The Exact Complexity of Star-Complement-Star

Jozef Jirásek[1] and Galina Jirásková[2(✉)]

[1] Institute of Computer Science, Faculty of Science, P. J. Šafárik University, Jesenná 5, 040 01 Košice, Slovakia
jozef.jirasek@upjs.sk

[2] Mathematical Institute, Slovak Academy of Sciences, Grešákova 6, 040 01 Košice, Slovakia
jiraskov@saske.sk

Abstract. We show that the state complexity of the star-complement-star operation is given by $\frac{3}{2}f(n-1) + 2f(n-2) + 2n - 5$, where $f(2) = 2$ and $f(n) = \sum_{i=1}^{n-2}\binom{n}{i}f(n-i) + 2$. The function $f(n)$ counts the number of distinct resistances possible for n arbitrary resistors each connected in series or parallel with previous ones, or the number of labeled threshold graphs on n vertices, and $f(n) \sim n!(1-\ln 2)/(\ln 2)^{n+1} = 2^{n\log n - 0.91n + o(n)}$. Our witness language is defined over a quaternary alphabet, and we strongly conjecture that the size of the alphabet cannot be decreased.

1 Introduction

The Kuratowski 14-theorem states that applying the operations of closure and complementation to a set in a topological space in any order and any number of times results in at most 14 distinct sets. The Kuratowski algebras in the settings of formal languages have been investigated by Brzozowski et al. [3]. They showed that at most 14 distinct languages may be produced by applying the star and complementation operations to a given language. Moreover, every such language can be expressed, up to inclusion of the empty string, as one of the following 5 languages and their complements: L, L^+, L^{c+}, L^{+c+}, and L^{+c+c}; here L^c denotes the complement and L^+ denotes the positive closure of L, and we use an exponent notation as follows: $L^{+c} = (L^+)^c$, $L^{+c+} = ((L^+)^c)^+$, etc.

While a language and its complement have the same complexity, and the complexity of L^+ is known to be $\frac{3}{4}2^n - 1$ since 70's [6], the only language in this chain which could possibly have a double-exponential state complexity was L^{+c+}. Surprisingly, as shown in [4], its state complexity is in $2^{\Theta(n\log n)}$, and lower bound

J. Jirásek—Research supported by VEGA grant 1/0056/18 and grant APVV-15-0091.

G. Jirásková—Research supported by VEGA grant 2/0084/15 and grant APVV-15-0091.

C. Câmpeanu (Ed.): CIAA 2018, LNCS 10977, pp. 223–235, 2018.
https://doi.org/10.1007/978-3-319-94812-6_19

example has been defined over a seven-letter alphabet. Nevertheless, since Θ is in the exponent, the gap between lower and upper bound remained large.

In this paper, we continue this research by a careful inspection of reachable and unreachable states in the resulting automaton, and we get the exact state complexity of the star-complement-star operation.

This complexity is given by $\frac{3}{2}f(n-1) + 2f(n-2) + 2n - 5$, where the function $f(n)$ counts, for example, the number of distinct resistances possible for n arbitrary resistors each connected in series or parallel with previous ones [1,12], or the number of labeled threshold graphs on n vertices [2], and $f(n) \sim n!(1 - \ln 2)/(\ln 2)^{n+1}$. Our witness language is defined over a quaternary alphabet, and we strongly conjecture that the size of alphabet is optimal.

2 Preliminaries

Let Σ be a finite non-empty alphabet of symbols. Then Σ^* denotes the set of strings over Σ including the empty string ε. A language is any subset of Σ^*. For a language L over Σ, the complement of L is the language $L^c = \Sigma^* \setminus L$. The (Kleene) star of a language L is the language $L^* = \bigcup_{i \geq 0} L^i$ where $L^0 = \{\varepsilon\}$ and $L^i = LL^{i-1}$. The positive closure of L is the language $L^+ = \bigcup_{i \geq 1} L^i$.

A *nondeterministic finite automaton* (NFA) is a quintuple $A = (Q, \Sigma, \cdot, I, F)$, where Q is a finite non-empty set of states, Σ is a finite non-empty input alphabet, the function $\cdot$ is the transition function that maps $Q \times \Sigma$ to 2^Q, $I \subseteq Q$ is the set of initial states, and $F \subseteq Q$ is the set of final (or accepting) states [9]. We say that (p, a, q) is a transition in the NFA A if $q \in p \cdot a$. The transition function is extended to the domain $2^Q \times \Sigma^*$ in the natural way. The language accepted by the NFA A is the set of strings $L(A) = \{w \in \Sigma^* \mid I \cdot w \cap F \neq \emptyset\}$.

An NFA A is a (complete) *deterministic finite automaton* (DFA) if $|I| = 1$ and for each state p and each input symbol a, the set $p \cdot a$ has exactly one element. In such a case, we write $p \cdot a = q$ instead of $p \cdot a = \{q\}$. We also use $p \xrightarrow{a} q$ to denote that $p \cdot a = q$. A DFA $A = (Q, \Sigma, \cdot, s, F)$ is minimal if all its states are reachable from the initial state, and every two distinct states are distinguishable.

The state complexity of a regular language L, $\mathrm{sc}(L)$, is the number of states of the minimal DFA recognizing the language L. The state complexity of the star-complement-star operation is the function from $\mathbb{N}$ to $\mathbb{N}$ defined as $n \mapsto \max\{\mathrm{sc}(L^{*c*}) \mid \mathrm{sc}(L) \leq n\}$.

The transitions on symbol a in a DFA A perform a transformation $a\colon Q \to Q$ defined by $qa = q \cdot a$. For a subset S of Q, we denote $Sa = \{qa \mid q \in S\}$ and $aS = \{q \in Q \mid qa \in S\}$. We say that a acts as a permutation on $S \subseteq Q$ if $Sa = S$; in such a case, for each $q \in S$ the set $aq = \{p \in Q \mid pa = q\}$ is nonempty.

A *cyclic permutation* of a set $\{q_1, q_2, \ldots, q_k\} \subseteq Q$ is a permutation a such that $q_i a = q_{i+1}$ if $1 \leq i \leq k-1$, $q_k a = q_1$, and $qa = q$ if $q \in Q \setminus \{q_1, q_2, \ldots, q_k\}$. We denote such a permutation as $a\colon (q_1, q_2, \ldots, q_k)$. For two states p and q, we use $(p \to q)$ to denote the transformation that maps p to q and fixes every state different from p. Each input string u also performs a transformation on Q given by the composition of its input symbols, that is, if $w = av$ with $a \in \Sigma$

and $v \in \Sigma^*$, then $w\colon Q \to Q$ is given by $qw = qav = (q \cdot a) \cdot v$. For $S \subseteq Q$ and $w \in \Sigma^*$, we denote $Sw = \{qw \mid q \in S\}$ and $wS = \{q \in Q \mid qw \in S\}$.

Every NFA $A = (Q, \Sigma, \cdot, I, F)$ can be converted to an equivalent deterministic automaton $\mathcal{D}(A) = (2^Q, \Sigma, \cdot, I, \{S \in 2^Q \mid S \cap F \neq \emptyset\})$ [8]. The DFA $\mathcal{D}(A)$ is called the *subset automaton* of the NFA A.

In what follows we use $[i, j]$ to denote the set of integers $\{\ell \mid i \leq \ell \leq j\}$ and we set $Q_n = [0, n-1]$.

3 Constructions of Automata for Plus-Complement-Plus

Let a language L be accepted by a DFA $A = (Q_n, \Sigma, \cdot, 0, F)$. Let $F^c = Q_n \setminus F$.

Let us construct the following automata.

- NFA A^+ for L^+ is constructed from the DFA A by adding the transition $(q, a, 0)$ whenever $q \cdot a \in F$. The set of final states of A^+ is F.
- DFA B for L^+ is the subset automaton $\mathcal{D}(A^+)$ restricted to reachable states. Thus, states of B are subsets of Q_n, its initial state is $\{0\}$, and its state S is final if $S \cap F \neq \emptyset$. Moreover, if a state of B contains a final state of A, then it also contains the state 0 since A^+ always has the transition $(q, a, 0)$ if $q \cdot a \in F$.
- DFA C for L^{+c} is constructed from the DFA B by interchanging the final and non-final states. Thus, states of C are subsets of Q_n, its initial state is $\{0\}$, and its state S is final if $S \subseteq F^c$, and it is non-final if $S \nsubseteq F^c$ in which case it also must contain the initial state 0 of A.
- NFA C^+ for L^{+c+} is constructed from the DFA C by adding the transition $(S, a, \{0\})$ whenever $S \cdot a$ is final, that is, whenever $S \cdot a \subseteq F^c$.

Then, the subset automaton $\mathcal{D}(C^+)$ is a DFA for $(L(A))^{+c+}$, and we are interested in the number of its reachable and distinguishable states. A DFA for L^{*c*} may require one more state to accept the empty string.
It follows from the constructions above that in the subset automaton $\mathcal{D}(C^+)$:

– each state $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ with $T_i \subseteq Q_n$ is a set of subsets of Q_n;
– if some $T_i \subseteq F^c$, then $\{0\} \in \mathcal{T}$;
– otherwise, for each i, we have $T_i \nsubseteq F^c$ and $0 \in T_i$;
– the transitions are given by

$$\mathcal{T} \xrightarrow{a} \bigcup_{\substack{T \in \mathcal{T} \\ T \cdot a \subseteq F^c}} \{T \cdot a, \{0\}\} \cup \bigcup_{\substack{T \in \mathcal{T} \\ T \cdot a \nsubseteq F^c}} \{T \cdot a \cup \{0\}\}.$$

It is known that each state of $\mathcal{D}(C^+)$ is equivalent to an antichain of subsets of Q_n [4, Lemma 1]; recall that a set of subsets of Q_n is an antichain if every two distinct elements in it are incomparable with respect to the set inclusion. This means that each final state of $\mathcal{D}(C^+)$ is equivalent to an antichain $\mathcal{T} = \{\{0\}, T_2, \ldots, T_k\}$ with $k \geq 2$ and all $T_i \subseteq F^c$ $(2 \leq i \leq k)$, while each its non-final state is equivalent to an antichain $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ with $k \geq 1$, $T_i \nsubseteq F^c$

and $0 \in T_i$ $(1 \leq i \leq k)$. Thus, the upper bound on the number of reachable and pairwise distinguishable states of $\mathcal{D}(C^+)$ is given by the number of antichains of subsets of an n-element set, known as the Dedekind number; see, for example, [5]. This number still grows double-exponentially. The next two propositions show that all these antichains can be distinguished using a growing alphabet.

Proposition 1. *Every two distinct antichains differ in a set S such that S is in one of them, and no subset of S is in the other one.*

Proof. Let $\mathcal{S}$ and $\mathcal{T}$ be two distinct antichains. Then without loss of generality, there is a set S with $S \in \mathcal{S} \setminus \mathcal{T}$. If no subset of S is in $\mathcal{T}$, then S is the desired set. Otherwise, $\mathcal{T}$ contains some subset T of S. Since $\mathcal{S}$ is an antichain, it cannot contain T and it cannot contain any of its subsets. Then T is the desired set. $\square$

Proposition 2. *There exists an n-state* DFA *A defined over an alphabet of size 2^n such that all the antichains are pairwise distinguishable in the subset automaton $\mathcal{D}(C^+)$.*

Proof. Define an n-state DFA $A = (Q_n, \{b_S \mid S \subseteq Q_n\}, \cdot, 0, \{0, 1\})$ where for each $q \in Q_n$ and $S \subseteq Q_n$, we have $q \cdot b_S = 2$ if $q \in S$, and $q \cdot b_S = 0$ otherwise. Let $\mathcal{S}$ and $\mathcal{T}$ be two distinct antichains. By Proposition 1, we may assume that there is a set S in $\mathcal{S} \setminus \mathcal{T}$ such that no subset of S is in $\mathcal{T}$. Then by b_S, the antichain $\mathcal{S}$ is sent to a final antichain containing $\{2\}$, while $\mathcal{T}$ is sent to a non-final antichain since each of its components is sent to a set containing the state 0. $\square$

Hence, all the antichains can be distinguished, and the question about how many of them are reachable was partially answered in [4]. It has been shown that each reachable state of $\mathcal{D}(C^+)$ is equivalent to some antichain $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ which satisfies (1) $1 \leq k \leq n$; (2) $T_i = \{q_i\} \cup S_i$ where $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_k$ and $q_1, q_2, \ldots, q_k$ are pairwise distinct states in $Q_n \setminus S_k$ [4, Lemma 2].

This reduced the number of reachable antichains from a double-exponential to at most $\sum_{k=1}^{n} \binom{n}{k} k!(k+1)^{n-k} \in 2^{O(n \log n)}$. Moreover, a language over a seven-letter alphabet was described in [4, Proof of Corollary 2] such that every DFA for its star-complement-star has at least $\lceil \frac{n}{2} \rceil^{n - \lceil \frac{n}{2} \rceil} \in 2^{\Omega(n \log n)}$ states. Thus the state complexity of the star-complement-star operation is in $2^{\Theta(n \log n)}$. However, since Θ is in an exponent, the gap between the lower and upper bound is large.

In what follows we aim to get the exact state complexity of this combined operation. We call an antichain valid if it satisfies the above mentioned two conditions, and we count the number of valid antichains in the next section.

4 The Number of Valid Antichains

The aim of this section is to count all the valid antichains. After giving their explicit definition, we first get the number of the valid antichains such that each element occurs in their union. Then we use this number to count all valid antichains. Recall that we denote by $[i, j]$ the set of integers $\{\ell \mid i \leq \ell \leq j\}$.

Definition 3. *An antichain* $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ *of subsets of* $[1, n]$ *is* valid *if*

(1) $1 \le k \le n$;
(2) for each i, *we have* $T_i = \{q_i\} \cup S_i$ *where*
- $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_k$,
- $q_1, q_2, \ldots, q_k$ *are pairwise distinct states in* $[1, n] \setminus S_k$.

For an antichain $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$, let $\cup\mathcal{T} = \bigcup_{i=1}^{k} T_i$ and $\cap\mathcal{T} = \bigcap_{i=1}^{k} T_i$.

Lemma 4. *Let* $n \ge 2$ *and* $f(n)$ *denote the number of valid antichains* $\mathcal{T}$ *of subsets of* $[1, n]$ *such that* $\cup\mathcal{T} = [1, n]$. *Then*

$$f(2) = 2 \text{ and } f(n) = \sum_{i=1}^{n-2} \binom{n}{i} f(n-i) + 2 \text{ if } n \ge 3. \tag{1}$$

Proof. Denote by $f_1(n)$ the number of valid antichains such that $\cup\mathcal{T} = [1, n]$ and $\{i\} \in \mathcal{T}$ for some i. Next, denote by $f_2(n)$ the number of valid antichains with $\cup\mathcal{T} = [1, n]$ and $i \in \cap\mathcal{T}$ for some i; notice that if $\mathcal{T}$ does not contain any singleton set, than $\cap\mathcal{T} \ne \emptyset$. A valid antichain with $\cup\mathcal{T} = [1, n]$ and containing a singleton set may contain either exactly one singleton set, or exactly two singleton sets, etc. Therefore,

$$f_1(n) = \sum_{i=1}^{n-2} \binom{n}{i} f_2(n-i) + 1; \tag{2}$$

notice that we cannot have exactly $n - 1$ one-element sets in an antichain. Similarly, if a valid antichain with $\cup\mathcal{T} = [1, n]$ does not contain any singleton set, then $\cap\mathcal{T}$ may have exactly one element, or exactly two elements, etc. Therefore,

$$f_2(n) = \sum_{i=1}^{n-2} \binom{n}{i} f_1(n-i) + 1; \tag{3}$$

First, let us prove that $f_1(n) = f_2(n)$. The proof is by induction on n. The basis, $n = 2$, holds true since $f_1(2) = 1$ because of the unique antichain $\{\{1\}, \{2\}\}$, and $f_2(2) = 1$ because of $\{\{1, 2\}\}$. Assume that $n \ge 3$ and that $f_1(i) = f_2(i)$ if $2 \le i \le n - 1$. Then $f_1(n) = f_2(n)$ follows from (2) and (3).

Next, consider all valid antichains $\mathcal{T}$ of subsets of $[1, n]$ with $\cup\mathcal{T} = [1, n]$, and denote the number of such antichains by $f(n)$. Notice that every such valid antichain either contains a singleton set or the intersection $\cap\mathcal{T}$ is non-empty. Therefore, $f(n) = f_1(n) + f_2(n)$, and we get

$$f(n) = \sum_{i=1}^{n-2} \binom{n}{i} (f_1(n-i) + f_2(n-i)) + 2 = \sum_{i=1}^{n-2} \binom{n}{i} f(n-i) + 2,$$

which concludes the proof. $\square$

Remark 5. The function $f(n)$ defines Sloane's sequence A005840 [10,11], and it counts the number of distinct resistances possible for n arbitrary resistors each connected in series or parallel with previous ones [1,12]. It also counts the number of labeled threshold graphs on n vertices [2]. The table of $f(n)$ for $n \leq 100$ can be found in [7]. These numbers are the coefficients of the generating function $(1-x)\mathrm{e}^x/(2-\mathrm{e}^x)$ [2], and

$$f(n) \sim n!(1-\ln 2)/(\ln 2)^{n+1} \in 2^{n\log n - n(\log\ln 2+\log \mathrm{e})+o(n)} \doteq 2^{n\log n-0.9139n+o(n)}.$$

Theorem 6. *Let $V(n)$ be the number of valid antichains of subsets of $[1,n]$. Then $V(n)=2f(n)+n-2$, where $f(n)$ is the function defined by (1).*

Proof. In $\cup\mathcal{T}$, nothing may be missing, or exactly one element may be missing, or exactly two elements may be missing, etc. Therefore,

$$V(n) = f(n) + \sum_{i=1}^{n-2}\binom{n}{i} f(n-i) + n,$$

where n is the number of antichains $\{\{i\}\}$ in which $n-1$ elements are missing in $\cup\mathcal{T}$. Hence $V(n)=2f(n)+n-2$, and the theorem follows. $\square$

5 Upper Bound

Our first aim is to show that some valid antichains are always unreachable. Recall that for an antichain $\mathcal{T}=\{T_1,T_2,\ldots,T_k\}$, $\cup\mathcal{T}=\bigcup_{i=1}^k T_i$ and $\cap\mathcal{T}=\bigcap_{i=1}^k T_i$.

Lemma 7 (Unreachable Antichains). *Let $n\geq 4$ and $A=(Q_n,\Sigma,\cdot,0,F)$ be a* DFA *with $|F|\geq 1$. Let $\mathcal{T}$ be a valid antichain as defined in Definition 3. If*

$$\{0,j\}\in\mathcal{T} \textit{ for some } j \textit{ in } Q_n\setminus\{0\}, \textit{ and } \cup\mathcal{T}=Q_n, \tag{4}$$

then $\mathcal{T}$ is unreachable in the subset automaton $\mathcal{D}(C^+)$.

Proof. Let $\mathcal{T}=(T_1,T_2,\ldots,T_k)$ be a valid antichain satisfying (4). Then $k\geq 2$. Assume that $\mathcal{T}$ can be reached from a reachable antichain $\mathcal{S}$ by reading a symbol a in Σ. Our aim is to show that $\mathcal{S}$ satisfies (4) as well. Then, the lemma follows.

Denote by $\mathrm{Im}(a)=\{qa \mid q\in Q_n\}$. Since each state is in $\cup\mathcal{T}$, each state, with a possible exception of the state 0, must be in $\mathrm{Im}(a)$. Thus $0\in\mathrm{Im}(a)$ implies that a performs a permutation on Q_n, and in such a case, we must have $\{a1,a2,\ldots,a(n-1)\}\subseteq\cup\mathcal{S}$. If $0\notin\mathrm{Im}(a)$, then there is exactly one state r in $Q_n\setminus\{0\}$ such that $r=pa=qa$ and $p\neq q$. Consider three cases:

(a) $\{0,j\}$ is reached from a state S in $\mathcal{S}$ with $|S|\geq 3$. Let p,q,r be three distinct states in S. Since we can have neither $pa=qa=ra$ nor $pa=qa=0$, we must have $pa=0$ and $qa=ra=j$. This is a contradiction since $0\in\mathrm{Im}(a)$ implies that a is a permutation.

(b) $\{0, j\}$ is reached from a state in a final antichain $\mathcal{S}$. Then we have $\cup\mathcal{S} \subseteq (Q_n \setminus F) \cup \{0\}$, so $|\cup\mathcal{S}| \leq n-1$. If $|F| \geq 2$, then $|\cup\mathcal{S}| \leq n-2$, and therefore

$$|\cup\mathcal{T}| \leq |(\cup\mathcal{S})a| \leq n-1,$$

a contradiction. If $|F| = 1$, then we must have $0a \in F$ to get a set $\{0, 0a\}$ in $\mathcal{T}$ from the set $\{0\}$ in $\mathcal{S}$. However, then the unique final state $0a$ cannot be in any other component of the antichain $\mathcal{T}$. Since $k \geq 2$, we must have $0 \in \mathrm{Im}(a)$ and $a0 \in \cup\mathcal{S}$. However, then a is a permutation and $\{a1, a2, \ldots, a(n-1)\} \subseteq \cup\mathcal{S}$. This is a contradiction since we have $|\cup\mathcal{S}| \leq n-1$. Notice that this case covers the reachability of $\{0, j\}$ from one-element sets, or from two-element sets $\{p, q\}$ with $p \neq 0$ and $q \neq 0$.

(c) $\{0, j\}$ is reached from a two-element state $\{0, q\}$ in $\mathcal{S}$ where $q \neq 0$. Then $\mathcal{S}$ is a non-final antichain of size at least 2, and therefore the state 0 is in each of its components. If $0a = j$, then we would have j in each component of $\mathcal{T}$, so $\mathcal{T}$ would not be an antichain, a contradiction. Therefore $0a = 0$ and $qa = j$. This means that $0 \in \mathrm{Im}(a)$, so a is a permutation, and $\{a1, a2, \ldots, a(n-1)\} \subseteq \cup\mathcal{S}$. Since $a0 = 0$ and $0 \in \cup\mathcal{S}$, we get $\cup\mathcal{S} = Q_n$, so $\mathcal{S}$ satisfies (4). □

Our next result provides an upper bound on the state complexity of star-complement-star. We discuss all possible choices of the final states in a given DFA, and show that the number of reachable valid antichains in the subset automaton for plus-complement-plus is maximal if the initial state of the given DFA is final, and if there is exactly one non-initial final state.

Theorem 8 (Star-Complement-Star: Upper Bound). *Let $n \geq 4$ and $A = (Q_n, \Sigma, \cdot, 0, F)$ be an n-state* DFA*. Then the language $L(A)^{*c*}$ is accepted by a* DFA *of at most $\frac{3}{2}f(n-1) + 2f(n-2) + 2n - 5$ states, where $f(n)$ is the function defined by (1).*

Proof. Recall that we denoted the number of valid antichains of subsets of $[1, n]$ by $V(n)$, and $V(n) = 2f(n) + n - 2$ by Theorem 6. We now discuss possible choices of the set of final states F of the DFA A. If $F = \emptyset$ then $L(A) = \emptyset$ and $L(A)^{*c*} = \Sigma^*$. If $F = \{0\}$, then $L(A) = L(A)^*$, so $L(A)^{*c} = L(A)^c$. Thus the state complexity of $L(A)^{*c}$ is n, and therefore the state complexity of $(L(A))^{*c*}$ is at most $\frac{3}{4}2^n$ [6,13], which is less than $\frac{3}{2}f(n-1) + 2f(n-2) + 2n - 5$ if $n \geq 4$.

Let $|F \cap [1, n-1]| = k$ where $k \geq 1$. By Lemma 7, the antichains containing $\{0, j\}$ for some j and with $\cup\mathcal{T} = Q_n$ are always unreachable in $\mathcal{D}(C^+)$, and there are $f_1(n-1) = f(n-1)/2$ of them. Next, only the following valid antichains may be reachable in $\mathcal{D}(C^+)$:

(i) the initial antichain $\{\{0\}\}$;
(ii) the final antichains $\{\{0\}, T_2, T_3, \ldots, T_k\}$ with $k \geq 2$ and $T_i \subseteq [1, n-1] \setminus F$;
(iii) the non-final antichains $\{T_1, T_2, \ldots, T_k\}$ with $k \geq 1$ and $T_i \cap F \neq \emptyset$, except for those containing $\{0, j\}$ for some j and with $\cup\mathcal{T} = Q_n$.

Notice that in case (ii), the antichain $\{T_2, T_3, \ldots, T_k\}$ is a valid antichain of subsets of $[1, n-1] \setminus F$, and that there are $V(n-1-k)$ such valid antichains.

Now consider antichains in case (iii). Since each T_i contains a final state of A, it also must contain the state 0, because in the construction of A^+ we added the transition $(p, a, 0)$ whenever $p \cdot a \in F$ in A.

If $0 \in F$, then each subset containing 0, is final in B, so non-final in C^+, so every antichain of the form $\{\{0\} \cup T_1, \{0\} \cup T_2, \ldots, \{0\} \cup T_k\}$ with $k \geq 1$ and $T_i \subseteq [1, n-1]$ $(1 \leq i \leq k)$ may possibly be reachable, except for those containing $\{0, j\}$ for some j and with $\cup\mathcal{T} = Q_n$. The number of such valid antichains is $V(n-1) - f(n-1)/2$.

However, if $0 \notin F$ and there is a state $q \in [1, n-1] \setminus F$, then the state $\{0, f\}$ is final in C^+, and therefore, in $\mathcal{D}(C^+)$ it only may be reached together with the initial state $\{0\}$ since in the construction of C^+, we added the transition $(S, a, \{0\})$ whenever $S \cdot a \subseteq F^c$. Thus the antichain $\{\{0, f\}\}$ considered in case (iii) is unreachable in this case. So the only way how to reach $V(n-1) - f(n-1)/2$ antichains in (iii) with $0 \notin F$ is to have $F = [1, n-1]$. However, in such a case, we do not have any final antichain.

Hence to get $V(n-1) - f(n-1)/2$ antichains in (iii) and at least one final antichain, we must have $0 \in F$. Finally, to get the maximal number of final antichains, we must have $k = 1$.

It follows that the number of reachable antichains in $\mathcal{D}(C^+)$ is maximal if $0 \in F$ and $|F \cap [1, n-1]| = 1$. In such a case, this number is equal to $1 + V(n-2) + V(n-1) - f(n-1)/2 = (3/2)f(n-1) + 2f(n-2) + 2n - 6$. Finally, to get a DFA for L^{*c*}, a new initial and final state may be required to accept the empty string. Our proof is complete. $\square$

6 Matching Lower Bound

Our next aim is to define a quaternary language such that the state complexity of its star-complement-star meets our upper bound given in Theorem 8. Recall that $[i, j] = \{\ell \mid i \leq \ell \leq j\}$ and $Q_n = [0, n-1]$.

Definition 9 (Quaternary Witness Language). *Let $n \geq 4$. Define an n-state* DFA $A = (Q_n, \{a, b, c, d\}, \cdot, 0, \{0, 1\})$, *where* a: $(0, 1, 2)$, b: $(1, 2, \ldots, n-1)$, c: $(2, 3, \ldots, n-1)$, *and* d: $(0 \to 2)$. *The* DFA A *is shown in Fig. 1.*

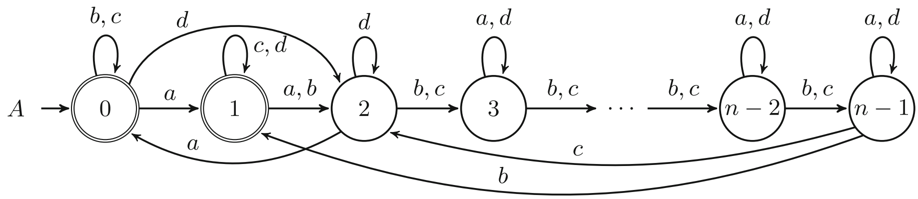

Fig. 1. A quaternary witness for star-complement-star meeting the upper bound $\frac{3}{2}f(n-1) + 2f(n-2) + 2n - 5$, where $f(n) = \sum_{i=1}^{n-2} \binom{n}{i} f(n-i) + 2$ and $f(2) = 2$.

Lemma 10. *Let $n \geq 4$ and A be an n-state* DFA *described in Definition 9. Let C^+ be the* NFA *for $L(A)^{+c+}$ described in Sect. 3. Let $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ be an antichain of subsets of Q_n such that*

(i) $1 \leq k \leq n$;
(ii) $T_i = \{q_i\} \cup S_i$ $(1 \leq i \leq k)$, *where* $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_k$ *and* $q_1, q_2, \ldots, q_k$ *are pairwise distinct states in* $Q_n \setminus S_k$;
(iii) either $T_1 = \{0\}$ *and* $T_i \subseteq [2, n-1]$ *if* $2 \leq i \leq k$, *or* $0 \in T_i$ *for each* i;
(iv) if $\{0, j\} \in \mathcal{T}$ *for some* $j \geq 1$, *then there is* q *in* $[1, n-1]$ *such that* $q \notin \cup\mathcal{T}$.

Then $\mathcal{T}$ is reachable in the subset automaton $\mathcal{D}(C^+)$. All these antichains are pairwise distinguishable, and there are $\frac{3}{2}f(n-1) + 2f(n-2) + 2n - 6$ of them.

Proof. To simplify the notation let us set $m := n - 1$. The proof is by induction on the size of an antichain $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$. Let $k = 1$, so $\mathcal{T} = \{T\}$. Then T must be a non-final state of the NFA C^+. This means that $0 \in T$. Let us show by induction on $|T|$ that the state $\{T\}$ is reachable in the subset automaton $\mathcal{D}(C^+)$. The set $\{\{0\}\}$ is the initial state of $\mathcal{D}(C^+)$. Let $T \subseteq Q_n$ and $0 \in T$. Let $j = \min(T \setminus \{0\})$. Then $T \setminus \{0, j\} \subseteq [j+1, m]$, $b^{j-1}(T \setminus \{0, j\}) \subseteq [2, m]$, and

$$\{\{0\} \cup ab^{j-1}(T \setminus \{0, j\})\} \xrightarrow{a} \{\{0, 1\} \cup b^{j-1}(T \setminus \{0, j\})\} \xrightarrow{b^{j-1}} \{\{0, j\} \cup (T \setminus \{0, j\})\} = \{T\},$$

where the starting set is reachable by the induction assumption; notice that $(0, a, 0)$ and $(0, b, 0)$ are transitions in the NFA A^+, so while reading any string over $\{a, b\}$, the NFA C^+ is always in a state containing 0, that is, in a rejecting state, therefore the initial state $\{0\}$ of C^+ cannot be added while reading such a string.

Now let $2 \leq k \leq n$ and assume that each antichain satisfying (i)–(iv) of size $k-1$ is reachable. Let $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ be an antichain odd size k satisfying (i)–(iv).

To simplify the exposition, let us use $q + S$ to denote the set $\{q\} \cup S$, where $q \in Q_n$ and $S \subseteq Q_n$. Then by (ii), $T_i = q_i + S_i$ where $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_k$, and $q_1, q_2, \ldots, q_k$ are pairwise distinct states in $Q_n \setminus S_k$. Consider several cases:

(1) Let $T_1 = \{0\}$. Then $T_i \subseteq [2, m]$ if $2 \leq i \leq k$, so $1 \notin \cup\mathcal{T}$.

 (a) First, let $T_2 = \{j\}$ for some j in $[2, m]$. Let u be a string in c^* such that $uj = 2$. Let $\mathcal{T}' = \{\{0\}, uT_3, \ldots, uT_k\}$. Then $\mathcal{T}'$ is an antichain of size $k-1$ which satisfies (i)–(iv). Therefore, $\mathcal{T}'$ is reachable by the induction assumption. Since

$$\mathcal{T}' = \{\{0\}, uT_3, \ldots, uT_k\} \xrightarrow{d} \{\{0\}, \{2 = uj\}, uT_3, \ldots, uT_k\} \xrightarrow{u} \{\{0\}, \{j\}, T_3, \ldots, T_k\} = \mathcal{T},$$

 the antichain $\mathcal{T}$ is reachable.

(b) Now let $|T_2| \geq 2$, so there is a state $j \in [2,m]$ such that $j \in \cap\mathcal{T}$. Let u be a string in c^* such that $uj = 2$, and therefore $2 \in \cap u\mathcal{T}$. Let $\mathcal{T}' = \{0+uT_2, 0+uT_3, \ldots, 0+uT_k\}$. Then $1 \notin \cup\mathcal{T}'$, so $\mathcal{T}'$ is an antichain of size $k-1$ satisfying (i)–(iv), and therefore it is reachable by the induction assumption. Since

$$\mathcal{T}' = \{0+uT_2, 0+uT_3, \ldots, 0+uT_k\} \xrightarrow{d} \{\{0\}, uT_2, uT_3, \ldots, uT_k\} \xrightarrow{u} \{\{0\}, T_2, T_3, \ldots, T_k\} = \mathcal{T},$$

the antichain $\mathcal{T}$ is reachable.

(2) Let $T_1 = \{0,1\}$. Then $0 \in T_i$ and $1 \notin T_i$ if $2 \leq i \leq k$. By (iv), there is a state q in $[2,m]$ such that $q \notin \cup\mathcal{T}$. Let u be the string in c^* such that $uq = 2$. Let $T_i' = T_i \setminus \{0\}$. Then

$$\{\{0\}, uq+uT_2', \ldots, uq+uT_k'\} \xrightarrow{a} \{\{0,1\}, 0+uT_2', \ldots, 0+uT_k'\} \xrightarrow{u} \{\{0,1\}, T_2, \ldots, T_k\} = \mathcal{T},$$

where the starting antichain satisfies (i)–(iv) and it is considered in case (1).

(3) Let $T_1 = \{0,j\}$ and $j \geq 2$. Then $0 \in T_i$ and $j \notin T_i$ if $2 \leq i \leq k$. By (iv), there is a state q in $[1,m]$ and $q \neq j$ such that $q \notin \cup\mathcal{T}$. Let u be a string in b^* such that $uj = 1$. Then $uq \neq 1$ and $uq \notin \cup u\mathcal{T}$. Next,

$$\{\{0,1\}, uT_2, \ldots, uT_k\} \xrightarrow{u} \{\{0,j\}, T_2, \ldots, T_k\},$$

where the starting antichain satisfies (i)–(iv), and it is considered in case (2).

(4) Let $|T_1| \geq 3$. We prove this case by induction on $|T_1|$. First, let $|T_1| = 3$. Then $0 \in \cap\mathcal{T}$ and there is a state $q \in T_1 \setminus \{0\}$ such that $q \in \cap\mathcal{T}$. Let u be a string in b^* such that $uq = 1$ and let $T_i' = T_i \setminus \{0,q\}$. Then $uT_i' \subseteq [2,m]$. Therefore,

$$\{0+auT_1', 0+auT_2', \ldots, 0+auT_k'\} \xrightarrow{a}$$
$$\{\{0,1\} \cup uT_1', \{0,1\} \cup uT_2', \ldots, \{0,1\} \cup uT_k'\} \xrightarrow{u} \mathcal{T}$$

since $0u = 0$ and $1u = uqu = q$. The starting antichain is considered in cases (2)–(3). The induction step is exactly the same, except that the starting set is reachable by induction on $|T_1|$.

To prove distinguishability, let $\mathcal{S}$ and $\mathcal{T}$ be two distinct antichains. By Proposition 1, we may assume that there is a set $S \in \mathcal{S} \setminus \mathcal{T}$ such that no subset of S is in $\mathcal{T}$. Notice that the set S must be different from $[0,m]$ because otherwise it is not true that no subset of $[0,m]$ is in $\mathcal{T}$. We also must have $S \neq [1,m]$ since $1 \in S$ implies $0 \in S$. Then S may be send to S' with $1 \notin S'$ using a string u in b^*; while still no subset of S' is in $\mathcal{T}u$. Thus, we may assume that $1 \notin S$.

First, let $\mathcal{S}$ and $\mathcal{T}$ be two final antichains. Then $S \subseteq [2,m]$. Let $i \in [2,m] \setminus S$. Then the string $u_i = c^{n-1-i}bc^{i-2}$ sends each state of S to itself, and the state i

to the state 1 in the DFA A. It follows that for each subset S of $[2, m]$, there is a string $u_S \in \{b, c\}^*$ (equal to the concatenation of strings u_i for $i \notin S$) by which S is sent to itself, while each set containing a state in $[2, m] \setminus S$ is sent to a set containing $\{0, 1\}$ in the NFA C^+; recall that $0 \cdot c = 0 \cdot b = 0$ and $1 \cdot c = 1$. It follows that the antichain $\mathcal{S}$ is send to a final antichain containing the state S by u_S. On the other hand, the antichain $\mathcal{T}$ is send to a non-final antichain equivalent to $\{\{0\}\}$ since $\{\{0\}\}$ remains in itself upon reading $u_S \in \{b, c\}^*$, while any other set in $\mathcal{T}$ is sent to a superset of $\{0, 1\}$ since it is not a subset of S.

If $\mathcal{S}$ and $\mathcal{T}$ are non-final, let $S' = S \setminus \{0\}$. Then $\mathcal{S}$ is sent to an antichain containing the set S by $u_{S'}$, while each set in $\mathcal{T}$ is sent to a superset of $\{0, 1\}$. Now we use the symbol d. Then $\mathcal{S} u_{S'} d$ is a final antichain containing the set $S' \cup \{2\}$, while $\mathcal{T} u_{S'} d$ is a non-final antichain of supersets of $\{0, 1\}$. ☐

Theorem 11 (Star-Complement-Star: Lower Bound; $|\Sigma| = 4$). *Let $n \geq 4$ and A be an n-state* DFA *from Definition 9. Then every* DFA *for the language $L(A)^{*c*}$ has at least $\frac{3}{2} f(n-1) + 2f(n-2) + 2n - 5$ states where $f(n)$ is the function defined in (1).*

Proof. First, notice that we have $L(A)^{*c*} = L(A)^{+c+} \cup \{\varepsilon\}$. To get an NFA C^* for $L(A)^{*c*}$, we add a new initial and final state q_0 to the NFA C^+ for $L(A)^{+c+}$. Thus C^* has two initial states, namely, q_0 and $\{0\}$, so the initial state of $\mathcal{D}(C^*)$ is $\{q_0, \{0\}\}$, and it is final. It is also the unique state of $\mathcal{D}(C^*)$ which contains the state q_0. By reading c it is sent to the initial state $\{\{0\}\}$ of the subset automaton $\mathcal{D}(C^+)$, which has $\frac{3}{2} f(n-1) + 2f(n-2) + 2n - 6$ reachable and pairwise distinguishable antichains by Lemma 10.

Let us show that the final state $\{q_0, \{0\}\}$ is distinguishable from any final antichain. To this aim, let $\mathcal{T} = \{\{0\}, T_2, \ldots, T_k\}$ be a final antichain where $k \geq 2$ and each T_i is a non-empty subset of $[2, n-1]$. Then, by c, the antichain $\mathcal{T}$ is sent to a final antichain, while $\{q_0, \{0\}\}$ is sent to the non-final antichain $\{\{0\}\}$. ☐

The next theorem summarizes our results.

Theorem 12 (State Complexity of Star-Complement-Star). *Let $n \geq 4$ and L be a language accepted by an n-state* DFA. *Then the language L^{*c*} is accepted by a* DFA *with $\frac{3}{2} f(n-1) + 2f(n-2) + 2n - 5$ states, where $f(2) = 2$ and $f(n) = \sum_{i=1}^{n-2} \binom{n}{i} f(n-i) + 2$, and $f(n) \sim n!(1 - \ln 2)/(\ln 2)^{n+1} \doteq 2^{n \log n - 0.91n + o(n)}$. This upper bound is tight, and it is met by the quaternary language recognized by the* DFA *$A = (\{0, 1, \ldots, n-1\}, \{a, b, c, d\}, \cdot, 0, \{0, 1\})$ where $a\colon (0, 1, 2)$, $b\colon (1, 2, \ldots, n-1)$, $c\colon (2, 3, \ldots, n-1)$, $d\colon (0 \to 2)$.* ☐

7 Conclusions

We proved that the exact state complexity of the star-complement-star operation is given by $\frac{3}{2} f(n-1) + 2f(n-2) + 2n - 5$ where $f(n)$ is the function that counts, for example, the number of distinct resistances possible for n arbitrary resistors, each connected in series or parallel with previous ones, or the number of labeled threshold graphs on n vertices. It defines Sloane's sequence A005840. These numbers are the coefficients of the generating function $(1-x)e^x/(2-e^x)$, and $f(n) \sim n!(1-\ln 2)/(\ln 2)^{n+1}$.

Our witness language is defined over a quaternary alphabet, we are most likely able to show that at least three symbols are necessary. Our computations, summarized in Table 1, show that the upper bound cannot be met by any ternary language if $n \geq 5$, but to prove this seems to be a challenging problem. A lower bound in the binary case is of interest too. On the other hand, the unary case is easy since L^{*c*} equals $\{\varepsilon\}$ if $a \in L$, and it equals a^* otherwise.

Table 1. Computations—the state complexity of plus-complement-plus: the binary and ternary case; lower bound from [4] with a witness over a seven-letter alphabet; the exact complexity with a quaternary lower bound example; the upper bound from [4].

n	$\|\Sigma\| = 2$	$\|\Sigma\| = 3$	$\lceil \frac{n}{2} \rceil^{n-\lceil \frac{n}{2} \rceil}$	State complexity of $L \mapsto L^{+c+}$ with a quaternary witness	$\sum_{k=1}^{n} \binom{n}{k} k!(k+1)^{n-k}$
4	11	18	4	18	260
5	29	77	9	89	2 300
6	134	468	27	596	24 342
7	826		64	4 983	300 454
8			256	49 294	4 238 152
9			625	560 533	67 255 272
10			3 125	7 194 216	1 185 860 330

Acknowledgement. We would like to thank Jeffrey Shallit for proposing such an interesting problem. The work on finding its solution was really funny for both of us, and it helped us to almost forget that our children eventually left our place.

References

1. Amengual, A.: The intriguing properties of the equivalent resistances of n equal resistors combined in series and in parallel. Am. J. Phys. **68**(2), 175–179 (2000). https://doi.org/10.1119/1.19396
2. Beissinger, J.S., Peled, U.N.: Enumeration of labelled threshold graphs and a theorem of Frobenius involving Eulerian polynomials. Graph. Comb. **3**(1), 213–219 (1987). https://doi.org/10.1007/BF01788543
3. Brzozowski, J.A., Grant, E., Shallit, J.: Closures in formal languages and Kuratowski's theorem. Int. J. Found. Comput. Sci. **22**(2), 301–321 (2011). https://doi.org/10.1142/S0129054111008052

4. Jirásková, G., Shallit, J.: The state complexity of star-complement-star. In: Yen, H.-C., Ibarra, O.H. (eds.) DLT 2012. LNCS, vol. 7410, pp. 380–391. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31653-1_34
5. Kleitman, D., Markowsky, G.: On Dedekind's problem: the number of isotone boolean functions. II. Trans. Amer. Math. Soc. **213**, 373–390 (1975)
6. Maslov, A.N.: Estimates of the number of states of finite automata. Sov. Math. Dokl. **11**(5), 1373–1375 (1970)
7. Noe, T.D.: Table of $f(n)$ for $n = 0..100$ (2018). https://oeis.org/A005840/b005840.txt
8. Rabin, M.O., Scott, D.S.: Finite automata and their decision problems. IBM J. Res. Dev. **3**(2), 114–125 (1959). https://doi.org/10.1147/rd.32.0114
9. Sipser, M.: Introduction to the Theory of Computation. Cengage Learning, Florence (2012). https://doi.org/10.1016/S0304-3975(81)80005-9
10. Sloane, N.J.A.: Online encyclopedia of integer sequences (2018). http://oeis.org
11. Sloane, N.J.A., Plouffe, S.: The Encyclopedia of Integer Sequences. Academic Press, San Diego (1995)
12. Weisstein, E.W.: "Resistor network." from mathworld-a wolfram web resource (2018). http://mathworld.wolfram.com/ResistorNetwork.html
13. Yu, S., Zhuang, Q., Salomaa, K.: The state complexities of some basic operations on regular languages. Theor. Comput. Sci. **125**(2), 315–328 (1994). https://doi.org/10.1016/0304-3975(92)00011-F