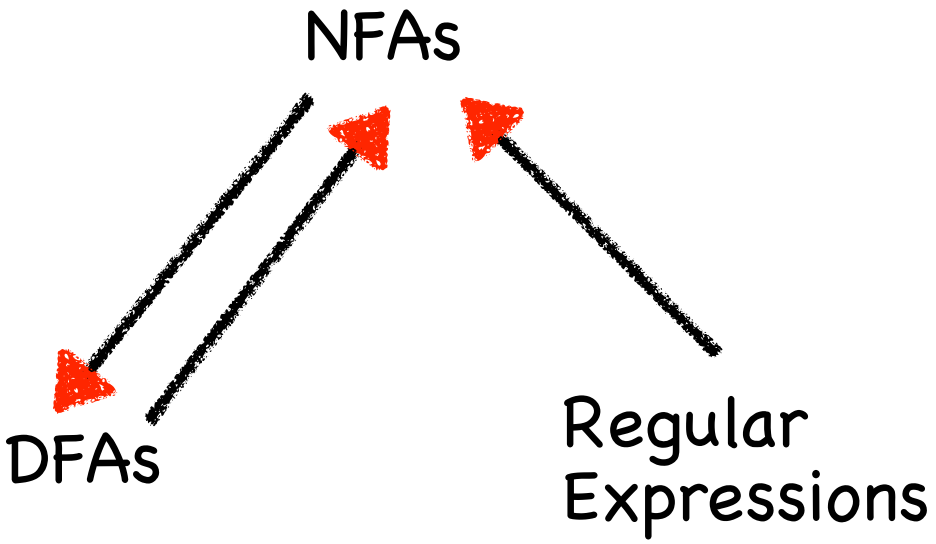
Regular Languages

CS 81: Computability and Logic

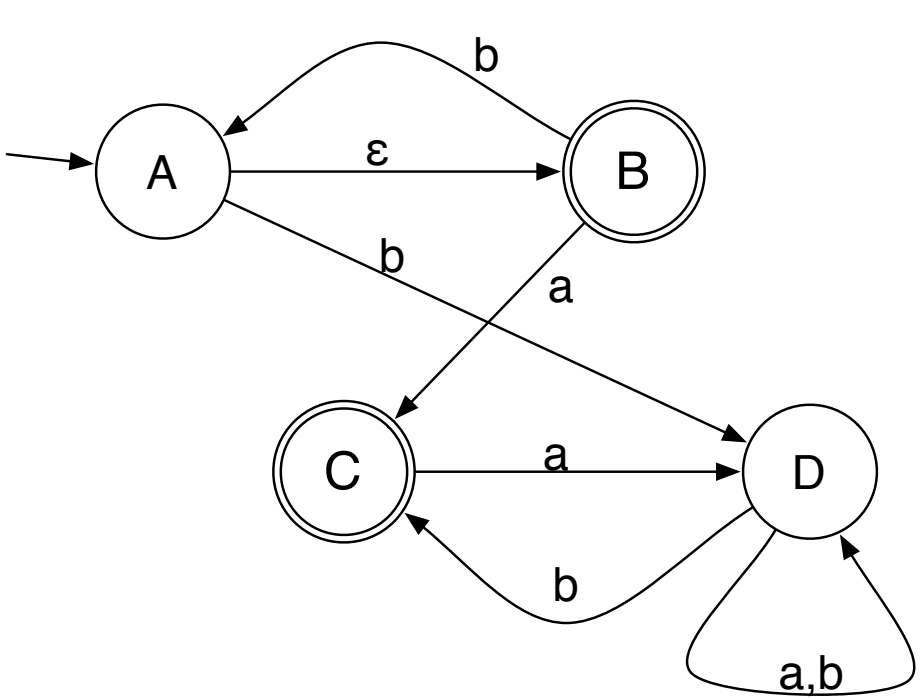
October 21, 2010

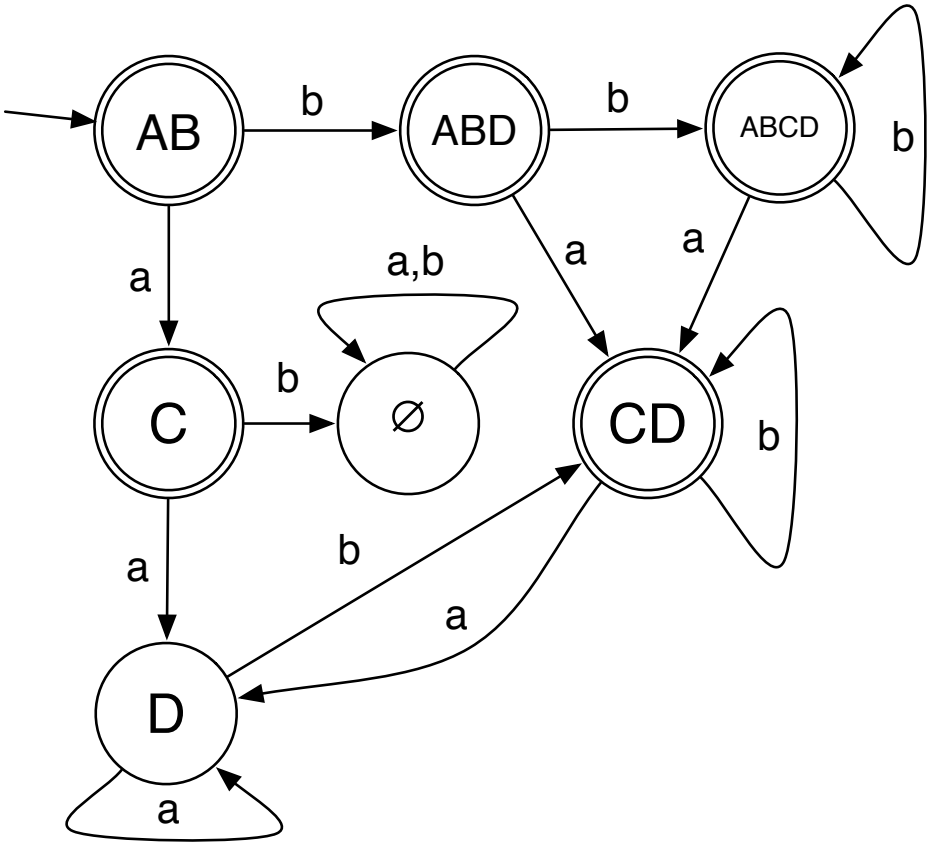
Representing Regular Languages: The Story So Far



Components of a State Machine?

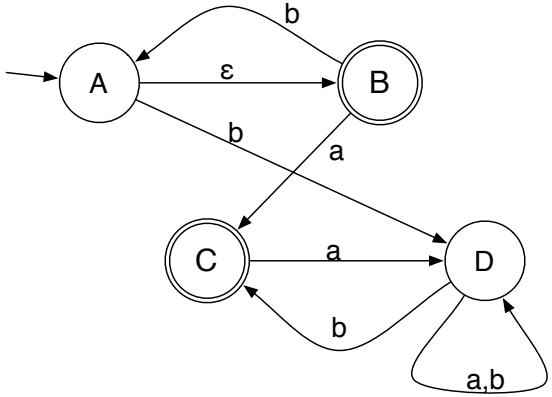
From NFA to DFA: the **Subset Construction**

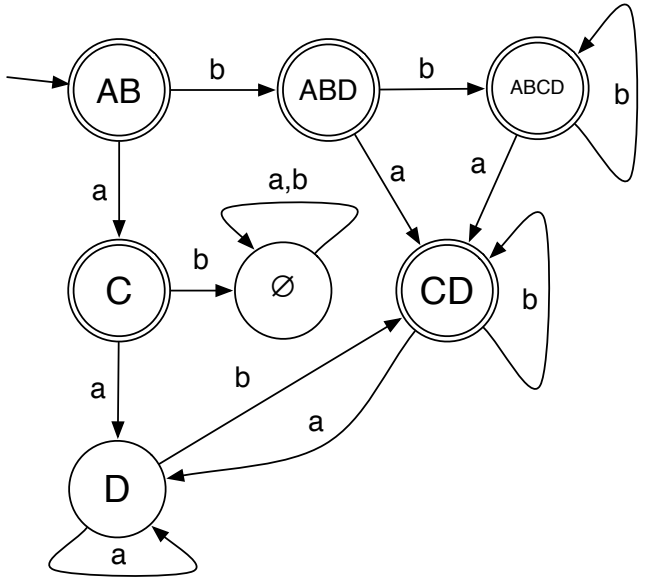




The Subset Construction Formalized

- Input: An NFA $(\Sigma, Q, \rightarrow, q_0, F)$
- Output: A DFA $(\Sigma, \mathcal{P}(Q), \rightarrow', q_0', F')$
  - Definition?





Regular Expressions

- What are the “primitive” regular expressions?
- What other regular expression forms have you seen?

M[ou]'?am+[ae]r .*([AEae]l[-])?[GKQ]h?[aeu]+([dtz][dhz]?)+af[iy]

- Muammar Qaddafi
- Mo'ammar Gadhafi
- Muammar Kaddafi
- Muammar Qadhafi
- Moammar El Kadhafi
- Muammar Gadafi
- Mu'ammar al-Qadafi
- Moamer El Kazzafi
- Moamar al-Gaddafi
- Mu'ammar Al Qathafi
- Muammar Al Qathafi
- Mo'ammar el-Gadhafi
- Moamar El Kadhafi
- Muammar al-Qadhafi
- Mu'ammar al-Qadhdhafi
- Mu'ammar Qadafi
- Moamar Gaddafi
- Mu'ammar Qadhdhafi
- Muammar al-Khaddafi
- Mu'amar al-Kadafi
- Muammar Ghaddafy
- Muammar Ghadafi
- Muammar Ghaddafi
- Muamar Kaddafi
- Muammar Quathafi
- Muammar Gheddafi
- Muamar Al-Kaddafi
- Moammar Khadafy
- Moammar Qudhafi
- Mu'ammar al-Qaddafi
- Mu'ammar Muhammad Abu Minyar al-Qadhafi

from the RX Library

Exercise

1. Give a regular expression for C identifiers:
   - Can contain letters, digits, and underscores
   - Must begin with a letter or underscore.

```
main
__Z6rotateiii
```

2. Give a regular expression for Ada identifiers:
   - Can contain letters, digits, and underscores (_)
   - Must begin with a letter
   - Have no consecutive underscores or an underscore at the end

```
woohoo32
Last_Nonzero_Row
```

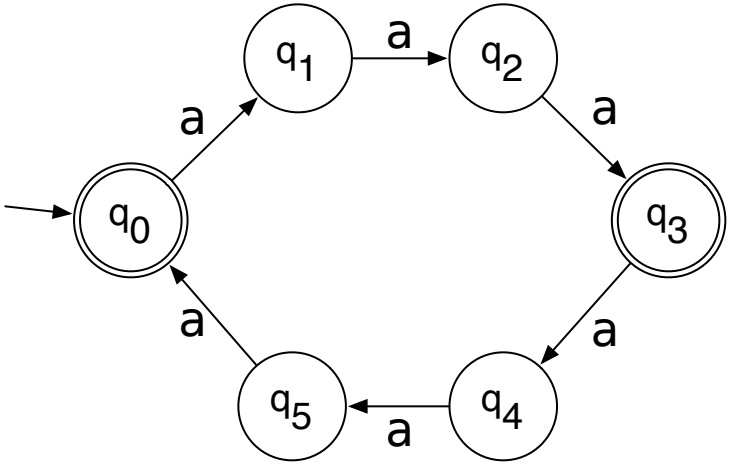
Exercise

- 6
- 9
- 2u
- 2L
- 2UL
- 2lu
- 2uu
- 02
- 002
- 09
- 02ul
- 0Xff90
- 0x0
- 0x0F
- 0x0FlU
- 0x0x3
- 0
- 0lu
- 0x
-

From Automata to Regexps

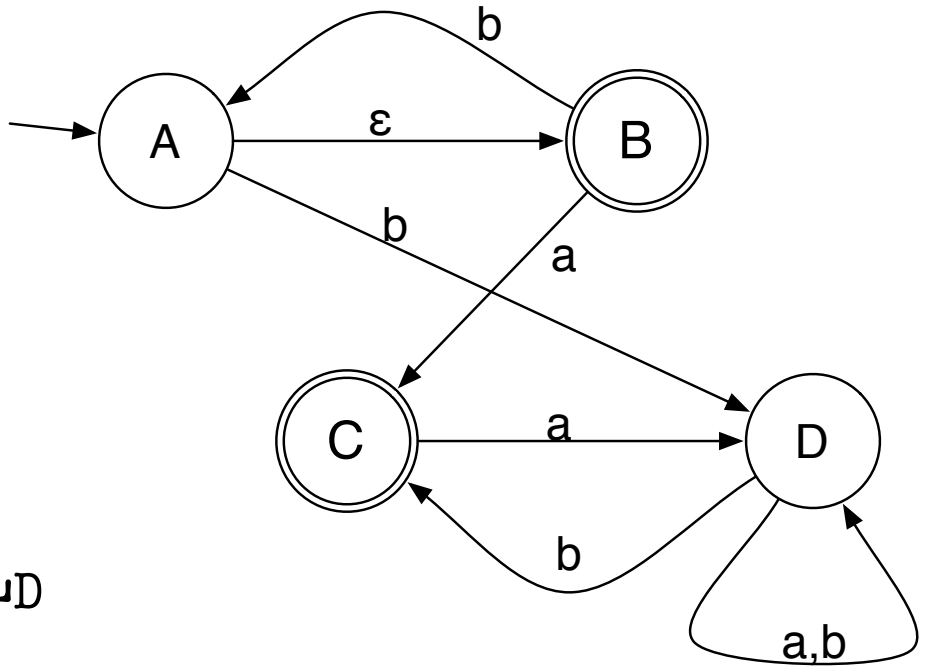
- Two approaches
  - Solving equations
  - Generalized NFAs

The Language of a State



- Let L_q be the set of strings are accepted when starting from state q.
- What is L_{q0}, L_{q1}, L_{q2}, ...?
- How is L_{q1} related to L_{q2}?

Example



- $L_A = \varepsilon L_B \mid bL_D$
- $L_B =$
- $L_C =$
- $L_D =$

Solving Equations using Arden's Rule

- The equation

  $L = AL \mid B$

  with A and B being languages and L an unknown has the solution

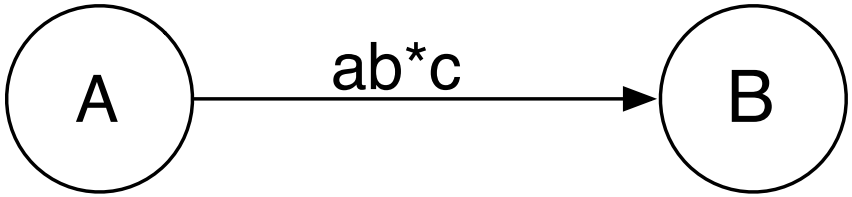
  $L = A^* B$

- This is the <u>smallest</u> solution.
  - If $\varepsilon \notin A$, this solution is unique.
  - Otherwise, A^*C is a solution for any $B \subseteq C$.

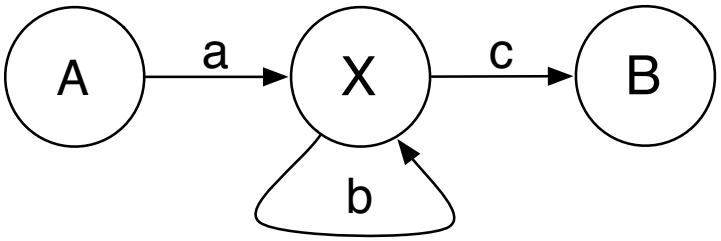
$$
\begin{aligned}
L_A &= L_B \mid bL_D \\
L_B &= \varepsilon \mid bL_A \mid aL_C \\
L_C &= \varepsilon \mid aL_D \\
L_D &= [ab]L_D \mid bL_C
\end{aligned}
$$

Generalized NFAs

- Just like an NFA, but edges have regular expressions rather than single symbols

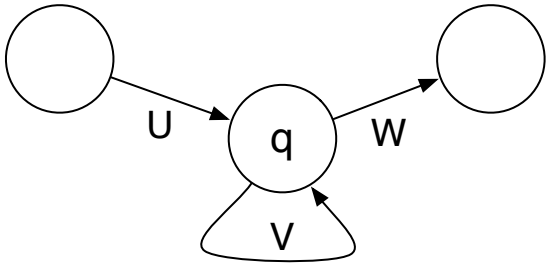


- Since regular expressions can be turned into NFAs, we aren't adding any extra power.

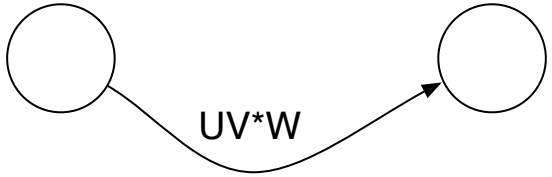


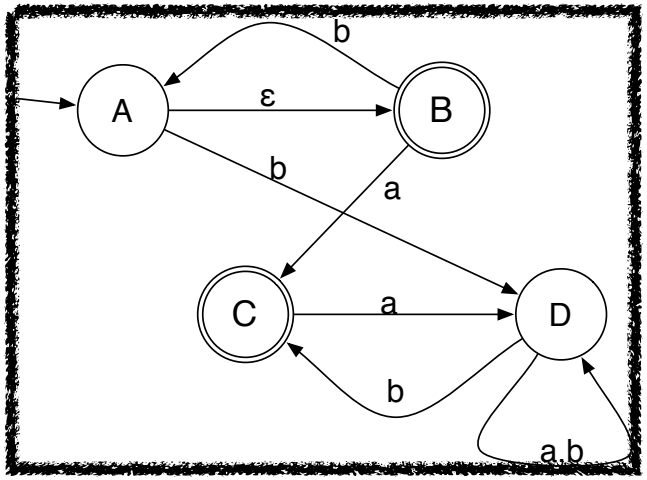
Removing States

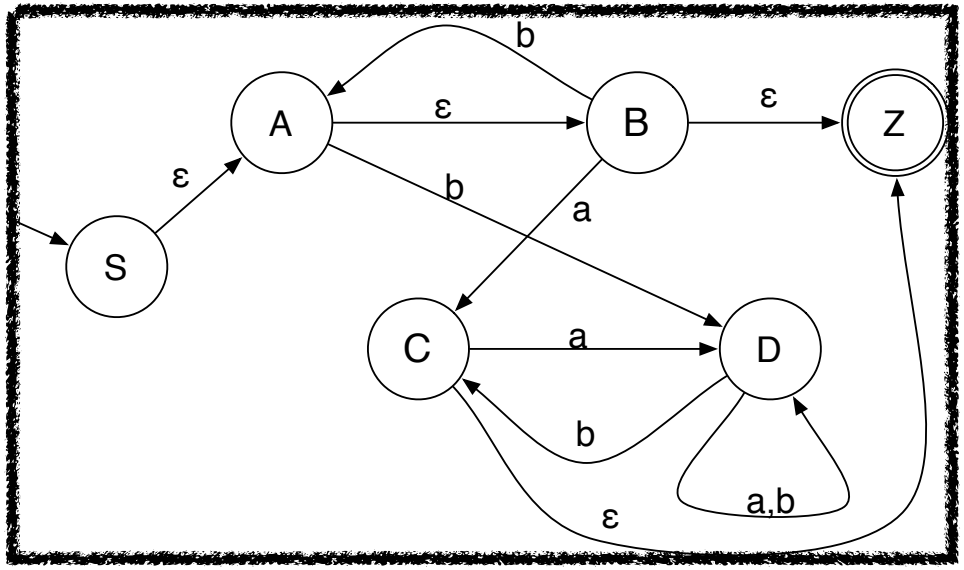
- Pick a state q to remove.
- For every incoming/outgoing pair



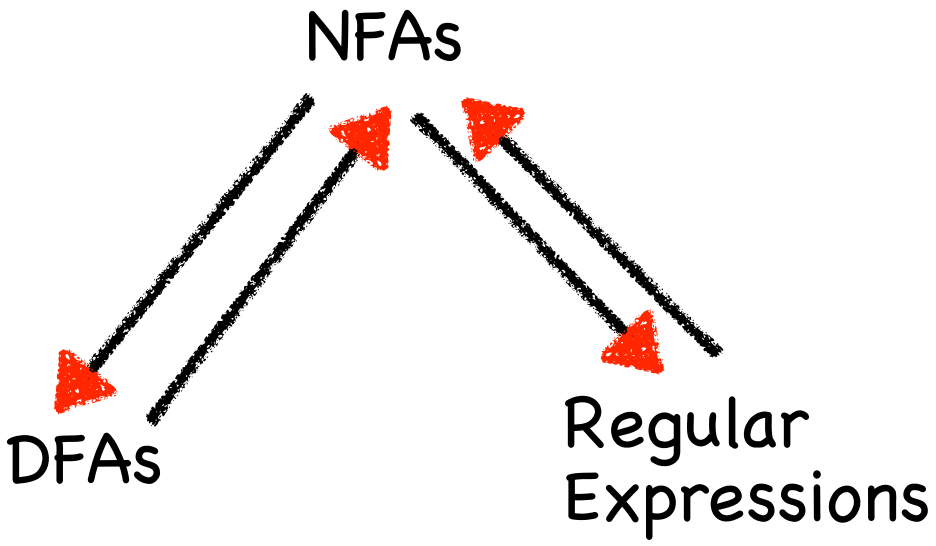
- replace by the direct edge







Representing Regular Languages: Conclusion



Closure Properties

- A family of languages is a set of languages.
  - The family of all finite languages
  - The family of all languages
  - The family of all regular languages
- A family F is closed under an operation if applying the operation to languages in F always produces a result in F.

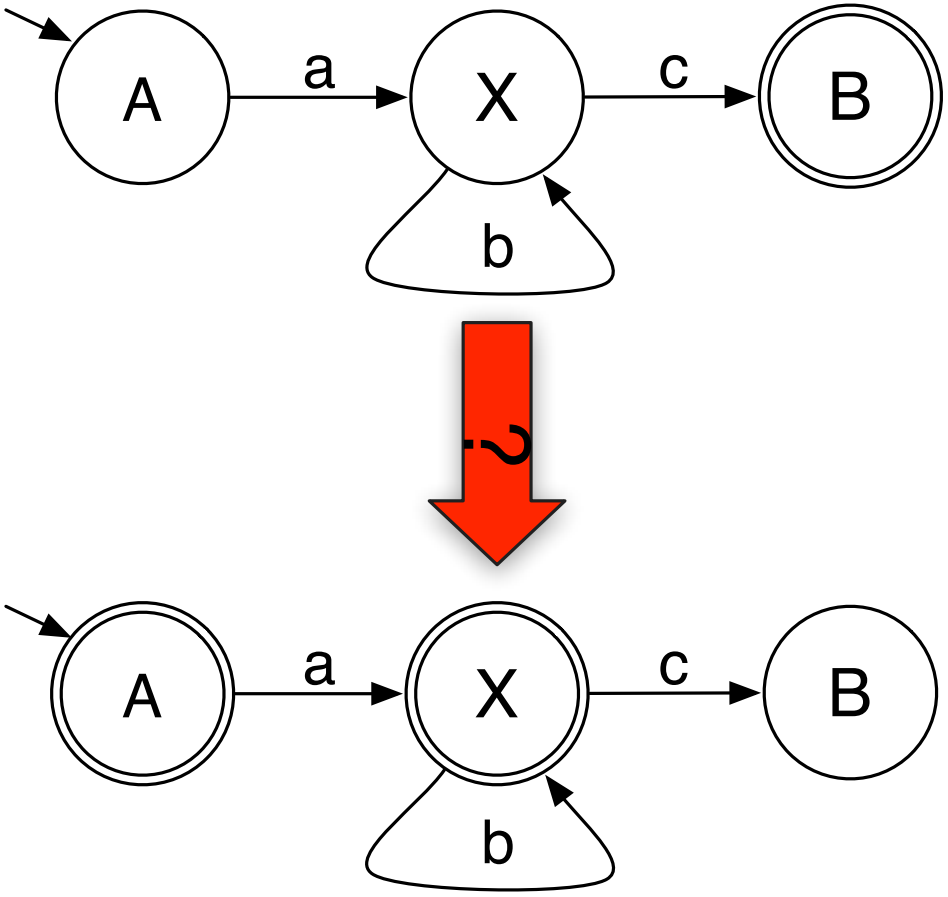
Example: Finite Languages

- Are the finite languages closed under:
  - Union? $(A \cup B)$
  - Intersection? $(A \cap B)$
  - Concatenation (AB)
  - Star? (A^*)
  - Complement? (A^c)

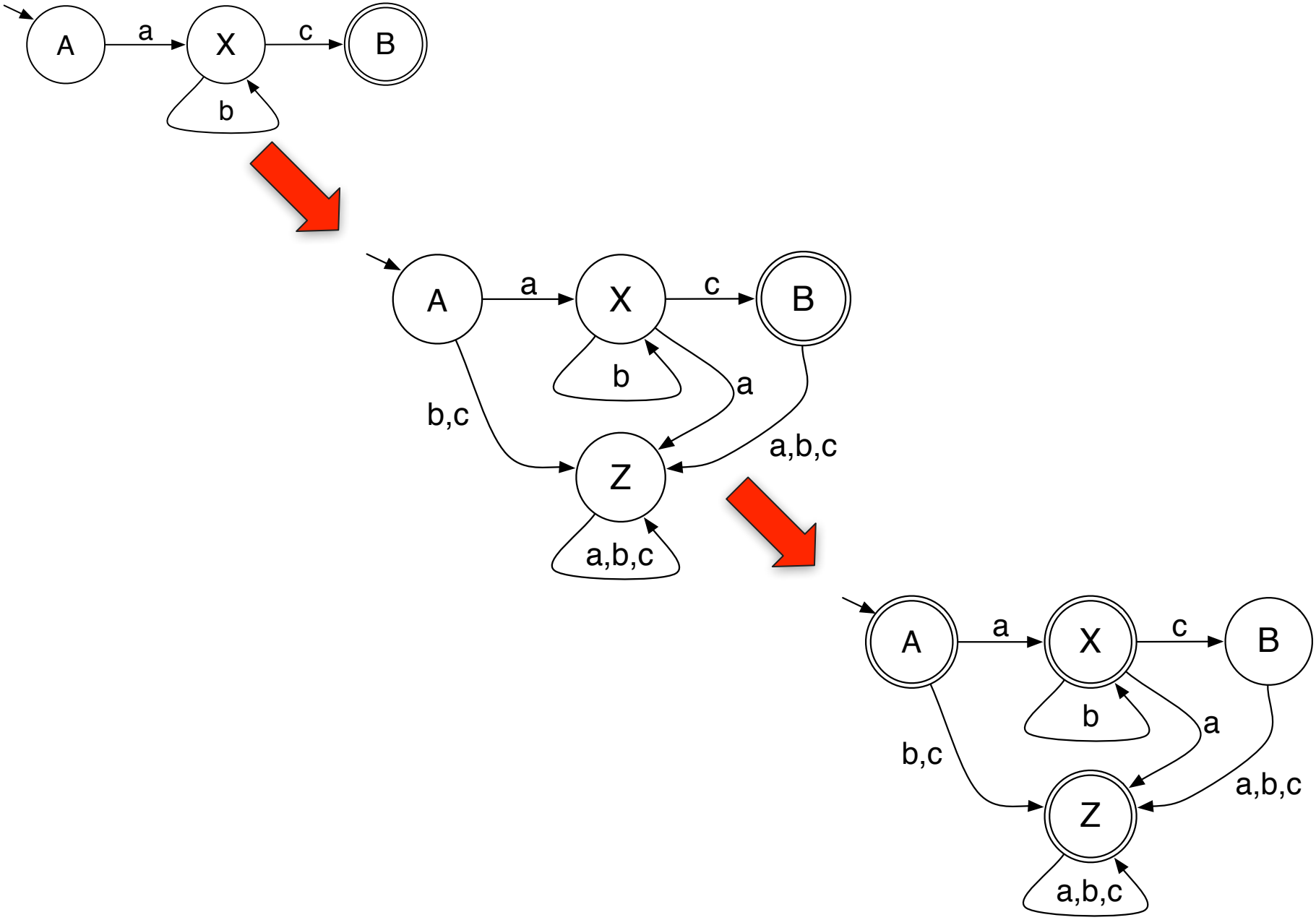
Closure Properties for Regular Languages

- Regular languages are closed under
  - Concatenation
  - Union
  - Star
  - Complement
  - Intersection
- Proofs: Consider the corresponding automata...

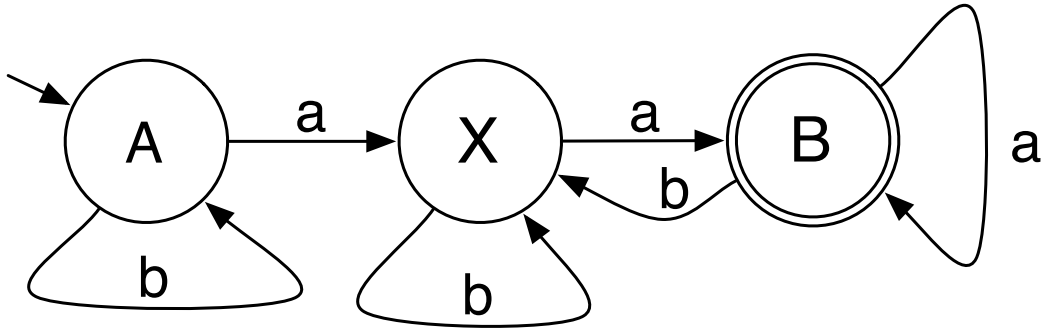
Complement?

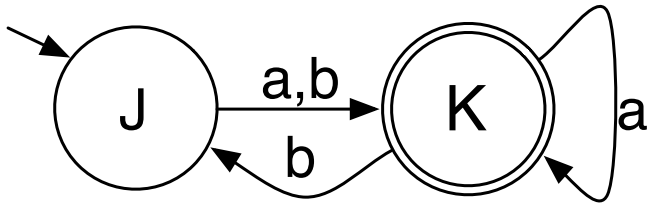


Complement



Intersection





Intersection