$ and $Q_B = \{0, 1, \ldots, n-1\}$. Notice that A^R and B are deterministic, so A and B are unambiguous. Construct an NFA N for KL from automata A and B as described above, that is, by adding transitions on a, b, c from state q_{m-2} to state 0, and by adding transitions on α, β, γ from state q_{m-1} to state 0. The initial state of N is q_0 and the unique final state is $n - 1$.

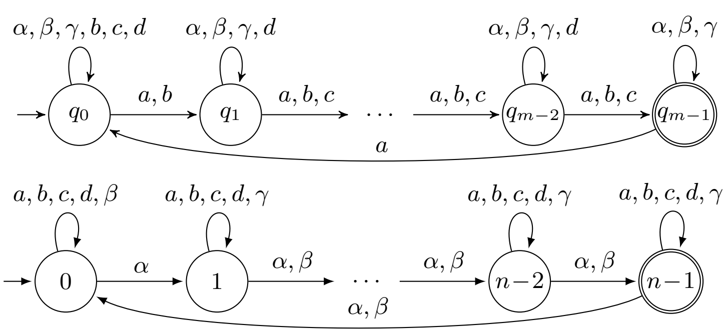

Fig. 4. Witness UFAs for concatenation meeting the upper bound $3/4 \cdot 2^{m+n} - 1$.

First, let us show that the family of all the non-empty sets that are reachable in N contains each subset $S \cup T$, where $S \subseteq Q_A$, $T \subseteq Q_B$ and if $q_{m-1} \in S$ then $0 \in T$. Consider several cases:

(1) Let $S = \{q_{m-1}\} \cup T$, where $T \subseteq Q_B$ and $0 \in T$. The set $\{q_{m-1}, 0\}$ is reached from the initial set $\{q_0\}$ by a^{m-1}. Next, each set $\{q_{m-1}, 0, j_2, \ldots, j_k\}$ of size $k+1$ is reached from the set $\{q_{m-1}, 0, j_3 - j_2 + 1, \ldots, j_k - j_2 + 1\}$ of size k by the string $\alpha\beta^{j_2-1}$. This proves case (1) by induction.

(2) Let $S \subseteq Q_A$, $T \subseteq Q_B$, and $0 \in T$. The set $\{q_0\} \cup T$ is reached by a from the set $\{q_{m-1}\} \cup T$ which is considered in case (1). Next, in the NFA A, each subset S of Q_A with $q_0 \notin S$ is reached from a set S' of the same size and with $0 \in S'$ by a string in a^*, and each set such set S' is reached from a smaller set by a string in cb^*. This proves case (2) by induction since there is a loop on a, b, c in each state of T, and we have $0 \in T$.

(3) Let $S \subseteq Q_A$ and $q_{m-1} \notin S$. Let $T \subseteq Q_B$ and $0 \notin T$. Then the set $S \cup T$ is reached by γ from the set $S \cup T \cup \{0\}$ which is considered in case (2). It follows that N has $3/4 \cdot 2^{m+n} - 1$ non-empty reachable sets.

Now let us show that each non-empty set is co-reachable in N. First, notice that in B^R, we can shift each set cyclically by using a string in α^*, and we can eliminate state 0 from each set containing 0 using γ. This means that each subset of Q_B is reached from Q_B in B^R. Similarly, each string in a^* performs a cyclic shift of any set in A^R. Next, we can eliminate a state q_i from any set S containing the state q_i and a state q_j with $j > i$: We shift S cyclically so that state q_i is moved to q_0, then we use a string in c^+ to "merge" original states q_i and q_j. After that reading a string in a^* results in eliminating the state q_i from the original set S. Therefore, each non-empty subset of Q_A is reachable from Q_A in A^R. The empty set is reached from $\{q_1\}$ by c in A^R. It follows that for each $S \subseteq Q_A$ with $q_{m-1} \in S$ and each $T \subseteq Q_B$, there exist a string u_S over $\{a, c\}$ and a string v_T over $\{\alpha, \gamma\}$ such that in N^R, we have

$$\to \{n-1\} \xrightarrow{\alpha^{n-1}a^m\alpha} Q_A \cup \{n-1\} \xrightarrow{u_S} S \cup \{n-1\} \xrightarrow{\alpha^{n-1}\beta^n} S \cup Q_B \xrightarrow{u_T} S \cup T;$$

recall that there is a loop on a, b, c in each state of Q_B, and there is a loop on α, β, γ in each state of Q_A. Next if $q_{m-1} \notin S$, then for each $T \subseteq Q_B$ the set $S \cup T$ is reached from $S \cup \{q_{m-1}\} \cup T$ by d since we have a loop on d in each state of A and B, except for state q_{m-1}. Hence each non-empty set is co-reachable in NFA N, and our proof is complete. □

Now we consider the Kleene closure (star) and positive closure operations. For a language L, the *star* of L is the language $L^* = \bigcup_{i \geq 0} L^i$, where $L^0 = \{\varepsilon\}$ and $L^{i+1} = L^i L$. The *positive closure* of L is $L^+ = \bigcup_{i \geq 1} L^i$. The state complexity of the star operation is $3/4 \cdot 2^n$ with binary witness languages [19,27]. In the unary case, the tight upper bound is $(n-1)^2 + 1$ [4,27]. The nondeterministic state complexity of star is $n+1$, with witnesses defined over a unary alphabet [9]. We first consider the positive closure of UFA languages, and we get a tight upper bound $3/4 \cdot 2^n - 1$ on its complexity. Our worst-case example is defined over a ternary alphabet.

Theorem 13 (Positive Closure). *Let L be a language over Σ with* $\mathrm{usc}(L) = n$*, where $n \geq 2$. Then* $\mathrm{usc}(L^+) \leq 3/4 \cdot 2^n - 1$*, and the bound is tight if $|\Sigma| \geq 3$.*

Proof. To get an upper bound, let $A = (Q, \Sigma, \cdot, I, F)$ be an n-state UFA for L. Construct an NFA $N = (Q, \Sigma, \cdot^+, I, F)$ for L^+ where the transition function $\cdot^+$ is defined as

$$q \cdot^+ a = \begin{cases} q \cdot a \cup I, & \text{if } q \cdot a \cap F \neq \emptyset; \\ q \cdot a, & \text{otherwise} \end{cases}$$

for each state q in Q and each symbol a in Σ. Notice that if a set S is reachable in N and $S \cap F \neq \emptyset$, then $I \subseteq S$. Hence a set containing some final state of A and not containing some initial state of A cannot be reachable in N. Denote this family of unreachable sets by $Unreach$. Let us count the number of such sets. To this aim, denote $|I \setminus F| = i$, $|I \cap F| = j$, and $|F \setminus I| = k$. Since A is unambiguous, we must have $j \leq 1$. Consider three cases:

(1) Let $j = 0$. Then we must have $i, k \geq 1$ since A has at least one initial and at least one final state. A set in $Unreach$ contains a non-empty subset of F, a non-universal subset of I, and any subset of the remaining states. Thus we have

$$|Unreach| = (2^i - 1)(2^k - 1)2^{n-i-k} = (1 - \frac{1}{2^i})(1 - \frac{1}{2^k})2^n \geq 1/4 \cdot 2^n.$$

(2) Let $j = 1$ and $i = k = 0$. Then A has one initial state, which is also the only final state of A. Then $L = L^+$, and so $\mathrm{usc}(L^+) = \mathrm{usc}(L)$.

(3) Let $j = 1$ and let at least one of i, k be positive. A set in $Unreach$ which contains the unique initial and final state must contain a non-universal subset of $I \setminus F$. A set which does not contain this state must contain a non-empty subset of $F \setminus I$. Thus we have $|Unreach| =$

$$\big((2^i - 1)2^k + 2^i(2^k - 1)\big)\, 2^{n-i-k-1} = \big(1 - \frac{1}{2^{k+1}} - \frac{1}{2^{i+1}}\big)\, 2^{n-1} \geq 1/4 \cdot 2^n.$$

In case (2), we have $n \leq 3/4 \cdot 2^n - 1$ since $n \geq 2$. Next, in cases (1) and (3), at least $1/4 \cdot 2^n$ subsets are unreachable. The empty subset is not in $Unreach$. It follows that there are at most $3/4 \cdot 2^n - 1$ reachable non-empty subsets in N. This proves the upper bound.

To prove tightness, let L be the language accepted by the ternary DFA A shown in Fig. 5 (top). Construct the NFA N for L^+ as described above. Notice that the DFA A restricted to the alphabet $\{a, b\}$ is the same as the witness DFA for the star operation from [27, Theorem 3.3, Fig. 4] In particular, this means that N has $3/4 \cdot 2^n - 1$ non-empty reachable subsets.

Let us show that each non-empty set is co-reachable in N. To this aim, we use only the transitions on a and c. Fig. 5 (bottom) shows these transitions in N^R. Notice, that we can shift each subset S of $\{0, 1, \ldots, n-1\}$ cyclically by one to the set $\{(s-1) \bmod n \mid s \in S\}$: we use a if $0 \notin S$ or if both 0 and $n-1$ are in S, and we use ac otherwise. Next, we can eliminate a state i from any set containing states i and $i+1$. It follows that each non-empty set is reachable from $\{0, 1, \ldots, n-1\}$ in N^R. To conclude the proof, notice that the initial set $\{n-1\}$ of N^R goes to $\{0, 1, \ldots, n-1\}$ by the string a^{n^2}. $\square$

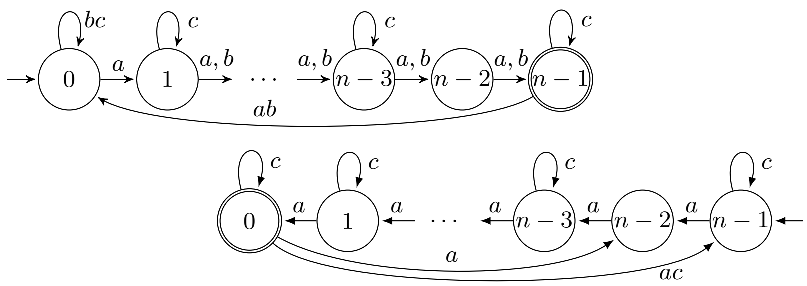

Fig. 5. The witness UFA for positive closure meeting the upper bound $3/4 \cdot 2^n - 1$ (top), and the transitions on a, c in the NFA N^R (bottom).

Finally, let us show that the tight upper bound for star on UFAs is $3/4 \cdot 2^n$, and that it is met by the language accepted by the UFA shown in Fig. 5 (top).

Theorem 14 (Star). *Let L be a regular language over Σ with $\operatorname{usc}(L) = n$, where $n \geq 2$. Then $\operatorname{usc}(L^*) \leq 3/4 \cdot 2^n$, and the bound is tight if $|\Sigma| \geq 3$.*

Proof. The upper bound follows from the previous theorem since if $\varepsilon \in L$, then $L^+ = L^*$, and otherwise we only need to add one more initial and final state to the UFA for L^+ to accept the empty string. The resulting automaton is unambiguous since the new state accepts only the empty string which is not accepted by UFA for L^+.

For tightness, consider the language L accepted by the UFA A shown in Fig. 5 (top). Construct an NFA N for L^* from UFA A by adding a new initial and final state q_0, and by adding the transitions on a, b from $n-2$ to 0, and the transition by c from $n-1$ to 0. As shown in [27, Theorem 3.3] the NFA N has $3/4 \cdot 2^n$ reachable sets: the initial set $\{q_0, 0\}$, all the subsets of $\{0, 1, \ldots, n-1\}$ containing state 0, and all the non-empty subsets of $\{1, 2, \ldots, n-2\}$.

Next, consider the NFA N^R. The initial set of N^R is $\{q_0, n-1\}$. Next, as we have shown in the proof of Theorem 13, each non-empty set is reachable in N^R. Now consider the $3/4 \cdot 2^n \times 2^n$ matrix M_N.

	$\{q_0, n-1\}$	$T \neq \{n-1\}$	$T = \{n-1\}$
$\{q_0, 0\}$	1		0
S $n-1 \in S$	1 1 ... 1		1 1 ... 1
S $n-1 \notin S$	0 0 ... 0		0 0 ... 0

Table 1. Matrix M_N for NFA N for the star of the language from Fig. 5 (top).

Assume that the first row is indexed by $\{q_0, 0\}$, then there are rows indexed by reachable sets containing $n-1$, and then by reachable sets not containing $n-1$. Next, assume that the first column is indexed by the initial co-reachable set $\{q_0, n-1\}$, the last column is indexed by $\{n-1\}$, and all the remaining columns are indexed by the remaining co-reachable sets. This is illustrated in Table 1.

Then the rank of the sub-matrix obtained from M_N by removing the first row and the first column is $3/4 \cdot 2^n - 1$. Next, notice that the first and the last columns differ only in the entry in the first row. Hence the first row cannot be expressed as a linear combination of any others rows. Therefore $\mathrm{rank}(M_N) = 3/4 \cdot 2^n$, and the theorem follows by Proposition 4. $\square$

4 Partial Results for Complementation and Union

In this section we present partial results for the complementation and union operations on UFA languages. The complement of a language L over Σ is the language $L^c = \Sigma^* \setminus L$. A language and its complement have the same state complexity since to get a DFA for the complement of L, we only need to interchange the sets of final and non-final states in a DFA for L. For NFAs, the tight upper bound for complementation is 2^n with witnesses defined over a binary alphabet [9,14]. For unary UFAs, the problem was studied by Okhotin who provided a lower bound $n^2 - o(1)$ for complementation of unary UFAs [20, Theorem 6]. In the next theorem we deal with an upper bound. Then we consider union.

Theorem 15 (Complementation: Upper Bound). *Let L be a regular language with $\mathrm{usc}(L) = n$, where $n \geq 7$. Then $\mathrm{usc}(L^c) \leq 2^{0.79n + \log n}$.*

Proof. Let A be an n-state UFA for L and $\mathcal{R}$ and $\mathcal{C}$ be the sets of non-empty reachable and co-reachable sets of A. First, we show that $\mathrm{usc}(L^c) \leq \min\{|\mathcal{R}|, |\mathcal{C}|\}$. We have $\mathrm{usc}(L^c) \leq |\mathcal{R}|$ since we can get a DFA for L^c by applying the subset construction to A and by interchanging the sets of final and non-final states in the resulting DFA that has $|\mathcal{R}|$ reachable states. Next, we have $\mathrm{usc}(L^c) \leq |\mathcal{C}|$ since the NFA A^R is unambiguous, so $\mathrm{usc}((L^R)^c) \leq |\mathcal{C}|$ which means that $\mathrm{usc}(L^c) \leq |\mathcal{C}|$ since complement and reversal commutes and the reverse of a UFA is a UFA.

Next, let $k = \max\{|X| \mid X \in \mathcal{R}\}$, and pick a set S in $\mathcal{R}$ of size k. Then each set in $\mathcal{R}$ has size at most k, and each set in $\mathcal{C}$ may have at most one element in S by Proposition 2. Thus

$$|\mathcal{R}| \leq \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k} \quad \text{and} \quad |\mathcal{C}| \leq (k+1)2^{n-k}.$$

If $k \geq n/2$, then $|\mathcal{C}| \leq (n/2 + 1) \cdot 2^{n/2} \leq 2^{0.5n + \log n}$, and the theorem follows. Now assume that $k < n/2$. Then $|\mathcal{R}| \leq k\binom{n}{k} \leq n(\frac{en}{k})^k$ and $|C| \leq n2^{n-k}$. Let $r(k) = n(\frac{en}{k})^k$ and $c(k) = n2^{n-k}$. Then $r(k)$ increases, while $c(k)$ decreases with k. It follows that if we pick a k_0 such that $k_0 < n/2$, then $usc(L^c) \leq r(k_0)$ if $k \leq k_0$, and $usc(L^c) \leq c(k_0)$ otherwise. By setting $k = nx$ and by solving $(\frac{en}{nx})^{nx} = 2^{n-nx}$, we get $x_0 = 0.2144$, $k_0 = 0.2144n$, $r(k_0) \leq 2^{0.7856n + \log n}$, and $c(k_0) \leq 2^{0.785629n + \log n}$. This completes our proof. $\square$

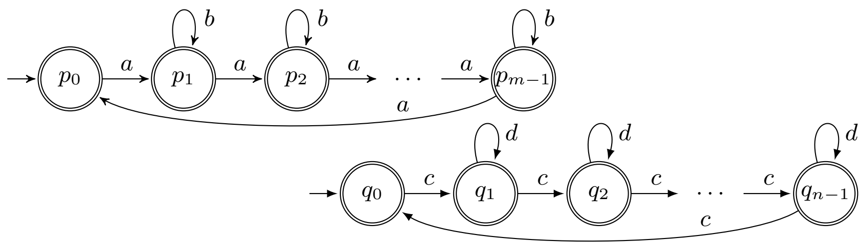

Fig. 6. The UFAs for union meeting the bound $mn + m + n$; the loops on c, d resp. on a, b in each state of A and B, respectively, are not displayed.

Proposition 16 (Union). *Let K and L be languages over Σ with* $\operatorname{usc}(K) = m$ *and* $\operatorname{usc}(L) = n$, *where* $1 \leq m \leq n$. *Then*

(a) $\operatorname{usc}(K \cup L) \leq m + n \cdot \operatorname{usc}(K^c) \leq m + n2^{0.79n + \log n}$;

(b) the bound $mn + m + n$ *is met if* $|\Sigma| \geq 7$.

Proof. (a) The claim follows from the equality $K \cup L = K \dot{\cup} (L \cap K^c)$, where $\dot{\cup}$ denotes a disjoint union, since we have $\operatorname{usc}(L \cap K^c) \leq n \cdot \operatorname{usc}(L^c)$ by Theorem 8, and, moreover, the NFA for a disjoint union of UFAs is unambiguous. The second inequality is given by Theorem 15.

(b) Let K and L be the languages over $\{a, b, c, d\}$ accepted by DFAs A and B, such that we have a loop on c, d in each state of A and a loop on a, b in each state of B; the remaining transitions are shown in Fig. 6. Construct an NFA N for $K \cup L$ from A and B so that the set of initial states is $\{p_0, q_0\}$ and the set of final states is $\{p_0, \ldots, p_{m-1}, q_0, \ldots, q_{n-1}\}$. Then the sets $\{p_i, q_j\}$, $\{p_i\}$, $\{q_j\}$ with $0 \leq i \leq m-1$ and $0 \leq j \leq n-1$ are reachable in N. To conclude the proof notice that each set is co-reachable in N since we can shift any set in A cyclically by reading a and eliminate one state by reading b, while each state of B remains in itself. The same can by done symmetrically with the sets in B. $\square$

5 Conclusions

We investigated the complexity of basic regular operations on languages represented by unambiguous finite automata. Since the reverse of an unambiguous automaton is unambiguous, a language and its reversal have the same complexity for UFAs. Next, we obtained tight upper bounds for intersection (mn), left and right quotients ($2^n - 1$), positive closure ($3/4 \cdot 2^n - 1$), star ($3/4 \cdot 2^n$), shuffle ($2^{mn} - 1$), and concatenation ($3/4 \cdot 2^{m+n} - 1$).

To get upper bounds, we constructed an NFA for the language resulting from an operation, and applied the (incomplete) subset construction to it. For lower bounds, we defined witness languages in such a way that we were able to assign a matrix to a resulting language. The rank of this matrix provided a lower bound on the unambiguous state complexity of the resulting language.

To prove tightness, we used a binary alphabet for intersection and left and right quotients, a ternary alphabet for star and positive closure, a five-letter alphabet for shuffle, and a seven-letter alphabet for concatenation.

For complementation and union, we provided upper bounds $2^{0.79\,n+\log n}$ and $m+n2^{0.79\,n+\log n}$, respectively. Finally, we got a lower bound $mn+m+n$ for union in the quaternary case. The exact complexity of complementation and union on UFAs remains open. All our result are summarized and compared to the known results on the state complexity and nondeterministic state complexity of the considered operations in Table 2.

In the case of complementation, we tried to use a fooling set lower bound method, but we were able to describe a fooling set for the complement of an n-state UFA language only of size $n+\log n$. Moreover, it seems that every such fooling set is of size which is linear in n [7]. Thus the fooling set technique cannot be used to get a larger lower bound. Neither the method based on the rank of matrices can be used here since the matrices of a language and its complement have the same rank, up to one. Therefore, to get a larger lower bound for complementation, some other techniques should be developed.

	sc	$\|\Sigma\|$	nsc	$\|\Sigma\|$	usc	$\|\Sigma\|$
reversal	2^n	2	$n+1$	2	n	1
intersection	mn	2	mn	2	mn	2
left quotient	2^n-1	2	$n+1$	2	2^n-1	2
right quotient	n	1	n	1	2^n-1	2
positive closure	$3/4\cdot 2^n-1$	2	n	1	$3/4\cdot 2^n-1$	3
star	$3/4\cdot 2^n$	2	$n+1$	1	$3/4\cdot 2^n$	3
shuffle	$\geq 2^{(m-1)(n-1)}$	5	mn	2	$2^{mn}-1$	5
concatenation	$(m-1/2)\cdot 2^n$	2	$m+n$	2	$3/4\cdot 2^{m+n}-1$	7
complement	n	1	2^n	2	$\leq 2^{0.79n+\log n}$	
					$\geq n^{2-\epsilon}$	1 [20]
union	mn	2	$m+n+1$	2	$\leq m+n2^{0.79n+\log n}$	
					$\geq mn+m+n$	4

Table 2. The complexity of regular operations: state complexity, nondeterministic state complexity, unambiguous state complexity, and the sizes of alphabets for worst-case examples.

References

1. Allauzen, C., Mohri, M., Rastogi, A.: General algorithms for testing the ambiguity of finite automata and the double-tape ambiguity of finite-state transducers. Int. J. Found. Comput. Sci. 22(4), 883–904 (2011), `http://dx.doi.org/10.1142/S0129054111008477`

2. Birget, J.: Partial orders on words, minimal elements of regular languages and state complexity. Theor. Comput. Sci. 119(2), 267–291 (1993), http://dx.doi.org/10.1016/0304-3975(93)90160-U
3. Câmpeanu, C., Salomaa, K., Yu, S.: Tight lower bound for the state complexity of shuffle of regular languages. Journal of Automata, Languages and Combinatorics 7(3), 303–310 (2002)
4. Čevorová, K.: Kleene star on unary regular languages. In: Jürgensen, H., Reis, R. (eds.) Descriptional Complexity of Formal Systems - 15th International Workshop, DCFS 2013, London, ON, Canada, July 22-25, 2013. Proceedings. Lecture Notes in Computer Science, vol. 8031, pp. 277–288. Springer (2013), http://dx.doi.org/10.1007/978-3-642-39310-5_26
5. Chan, T., Ibarra, O.H.: On the finite-valuedness problem for sequential machines. Theor. Comput. Sci. 23, 95–101 (1983), http://dx.doi.org/10.1016/0304-3975(88)90012-6
6. Colcombet, T.: Unambiguity in automata theory. In: Shallit, J., Okhotin, A. (eds.) Descriptional Complexity of Formal Systems - 17th International Workshop, DCFS 2015, Waterloo, ON, Canada, June 25-27, 2015. Proceedings. Lecture Notes in Computer Science, vol. 9118, pp. 3–18. Springer (2015), http://dx.doi.org/10.1007/978-3-319-19225-3_1
7. Eliáš, P.: Fooling sets for complements of UFA languages. Unpublished manuscript (2016)
8. Glaister, I., Shallit, J.: A lower bound technique for the size of nondeterministic finite automata. Inf. Process. Lett. 59(2), 75–77 (1996), http://dx.doi.org/10.1016/0020-0190(96)00095-6
9. Holzer, M., Kutrib, M.: Nondeterministic descriptional complexity of regular languages. Int. J. Found. Comput. Sci. 14(6), 1087–1102 (2003), http://dx.doi.org/10.1142/S0129054103002199
10. Hopcroft, J.E., Ullman, J.D.: Introduction to Automata Theory, Languages and Computation. Addison-Wesley (1979)
11. Hromkovič, J., Schnitger, G.: Ambiguity and communication. Theory Comput. Syst. 48(3), 517–534 (2011), http://dx.doi.org/10.1007/s00224-010-9277-4
12. Hromkovič, J., Seibert, S., Karhumäki, J., Klauck, H., Schnitger, G.: Communication complexity method for measuring nondeterminism in finite automata. Inf. Comput. 172(2), 202–217 (2002), http://dx.doi.org/10.1006/inco.2001.3069
13. Ibarra, O.H., Ravikumar, B.: On sparseness, ambiguity and other decision problems for acceptors and transducers. In: Monien, B., Vidal-Naquet, G. (eds.) STACS 86, 3rd Annual Symposium on Theoretical Aspects of Computer Science, Orsay, France, January 16-18, 1986, Proceedings. Lecture Notes in Computer Science, vol. 210, pp. 171–179. Springer (1986), http://dx.doi.org/10.1007/3-540-16078-7_74
14. Jirásková, G.: State complexity of some operations on binary regular languages. Theor. Comput. Sci. 330(2), 287–298 (2005), http://dx.doi.org/10.1016/j.tcs.2004.04.011
15. Leiss, E.L.: Succint representation of regular languages by boolean automata. Theor. Comput. Sci. 13, 323–330 (1981), http://dx.doi.org/10.1016/S0304-3975(81)80005-9
16. Leung, H.: Separating exponentially ambiguous finite automata from polynomially ambiguous finite automata. SIAM J. Comput. 27(4), 1073–1082 (1998), http://dx.doi.org/10.1137/S0097539793252092

17. Leung, H.: Descriptional complexity of nfa of different ambiguity. Int. J. Found. Comput. Sci. 16(5), 975–984 (2005), http://dx.doi.org/10.1142/S0129054105003418
18. Lupanov, O.B.: A comparison of two types of finite automata. Problemy Kibernetiki 9, Kibernetiki (1963), (in Russian) German translation: Über den Vergleich zweier Typen endlicher Quellen. Probleme der Kybernetik 6, 328–335 (1966)
19. Maslov, A.N.: Estimates of the number of states of finite automata. Soviet Math. Doklady 11, 13731375 (1970)
20. Okhotin, A.: Unambiguous finite automata over a unary alphabet. Inf. Comput. 212, 15–36 (2012), http://dx.doi.org/10.1016/j.ic.2012.01.003
21. Ravikumar, B., Ibarra, O.H.: Relating the type of ambiguity of finite automata to the succinctness of their representation. SIAM J. Comput. 18(6), 1263–1282 (1989), http://dx.doi.org/10.1137/0218083
22. Schmidt, E.M.: Succinctness of description of context-free, regular, and finite languages. Ph. D. thesis. Cornell University (1978)
23. Sipser, M.: Introduction to the theory of computation. PWS Publishing Company (1997)
24. Stearns, R.E., Hunt, H.B.: On the equivalence and containment problems for unambiguous regular expressions, regular grammars and finite automata. SIAM J. Comput. 14(3), 598–611 (1985), http://dx.doi.org/10.1137/0214044
25. Weber, A., Seidl, H.: On the degree of ambiguity of finite automata. Theor. Comput. Sci. 88(2), 325–349 (1991), http://dx.doi.org/10.1016/0304-3975(91)90381-B
26. Yu, S.: Handbook of Formal Languages: Volume 1 Word, Language, Grammar, chap. Regular Languages, pp. 41–110. Springer Berlin Heidelberg, Berlin, Heidelberg (1997), http://dx.doi.org/10.1007/978-3-642-59136-5_2
27. Yu, S., Zhuang, Q., Salomaa, K.: The state complexities of some basic operations on regular languages. Theor. Comput. Sci. 125(2), 315–328 (1994), http://dx.doi.org/10.1016/0304-3975(92)00011-F