

# State Complexity of Pseudocatenation\*

Lila Kari and Timothy Ng

School of Computer Science, University of Waterloo  
Waterloo, Ontario N2L 3G1, Canada  
{lila.kari, tim.ng}@uwaterloo.ca

**Abstract.** The state complexity of a regular language  $L_m$  is the number  $m$  of states in a minimal deterministic finite automaton (DFA) accepting  $L_m$ . The state complexity of a regularity-preserving binary operation on regular languages is defined as the maximal state complexity of the result of the operation, where the two operands range over all languages of state complexities  $\leq m$  and  $\leq n$ , respectively. We consider the deterministic and nondeterministic state complexity of pseudocatenation. The pseudocatenation of two words  $x$  and  $y$  with respect to an antimorphic involution  $\theta$  is the set  $\{xy, x\theta(y)\}$ . This operation was introduced in the context of DNA computing as the generator of pseudopowers of words (a pseudopower of a word  $u$  is a word in  $u\{u, \theta(u)\}^*$ ). We prove that the state complexity of the pseudocatenation of languages  $L_m$  and  $L_n$ , where  $m, n \geq 3$ , is at most  $(m-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$ . Moreover, for  $m, n \geq 3$  there exist languages  $L_m$  and  $L_n$  over an alphabet of size 4, whose pseudocatenation meets the upper bound. We also prove that the state complexity of the positive pseudocatenation closure of a regular language  $L_n$  has an upper bound of  $2^{2n-1} - 2^n + 1$ , and that this bound can be reached, with the witness being a language over an alphabet of size 4.

## 1 Introduction

In the context of DNA computing, the fact that one can consider a DNA strand and its Watson-Crick complement “equivalent” from the point of view of their information content led to several natural, as well as theoretically interesting, extensions of notions in combinatorics of words and formal language theory such as the pseudo-palindrome [21], pseudo-commutativity [18], or pseudoknot-bordered words [19]. In this context, Watson-Crick complementarity has been modelled mathematically by an antimorphic involution  $\theta$ , i.e., a function that is an antimorphism,  $\theta(uv) = \theta(v)\theta(u)$ ,  $\forall u, v \in \Sigma^*$ , and an involution,  $\theta(\theta(x)) = x$ ,  $\forall x \in \Sigma^*$ . For example, in [10], a word  $w$  is called a  $\theta$ -power or pseudopower if it is of the form  $w \in u\{u, \theta(u)\}^*$ , and the related notions of  $\theta$ -periodicity and  $\theta$ -primitivity can be analogously defined. The static notions of the power and

---

\* This research was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant R2824A01, and a University of Waterloo School of Computer Science Grant to L.K.

period of word are intrinsically connected to the word operation that dynamically generates that power. In the case of the classical notion of power and period of a word, that operation is catenation, and in the case of  $\theta$ -power and  $\theta$ -periodicity, that operation is  $\theta$ -catenation, defined and studied in [17]. Here we continue the investigation of  $\theta$ -catenation, defined by  $x \odot^\theta y = \{xy, x\theta(y)\}$ , by studying its state complexity.

The state complexity of a language operation is a complexity measure based on the number of states of the machine that recognizes the result of the language operation, expressed as a function of the size of the machines recognizing the operand languages. Operational state complexity has been studied since the early 90s and continues to be an active area of research [12, 26]. Recently, there have been several investigations of state complexity for operations modelling biological phenomena, such as hairpin completion [16], inversion [6], duplication [5], and overlap assembly [3].

The state complexity of combinations of operations has also been studied extensively, as many language operations can be expressed as a combination of several basic operations. While one can obtain an upper bound for the state complexity of multiple operations by simply composing the state complexities of each operation, in many cases, the exact state complexity of the combination of operations is much lower than the bound obtained in this fashion [23]. Furthermore, the exact state complexity of a combination of operations is undecidable [24], thus motivating further study in this direction [1, 2, 7–9, 13, 14, 20].

In this paper, we consider the deterministic and nondeterministic state complexity of the pseudocatenation and positive pseudocatenation closure operations with respect to an antimorphism  $\theta$ . We note that for our constructions,  $\theta$  need not be an involution. We fix notation and definitions in Section 2. In Section 3, we consider the state complexity of the pseudocatenation operation. In Section 4, we consider the positive closure of a language with respect to pseudocatenation. We conclude in Section 5.

## 2 Preliminaries

Let  $\Sigma$  be a finite alphabet. We denote by  $\Sigma^*$  the set of all finite words over  $\Sigma$ , including the empty word, which we denote by  $\varepsilon$ . We denote the length of a word  $w = a_1a_2 \cdots a_n$  by  $|w| = n$ . The reversal of a word  $w = a_1a_2 \cdots a_n$  is denoted by  $w^R = a_n \cdots a_2a_1$ . If  $w = xyz$ , then we say that  $x$  is a prefix of  $w$ ,  $y$  is a factor or subword of  $w$ , and  $z$  is a suffix of  $w$ . For a word  $u \in \Sigma^*$ , we denote the number of occurrences of  $u$  as a factor of  $w$  by  $|w|_u$ .

A deterministic finite automaton (DFA) is a tuple  $A = (Q, \Sigma, \delta, s, F)$  where  $Q$  is a finite set of states,  $\Sigma$  is an alphabet,  $\delta$  is a function  $\delta : Q \times \Sigma \rightarrow Q$ ,  $s \in Q$  is the initial state, and  $F \subset Q$  is a set of final states. We extend the transition function  $\delta$  to a function  $Q \times \Sigma^* \rightarrow Q$  in the usual way. A DFA  $A$  is complete if  $\delta$  is defined for all  $q \in Q$  and  $a \in \Sigma$ . We will also make use of the notation  $q \xrightarrow{w} q'$  for  $\delta(q, w) = q'$  whenever convenient.

A word  $w \in \Sigma^*$  is accepted by  $A$  if  $\delta(s, w) \in F$ . The language recognized by  $A$  is  $L(A) = \{w \in \Sigma^* \mid \delta(s, w) \in F\}$ . A state  $q$  is reachable if there exists a string  $w \in \Sigma^*$  such that  $\delta(s, w) = q$ . Two states  $p$  and  $q$  of  $A$  are equivalent if  $\delta(p, w) \in F$  if and only if  $\delta(q, w) \in F$  for every word  $w \in \Sigma^*$ . A DFA  $A$  is minimal if each state  $q \in Q$  is reachable from the initial state and no two states are equivalent. The state complexity of a regular language  $L$ , denoted  $\text{sc}(L)$  is the number of states of the minimal complete DFA recognizing  $L$  [25].

A nondeterministic finite automaton (NFA) is a tuple  $A = (Q, \Sigma, \delta, I, F)$  where  $Q$  is a finite set of states,  $\Sigma$  is an alphabet,  $\delta$  is a function  $\delta : Q \times \Sigma \rightarrow 2^Q$ ,  $I \subseteq Q$  is a set of initial states, and  $F$  is a set of final states. The language recognized by an NFA  $A$  is  $L(A) = \{w \in \Sigma^* \mid \bigcup_{q \in I} \delta(q, w) \cap F \neq \emptyset\}$ . The nondeterministic state complexity of a regular language is the minimum number of states for any NFA which accepts  $L$ . We denote the nondeterministic state complexity of  $L$  by  $\text{nsc}(L)$ .

A set of pair of strings  $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$  with  $x_i, y_i \in \Sigma^*$  for  $1 \leq i \leq m$  is a fooling set for a regular language  $L$  if  $x_i y_i \in L$  for  $1 \leq i \leq m$  and for all  $1 \leq i < j \leq m$ , either  $x_i y_j \notin L$  or  $x_j y_i \notin L$ . If  $L$  has a fooling set  $S$ , then  $\text{nsc}(L) \geq |S|$  [15].

Let  $\theta : \Sigma \rightarrow \Sigma$  be a mapping. We say  $\theta$  is a morphism if for  $u, v \in \Sigma^*$ , we have  $\theta(uv) = \theta(u)\theta(v)$ . We say  $\theta$  is an antimorphism if we have  $\theta(uv) = \theta(v)\theta(u)$ . The mapping  $\theta$  is an involution if for all words  $u \in \Sigma^*$ , we have  $\theta(\theta(u)) = u$ . For example, if  $\Sigma = \{A, C, G, T\}$  we can define Watson-Crick complementarity for DNA as an antimorphic involution  $\theta$  by  $\theta(A) = T$ ,  $\theta(C) = G$ ,  $\theta(G) = C$ , and  $\theta(T) = A$ . Then the Watson-Crick complement of a DNA string  $w$  is given by  $\theta(w)$ .

**Definition 1.** Let  $\theta$  be an antimorphic involution and  $x, y \in \Sigma^*$ . We define the  $\theta$ -catenation operation  $\odot^\theta$ , also called pseudocatenation with respect to  $\theta$ , by

$$x \odot^\theta y = \{xy, x\theta(y)\}.$$

We can define  $\theta$ -catenation for languages by

$$L_1 \odot^\theta L_2 = \{xy, x\theta(y) \mid x \in L_1, y \in L_2\}.$$

This operation can be extended to an iterated variant by  $L^{\odot_0^\theta} = \{\varepsilon\}$ ,  $L^{\odot_1^\theta} = L$ , and  $L^{\odot_n^\theta} = L^{\odot_{n-1}^\theta} \odot^\theta L$ . Then we can take the positive  $\theta$ -catenation closure by

$$L^{\odot_+^\theta} = \bigcup_{i \geq 1} L^{\odot_i^\theta}.$$

Although  $\theta$ -catenation is defined for both morphisms and antimorphisms, we will consider only the state complexity for antimorphisms. For morphic  $\theta$ , many state complexity results are the same as the state complexity of combined operations studied previously. Furthermore, we note that the condition that  $\theta$  be involutive is not strictly necessary in our constructions.

We will make use of the following notation for the NFA recognizing  $\theta(L(A))$  for a given DFA  $A$  and antimorphism  $\theta$ . Let  $A = (Q, \Sigma, \delta, s, F)$  be a DFA. Let

$P \subseteq Q$  be a set of states of  $Q$ . We denote by  $\bar{P} = \{\bar{q} \mid q \in P\}$ . Then define the DFA  $\bar{A} = (\bar{Q}, \Sigma, \delta^{-1}, \bar{F}, \{\bar{s}\})$ , where the transition function  $\delta^{-1} : \bar{Q} \times \Sigma \rightarrow 2^{\bar{Q}}$  is defined for  $\bar{q} \in \bar{Q}$  and  $a \in \Sigma$  by  $\delta^{-1}(\bar{q}, a) = \{\bar{q}' \mid \delta(q', \theta(a)) = q\}$ . In other words, every transition of  $A$  is reversed and relabeled according to  $\theta$  in  $\bar{A}$ . Then  $L(\bar{A}) = \theta(L(A))$ .

### 3 State Complexity of $\theta$ -Catenation

We will consider the state complexity of the  $\theta$ -catenation of two regular languages. It was shown in [17] that the class of regular languages is closed under  $\theta$ -catenation. This is easy to see from the following expression for the  $\theta$ -catenation of  $L_1$  and  $L_2$ , which follows directly from the definition.

**Proposition 2.** *Let  $L_1, L_2 \subseteq \Sigma^*$  be languages and  $\theta$  an antimorphism. Then  $L_1 \odot^\theta L_2 = L_1(L_2 \cup \theta(L_2))$ .*

First, we consider an NFA for recognizing  $L_1 \odot^\theta L_2$  and its nondeterministic state complexity.

**Proposition 3.** *For  $m, n \geq 1$ , let  $A$  and  $B$  be NFAs defined over an alphabet  $\Sigma$  with  $m$  and  $n$  states and let  $\theta$  be an antimorphism. Then there exists an NFA that recognizes  $L(A) \odot^\theta L(B)$  with at most  $m + 2n$  states and this bound can be reached.*

The proof of Proposition 3 makes use of the following construction for an NFA  $C$  that recognizes  $L(A) \odot^\theta L(B)$ . Let  $A = (Q_A, \Sigma, \delta_A, I_A, F_A)$  and  $B = (Q_B, \Sigma, \delta_B, I_B, F_B)$ . We denote by  $\bar{B}$  the NFA for  $\theta(L(B))$ , defined  $\bar{B} = (\bar{Q}_B, \Sigma, \delta_B^{-1}, \bar{F}_B, \bar{I}_B)$ . We define an NFA  $C = (Q_C, \Sigma, \delta_C, I_C, F_C)$  where  $Q_C = Q_A \cup Q_B \cup \bar{Q}_B$ ,  $I_C = I_A$ ,  $F_C = F_B \cup \bar{F}_B$ , and the transition function  $\delta_C : Q_C \times \Sigma \rightarrow 2^{Q_C}$  is defined for  $q \in Q_C$  and  $a \in \Sigma$  by

$$\delta_C(q, a) = \begin{cases} \delta_A(q, a) & \text{if } q \in Q_A, \\ \delta_B(q, a) & \text{if } q \in Q_B, \\ \delta_B^{-1}(q, a) & \text{if } q \in \bar{Q}_B, \\ \delta_A(q, a) \cup I_B \cup \bar{F}_B & \text{if } (\delta_A(q, a) \cap F_A) \neq \emptyset. \end{cases}$$

From this construction, it follows that  $C$  has most  $m + 2n$  states, and this bound is also reachable.

We will now consider the deterministic state complexity of  $\theta$ -catenation. We note again that  $L(A) \odot^\theta L(B) = L(A)(L(B) \cup \theta(L(B)))$ . By directly computing the state complexity of the union  $(L(B) \cup \theta(L(B)))$  and composing it with the state complexity for the catenation  $L(A)(L(B) \cup \theta(L(B)))$ , we obtain an upper bound of  $m2^{n2^n} - 2^{n2^n - 1}$  states for  $L(A) \odot^\theta L(B)$ . This is clearly incorrect, since determinizing the NFA from Proposition 3 gives at most  $2^{m+2n}$  states. Instead, we apply a construction similar to the one from [7] to  $L(A)(L(B) \cup \theta(L(B)))$ .

**Proposition 4.** *Let  $m, n \geq 3$ ,  $\theta$  be an antimorphism, and  $A$  and  $B$  be DFAs defined over an alphabet  $\Sigma$  with  $m$  and  $n$  states, respectively. Then there exists a DFA that recognizes  $L(A) \odot^\theta L(B)$  with at most  $(m-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$  states.*

*Proof.* We will define a DFA  $C$  that recognizes  $L(A) \odot^\theta L(B)$  given two DFAs  $A$  and  $B$ . Let  $A = (Q_A, \Sigma, \delta_A, s_A, F_A)$  and  $B = (Q_B, \Sigma, \delta_B, s_B, F_B)$ . We define the DFA  $C = (Q_C, \Sigma, \delta_C, s_C, F_C)$  by the set of states

$$\begin{aligned} Q_C = & \{ \langle q, P, \bar{R} \rangle \mid q \in Q_A - F_A, P \in 2^{Q_B} - \{\emptyset\}, \bar{R} \in 2^{\overline{Q_B}} - \{\emptyset\} \} \\ & \cup \{ \langle q, \emptyset, \emptyset \rangle \mid q \in Q_A - F_A \} \\ & \cup \{ \langle q, P \cup \{s_B\}, \bar{R} \cup \overline{F_B} \rangle \mid q \in F_A, P \in 2^{Q_B - \{s_B\}}, \bar{R} \in 2^{\overline{Q_B - \overline{F_B}}} \}, \end{aligned}$$

the initial state

$$s_C = \begin{cases} \langle s_A, \emptyset, \emptyset \rangle & \text{if } s_A \notin F_A, \\ \langle s_A, \{s_B\}, \overline{F_B} \rangle & \text{otherwise,} \end{cases}$$

the set of final states  $F_C = \{ \langle q, P, \bar{R} \rangle \in Q_C \mid (P \cup \bar{R}) \cap (F_B \cup \{\bar{s}_B\}) \neq \emptyset \}$ , and the transition function  $\delta_C(\langle q, P, \bar{R} \rangle, a) = \langle q', P', \bar{R}' \rangle$  for  $a \in \Sigma$  where  $q' = \delta_A(q, a)$ ,

$$\begin{aligned} P' &= \begin{cases} \bigcup_{p \in P} \delta_B(p, a) \cup \{s_B\} & \text{if } q' \in F_A, \\ \bigcup_{p \in P} \delta_B(p, a) & \text{otherwise,} \end{cases} \\ \bar{R}' &= \begin{cases} \delta_B^{-1}(\bar{R}, a) \cup \overline{F_B} & \text{if } q' \in F_A, \\ \delta_B^{-1}(\bar{R}, a) & \text{otherwise.} \end{cases} \end{aligned}$$

Informally, the DFA  $C$  operates as follows. The states of  $C$  are 3-tuples  $\langle q, P, \bar{R} \rangle$ , where  $q$  is a state of  $A$ , and  $P$  and  $\bar{R}$  are subsets of states of  $B$ . The first component  $q$  denotes the current state of a computation on  $A$ , the second component  $P$  denotes a set of states corresponding to the current states of computations on  $B$ , and the third component  $\bar{R}$  denotes a set of states corresponding to the current states of computations on  $\bar{B}$ , the NFA recognizing  $\theta(L(B))$ .

Upon reading a symbol  $a \in \Sigma$ , the computations advance one step to  $\langle q', P', \bar{R}' \rangle$ . If  $q'$  is a final state of  $A$ , then in addition to updating the sets  $P$  and  $R'$  to advance one step in computation, the initial state  $s_B$  of  $B$  is added to  $P'$  and the set of initial states  $\overline{F_B}$  of  $\bar{B}$ , the NFA recognizing  $\theta(L(B))$ , is added to  $\bar{R}'$ .

We will now consider the size of  $Q_C$ , the state set of  $C$ . Let  $k_A = |F_A|$  and  $k_B = |F_B|$ . We have

$$|Q_C| = (m - k_A)(2^n - 1)(2^n - 1) + (m - k_A) + k_A(2^{n-1})(2^{n-k_B}).$$

However, note that since  $B$  is a complete DFA, we have  $\delta_B^{-1}(\overline{Q_B}, \sigma) = \overline{Q_B}$  for all  $\sigma \in \Sigma$ . Then for all states  $q \in Q_A$ ,  $P \subseteq Q_B$ , and symbols  $\sigma \in \Sigma$ , we have  $\delta_C(\langle q, P, \overline{Q_B} \rangle, \sigma) = \langle q', P', \overline{Q_B} \rangle$ . Since  $\bar{s} \in \overline{Q_B}$ , any state of the form  $\langle q, P, \overline{Q_B} \rangle$  is a final state. Thus, for all states  $q \in Q_A$ ,  $P \subseteq Q_B$ , and words  $w \in \Sigma^*$ , we have

$\delta_C(\langle q, P, \overline{Q_B} \rangle, w) \in F_C$ . Therefore, all states with the third component  $\overline{R} = \overline{Q_B}$  are equivalent and indistinguishable and we revise our upper bound down to

$$(m - k_A)(2^n - 1)(2^n - 1) + (m - k_A) + k_A(2^{n-1})(2^{n-k_B} - 1) + 1.$$

This value is maximized when  $k_A = 1$  and  $k_B = 1$ , giving a total of  $(m - 1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$  states.  $\square$

We will now show that this bound is reachable.

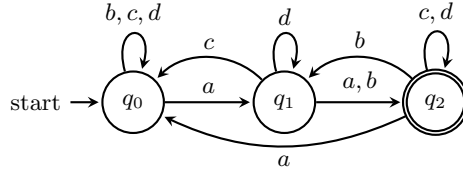
**Lemma 5.** *For  $m, n \geq 3$ , there exist an  $m$ -state DFA  $A$ , an  $n$ -state DFA  $B$ , and an antimorphism  $\theta$  over an alphabet of size 4 such that*

$$\text{sc}(L(A) \odot^\theta L(B)) \geq (m - 1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1.$$

The main idea of the proof of Lemma 5 is to demonstrate that the bound from Proposition 4 is reachable by using the witness  $\mathcal{W}_n(a, b, c, d)$  defined by Brzozowski [2]. Let  $\Sigma = \{a, b, c, d\}$  and let  $\theta : \Sigma^* \rightarrow \Sigma^*$  be the Watson-Crick antimorphism defined by

$$\theta(a) = d \quad \theta(b) = c \quad \theta(c) = b \quad \theta(d) = a.$$

We set  $A = \mathcal{W}_m(a, b, c, d)$  with  $m$  states and  $B = \mathcal{W}_n(a, b, c, d)$  with  $n$  states. Then we define  $\overline{B} = \overline{\mathcal{W}_n(a, b, c, d)}$ . That is,  $L(\overline{B}) = \theta(L(\mathcal{W}_n(a, b, c, d))) = L(\mathcal{W}_n(d, c, b, a))^R$ . The DFA  $\mathcal{W}_3(a, b, c, d)$  is shown in Figure 1 and the DFA  $B$  and the NFA  $\overline{B}$  are shown in Figure 2.



**Fig. 1.** The DFA  $\mathcal{W}_3(a, b, c, d)$

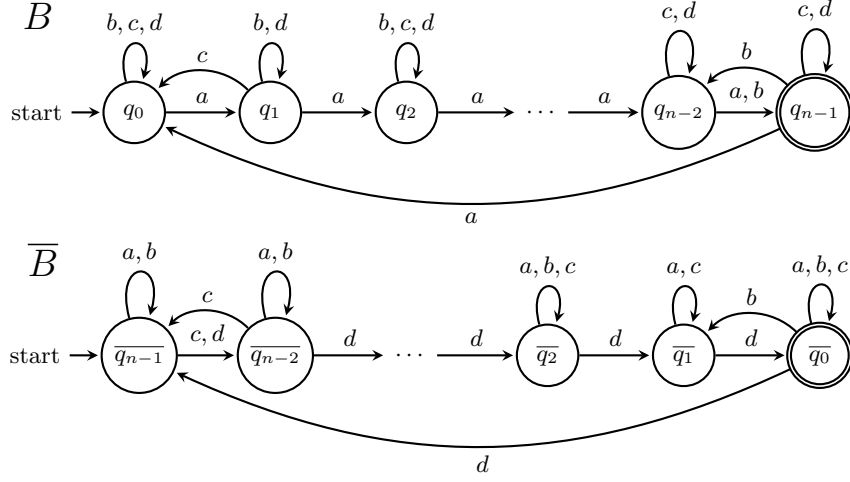
Proposition 4 and Lemma 5 are summarized in the following theorem.

**Theorem 6.** *For  $m, n \geq 3$ , regular languages  $L_m$  and  $L_n$  with  $\text{sc}(L_m) = m$  and  $\text{sc}(L_n) = n$ , and antimorphism  $\theta$ ,*

$$\text{sc}(L_m \odot^\theta L_n) \leq (m - 1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$$

and this bound can be reached in the worst case.

Furthermore, observe that the witnesses used in Lemma 5 belong to the same family of DFAs  $\mathcal{W}_n(a, b, c, d)$ . Setting  $m = n$  gives us the same DFA and we obtain a tight bound for the state complexity of the pseudosquare of  $L$ ,  $L^{\odot_2^\theta}$ , via Lemma 5.



**Fig. 2.** The DFA  $B = \mathcal{W}_n(a, b, c, d)$  and the NFA  $\overline{B} = \overline{\mathcal{W}_n(a, b, c, d)}$

**Corollary 7.** For  $n \geq 3$ , let  $L_n$  be a regular language with  $\text{sc}(L_n) = n$  and let  $\theta$  be an antimorphism. Then

$$\text{sc}(L_n^{\odot_\theta^2}) \leq (n-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$$

and this bound can be reached in the worst case.

## 4 State Complexity of $\theta$ -Catenation Closure

In this section, we consider the  $\theta$ -catenation closure of a regular language. This is analogous to the positive Kleene closure, but with respect to  $\theta$ -catenation. It was shown in [17] that the positive closure of a regular language with respect to  $\theta$ -catenation is also regular. The following equality follows from the definition.

**Proposition 8.** Let  $L$  be a language and let  $\theta$  be an antimorphism. Then the positive  $\theta$ -catenation closure is  $L^{\odot_\theta^+} = L(L \cup \theta(L))^*$ .

It is important to note that the positive closure with respect to  $\theta$ -catenation is not  $(L \cup \theta(L))^+$ , as words  $w \in L^{\odot_\theta^+}$  have the form  $w = uv_1v_2 \cdots v_{k-1}$  where  $u \in L$  and  $v_i \in L \cup \theta(L)$  for  $1 \leq i \leq k$  [17], whereas  $(L \cup \theta(L))^+$  also contains words of the form  $\theta(u)v_1v_2 \cdots v_{k-1}$ .

We will first consider an NFA for recognizing  $L^{\odot_\theta^+}$  and its nondeterministic state complexity.

**Proposition 9.** For  $n \geq 1$ , let  $A$  be an NFA with  $n$  states defined over an alphabet  $\Sigma$  and let  $\theta$  be an antimorphism. Then there exists an NFA that recognizes  $L(A)^{\odot_\theta^+}$  with at most  $2n$  states. Furthermore, this bound can be reached in the worst case.

The proof of Proposition 9 defines an NFA  $A'$  that recognizes  $L(A)^{\circledast}$ . Let  $A = (Q, \Sigma, \delta, I, F)$  be an  $n$ -state NFA. We denote by  $\bar{A}$  the NFA recognizing  $\theta(L(A))$ ,  $\bar{A} = (\bar{Q}, \Sigma, \delta^{-1}, \bar{F}, \bar{I})$ . We will define an NFA  $A'$  which recognizes  $A^{\circledast}$  with respect to an antimorphism  $\theta$  by  $A' = (Q', \Sigma, \delta', I', F')$ , where  $Q' = Q \cup \bar{Q}$ ,  $I' = I$ ,  $F' = F \cup \bar{I}$ , and the transition function  $\delta' : Q' \times \Sigma \rightarrow 2^{Q'}$  is defined for  $q \in Q'$  and  $a \in \Sigma$  by

$$\delta'(q, a) = \begin{cases} \delta(q, a) & \text{if } q \in Q, \\ \delta^{-1}(q, a) & \text{if } q \in \bar{Q}, \\ \delta(q, a) \cup I \cup \bar{F} & \text{if } q \in Q \text{ and } (\delta(q, a) \cap F) \neq \emptyset, \\ \delta^{-1}(q, a) \cup I \cup \bar{F} & \text{if } q \in \bar{Q} \text{ and } (\delta^{-1}(q, a) \cap \bar{I}) \neq \emptyset. \end{cases}$$

From this construction, it follows that  $A'$  has at most  $2n$  states, and this bound is also reachable.

We will now consider the deterministic state complexity of the positive  $\theta$ -catenation closure. In [23], it was shown that the state complexity of  $(L_1 \cup L_2)^*$  was much lower than the straightforward upper bound of  $2^{mn-1} + 2^{mn-2}$ . Indeed, the bound obtained from the NFA of Proposition 9 is already  $2^{2n}$  states. We will show that the state complexity of  $\theta$ -catenation closure is still lower than this.

**Proposition 10.** *For  $n \geq 3$ , let  $A$  be a DFA defined over an alphabet  $\Sigma$  with  $n$  states and let  $\theta$  be an antimorphism. Then there exists a DFA that recognizes  $L(A)^{\circledast}$  with at most  $2^{2n-1} - 2^n + 1$  states.*

*Proof.* We define a DFA  $A'$  that recognizes  $L(A)^{\circledast}$  given a DFA  $A$ . Let  $A = (Q, \Sigma, \delta, s, F)$ . We define the DFA  $A' = (Q', \Sigma, \delta', s', F')$  with the set of states

$$Q' = \{ \langle P, \bar{R} \rangle \mid \emptyset \neq P \subseteq Q - F, \bar{R} \subseteq \bar{Q} - \{\bar{s}\} \} \\ \cup \{ \langle P \cup \{s\}, \bar{R} \cup \bar{F} \rangle \subseteq Q \times \bar{Q} \mid (P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) \neq \emptyset \},$$

the initial state

$$s' = \begin{cases} \langle \{s\}, \emptyset \rangle & \text{if } s \notin F, \\ \langle \{s\}, \bar{F} \rangle & \text{if } s \in F, \end{cases}$$

the set of final states  $F' = \{ \langle P, \bar{R} \rangle \subseteq Q \times \bar{Q} \mid (P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) \neq \emptyset \}$ , and the transition function for a state  $\langle P, \bar{R} \rangle$  and symbol  $a \in \Sigma$  with  $P' = \delta(P, a)$  and  $\bar{R}' = \delta^{-1}(\bar{R}, a)$  is defined by

$$\delta'(\langle P, \bar{R} \rangle, a) = \begin{cases} \langle P' \cup \{s\}, \bar{R}' \cup \bar{F} \rangle & \text{if } (P' \cup \bar{R}') \cap (F \cup \{\bar{s}\}) \neq \emptyset, \\ \langle P', \bar{R}' \rangle & \text{otherwise.} \end{cases}$$

Informally, DFA  $A'$  operates by first simulating a computation of  $A$ , since by definition, we have  $L(A)^{\circledast} = L(A)(L(A) \cup \theta(L(A)))^*$ . Once the computation reaches a final state of  $A$ , an initial state for  $A$  and  $\bar{A}$  is added to the current state set and the computation continues. Whenever the current state of  $A'$



contains a final state of  $A$  or  $\bar{A}$ , the initial states of both machines are added. The computation continues until the input is read and accepts if and only if a final state of  $A$  or  $\bar{A}$  is contained in the state of  $A'$  when the input has been read.

Now let us consider the state set  $Q'$  of  $A'$ ,

$$\begin{aligned} Q' &= Q_1 \cup Q_2, \\ Q_1 &= \{\langle P, \bar{R} \rangle \mid \emptyset \neq P \subseteq Q - F, \bar{R} \subseteq \bar{Q} - \{\bar{s}\}\}, \\ Q_2 &= \{\langle P \cup \{s\}, \bar{R} \cup \bar{F} \rangle \subseteq Q \times \bar{Q} \mid (P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) \neq \emptyset\}. \end{aligned}$$

The size of  $Q'$  will depend on  $k$  and whether or not  $s \in F$ . We will consider each term. Let  $k = |F|$ .

First,  $Q_1$  is the set of states with components that do not contain any final state of  $A$  or  $\bar{A}$ . There are  $2^{n-k} - 1$  nonempty subsets of  $Q - F$  and there are  $2^{n-1}$  subsets of  $\bar{Q} - \{\bar{s}\}$ . This gives us  $|Q_1| = (2^{n-k} - 1)(2^{n-1})$ .

Then,  $Q_2$  is the set of states of the form  $\langle P, \bar{R} \rangle$  with  $P \subseteq Q$  and  $\bar{R} \subseteq \bar{Q}$  such that  $(P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) \neq \emptyset$ . That is at least one of a final state of  $A$  is in  $P$  or  $\bar{s}$  is in  $\bar{R}$ . Then  $s \in P$  and  $\bar{F} \subseteq \bar{R}$ . This gives up to  $(2^{n-1})(2^{n-k})$  states.

This count may include states such that  $s \in P$  and  $\bar{F} \subseteq \bar{R}$  but  $(P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) = \emptyset$ , depending on whether or not  $s \in F$ . If  $s \in F$ , then there are no such states, since  $s \in F$  and  $s \in P$  implies that  $(P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) \neq \emptyset$ .

However, if  $s \notin F$ , there are up to  $(2^{n-1-k})^2$  states  $\langle P, \bar{R} \rangle$  such that  $s \in P$ ,  $\bar{F} \subseteq \bar{R}$ , and  $(P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) = \emptyset$  which must be removed from the total, resulting in at most  $(2^{n-1})(2^{n-k}) - (2^{n-1-k})^2$  states when  $s \notin F$ .

Finally, we must account for states of the form  $\langle P, \bar{Q} \rangle$ . Since  $A$  is a complete DFA, we have  $\delta^{-1}(\bar{Q}, \sigma) = \bar{Q}$  for all  $\sigma \in \Sigma$ . Since  $\bar{s} \in \bar{Q}$ , we have  $\langle P, \bar{Q} \rangle$  for all  $P \subseteq Q$  and therefore  $\delta'(\langle P, \bar{Q} \rangle, w) \in F'$  for all  $P \subseteq Q$  and  $w \in \Sigma^*$ . Thus, all such states are equivalent and indistinguishable. Since  $\bar{s} \in \bar{Q}$ , for all states  $\langle P, \bar{Q} \rangle$ , we have  $s \in P$  and thus there are  $2^{n-1}$  such states to be merged into a single state.

Thus, in total, we have

$$|Q'| \leq \begin{cases} (2^{n-k} - 1)(2^{n-1}) + (2^{n-1})(2^{n-k}) - 2^{n-1} + 1 & \text{if } s \in F, \\ (2^{n-k} - 1)(2^{n-1}) + (2^{n-1})(2^{n-k}) - (2^{n-1-k})^2 - 2^{n-1} + 1 & \text{if } s \notin F. \end{cases}$$

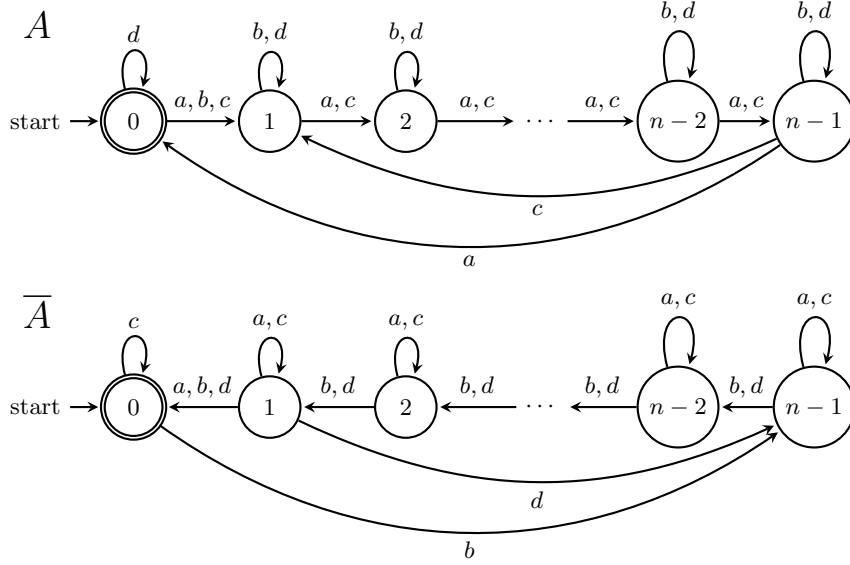
From this, we can see that the size of  $Q'$  is maximized when  $k = 1$  and  $s \in F$ . Thus,  $Q'$  has size at most  $(2^{n-1} - 1)2^{n-1} + (2^{n-1})^2 - 2^{n-1} + 1 = 2^{2n-1} - 2^n + 1$ .  $\square$

**Lemma 11.** *Let  $n \geq 3$ . Then there exists an  $n$ -state DFA  $A$  and an antimorphism  $\theta$  over an alphabet of size 4 such that*

$$\text{sc}(L(A)^{\odot_+^{\theta}}) \geq 2^{2n-1} - 2^n + 1.$$

To prove Lemma 11, we demonstrate that the upper bound from Proposition 10 is reachable via the following witness. Let  $\Sigma = \{a, b, c, d\}$  and let  $\theta : \Sigma \rightarrow \Sigma$  be the antimorphism defined by

$$\theta(a) = b \quad \theta(b) = a \quad \theta(c) = d \quad \theta(d) = c.$$



**Fig. 3.** The DFA  $A$  and the NFA  $\bar{A}$

We define a DFA  $A$ , shown in Figure 3 together with the NFA  $\bar{A}$  which recognizes the language  $\theta(L(A))$ .

From Proposition 10 and Lemma 11, we can summarize our results in the following theorem.

**Theorem 12.** *For  $n \geq 3$ , a regular language  $L$  with  $sc(L) = n$ , and an antimorphism  $\theta$ ,*

$$sc(L^{\odot_+^\theta}) \leq 2^{2n-1} - 2^n + 1$$

*and this bound can be reached in the worst case.*

## 5 Conclusion

We have given tight bounds for the deterministic and nondeterministic state complexity of pseudocatenation and positive pseudocatenation closure. The deterministic state complexity bounds for each operation differ from those for the corresponding classical operations, catenation and star, and the bounds derived from combined operations. A comparison between the bounds is given in Table 1.

One question that arises is to consider variants of the pseudocatenation operation. The definition of  $\odot^\theta$  on two words  $u$  and  $v$  was defined by Kari and Kulkarni [17] to be the set comprising  $uv$  and  $u\theta(v)$ . This definition coincides with  $\theta$ -powers and  $\theta$ -primitivity as defined by Czeizler et al. [10]. However, a definition of  $\theta$ -catenation that incorporates  $\theta(u)v$  also makes sense to consider from the biological point of view, since it is the Watson-Crick complement of  $u\theta(v)$ .

Operation	State complexity	
$L_m \odot^\theta L_n$	$(m-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$	Thm. 6
$L_m L_n$	$m2^n - 2^{n-1}$	[26]
$L_m(L_n \cup L_p)$	$(m-1)(2^{n+p} - 2^n - 2^p + 2) + 2^{n+p-2}$	[7]
$L_n^{\odot_2^\theta}$	$(n-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$	Cor. 7
$L_n^2$	$n2^n - 2^{n-1}$	[22]
$L_n^{\odot_+^\theta}$	$2^{2n-1} - 2^n + 1$	Thm. 12
$L_n^*$	$2^{n-1} + 2^{n-2}$	[26]
$(L_m \cup L_n)^*$	$2^{m+n-1} - 2^{m-1} - 2^{n-1} + 1$	[23]

**Table 1.** A comparison of the deterministic state complexity for each operation

There are also further questions considering the state complexity of the current pseudocatenation operation  $\odot^\theta$ . We can consider the state complexity of pseudocatenation for sub-regular language classes, such as finite languages. We also noted earlier that as a result of our choice of witnesses in Lemma 5, we were also able to obtain the state complexity for the pseudosquare  $L^{\odot_2^\theta}$  (Corollary 7). Domaratzki and Okhotin [11] gave a tight state complexity bound for the cube of a language  $L^3$ , which was improved by Caron et al. [4]. Asymptotic state complexity bounds for the  $k$ th power of a language  $L^k$  are also given in [11]. A natural next question to consider is the state complexity of pseudocubes and pseudopowers with respect to  $\theta$ .

## References

1. Brzozowski, J., Liu, D.: Universal Witnesses for State Complexity of Basic Operations Combined with Reversal. In: Implementation and Application of Automata (CIAA 2013). LNCS, vol. 7982, pp. 72–83 (2013)
2. Brzozowski, J., Liu, D.: Universal Witnesses for State Complexity of Boolean Operations and Concatenation Combined with Star. In: Descriptive Complexity of Formal Systems (DCFS 2013). LNCS, vol. 8031, pp. 30–41 (2013)
3. Brzozowski, J.A., Kari, L., Li, B., Szykuła, M.: State Complexity of Overlap Assembly. In: Implementation and Application of Automata (CIAA 2018). LNCS, vol. 10977, pp. 109–120 (2018)
4. Caron, P., Luque, J.G., Patrou, B.: State complexity of multiple catenation. arXiv **1607.04031** (2016)
5. Cho, D.J., Han, Y.S., Kim, H., Palioudakis, A., Salomaa, K.: Duplications and pseudo-duplications. International Journal of Unconventional Computing **12**(2-3), 157–168 (2016)
6. Cho, D.J., Han, Y.S., Ko, S.K., Salomaa, K.: State complexity of inversion operations. Theoretical Computer Science **610**, 2–12 (2016)

7. Cui, B., Gao, Y., Kari, L., Yu, S.: State Complexity of Two Combined Operations: Catenation-Union and Catenation-Intersection. *International Journal of Foundations of Computer Science* **22**(08), 1797–1812 (2011)
8. Cui, B., Gao, Y., Kari, L., Yu, S.: State complexity of combined operations with two basic operations. *Theoretical Computer Science* **437**, 82–102 (2012)
9. Cui, B., Gao, Y., Kari, L., Yu, S.: State Complexity of Two Combined Operations: Catenation-Star and Catenation-Reversal. *International Journal of Foundations of Computer Science* **23**(01), 51–66 (2012)
10. Czeizler, E., Kari, L., Seki, S.: On a special class of primitive words. *Theoretical Computer Science* **411**(3), 617–630 (2010)
11. Domaratzki, M., Okhotin, A.: State complexity of power. *Theoretical Computer Science* **410**(24-25), 2377–2392 (2009)
12. Gao, Y., Moreira, N., Reis, R., Yu, S.: A Survey on Operational State Complexity. *Journal of Automata, Languages and Combinatorics* **21**(4), 251–310 (2016)
13. Gao, Y., Yu, S.: State Complexity of Four Combined Operations Composed of Union, Intersection, Star and Reversal. In: *Descriptive Complexity of Formal Systems (DCFS 2011)*. LNCS, vol. 6808, pp. 158–171 (2011)
14. Gao, Y., Yu, S.: State Complexity of Combined Operations with Union, Intersection, Star and Reversal. *Fundamenta Informaticae* **116**, 79–92 (2012)
15. Glaister, I., Shallit, J.: A lower bound technique for the size of nondeterministic finite automata. *Information Processing Letters* **59**(2), 75–77 (1996)
16. Kari, L., Konstantinidis, S., Losseva, E., Sosik, P., Thierrin, G.: A formal language analysis of DNA hairpin structures. *Fundamenta Informaticae* **71**, 453–475 (2006)
17. Kari, L., Kulkarni, M.: Generating the Pseudo-powers of a Word. *Journal of Automata, Languages and Combinatorics* **19**(1-4), 157–171 (2014)
18. Kari, L., Mahalingam, K.: Watson-Crick conjugate and commutative words. In: *DNA Computing (DNA 13)*. LNCS, vol. 4848, pp. 273–283 (2008)
19. Kari, L., Seki, S.: On pseudoknot-bordered words and their properties. *Journal of Computer and Systems Sciences* **75**, 113–121 (2009)
20. Liu, G., Martin-Vide, C., Salomaa, A., Yu, S.: State complexity of basic language operations combined with reversal. *Information and Computation* **206**(9-10), 1178–1186 (2008)
21. de Luca, A., Luca, A.D.: Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science* **362**(1-3), 282–300 (2006)
22. Rampersad, N.: The state complexity of  $L^2$  and  $L^k$ . *Information Processing Letters* **98**(6), 231–234 (2006)
23. Salomaa, A., Salomaa, K., Yu, S.: State complexity of combined operations. *Theoretical Computer Science* **383**(2-3), 140–152 (2007)
24. Salomaa, A., Salomaa, K., Yu, S.: Undecidability of state complexity. *International Journal of Computer Mathematics* **90**(6), 1310–1320 (2013)
25. Yu, S.: Regular languages. In: Rozenberg, G., Salomaa, A. (eds.) *Handbook of Formal Languages*, pp. 41–110. Springer-Verlag, Berlin, Heidelberg (1997)
26. Yu, S., Salomaa, K., Zhuang, Q.: The state complexities of some basic operations on regular languages. *Theoretical Computer Science* **125**(2), 315–328 (1994)

## A Appendix

Here we include proofs that were omitted in the paper due to the limitation on the number of pages.

**Proposition 3.** *For  $m, n \geq 1$ , let  $A$  and  $B$  be NFAs defined over an alphabet  $\Sigma$  with  $m$  and  $n$  states and let  $\theta$  be an antimorphism. Then there exists an NFA that recognizes  $L(A) \odot^\theta L(B)$  with at most  $m + 2n$  states and this bound can be reached.* (p. 4)

*Proof.* Let  $A = (Q_A, \Sigma, \delta_A, I_A, F_A)$  and  $B = (Q_B, \Sigma, \delta_B, I_B, F_B)$ . We denote by  $\overline{B}$  the NFA for  $\theta(L(B))$ ,  $\overline{B} = (Q_B, \Sigma, \delta_B^{-1}, \overline{F_B}, \overline{I_B})$ . We define an NFA  $C = (Q_C, \Sigma, \delta_C, I_C, F_C)$  where  $Q_C = Q_A \cup Q_B \cup \overline{Q_B}$ ,  $I_C = I_A$ ,  $F_C = F_B \cup \overline{I_B}$ , and the transition function  $\delta_C : Q_C \times \Sigma \rightarrow 2^{Q_C}$  is defined for  $q \in Q_C$  and  $a \in \Sigma$  by

$$\delta_C(q, a) = \begin{cases} \delta_A(q, a) & \text{if } q \in Q_A, \\ \delta_B(q, a) & \text{if } q \in Q_B, \\ \delta_B^{-1}(q, a) & \text{if } q \in \overline{Q_B}, \\ \delta_A(q, a) \cup I_B \cup \overline{F_B} & \text{if } (\delta_A(q, a) \cap F_A) \neq \emptyset. \end{cases}$$

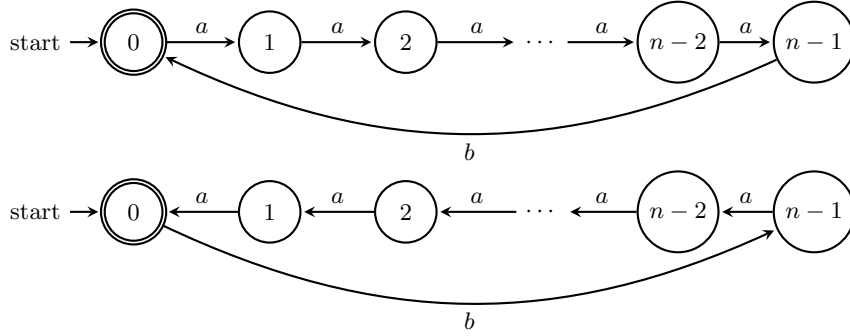
We claim that  $L(C) = L(A) \odot^\theta L(B)$ . Indeed, suppose we have a word  $w \in L(A) \odot^\theta L(B)$ . We can write  $w = uv$  where  $u \in L(A)$  and  $v \in L(B) \cup \theta(L(B))$ . Then  $\delta_C(I_C, w) = \delta_C(\delta_C(I_C, u), v)$ . Since  $u \in L(A)$ , we have  $(\delta_A(I_A, u) \cap F_A) \neq \emptyset$  and therefore  $I_B \cup \overline{F_B} \subseteq \delta_C(I_C, u)$ . Since  $v \in L(B) \cup \theta(L(B))$ , we then have  $(\delta_C(\delta_C(I_C, u), v) \cap (F_B \cup \overline{I_B})) \neq \emptyset$  and therefore  $w \in L(C)$ .

Suppose now that  $w \in L(C)$ . This means that a computation of  $w$  on  $C$  reaches a final state of  $B$  or  $\overline{B}$ . However, these states are only reachable from an initial state of  $B$  or  $\overline{B}$  and such states can only be reached when a final state of  $A$  is reached. Therefore, there must be a prefix of  $w$ , say  $u$  such that  $\delta_C(I_C, u) \cap F_A \neq \emptyset$ . But this means that  $u \in L(A)$ . Then the suffix of  $w$  from this state, say  $v$  is an accepting path on  $B$  or  $\overline{B}$  by definition. Therefore, we have  $w = uv$  with  $u \in L(A)$  and  $v \in L(B) \cup L(\overline{B})$  and thus  $w \in L(A) \odot^\theta L(B)$ .

From this construction,  $C$  has most  $m + 2n$  states. To see that this bound is reachable, we consider languages  $L_m = b^{m-1}(b^m)^*$  and  $L_n = (a^{n-1}b)^*$  over a binary alphabet  $\{a, b\}$  and let  $\theta$  be the identity antimorphism  $\theta(a) = a$  for all  $a \in \Sigma$ . The NFA that recognizes  $L_n$  is shown in Figure 4. We will define a fooling set  $S$  for  $L_m \odot^\theta L_n$ . Let  $S = S_1 \cup S_2 \cup S_3$ , where

- $S_1 = \{(b^i, b^{2m-1-i}) \mid 0 \leq i \leq m-1\}$ ,
- $S_2 = \{(b^{m-1}a^{n-1}ba^{n-1-i}, a^i b) \mid 0 \leq i \leq n-1\}$ ,
- $S_3 = \{(b^{m-1}ba^{n-1}ba^{n-1-i}, a^i) \mid 0 \leq i \leq n-1\}$ .

Let  $j > i$  and consider pairs  $(x_i, y_i)$  and  $(x_j, y_j)$  in  $S_1$ . We have  $x_i y_j = b^{2m-1-i+j} \notin L_m \odot^\theta L_n$ , since  $j-i > 0$ . Similarly, for pairs  $(x_i, y_i)$  and  $(x_j, y_j)$  in  $S_2$ , we have  $x_i y_j = b^{m-1}a^{n-1}ba^{n-1-i+j} \notin L_m \odot^\theta L_n$  and for pairs  $(x_i, y_i)$  and  $(x_j, y_j)$  in  $S_3$ , we have  $x_i y_j = b^{m-1}ba^{n-1}ba^{n-1-i+j} \notin L_m \odot^\theta L_n$ .



**Fig. 4.** The NFA recognizing  $L_n = (a^{n-1}b)^*$  and its reverse

Now let  $(x_i, y_i) \in S_1$  and  $(x_j, y_j) \in S_2$ . Then

$$x_j y_i = b^{m-1} a^{n-1} b a^{n-1-j} b^{2m-1-i} \notin L_m \odot^\theta L_n.$$

Similarly, if  $(x_j, y_j) \in S_3$ , we have

$$x_j y_i = b^{m-1} b a^{n-1} b a^{n-1-j} b^{2m-1-i} \notin L_m \odot^\theta L_n.$$

Finally, if  $(x_i, y_i) \in S_2$  and  $(x_j, y_j) \in S_3$ , we have

$$x_i y_j = b^{m-1} a^{n-1} b a^{n-1-i} a^j \notin L_m \odot^\theta L_n.$$

Thus, we have shown that  $S$  is a fooling set of size  $m + 2n$  for  $L_m \odot^\theta L_n$ .  $\square$

(p. 5)

**Proposition 4.** *Let  $m, n \geq 3$ ,  $\theta$  be an antimorphism, and  $A$  and  $B$  be DFAs defined over an alphabet  $\Sigma$  with  $m$  and  $n$  states, respectively. Then there exists a DFA that recognizes  $L(A) \odot^\theta L(B)$  with at most  $(m-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$  states.*

*Proof (Correctness of construction).* We will define a DFA  $C$  that recognizes  $L(A) \odot^\theta L(B)$  given two DFAs  $A$  and  $B$ . Let  $A = (Q_A, \Sigma, \delta_A, s_A, F_A)$  and  $B = (Q_B, \Sigma, \delta_B, s_B, F_B)$ . We define the DFA  $C = (Q_C, \Sigma, \delta_C, s_C, F_C)$  by the set of states

$$\begin{aligned} Q_C = & \{ \langle q, P, \bar{R} \rangle \mid q \in Q_A - F_A, P \in 2^{Q_B} - \{\emptyset\}, \bar{R} \in 2^{\overline{Q_B}} - \{\emptyset\} \} \\ & \cup \{ \langle q, \emptyset, \emptyset \rangle \mid q \in Q_A - F_A \} \\ & \cup \{ \langle q, P \cup \{s_B\}, \bar{R} \cup \overline{F_B} \rangle \mid q \in F_A, P \in 2^{Q_B - \{s_B\}}, \bar{R} \in 2^{\overline{Q_B - F_B}} \}, \end{aligned}$$

the initial state

$$s_C = \begin{cases} \langle s_A, \emptyset, \emptyset \rangle & \text{if } s_A \notin F_A, \\ \langle s_A, \{s_B\}, \overline{F_B} \rangle & \text{otherwise,} \end{cases}$$

the set of final states  $F_C = \{\langle q, P, \bar{R} \rangle \in Q_C \mid (P \cup \bar{R}) \cap (F_B \cup \{\bar{s}_B\}) \neq \emptyset\}$ , and the transition function  $\delta_C(\langle q, P, \bar{R} \rangle, a) = \langle q', P', \bar{R}' \rangle$  for  $a \in \Sigma$  where  $q' = \delta_A(q, a)$ ,

$$P' = \begin{cases} \bigcup_{p \in P} \delta_B(p, a) \cup \{s_B\} & \text{if } q' \in F_A, \\ \bigcup_{p \in P} \delta_B(p, a) & \text{otherwise,} \end{cases}$$

$$\bar{R}' = \begin{cases} \delta_B^{-1}(\bar{R}, a) \cup \bar{F}_B & \text{if } q' \in F_A, \\ \delta_B^{-1}(\bar{R}, a) & \text{otherwise.} \end{cases}$$

Informally, the DFA  $C$  operates as follows. The states of  $C$  are 3-tuples  $\langle q, P, \bar{R} \rangle$ , where  $q$  is a state of  $A$ , and  $P$  and  $\bar{R}$  are subsets of states of  $B$ . The first component  $q$  denotes the current state of a computation on  $A$ , the second component  $P$  denotes a set of states corresponding to the current states of computations on  $B$ , and the third component  $\bar{R}$  denotes a set of states corresponding to the current states of computations on  $\bar{B}$ , the NFA recognizing  $\theta(L(B))$ .

Upon reading a symbol  $a \in \Sigma$ , the computations advance one step to  $\langle q', P', \bar{R}' \rangle$ . If  $q'$  is a final state of  $A$ , then in addition to updating the sets  $P$  and  $R'$  to advance one step in computation, the initial state  $s_B$  of  $B$  is added to  $P'$  and the set of initial states  $\bar{F}_B$  of  $\bar{B}$ , the NFA recognizing  $\theta(L(B))$ , is added to  $\bar{R}'$ .

We will now formally show that  $L(C) = L(A) \odot^\theta L(B)$ . First, we will show that  $L(A) \odot^\theta L(B) \subseteq L(C)$ . Consider a word  $w = uv \in L(A) \odot^\theta L(B)$ , with  $u \in L(A)$  and  $v \in L(B) \cup \theta(L(B))$ . We claim that there exists a computation path in  $C$

$$s_C \xrightarrow{u} \langle q, P, \bar{R} \rangle \xrightarrow{v} \langle q', P', \bar{R}' \rangle.$$

Since  $u \in L(A)$ , there exists a path in  $A$  from  $s_A$  to  $q \in F_A$ . Then by definition, we have  $s_B \in P$  and  $\bar{F}_B \subseteq \bar{R}$ . Now, if  $v \in L(B)$ , there exists a path in  $B$  from  $s_B$  to a state in  $F_B$  and we have  $F_B \subseteq P'$ . Similarly, if  $v \in \theta(L(B))$ , then there exists a path in  $\bar{B}$  from  $\bar{F}_B$  to a state containing  $\bar{s}_B$  and we must have  $\bar{s}_B \in \bar{R}'$ . In either case, such a computation path in  $C$  exists and we have that  $w = uv$  is accepted by  $C$  and  $L(A) \odot^\theta L(B) \subseteq L(C)$ .

Next, we will show that  $L(C) \subseteq L(A) \odot^\theta L(B)$ . Consider a word  $w \in L(C)$ . Then there is a computation path in  $C$  on the input word  $w$  from  $s_C$  to a state  $\langle q', P', \bar{R}' \rangle$  where either  $F_B \cap P' \neq \emptyset$  or  $\bar{s}_B \in \bar{R}'$ . However, this implies that there exists a state  $\langle q, P, \bar{R} \rangle$  with  $q \in F_A$ ,  $s_B \in P$ , and  $\bar{F}_B \subseteq \bar{R}$  on the path. Then we can write  $w = w_1 w_2$ , with  $\delta_C(s_A, w_1) = \langle q, P, \bar{R} \rangle$  and  $\delta_C(\langle q, P, \bar{R} \rangle, w_2) = \langle q', P', \bar{R}' \rangle$ . But this means that  $\delta_A(s_A, w_1) \in F_A$  and either  $\delta_B(s_B, w_2) \in F_B$  or  $\delta_B^{-1}(\bar{F}_B, w_2) \cap \{\bar{s}_B\} \neq \emptyset$ . Thus,  $w_1 \in L(A)$  and  $w_2 \in L(B) \cup L(\bar{B})$ , and we have  $w \in L(C) \subseteq L(A) \odot^\theta L(B)$ .

Thus,  $L(C) = L(A) \odot^\theta L(B)$ . □

**Lemma 5.** *For  $m, n \geq 3$ , there exist an  $m$ -state DFA  $A$ , an  $n$ -state DFA  $B$ , and an antimorphism  $\theta$  over an alphabet of size 4 such that* (p. 6)

$$\text{sc}(L(A) \odot^\theta L(B)) \geq (m-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1.$$

*Proof.* To demonstrate that the bound from Proposition 4 is reachable, we will use the witness  $\mathcal{W}_n(a, b, c, d)$  from [2]. Let  $\Sigma = \{a, b, c, d\}$  and let  $\theta : \Sigma \rightarrow \Sigma$  be the Watson-Crick antimorphic defined by

$$\theta(a) = d \quad \theta(b) = c \quad \theta(c) = b \quad \theta(d) = a.$$

We set  $A = \mathcal{W}_m(a, b, c, d)$  with  $m$  states and  $B = \mathcal{W}_n(a, b, c, d)$  with  $n$  states. Then we define  $\overline{B} = \overline{\mathcal{W}_n(a, b, c, d)}$ . That is,  $L(\overline{B}) = \theta(L(\mathcal{W}_n(a, b, c, d))) = L(\mathcal{W}_n(d, c, b, a))^R$ . We will define  $A$ ,  $B$ , and  $\overline{B}$  explicitly.

Let  $A = (Q_A, \Sigma, \delta_A, s_A, F_A)$  with  $Q_A = \{p_0, \dots, p_{m-1}\}$ ,  $s_A = p_0$ ,  $F_A = \{p_{m-1}\}$ , and the transition function  $\delta_A$  is defined by

$$\begin{aligned} - \delta_A(p_i, a) &= p_{i+1 \bmod m} \text{ for } 0 \leq i \leq m-1, \\ - \delta_A(p_i, b) &= p_i \text{ for } 0 \leq i \leq m-3, \delta_A(p_{m-2}, b) = p_{m-1}, \delta_A(p_{m-1}, b) = p_{m-2}, \\ - \delta_A(p_i, c) &= p_i \text{ for } i = 0, 2, \dots, m-1, \delta_A(p_1, c) = p_0, \\ - \delta_A(p_i, d) &= p_i \text{ for } 0 \leq i \leq m-1. \end{aligned}$$

Let  $B = (Q_B, \Sigma, \delta_B, s_B, F_B)$  with  $Q_B = \{q_0, \dots, q_{n-1}\}$ ,  $s_B = q_0$ ,  $F_B = \{q_{n-1}\}$ , and  $\delta_B$  is defined by

$$\begin{aligned} - \delta_B(q_i, a) &= q_{i+1 \bmod n} \text{ for } 0 \leq i \leq n-1, \\ - \delta_B(q_i, b) &= q_i \text{ for } 0 \leq i \leq n-3, \delta_B(q_{n-2}, b) = p_{n-1}, \delta_B(q_{n-1}, b) = q_{n-2}, \\ - \delta_B(q_i, c) &= q_i \text{ for } i = 0, 2, \dots, n-1, \delta_B(q_1, c) = q_0, \\ - \delta_B(q_i, d) &= q_i \text{ for } 0 \leq i \leq n-1. \end{aligned}$$

Finally, let  $\overline{B} = (\overline{Q_B}, \Sigma, \delta_B^{-1}, \overline{F_B}, \{\overline{s_B}\})$ , which recognizes  $\theta(L(B))$ , where the transition function  $\delta_B^{-1} : \overline{Q} \times \Sigma \rightarrow 2^{\overline{Q}}$  is defined by

$$\begin{aligned} - \delta_B^{-1}(\overline{q_i}, d) &= \{\overline{q_{i-1 \bmod n}}\} \text{ for } 0 \leq i \leq n-1, \\ - \delta_B^{-1}(\overline{q_i}, c) &= \{\overline{q_i}\} \text{ for } 0 \leq i \leq n-3, \delta_B^{-1}(\overline{q_{n-2}}, c) = \overline{q_{n-1}}, \delta_B^{-1}(\overline{q_{n-1}}, c) = \overline{q_{n-2}}, \\ - \delta_B^{-1}(\overline{q_i}, b) &= \{\overline{q_i}\} \text{ for } 1 \leq i \leq n-1, \delta_B^{-1}(\overline{q_0}, b) = \{\overline{q_0}, \overline{q_1}\}, \\ - \delta_B^{-1}(\overline{q_i}, a) &= \overline{q_i} \text{ for } 0 \leq i \leq n-1. \end{aligned}$$

The DFA  $\mathcal{W}_3(a, b, c, d)$  is shown in Figure 1 and the DFA  $B$  and the NFA  $\overline{B}$  are shown in Figure 2.

We begin by showing these states are all reachable. First, we consider states of the form  $\langle p_i, \emptyset, \emptyset \rangle$ . We have  $\langle p_0, \emptyset, \emptyset \rangle \xrightarrow{a^i} \langle p_i, \emptyset, \emptyset \rangle$  for  $1 \leq i < m-2$ . Then

$$\langle p_{m-2}, \emptyset, \emptyset \rangle \xrightarrow{a} \langle p_{m-1}, \{q_0\}, \{\overline{q_{n-1}}\} \rangle \xrightarrow{a(ac^2)^{n-2}} \langle p_0, \{q_{n-1}\}, \{\overline{q_{n-1}}\} \rangle.$$

There are  $m-1$  such states.

Next, we consider states  $\langle p, P, \overline{R} \rangle$ , where  $P$  and  $\overline{R}$  are not empty. We first consider states  $\langle p_0, \{q_{n-1}\}, \overline{R} \rangle$  with  $\overline{R} \subseteq \overline{Q_B}$  and show that all such states are reachable by induction on the size of  $\overline{R}$ . First, we have

$$\langle p_0, \{q_{n-1}\}, \{\overline{q_{n-1}}\} \rangle \xrightarrow{d^{n-1-i}} \langle p_0, \{q_{n-1}\}, \{\overline{q_i}\} \rangle$$

for  $0 \leq i \leq n-1$  and thus all states with  $|\overline{R}| \leq 1$  are reachable.



Now, suppose that every state  $\langle p_0, \{q_{n-1}\}, \overline{T} \rangle$  with  $|\overline{T}| \leq \ell$  is reachable. We will show how to reach  $\langle p_0, \{q_{n-1}\}, \overline{R} \rangle$  with  $|\overline{R}| = \ell + 1$ . Let  $\overline{R} = \{\overline{q_{j_0}}, \dots, \overline{q_{j_\ell}}\} \subseteq \overline{Q_B}$  with  $n-1 \geq j_0 > \dots > j_\ell \geq 0$ .

If  $j_\ell = 0$  and  $j_{\ell-1} = 1$ , then  $\overline{R}$  is reachable from a set of size  $\ell$  as follows,

$$\langle p_0, \{q_{n-1}\}, \{\overline{q_{j_0}}, \dots, \overline{q_{j_{\ell-2}}}, \overline{q_0}\} \rangle \xrightarrow{b^2} \langle p_0, \{q_{n-1}\}, \{\overline{q_{j_0}}, \dots, \overline{q_{j_{\ell-2}}}, \overline{q_0}, \overline{q_1}\} \rangle.$$

If  $j_0 = n-1$  and  $j_1 = n-2$ , then  $\overline{R}$  is reachable from a state of size  $\ell + 1$  containing  $\overline{q_0}$  and  $\overline{q_1}$ , which we have just shown is reachable, as follows,

$$\begin{aligned} & \langle p_0, \{q_{n-1}\}, \{\overline{q_0}, \overline{q_1}, \overline{q_{j_0-2}}, \dots, \overline{q_{j_{\ell-2}-2}}\} \rangle \\ & \xrightarrow{d^2} \langle p_0, \{q_{n-1}\}, \{\overline{q_{n-1}}, \overline{q_{n-2}}, \overline{q_{j_0}}, \dots, \overline{q_{j_{\ell-2}}}\} \rangle. \end{aligned}$$

Then all other states with sets  $\overline{R}$  of size  $\ell + 1$  are reachable from a set of size  $\ell + 1$  containing  $\overline{q_{n-1}}$  and  $\overline{q_{n-2}}$ , which we have just shown is reachable, as follows,

$$\begin{aligned} & \langle p_0, \{q_{n-1}\}, \{\overline{q_{n-1}}, \overline{q_{n-2}}, \overline{q_{j_2+n-2-j_1}}, \dots, \overline{q_{j_\ell+n-2-j_1}}\} \rangle \\ & \xrightarrow{(dc)^{j_0-j_1-1}} \langle p_0, \{q_{n-1}\}, \{\overline{q_{n-1}}, \overline{q_{j_1+n-1-j_0}}, \overline{q_{j_2+n-1-j_0}}, \dots, \overline{q_{j_\ell+n-1-j_0}}\} \rangle \\ & \xrightarrow{d^{n-1-j_0}} \langle p_0, \{q_{n-1}\}, \{\overline{q_{j_0}}, \dots, \overline{q_{j_\ell}}\} \rangle \end{aligned}$$

Thus, we have shown that all states  $\langle p_0, \{q_{n-1}\}, \overline{R} \rangle$  are reachable.

Now, we consider states of the form  $\langle p_0, P, \overline{R} \rangle$  for  $P \subseteq Q_B$  and  $\overline{R}$  fixed. First, we must set  $\overline{R}$  appropriately. To do this, set

$$\overline{R_0} = \{\overline{q_{n-1}}, \overline{q_{j_1+(n-1-j_0)}}, \dots, \overline{q_{j_\ell+(n-1-j_0)}}\}.$$

From above, we know that  $\langle p_0, \{q_{n-1}\}, \overline{R_0} \rangle$  is reachable. Then we have

$$\langle p_0, \{q_{n-1}\}, \overline{R_0} \rangle \xrightarrow{a(ac^2)^{m-2}} \langle p_{m-1}, \{q_0\}, \overline{R_0} \rangle.$$

We will show that states  $\langle p_0, P, \overline{R_0} \rangle$  are reachable by induction on the size of  $P$ . First, for  $2 \leq i \leq n-1$ , we have

$$\langle p_{m-1}, \{q_0\}, \overline{R_0} \rangle \xrightarrow{a} \langle p_0, \{q_1\}, \overline{R_0} \rangle \xrightarrow{(ac^2)^{i-1}} \langle p_0, \{q_i\}, \overline{R_0} \rangle.$$

Thus, we have reached states  $\langle p_0, P, \overline{R_0} \rangle$  with  $|P| = 1$ .

Next, we will show that if states  $\langle p_{m-1}, \{q_0\} \cup T, \overline{R_0} \rangle$  with  $T \subseteq \{q_1, \dots, q_{n-1}\}$  and  $|T| = k$  are reachable, then the state  $\langle p_0, P, \overline{R_0} \rangle$  with  $P \subseteq Q_B$  and  $|P| = k+1$  is reachable. Let  $P = \{q_{i_0}, q_{i_1}, \dots, q_{i_k}\}$  with  $i_0 < i_1 < \dots < i_k$ . There are two cases to consider. First, if  $i_0 = 0$ , we have

$$\langle p_{m-1}, \{q_0, q_{i_2-i_1}, \dots, q_{i_k-i_1}\}, \overline{R_0} \rangle \xrightarrow{a(ac^2)^{i_1-1}} \langle p_0, P, \overline{R_0} \rangle.$$

Secondly, if  $i_0 > 0$ , then

$$\langle p_{m-1}, \{q_0, q_{i_1-(i_0-1)}, \dots, q_{i_k-(i_0-1)}\}, \overline{R_0} \rangle \xrightarrow{a(ac^2)^{i_0-1}} \langle p_0, P, \overline{R_0} \rangle.$$

Now, we will show that states  $\langle p_{m-1}, \{q_0\} \cup T, \overline{R_0} \rangle$  are reachable. Recall that  $\langle \{p_{m-1}, \{q_0\}, \overline{R_0}\} \rangle$  was shown to be reachable above. Then we consider  $|T| \geq 1$  with  $q_0 \notin T$  and we have

$$\langle p_0, \{q_{i_1-(m-1)}, \dots, q_{i_k-(m-1)}\}, \overline{R_0} \rangle \xrightarrow{a^{m-1}} \langle p_{m-1}, \{q_0, q_{i_1}, \dots, q_{i_k}\}, \overline{R_0} \rangle.$$

There are  $(2^{n-1})^2$  such states.

Finally, for  $0 \leq i \leq m-1$  and  $0 \leq i_1 < \dots < i_k \leq n-1$ , we have

$$\langle p_0, \{q_{i_0-i}, q_{i_1-i}, \dots, q_{i_k-i}\}, \overline{R_0} \rangle \xrightarrow{a^i} \langle p_i, P, \overline{R_0} \rangle.$$

Then to reach  $\overline{R}$  as originally intended, we have

$$\langle p_i, P, \overline{R_0} \rangle \xrightarrow{d^{n-1-j_0}} \langle p_i, P, \overline{R} \rangle.$$

There are  $(m-1)(2^n-1)^2$  such states. Thus, we have shown that all states of the form  $\langle p_i, P, \overline{R} \rangle$  are reachable and there are  $(m-1) + (2^{n-1})^2 + (2^n-1)^2$  such states.

Now, we will show that these states are pairwise distinguishable. Consider two states  $\langle q, P, \overline{R} \rangle$  and  $\langle q', P', \overline{R}' \rangle$ .

We first note that for  $\overline{R} = \overline{Q_B}$ , we have that  $\delta_C(\langle q, P, \overline{Q_B} \rangle, w) \in F_C$  for all  $q \in Q_A$ ,  $P \subseteq Q_B$ , and words  $w \in \Sigma^*$ . Thus, any states with  $\overline{R} = \overline{R}' = \overline{Q_B}$  are indistinguishable. There are  $2^{n-1}$  such states which can be merged into a single state.

Then, we consider when  $\overline{R} \neq \overline{R}'$ . There exists some state  $\overline{q_j} \in \overline{R}$  and  $\overline{q_j} \notin \overline{R}'$ . Then  $\overline{R}$  and  $\overline{R}'$  are distinguishable via the word  $d^j$ . Similarly, suppose that  $P \neq P'$ . Again, there exists a state  $q_i \in P$  and  $q_i \notin P'$ . Then  $P$  and  $P'$  are distinguishable by the word  $a^{n-1-i}$ . Note that these two cases can be dealt with independently, since  $a$  does not change  $\overline{R}$  and  $\overline{R}'$  and  $d$  does not change  $P$  and  $P'$ .

Finally, suppose  $q \neq q'$  and let  $q = p_i$  and without loss of generality let  $q' = p_j$  with  $i < j$  and consider the states  $\langle p_i, P, \overline{R} \rangle$  and  $\langle p_j, P, \overline{R} \rangle$ .

– If  $j < m-1$ , we have

$$\begin{aligned} \langle p_i, P, \overline{R} \rangle &\xrightarrow{(ba)^{m-1-j}} \langle p_{i-j+m-1}, P', \overline{R}' \rangle, \\ \langle p_j, P, \overline{R} \rangle &\xrightarrow{(ba)^{m-1-j}} \langle p_{m-1}, \{q_0\} \cup P', \{\overline{q_{n-1}}\} \cup \overline{R}' \rangle. \end{aligned}$$

These states can be distinguished because the second component of each state are distinguishable as previously shown.

– Similarly, if  $i < m-2$  and  $j = m-1$ , we have

$$\begin{aligned} \langle p_i, P, \overline{R} \rangle &\xrightarrow{a(ba)^{m-2-i}} \langle p_{m-1}, \{q_0\} \cup P', \{\overline{q_{n-1}}\} \cup \overline{R}' \rangle, \\ \langle p_{m-1}, P, \overline{R} \rangle &\xrightarrow{a(ba)^{m-2-i}} \langle p_k, P', \overline{R}' \rangle \end{aligned}$$

for some  $k < m-1$ . Again, the second components are different and may be distinguished.

- Finally, if  $i = m - 2$  and  $j = m - 1$ , applying the word  $a$  sends both states to states which fall under the case immediately above.

Thus, there are  $(m-1) + (2^{n-1})^2 + (2^n - 1)^2$  reachable states, of which  $2^{n-1}$  are indistinguishable. This gives a total of  $(m-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$  states which are reachable and pairwise distinguishable as desired.  $\square$

**Proposition 9.** *For  $n \geq 1$ , let  $A$  be an NFA with  $n$  states defined over an alphabet  $\Sigma$  and let  $\theta$  be an antimorphism. Then there exists an NFA that recognizes  $L(A)^{\circlearrowleft}_+$  with at most  $2n$  states. Furthermore, this bound can be reached in the worst case.* (p. 7)

*Proof.* Let  $A = (Q, \Sigma, \delta, I, F)$  be an  $n$ -state NFA. We denote by  $\bar{A}$  the NFA recognizing  $\theta(L(A))$ ,  $\bar{A} = (\bar{Q}, \Sigma, \delta^{-1}, \bar{F}, \bar{I})$ . We will define an NFA  $A'$  which recognizes  $A^{\circlearrowleft}_+$  with respect to an antimorphism  $\theta$  by  $A' = (Q', \Sigma, \delta', I', F')$ , where  $Q' = Q \cup \bar{Q}$ ,  $I' = I$ ,  $F' = F \cup \bar{I}$ , and the transition function  $\delta' : Q' \times \Sigma \rightarrow 2^{Q'}$  is defined for  $q \in Q'$  and  $a \in \Sigma$  by

$$\delta'(q, a) = \begin{cases} \delta(q, a) & \text{if } q \in Q, \\ \delta^{-1}(q, a) & \text{if } q \in \bar{Q}, \\ \delta(q, a) \cup I \cup \bar{F} & \text{if } q \in Q \text{ and } (\delta(q, a) \cap F) \neq \emptyset, \\ \delta^{-1}(q, a) \cup I \cup \bar{F} & \text{if } q \in \bar{Q} \text{ and } (\delta^{-1}(q, a) \cap \bar{I}) \neq \emptyset. \end{cases}$$

We claim that  $L(A') = L(A)^{\circlearrowleft}_+$ . To see this, suppose we have a word  $w \in L(A)^{\circlearrowleft}_+$ . We can write  $w = u_1 \cdots u_k$  for  $u_1 \in L(A)$  and  $u_i \in L(A) \cup \theta(L(A))$  for  $2 \leq i \leq k$ . Since  $u_1 \in L(A)$ , there is a path in  $A$  from  $I$  to a final state of  $A$  and so there is also such a path in  $A'$ . But every transition to a final state also reaches the initial states  $I$  and  $\bar{F}$  of  $A$  and  $\bar{A}$ , respectively. We can repeat the process for each  $u_i$  and the computation on  $w$  ends on a final state of  $A$  or  $\bar{A}$ . Thus,  $w \in L(A')$ .

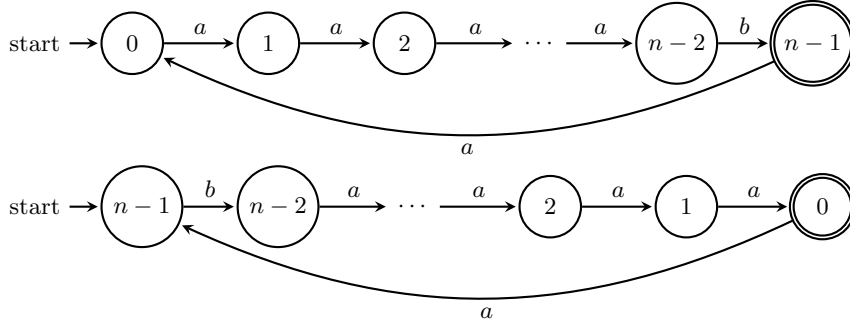
Now suppose that  $w \in L(A')$ . We claim that since  $w$  reaches a final state of  $A'$ , there must be a prefix  $u$  of  $w$  which reaches a state in  $F$  on  $A$ . Suppose otherwise. Since  $w$  cannot reach a state in  $F$ , it must reach a state in  $\bar{I}$ , a final state of  $\bar{A}$ . However, such states are unreachable in  $A'$  from an initial state in  $I'$ , since  $I' = I$  and only contains initial states of  $A$ . Thus we can write  $w = uv$  with  $u \in L(A)$ .

Now, we will show that  $v \in (L(A) \cup L(\bar{A}))^*$  by induction on  $|v|$ . If  $v = \varepsilon$ , then it is clear that this holds and we have  $w = u \in L(A) \subseteq L(A)^{\circlearrowleft}_+$ . Now let  $|v| = n$  and assume that for all words  $v' \in \Sigma^*$  with  $|v'| < n$ , we have  $v' \in (L(A) \cup L(\bar{A}))^* \text{pref}(L(A) \cup L(\bar{A}))$ . We can factor  $v = v_1 v_2$ . Since  $|v_1| < |v| = n$ ,  $v_1 \in (L(A) \cup L(\bar{A}))^* \cdot \text{pref}(L(A) \cup L(\bar{A}))$  by our assumption. However, we can clearly choose  $v_1$  to be shorter so that  $v_1 \in (L(A) \cup L(\bar{A}))^*$ . Then there is a computation on  $v_1$  which ends in a final state of  $A$  or  $\bar{A}$ . Then by definition, this computation also ends on an initial state of  $A$  or  $\bar{A}$ . Since  $v_2$  continues a computation from these states, it is a prefix of a word recognized by either  $A$  or

$\bar{A}$ . Then if  $v_2$  reaches a final state of  $A$  or  $\bar{A}$ , it must be a word recognized by either  $A$  or  $\bar{A}$ . Thus, we have  $v_2 \in L(A) \cup L(\bar{A})$ . This gives us  $v \in (L(A) \cup L(\bar{A}))^*$  and therefore,  $w \in L(A)^{\circ_+^{\theta}}$ .

From this construction,  $A'$  has at most  $2n$  states. Now, we will show this bound is reachable. Let  $L_n = a^{n-2}b(a^{n-1}b)^*$  and let  $\theta$  be the identity antimorphism with  $\theta(a) = a$  for all  $a \in \Sigma$ . This language is recognized by the  $n$ -state NFA shown in Figure 5. We will show that a fooling set  $S$  of size  $2n$  exists for  $L_n^{\circ_+^{\theta}}$  with respect to  $\theta$ . Let  $S = S_1 \cup S_2 \cup S_3$  where

- $S_1 = \{(a^i, a^{n-2-i}b) \mid 0 \leq i \leq n-2\}$ ,
- $S_2 = \{(a^{n-2}bba^i, a^{n-2-i}) \mid 0 \leq i \leq n-2\}$ ,
- $S_3 = \{(a^{n-2}b, \epsilon), (a^{n-2}bba^{n-1}, ba^{n-2})\}$ .



**Fig. 5.** The NFA recognizing  $L_n = a^{n-2}b(a^{n-1}b)^*$  and its reverse

Let  $i < j$  and consider pairs  $(x_i, y_i)$  and  $(x_j, y_j)$  in  $S_1$ . Then  $x_j y_i = a^{n-2-i+j}b \notin L_n^{\circ_+^{\theta}}$  since  $j - i > 0$ . Similarly, for pairs  $(x_i, y_i)$  and  $(x_j, y_j)$  in  $S_2$ , we have  $x_j y_i = a^{n-2}bba^{n-2-i+j} \notin L_n^{\circ_+^{\theta}}$ . Next, consider pairs  $(x_i, y_i) \in S_1$  and  $(x_j, y_j) \in S_2$ . Then  $x_i y_j = a^{i+n-2-j} \notin L_n^{\circ_+^{\theta}}$ .

Now, consider the two pairs in  $S_3$ . First, it is clear that  $a^{n-2}bba^{n-1} \notin L_n^{\circ_+^{\theta}}$ . Then for pairs  $(x_i, y_i) \in S_1$  and  $(a^{n-2}b, \epsilon)$ , we have  $x_i \cdot \epsilon = a^i \notin L_n^{\circ_+^{\theta}}$ . For pairs  $(x_i, y_i) \in S_1$  and  $(a^{n-2}bba^{n-1}, ba^{n-2})$ , we have  $x_i \cdot ba^{n-2} = a^i ba^{n-2} \notin L_n^{\circ_+^{\theta}}$ . Similarly, taking  $(x_i, y_i) \in S_2$  and  $(a^{n-2}b, \epsilon)$ , we have  $a^{n-2}b \cdot y_i = a^{n-2}ba^{n-2-i} \notin L_n^{\circ_+^{\theta}}$  and with  $(a^{n-2}bba^{n-1}, ba^{n-2})$ , we have  $a^{n-2}bba^{n-1} \cdot y_i = a^{n-2}bba^{n-1}a^{n-2-i} \notin L_n^{\circ_+^{\theta}}$ .

Thus,  $S$  is a fooling set of size  $2n$  for  $L_n^{\circ_+^{\theta}}$ .  $\square$

**Proposition 10.** For  $n \geq 3$ , let  $A$  be a DFA defined over an alphabet  $\Sigma$  with

$n$  states and let  $\theta$  be an antimorphism. Then there exists a DFA that recognizes  $L(A)^{\circledast}$  with at most  $2^{2n-1} - 2^n + 1$  states.

*Proof (Correctness of construction).* Let  $A = (Q, \Sigma, \delta, s, F)$ . We define the DFA  $A' = (Q', \Sigma, \delta', s', F')$  with the set of states

$$Q' = \{\langle P, \bar{R} \rangle \mid \emptyset \neq P \subseteq Q - F, \bar{R} \subseteq \bar{Q} - \{\bar{s}\}\} \\ \cup \{\langle P \cup \{s\}, \bar{R} \cup \bar{F} \rangle \subseteq Q \times \bar{Q} \mid (P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) \neq \emptyset\},$$

the initial state

$$s' = \begin{cases} \langle \{s\}, \emptyset \rangle & \text{if } s \notin F, \\ \langle \{s\}, \bar{F} \rangle & \text{if } s \in F, \end{cases}$$

the set of final states  $F' = \{\langle P, \bar{R} \rangle \subseteq Q \times \bar{Q} \mid (P \cup \bar{R}) \cap (F \cup \{\bar{s}\}) \neq \emptyset\}$ , and the transition function for a state  $\langle P, \bar{R} \rangle$  and symbol  $a \in \Sigma$  with  $P' = \delta(P, a)$  and  $\bar{R}' = \delta^{-1}(\bar{R}, a)$  is defined by

$$\delta'(\langle P, \bar{R} \rangle, a) = \begin{cases} \langle P' \cup \{s\}, \bar{R}' \cup \bar{F} \rangle & \text{if } (P' \cup \bar{R}') \cap (F \cup \{\bar{s}\}) \neq \emptyset, \\ \langle P', \bar{R}' \rangle & \text{otherwise.} \end{cases}$$

Informally, DFA  $A'$  operates by first simulating a computation of  $A$ , since by definition, we have  $L(A)^{\circledast} = L(A)(L(A) \cup \theta(L(A)))^*$ . Once the computation reaches a final state of  $A$ , an initial state for  $A$  and  $\bar{A}$  is added to the current state set and the computation continues. Whenever the current state of  $A'$  contains a final state of  $A$  or  $\bar{A}$ , the initial states of both machines are added. The computation continues until the input is read and accepts if and only if a final state of  $A$  or  $\bar{A}$  is contained in the state of  $A'$  when the input has been read.

First, we will show that  $L(A)^{\circledast} \subseteq L(A')$ . Let  $w \in L(A)^{\circledast}$ . We can write  $w = u_1 u_2 \cdots u_k$  for  $k \geq 1$ , where  $u_1 \in L(A)$  and  $u_i \in L(A) \cup \theta(L(A))$  for  $2 \leq i \leq k$ . We claim that there exists a computation path for  $w$  in  $A'$ ,

$$s' \xrightarrow{u_1} \langle P_1, \bar{R}_1 \rangle \xrightarrow{u_2} \langle P_2, \bar{R}_2 \rangle \xrightarrow{u_3} \cdots \xrightarrow{u_k} \langle P_k, \bar{R}_k \rangle.$$

Since  $u_1 \in L(A)$ , there must exist a path in  $A$  from  $s$  to a final state  $q \in F$ . Then  $q \in P_1$  and by definition, we have  $s \in P_1$  and  $\bar{F} \subseteq \bar{R}_1$ . Then we observe the same argument holds for  $2 \leq i \leq k$ : since  $s \in P_{i-1}$  and  $\bar{F} \subseteq \bar{R}_{i-1}$ , if  $u_i \in L(A)$ , there exists a path in  $A$  beginning at  $s$  such that  $F \cap P_i \neq \emptyset$  and if  $u_i \in \theta(L(A))$ , then there exists a path in  $\bar{A}$  starting at  $\bar{F}$  such that  $\bar{s} \in \bar{R}_i$ . Then by definition  $s \in P_i$  and  $\bar{F} \subseteq \bar{R}_i$ . Furthermore, we have  $(P_k \cup \bar{R}_k) \cap (F \cup \{\bar{s}\}) \neq \emptyset$ . Thus,  $\langle P_k, \bar{R}_k \rangle$  is a final state of  $A'$  and the above computation path exists and therefore,  $w \in L(A')$ .

Next, we will show that  $L(A') \subseteq L(A)^{\circledast}$ . Let  $w$  be a word that reaches a state  $\langle P, \bar{R} \rangle$  in  $A'$ . We claim that we can factor  $w = uv$  such that  $u \in L(A)^{\circledast} \cup \{\varepsilon\}$  and  $v \in \text{pref}(L(A) \cup \theta(L(A)))$ . We will consider two cases.

First, suppose that we have  $u = \varepsilon$ . We claim that in this case,  $v \in \text{pref}(L(A))$ . To see this, we observe that if  $\bar{R} \neq \emptyset$ , then there must be a state  $\langle P', \bar{R}' \rangle$  with  $\bar{F} \subseteq \bar{R}'$  reachable on some prefix of  $v$ . However, such a state is only

reachable from  $s'$  if it is a final state  $A'$ . In other words, we have  $v = v_1v_2$  with  $v_1 \in L(A) \subseteq L(A)^{\circ+}$ , a contradiction. Then the only states reachable from  $s'$  are states  $\langle \{q\}, \emptyset \rangle$  with  $q \in Q$  and thus,  $v \in \text{pref}(L(A)) \subseteq \text{pref}(L(A) \cup \theta(L(A)))$  as desired.

Now consider  $u \neq \varepsilon$ . Then there exists a computation path  $s' \xrightarrow{u} \langle P', \bar{R}' \rangle \xrightarrow{v} \langle P, \bar{R} \rangle$  such that  $\langle P', \bar{R}' \rangle$  is a final state of  $A'$ . To see this, we observe that if  $\bar{R}' \neq \emptyset$ , some prefix of  $w$  must have entered a final state of  $A'$ , as argued above. Otherwise, we have the case  $u = \varepsilon$ . Therefore,  $w \in L(A)^{\circ+}$  and  $\langle P', \bar{R}' \rangle$  is a final state of  $A'$ . Then by definition, we have  $s \in P'$  and  $\bar{F} \subseteq \bar{R}'$  and thus,  $v \in \text{pref}(L(A) \cup \theta(L(A)))$ .

Then since  $w \in L(A)^{\circ+} \cdot \text{pref}(L(A) \cup \theta(L(A)))$ , if  $w$  reaches a final state of  $A'$ , we can write  $w = uv$  such that  $u \in L(A)^{\circ+} \cup \{\varepsilon\}$  and  $v \in L(A) \cup \theta(L(A))$ . Thus, we have  $w \in L(A)^{\circ+}$  and we have shown  $L(A') \subseteq L(A)^{\circ+}$  and therefore,  $L(A') = L(A)^{\circ+}$ .  $\square$

(p. 9)

**Lemma 11.** *Let  $n \geq 3$ . Then there exists an  $n$ -state DFA  $A$  and an antimorphism  $\theta$  over an alphabet of size 4 such that*

$$\text{sc}(L(A)^{\circ+}) \geq 2^{2n-1} - 2^n + 1.$$

*Proof.* Let  $\Sigma = \{a, b, c, d\}$  and let  $\theta : \Sigma \rightarrow \Sigma$  be the antimorphism defined by

$$\theta(a) = b \quad \theta(b) = a \quad \theta(c) = d \quad \theta(d) = c.$$

Let  $A = (Q, \Sigma, \delta, s, F)$ , with  $Q = \{0, 1, \dots, n-1\}$ ,  $s = 0$ ,  $F = \{0\}$  and the transition function  $\delta : Q \times \Sigma \rightarrow Q$  is defined by

- $\delta(i, a) = i + 1 \pmod n$  for  $0 \leq i \leq n-1$ ,
- $\delta(i, b) = i$  for  $1 \leq i \leq n-1$ ,  $\delta(0, b) = 1$ ,
- $\delta(i, c) = i + 1$  for  $0 \leq i \leq n-1$ ,  $\delta(n-1, c) = 1$ ,
- $\delta(i, d) = i$  for  $0 \leq i \leq n-1$ .

Then we have the NFA  $\bar{A} = (\bar{Q}, \Sigma, \delta^{-1}, \bar{F}, \{\bar{s}\})$  which recognizes  $\theta(L(A))$ , where the transition function  $\delta^{-1} : \bar{Q} \times \Sigma \rightarrow 2^{\bar{Q}}$  is defined by

- $\delta^{-1}(\bar{i}, a) = \{\bar{i}\}$  for  $i = 2 \leq i \leq n-1$ ,  $\delta^{-1}(\bar{1}, a) = \{\bar{0}, \bar{1}\}$ ,
- $\delta^{-1}(\bar{i}, b) = \{\bar{i} - 1 \pmod n\}$  for  $0 \leq i \leq n-1$ ,
- $\delta^{-1}(\bar{i}, c) = \{\bar{i}\}$  for  $0 \leq i \leq n-1$ ,
- $\delta^{-1}(\bar{i}, d) = \{\bar{i} - 1\}$  for  $i = 2 \leq i \leq n-1$ ,  $\delta^{-1}(\bar{1}, d) = \{\bar{0}, \bar{n} - 1\}$ .

The DFA  $A$  and the NFA  $\bar{A}$  are shown in Figure 3.

Let  $A'$  be the DFA which recognizes  $L(A)^{\circ+}$ , which we obtain by following the construction from Proposition 10. We will show that all states of  $A'$  are reachable.

First, we will show that all states  $\langle P, \emptyset \rangle$  with  $0 \notin P \subseteq Q$  are reachable by induction on the size of  $P$ . For  $|P| = 1$ , from the initial state  $s' = \langle \{0\}, \{\bar{0}\} \rangle$ , we have

$$s' \xrightarrow{a} \langle \{1\}, \emptyset \rangle \xrightarrow{a^{i-1}} \langle \{i\}, \emptyset \rangle$$

for  $1 \leq i \leq n-1$  and thus every state with  $|P| = 1$  and  $0 \notin P$  is reachable. Now, let  $S = \langle P, \emptyset \rangle$  with  $P = \{i_1, \dots, i_k\}$  where  $1 \leq i_1 < \dots < i_k \leq n-1$  and assume that all states  $\langle P', \emptyset \rangle$  with  $|P'| < k$  are reachable. We will show that  $P$  is reachable. First, observe that for any state  $\langle P', \emptyset \rangle$ , if  $t = \max P'$ , then

$$\langle P', \emptyset \rangle \xrightarrow{a^{n-t}} \langle P'', \{\bar{0}\} \rangle$$

such that  $P'' \subseteq Q$  with  $0 \in P''$  and  $|P''| = |P'|$ . Consider the set  $\{0, i_3 - (i_2 - i_1 - 1), \dots, i_k - (i_2 - i_1 - 1)\}$  of size  $k-1$ . We have

$$\begin{aligned} & \langle \{0, i_3 - (i_2 - i_1 - 2), \dots, i_k - (i_2 - i_1 - 2)\}, \{\bar{0}\} \rangle \\ & \xrightarrow{cb} \langle \{1, i_3 - (i_2 - i_1 - 1), \dots, i_k - (i_2 - i_1 - 1)\}, \{\overline{n-1}\} \rangle \\ & \xrightarrow{c^{i_2-1}} \langle \{i_2 - i_1, i_3 - i_1, \dots, i_k - i_1\}, \{\overline{n-1}\} \rangle \\ & \xrightarrow{b^{n-1}} \langle \{0, i_2 - i_1, \dots, i_k - i_1\}, \{\bar{0}\} \rangle \\ & \xrightarrow{a^{i_1}} \langle \{i_1, i_2, \dots, i_k\}, \emptyset \rangle. \end{aligned}$$

Thus, we have shown that  $P$  is reachable and therefore all states of the form  $\langle P, \emptyset \rangle$  with  $0 \notin P$  are reachable. There are  $2^{n-1} - 1$  such states. This also shows that all states of the form  $\langle P, \{\bar{0}\} \rangle$  with  $0 \in P$  are reachable.

We will now show that all states  $\langle P, \bar{R} \rangle$  with  $P \subseteq Q$  and  $\bar{R} \subseteq \bar{Q}$  are reachable by induction on the size of  $|\bar{R}|$ . First, we will show that states  $\langle P, \bar{R} \rangle$  with  $|\bar{R}| = 1$  are reachable. From above, all states  $\langle P, \{\bar{0}\} \rangle$  with  $0 \in P$  are reachable. Let  $P = \{i_1, i_2, \dots, i_k\}$  be a set of size  $k$  with  $1 \leq i_1 < i_2 < \dots < i_k \leq n-1$ . Then for  $1 \leq j \leq n-1$ ,

$$\begin{aligned} & \langle \{0, i_2 - i_1, \dots, i_k - i_1\}, \{\bar{0}\} \rangle \\ & \xrightarrow{b} \langle \{1, i_2 - i_1 + 1, \dots, i_k - i_1 + 1\}, \{\overline{n-1}\} \rangle \\ & \xrightarrow{a^{i_1-1}} \langle \{i_1, i_2, \dots, i_k\}, \{\overline{n-1}\} \rangle \\ & \xrightarrow{b^{n-1-j}} \langle \{i_1, i_2, \dots, i_k\}, \{\bar{j}\} \rangle. \end{aligned}$$

Thus, every state with  $|\bar{R}| = 1$  is reachable.

Now, let  $S = \langle P, \bar{R} \rangle$  with  $|\bar{R}| = \ell$  and suppose that all states  $\langle P, \bar{R}' \rangle$  with  $|\bar{R}'| < \ell$  are reachable. Let  $P = \{i_1, \dots, i_k\}$  where  $0 \leq i_1 < \dots < i_k \leq n-1$  and  $\bar{R} = \{\bar{j}_1, \dots, \bar{j}_\ell\}$  where  $n-1 \geq j_1 > \dots > j_\ell \geq 0$ .

First, we consider how to reach states with  $0 \in P$  and  $\bar{0} \in \bar{R}$ . That is, we set  $i_1 = 0$  and  $\bar{j}_\ell = \bar{0}$ . There are two cases. First, if  $i_2 > 1$ , we can reach  $S$  from a state with  $|P'| = k$ ,  $|\bar{R}'| = \ell - 1$ ,

$$\langle P', \bar{R}' \rangle = \langle \{n-1, i_2-1, \dots, i_k-1\}, \{\bar{j}_1, \dots, \bar{j}_{\ell-1}\} \rangle \xrightarrow{a} S.$$

If  $i_2 = 1$ , we can reach  $S$  from a state with  $|P'| = k$ ,  $|\bar{R}'| = \ell - 1$  by

$$\langle P', \bar{R}' \rangle = \langle \{n-1, i_2-2, \dots, i_k-2\}, \{\bar{j}_1, \dots, \bar{j}_{\ell-1}\} \rangle \xrightarrow{ac} S.$$

Thus, we have shown how to reach states  $\langle P, \bar{R} \rangle$  with  $0 \in P$  and  $\bar{0} \in \bar{R}'$ . There are  $(2^{n-1})^2$  such states.

Now, we will show how to reach states  $\langle P, \bar{R} \rangle$  with  $0 \notin P$  and  $\bar{0} \notin \bar{R}'$ . That is,  $i_1 > 0$  and  $\bar{j}_\ell < 0$ . We can reach a state  $\langle P', \bar{R}' \rangle$  with  $0 \in P'$ ,  $\bar{0} \in \bar{R}'$ ,  $|P'| = k$  and  $|\bar{R}'| = \ell$  as shown above. Then we can reach  $S$  by

$$\langle P', \bar{R}' \rangle = \langle \{0, i_2 - i_1, \dots, i_k - i_1\}, \{\bar{0}, \bar{j}_\ell - \bar{j}_2, \dots, \bar{j}_\ell - \bar{j}_{\ell-1}\} \rangle \xrightarrow{bc^{i_1-1}b^{n-1-j_1}} S.$$

Thus, we have shown how to reach states  $\langle P, R \rangle$  such that  $P$  and  $R$  do not contain any final states. There are  $(2^{n-1} - 1)^2$  such states. Together, we have shown that there are  $(2^{n-1})^2 + (2^{n-1} - 1)^2 + 2^{n-1} - 1$  reachable states in  $A'$ .

We will now show that each of these states is pairwise distinguishable. First, we note that any state  $\langle P, \bar{Q} \rangle$  is indistinguishable, since for every  $P \subseteq Q$  and  $w \in \Sigma^*$ , we have  $\delta'(\langle P, \bar{Q} \rangle, w) \in F'$ . There are  $2^{n-1}$  such states which can be merged into a single state.

Now consider two states  $\langle P, \bar{R} \rangle$  and  $\langle P', \bar{R}' \rangle$  with  $\bar{R}, \bar{R}' \neq \bar{Q}$ . If  $\bar{R} \neq \bar{R}'$ , then there exists a state  $\bar{j} \in \bar{R}$  and  $\bar{j} \notin \bar{R}'$ . These two states can be distinguished by the word  $b^{\bar{j}}$ .

Next, suppose  $P \neq P'$  and assume that  $\bar{R} = \bar{R}'$ , since otherwise, the two states can be distinguished as above. Then there exists a state  $i \in P$  and  $i \notin P'$ . Without loss of generality, if  $0 \in P$  and  $0 \notin P'$ , then the two states are distinguishable by the empty word  $\varepsilon$ .

Otherwise, either  $0$  is in both  $P$  and  $P'$  or it is in neither. In this case, if  $\bar{1} \notin \bar{R}$ , then the two states can be distinguished by  $a^{n-i}$ . Otherwise, if  $\bar{1} \in \bar{R}$ , then there is some state  $j \in \bar{Q} - \bar{R}$  which is not in  $\bar{R}$ , since  $\bar{R} \neq \bar{Q}$ . We have

$$\begin{aligned} \langle P, \bar{R} \rangle &\xrightarrow{a^{j-1}} \langle P \cup \{0\}, \bar{R}'' \rangle \\ \langle P', \bar{R} \rangle &\xrightarrow{a^{j-1}} \langle P' \cup \{0\}, \bar{R}'' \rangle \end{aligned}$$

such that  $\bar{1} \notin \bar{R}''$ . Then the two states can be distinguished by  $a^{n-i}$  as above.

Then there are  $(2^{n-1})^2 + (2^{n-1} - 1)^2 + 2^{n-1} - 1$  reachable states, of which  $2^{n-1}$  are indistinguishable. This gives a total of  $2^{2n-1} - 2^n + 1$  reachable and pairwise distinguishable states as desired.  $\square$