# State Complexity of Pseudocatenation

Lila Kari and Timothy Ng
School of Computer Science, University of Waterloo
LATA 2019, Saint Petersburg, Russia

Adenine

Cytosine

Guanine

Thymine

ACGTACGTAGCTGCATCGATCAG

ACGTACGTAGCTGCATCGATCAG
TGCATGCATCGACGTAGCTAGTC

ACGTACGTAGCTGCATCGATCAG

TGCATGCATCGACGTAGCTAGTC

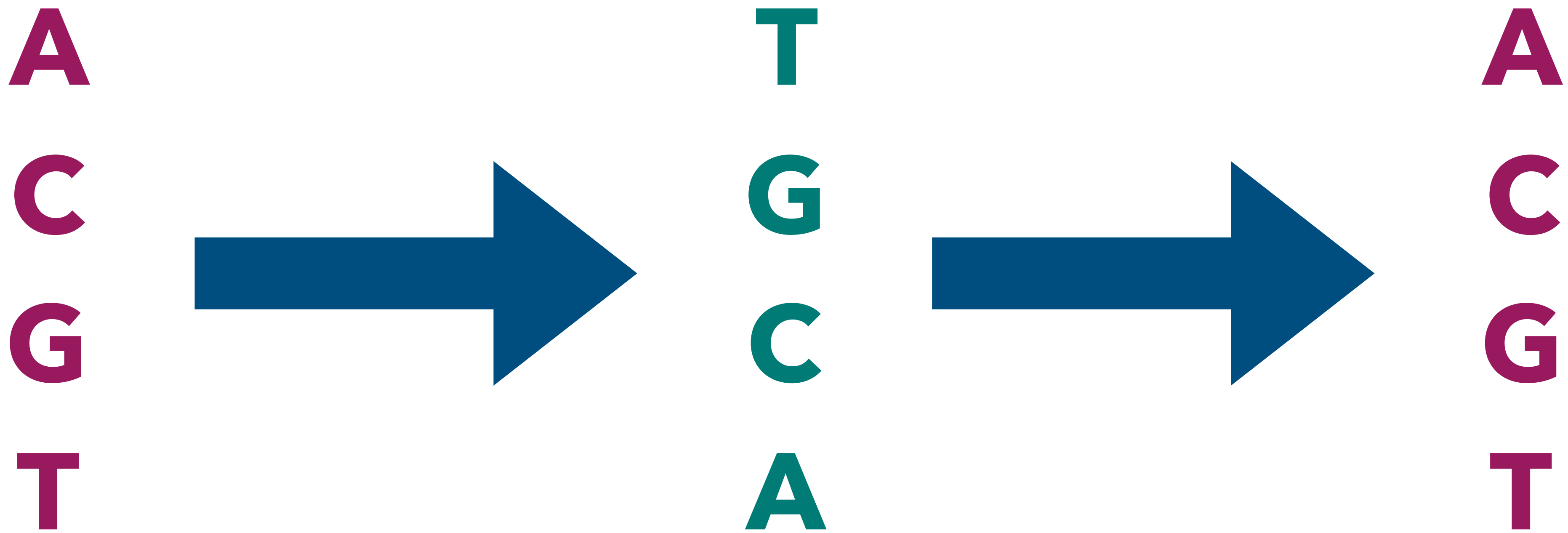

ACGTACGTAGCTGCATCGATCAG
TGCATGCATCGACGTAGCTAGTC

ACGTACGTAGCTGCATCGATCAG

TGCATGCATCGACGTAGCTAGTC

CTGATCGATGCAGCTACGTACGT


ACGTACGTAGCTGCATCGATCAG
TGCATGCATCGACGTAGCTAGTC
CTGATCGATGCAGCTACGTACGT




# Involution

ACGTACGTAGCTGCATCGATCAG

TGCATGCATCGACGTAGCTAGTC

$$\theta(a_1 a_2 \cdots a_{n-1} a_n) = \theta(a_n)\theta(a_{n-1}) \cdots \theta(a_2)\theta(a_1)$$

# Antimorphism


# Palindromicity

A. de Luca and A. De Luca (2006)

**AGGTACGTACCTGTACCTAGGTACAGGTAC**

$$w = x_1 x_2 \cdots x_k, x_i \in \{x_1, \theta(x_1)\}$$

$$w \in u\{u, \theta(u)\}^*$$

# Primitivity and repetition

E. Czeizler, L. Kari, S. Seki (2010)

AGT CAT

AGTCAT

AGTATG

$$u \odot^{\theta} v = \{uv, u\theta(v)\}$$

# Pseudocatenation

L. Kari and M. Kulkarni (2014)

AGGTACGTACCTGTACCTAGGTACAGGTAC

$$w = x_1 x_2 \cdots x_k, x_i \in \{x_1, \theta(x_1)\}$$

$$w \in u\{u, \theta(u)\}^*$$

$$w \in u^{\odot_k^\theta}$$

The **state complexity of a regular language** is the number of states in its minimal deterministic finite automaton.

The **state complexity of an operation** is the worst-case state complexity of the language resulting from the operation, as a function of the state complexity of the operands.

# Prior state complexity results

# Prior state complexity results

- Introduced (A. Maslov, 1970)

# Prior state complexity results

- Introduced (A. Maslov, 1970)
- Re-discovered (S. Yu, K. Salomaa, Q. Zhuang, 1994)

# Prior state complexity results

- Introduced (A. Maslov, 1970)
- Re-discovered (S. Yu, K. Salomaa, Q. Zhuang, 1994)
- Combined operations (A. Salomaa, K. Salomaa, S. Yu 2007)

# Prior state complexity results

- Introduced (A. Maslov, 1970)
- Re-discovered (S. Yu, K. Salomaa, Q. Zhuang, 1994)
- Combined operations (A. Salomaa, K. Salomaa, S. Yu 2007)
- Undecidability (A. Salomaa, K. Salomaa, S. Yu, 2013)

# Prior state complexity results

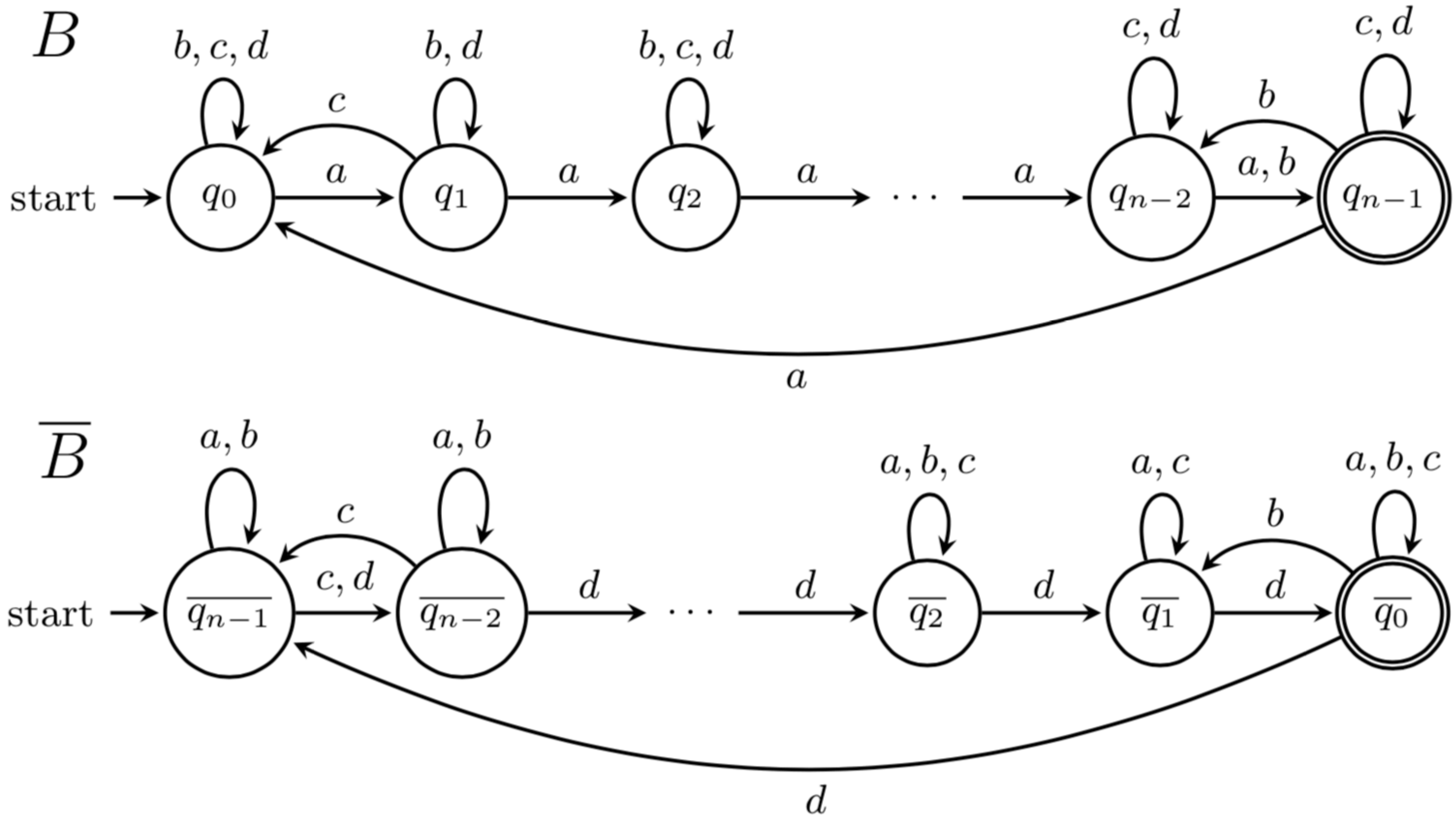
- Introduced (A. Maslov, 1970)
- Re-discovered (S. Yu, K. Salomaa, Q. Zhuang, 1994)
- Combined operations (A. Salomaa, K. Salomaa, S. Yu 2007)
- Undecidability (A. Salomaa, K. Salomaa, S. Yu, 2013)
- Universal witnesses (J.A. Brzozowski, 2012)

# Prior state complexity results

- Hairpin-free languages (L. Kari et al. 2006)
- (Pseudo-)Inversion (D.-J. Cho et al. 2016)
- (Pseudo-)Duplication (D.-J. Cho et al. 2016)
- Overlap assembly (J.A. Brzozowski et al. 2018)

# Regular languages are closed under antimorphism

- Regular languages are closed under morphism
- Regular languages are closed under reversal
- Antimorphism is morphism composed with reversal



B
start
$q_0$
$q_1$
$q_2$
$q_{n-2}$
$q_{n-1}$
$b, c, d$
$b, d$
$b, c, d$
$c, d$
$c, d$
$c$
$a$
$a$
$a$
$a$
$b$
$a, b$
$a$
$\overline{B}$
start
$\overline{q_{n-1}}$
$\overline{q_{n-2}}$
$\overline{q_2}$
$\overline{q_1}$
$\overline{q_0}$
$a, b$
$a, b$
$a, b, c$
$a, c$
$a, b, c$
$c$
$c, d$
$d$
$d$
$d$
$b$
$d$
$d$

$$L_1 \odot^{\theta} L_2 = L_1(L_2 \cup \theta(L_2))$$

$$L_1 \odot^{\theta} L_2 = L_1(L_2 \cup \theta(L_2))$$

$$L_m \cdot (L_n \cup L_p)$$

$$L_1 \odot^{\theta} L_2 = L_1(L_2 \cup \theta(L_2))$$

$$L_m \cdot (L_n \cup L_p)$$

**B. Cui et al. 2011**

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

**Machine A non-final**

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

**Machine A non-final**

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

**Machine A first pass**

Machine A non-final

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

Machine A first pass

Machine A final

$(m\text{-}k^A)(2^n\text{-}1)(2^n\text{-}1)$

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

$m\text{-}k_A$

$k_A(2^{n\text{-}1})(2^{n\text{-}kB})$

$(m\text{-}k^A)(2^n\text{-}1)(2^n\text{-}1)$

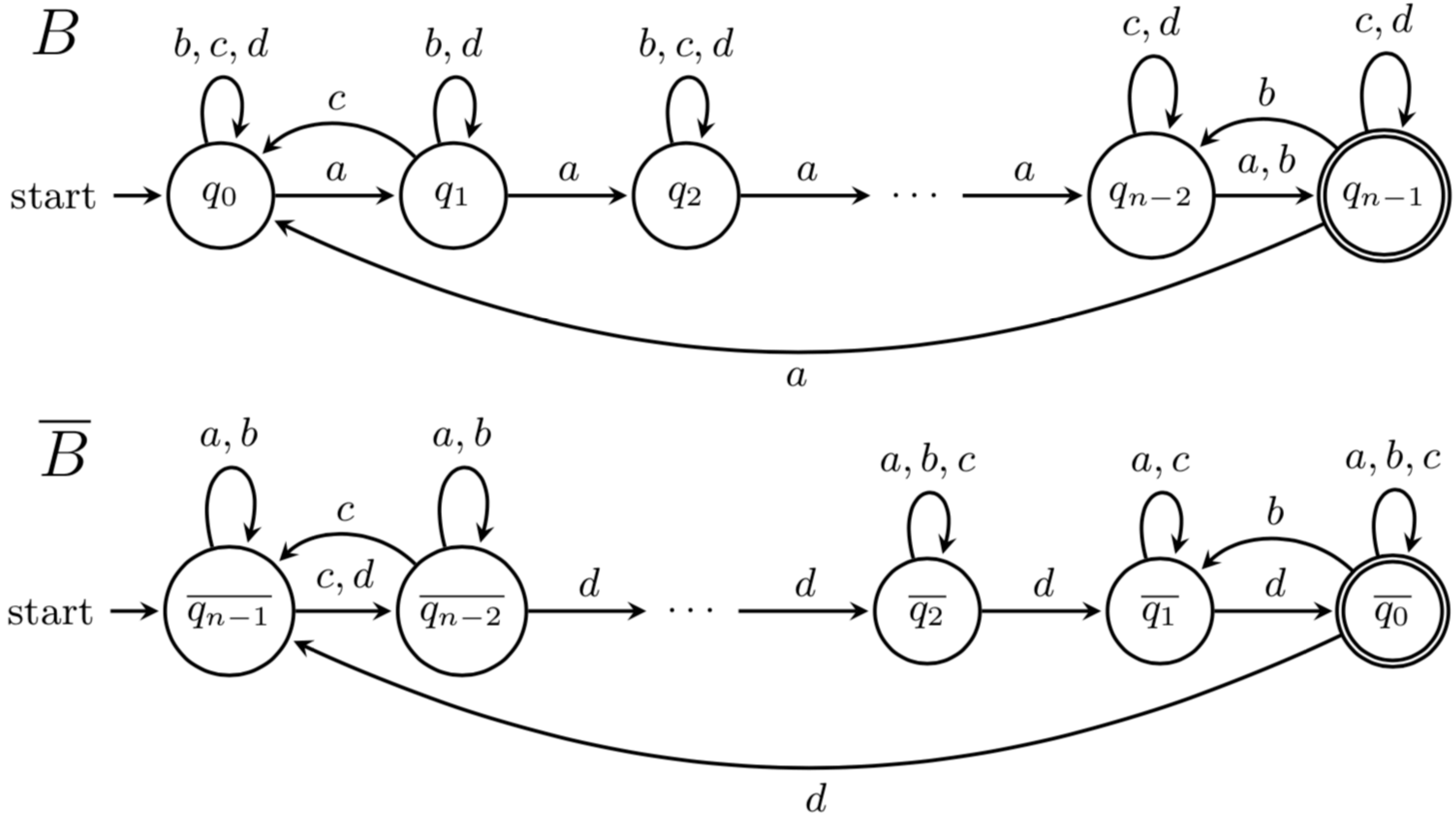
$$Q_C = \{\langle q, P, \overline{R}\rangle \mid q \in Q_A - F_A, P \in 2^{Q_B} - \{\emptyset\}, \overline{R} \in 2^{\overline{Q_B}} - \{\emptyset\}\}$$

$$\cup \{\langle q, \emptyset, \emptyset\rangle \mid q \in Q_A - F_A\}$$

$m\text{-}k_A$

$$\cup \{\langle q, P \cup \{s_B\}, \overline{R} \cup \overline{F_B}\rangle \mid q \in F_A, P \in 2^{Q_B - \{s_B\}}, \overline{R} \in 2^{\overline{Q_B} - \overline{F_B}}\},$$

$k_A(2^{n\text{-}1})(2^{n\text{-}kB}\text{-}1)+1$




# Lower bound

J.A. Brzozowski and D. Liu, 2013

| Operation | State complexity |
|---|---|
| $L_m \odot^{\theta} L_n$ | $(m-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$ |
| $L_m L_n$ | $m2^n - 2^{n-1}$ |
| $L_m(L_n \cup L_p)$ | $(m-1)(2^{n+p} - 2^n - 2^p + 2) + 2^{n+p-2}$ |

| Operation | State complexity |
| --- | --- |
| $L_n^{\odot_2^\theta}$ | $(n-1)(2^{2n} - 2^{n+1} + 2) + 2^{2n-2} - 2^{n-1} + 1$ |
| $L_n^2$ | $n2^n - 2^{n-1}$ |

# Pseudosquare

(Square: N. Rampersad, 2006)

AGGTACGTACCTGTACCTAGGTACAGGTAC

$$w = x_1 x_2 \cdots x_k, x_i \in \{x_1, \theta(x_1)\}$$

**AGGTACGTACCTGTACCTAGGTACAGGTAC**

$$w = x_1 x_2 \cdots x_k, x_i \in \{x_1, \theta(x_1)\}$$

$$L^{\odot^{\theta}_{+}}, L = \{AG, AGT, TG, CGT, TT\}$$

**AGGTACGTACCTGTACCTAGGTACAGGTAC**

$$w = x_1 x_2 \cdots x_k, x_i \in \{x_1, \theta(x_1)\}$$

**AGAGTACTCGTAAACGCTAGT**

$$L^{\odot^{\theta}_{+}}, L = \{AG, AGT, TG, CGT, TT\}$$

$$L^{\odot^{\theta}_{+}} = L(L \cup \theta(L))^*$$

$$L^{\odot^{\theta}_{+}} = L(L \cup \theta(L))^*$$

$$(L_m \cup L_n)^*$$

$$L^{\odot^{\theta}_{+}} = L(L \cup \theta(L))^*$$

$$(L_m \cup L_n)^*$$

**A. Salomaa, K. Salomaa, S. Yu (2007)**

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

**Non-final states**

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

$$Q' = Q_1 \cup Q_2,$$

**Non-final states**

$$Q_1 = \{\langle P, \overline{R}\rangle \mid \emptyset \neq P \subseteq Q - F, \overline{R} \subseteq \overline{Q} - \{\overline{s}\}\},$$

$$Q_2 = \{\langle P \cup \{s\}, \overline{R} \cup \overline{F}\rangle \subseteq Q \times \overline{Q} \mid (P \cup \overline{R}) \cap (F \cup \{\overline{s}\}) \neq \emptyset\}.$$

**Including final states**

$(2^{n\text{-}k}\text{-}1)(2^{n\text{-}1})$

$$Q' = Q_1 \cup Q_2,$$

$$Q_1 = \{\langle P, \overline{R}\rangle \mid \emptyset \neq P \subseteq Q - F, \overline{R} \subseteq \overline{Q} - \{\overline{s}\}\},$$

$$Q_2 = \{\langle P \cup \{s\}, \overline{R} \cup \overline{F}\rangle \subseteq Q \times \overline{Q} \mid (P \cup \overline{R}) \cap (F \cup \{\overline{s}\}) \neq \emptyset\}.$$

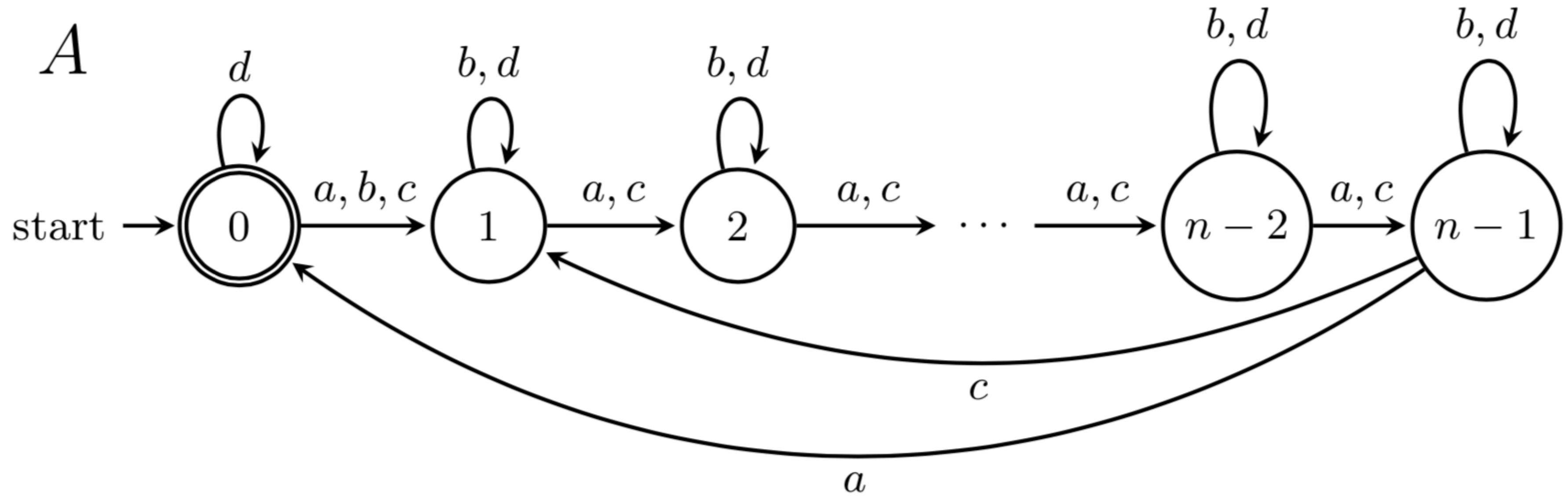
$(2^{n\text{-}1})(2^{n\text{-}k})$

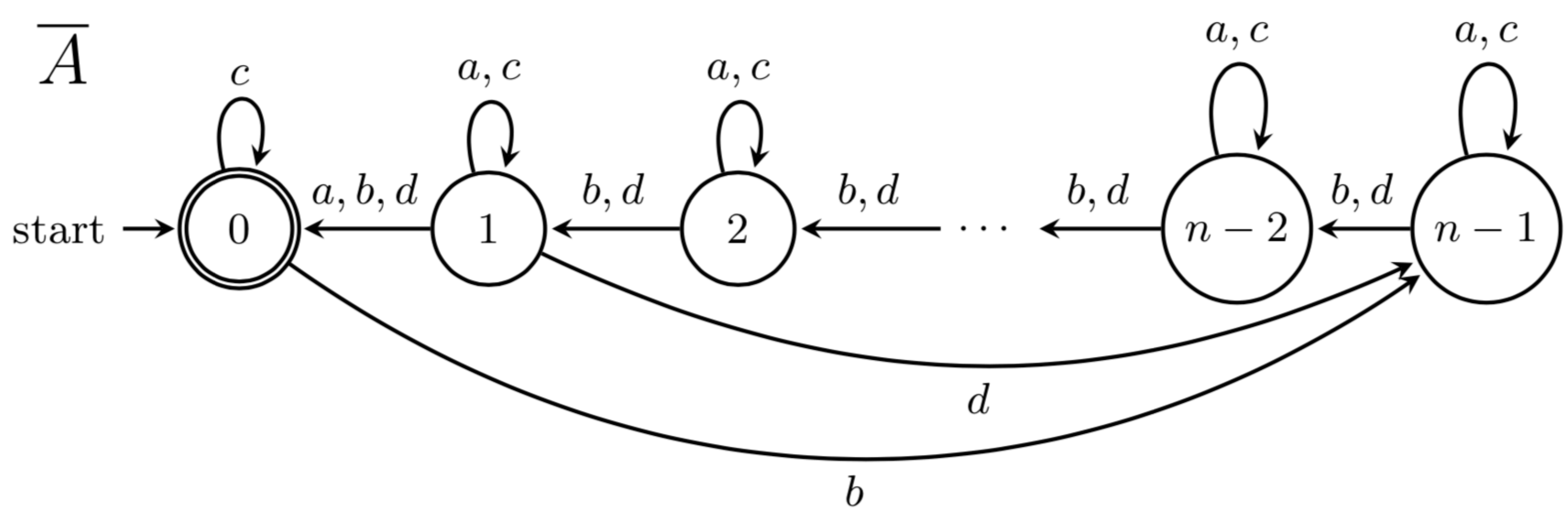
$$(2^{n-k}-1)(2^{n-1})$$

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

$$(2^{n-1})(2^{n-k})$$

**This changes further depending on whether or not $s \in F$.**


$A$
start
0
1
2
$n-2$
$n-1$
$\cdots$
$d$
$b, d$
$b, d$
$b, d$
$b, d$
$a, b, c$
$a, c$
$a, c$
$a, c$
$a, c$
$c$
$a$


$\overline{A}$
start
0
1
2
$n-2$
$n-1$
$\cdots$
$c$
$a, c$
$a, c$
$a, c$
$a, c$
$a, b, d$
$b, d$
$b, d$
$b, d$
$b, d$
$d$
$b$

# Lower bound

| Operation | State complexity |
| --- | --- |
| $L_n^{\odot_+^\theta}$ | $2^{2n-1} - 2^n + 1$ |
| $L_n^*$ | $2^{n-1} + 2^{n-2}$ |
| $(L_m \cup L_n)^*$ | $2^{m+n-1} - 2^{m-1} - 2^{n-1} + 1$ |

# Future work

$k^{th}$ pseudopower and pseudocube,
finite languages, universal witnesses

# Finite languages

- State complexity of operations on finite languages is often much lower (C. Câmpeanu et al. 1999)
- Starting with a finite set and generating sets of words makes sense from the biological perspective

# Universal witnesses

- Universal witnesses used in pseudocatenation are not the ones for combinations of operations with reversal!
- Lower bound witness for iterated pseudocatenation is not a universal witness as defined by Brzozowski
- Are there universal witnesses for combinations of operations involving antimorphisms?

# $k^{th}$ pseudopower

- State complexity of power and cube (M. Domaratzki and A. Okhotin, 2009)
- State complexity of square and power for unary languages (N. Rampersad, 2006)
- State complexity of k-catenation (Ésik et al. 2009, Caron et al. 2016)

# Thank you