CSE443 Compilers

Dr. Carl Alphonce
alphonce@buffalo.edu
343 Davis Hall

Announcements

- Initial attendance sheet
- How did PM meetings go yesterday?

Phases of a compiler

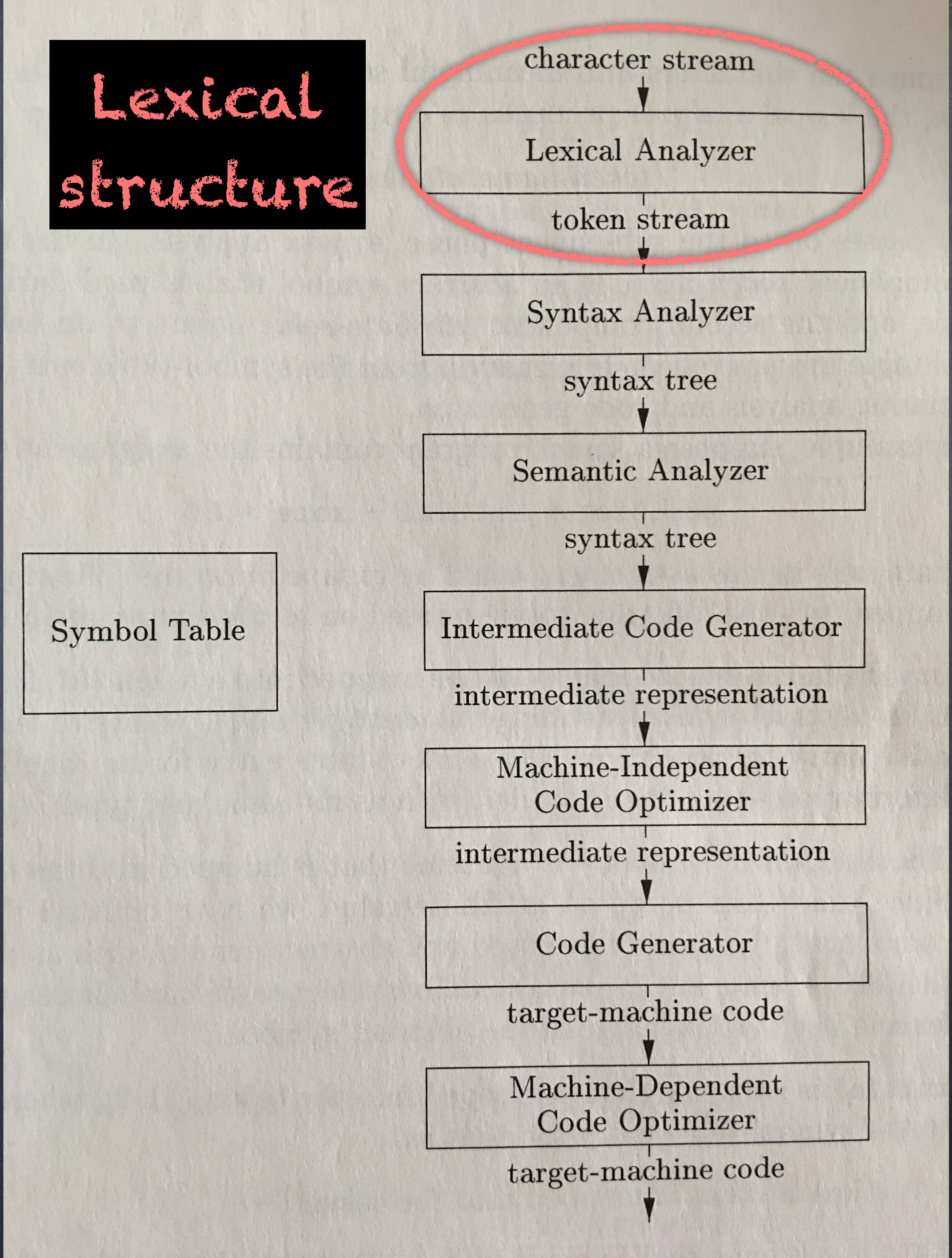
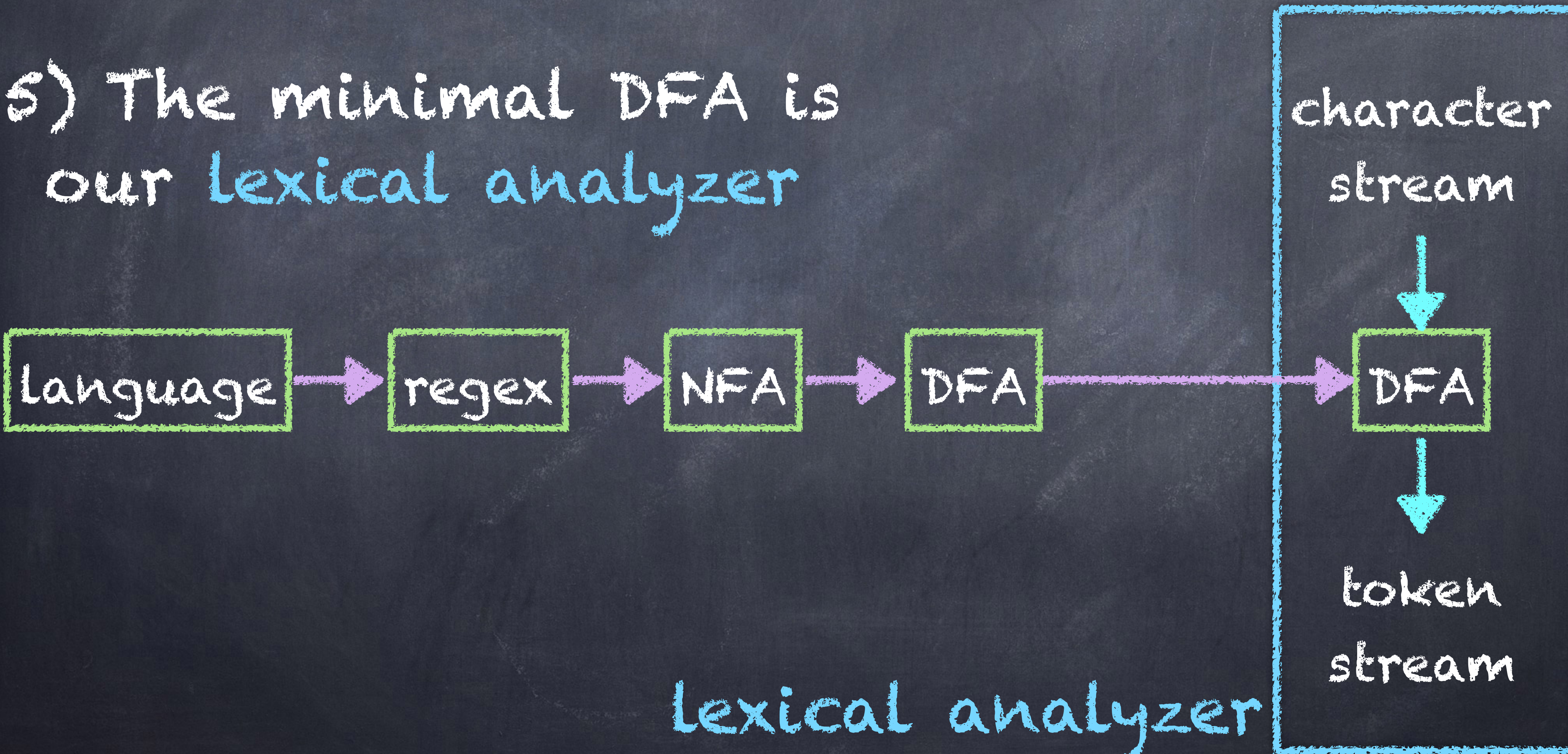


Figure 1.6, page 5 of text

Process of building lexical analyzer

5) The minimal DFA is our lexical analyzer



Focus last time

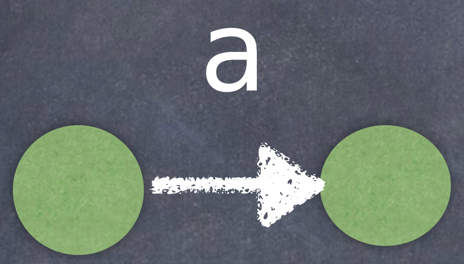
focus today

(a|b)*abb

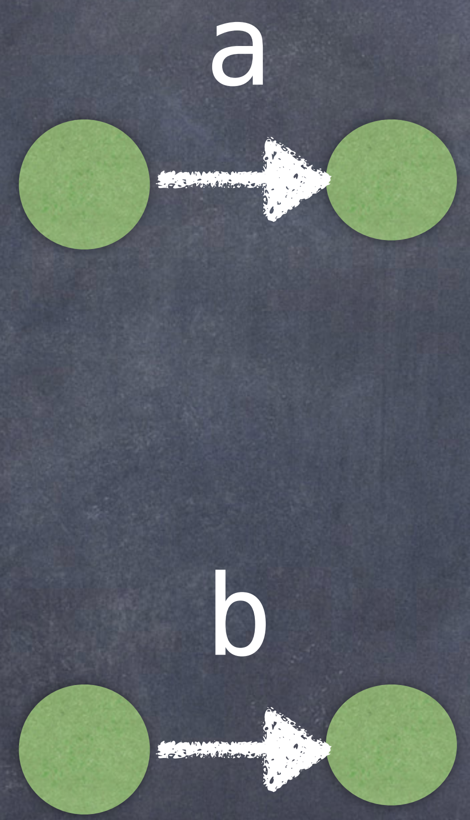
first we construct an NFA

from this regular expression

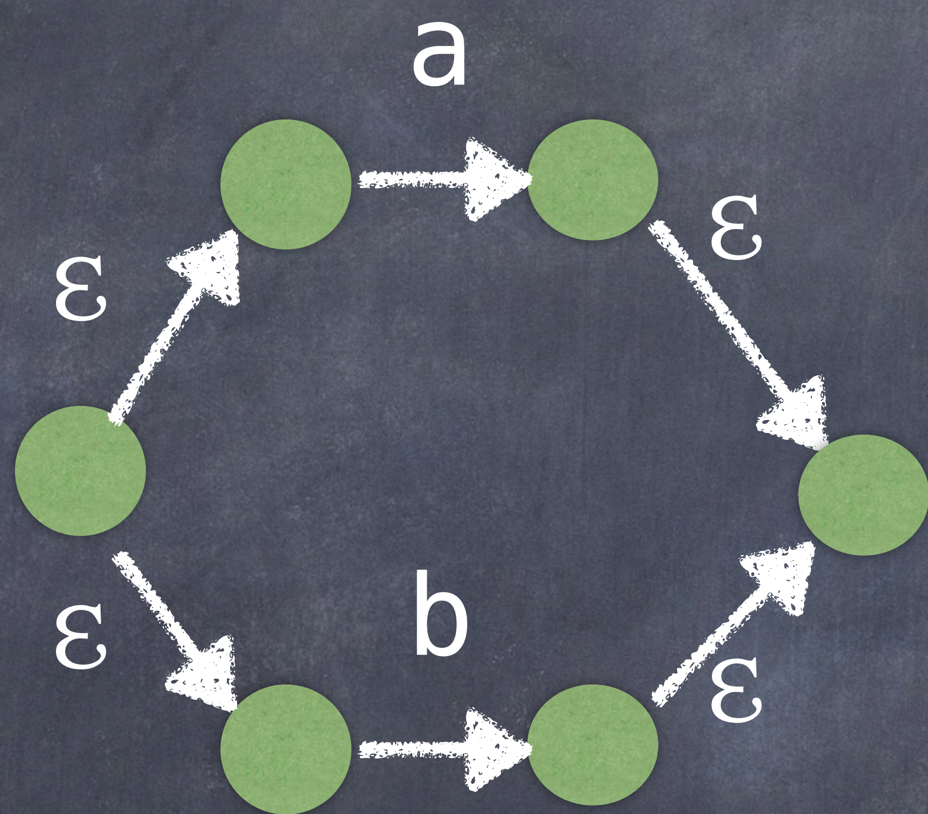
(a|b)*abb



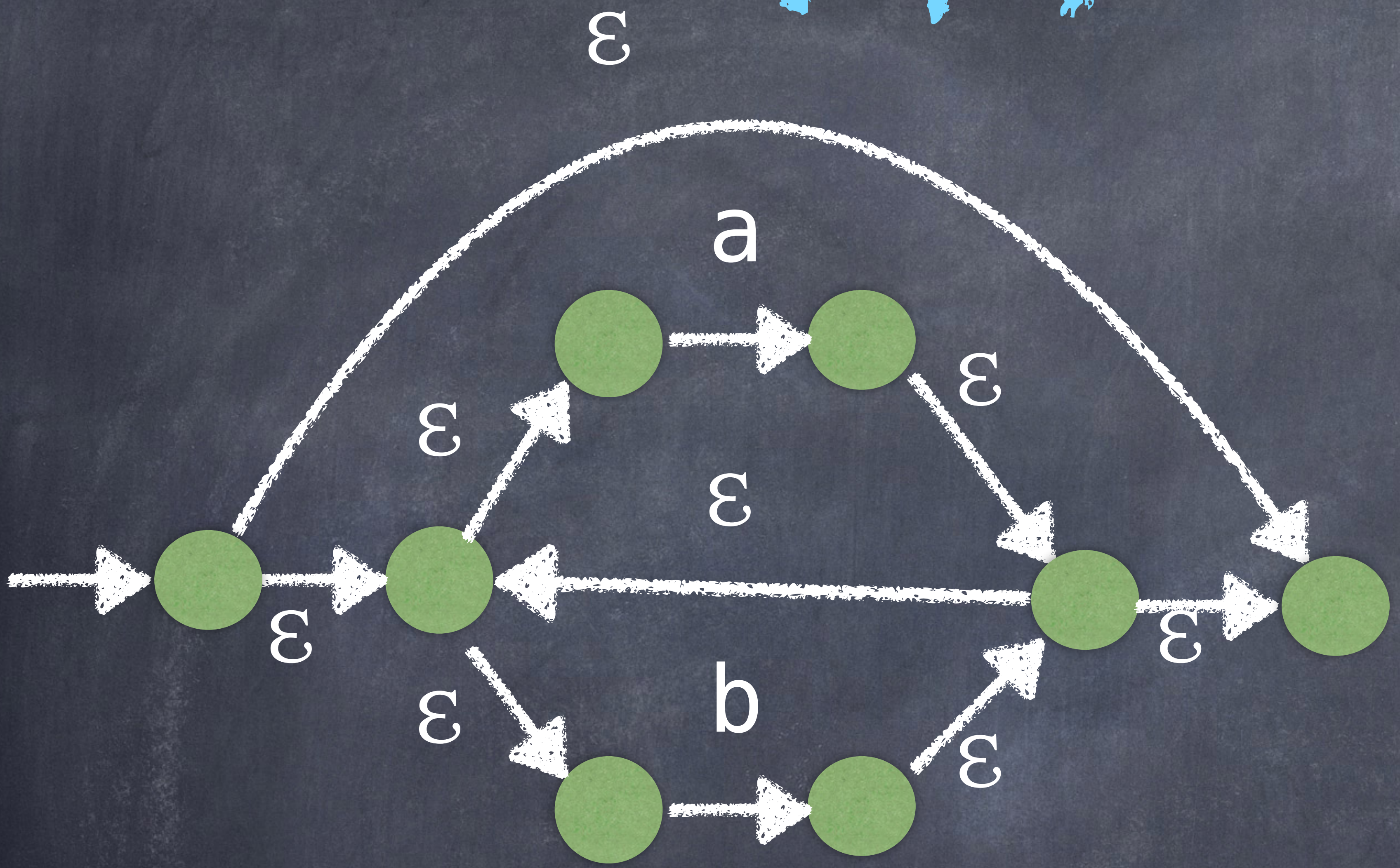
(a|b)*abb



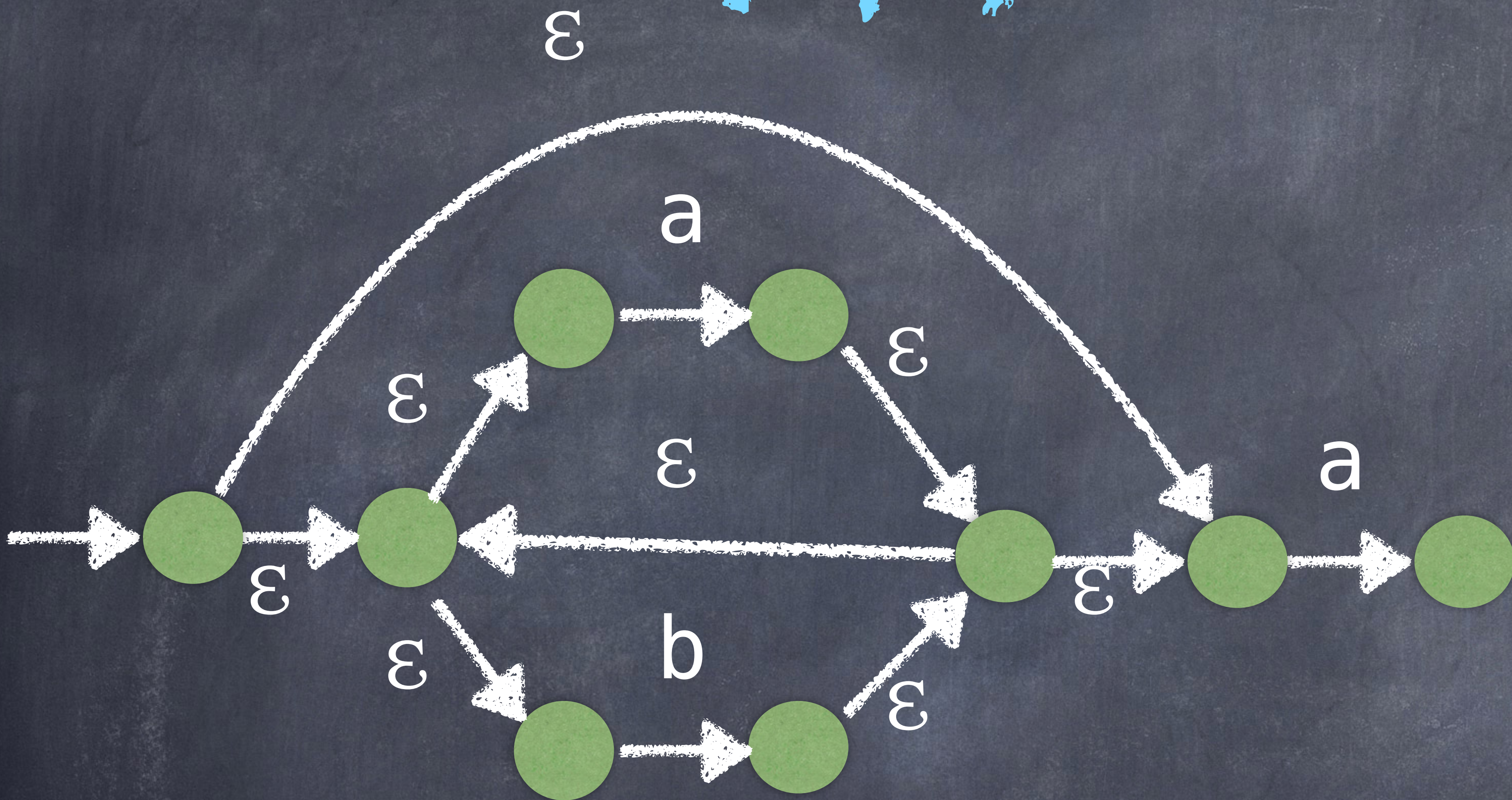
(a|b)*abb



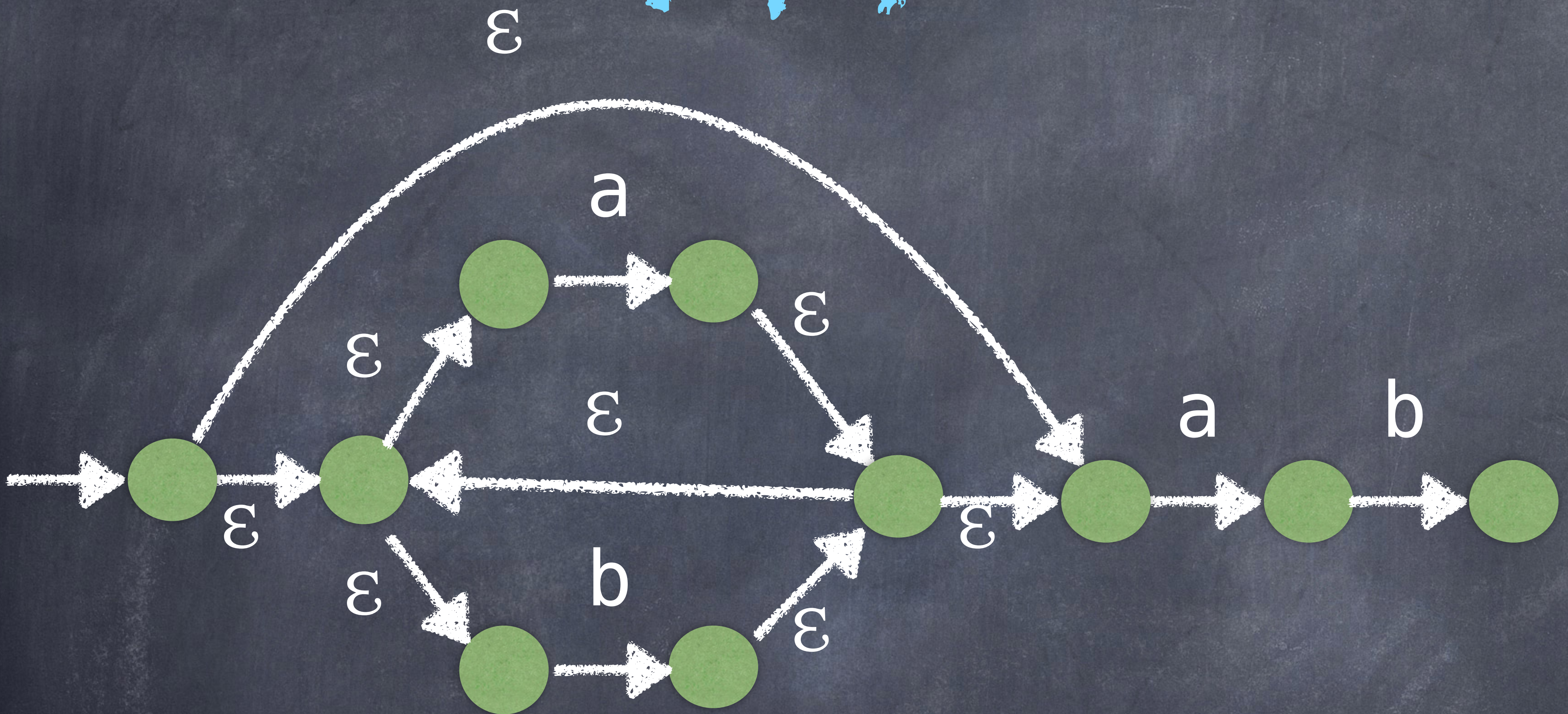
(a|b)*abb



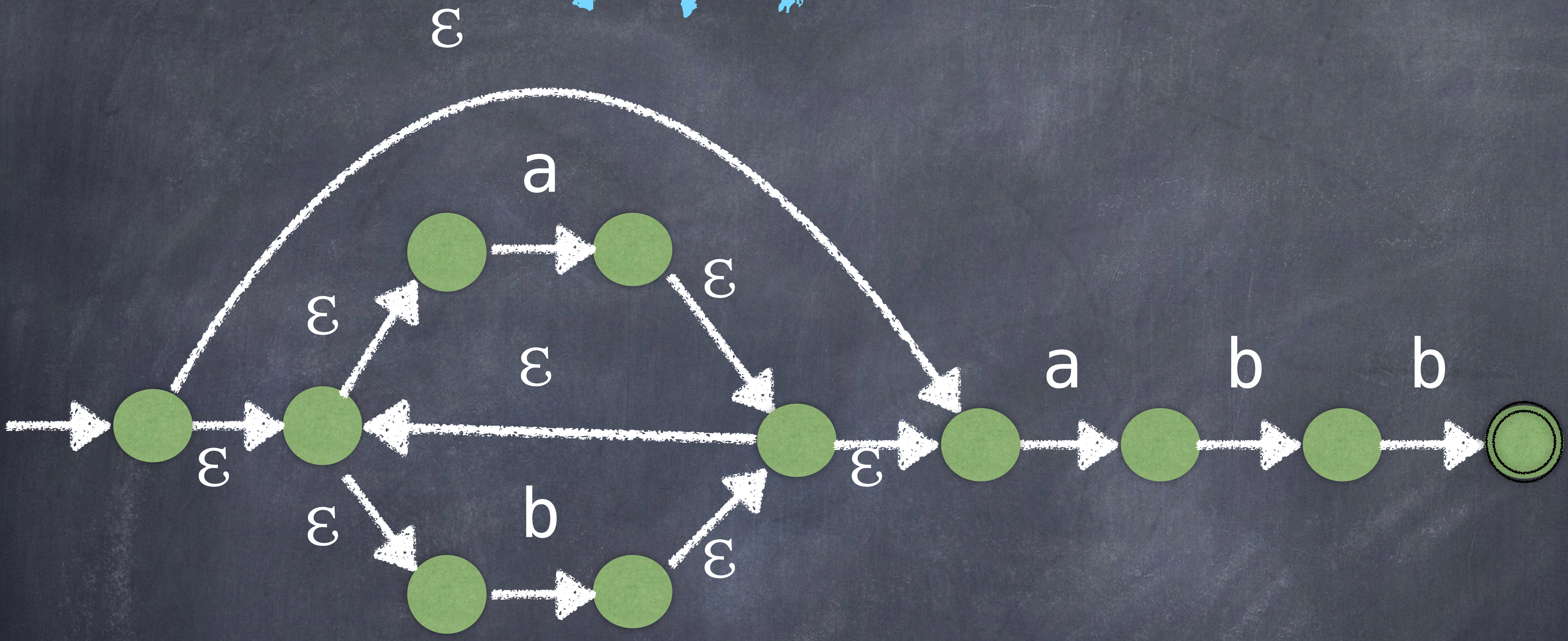
(a|b)*abb



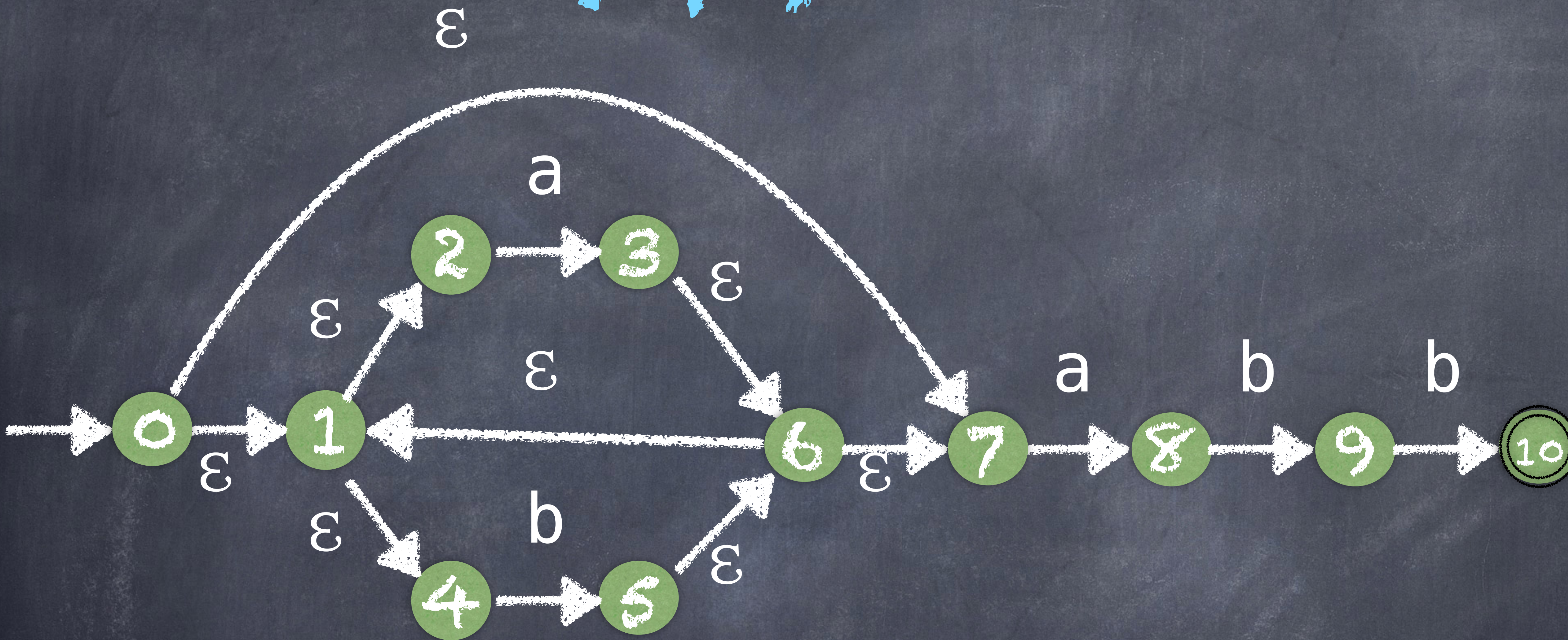
(a|b)*abb



(a|b)*abb



(a|b)*abb



Operations

- ε-closure(t) is the set of states reachable from state t using only ε-transitions.
- ε-closure(T) is the set of states reachable from any state $t \in T$ using only ε-transitions.
- move(T,a) is the set of states reachable from any state $t \in T$ following a transition on symbol $a \in \Sigma$.

NFA -> DFA algorithm

(set of states construction - page 153 of text)

- INPUT: An NFA $N = (S, \Sigma, \delta, s_0, F)$
- OUTPUT: A DFA $D = (S', \Sigma, \delta', s_0', F')$ such that $\mathcal{L}(D)=\mathcal{L}(N)$
- ALGORITHM:

  Compute $s_0' = \varepsilon\text{-closure}(s_0)$, an unmarked set of states

  Set $S' = \{ s_0' \}$

  while there is an unmarked $T \in S'$

      mark T

      for each symbol $a \in \Sigma$

          let $U = \varepsilon\text{-closure}(move(T,a))$

          if $U \notin S'$, add unmarked U to S'

          add transition: $\delta'(T,a) = U$

  F' is the subset of S' all of whose members contain a state in F.

NFA -> DFA algorithm

(set of states construction - page 153 of text)

$S_0' = \{ A = \{0,1,2,4,7\} \}$

Pick an unmarked set from S_0', A, mark it, and $\forall x \in \Sigma$ let $U = \varepsilon\text{-closure}(move(A,x))$, if $U \notin S'$, add unmarked U to S' and add transition: $\delta'(A,x) = U$

$S_1' = \{ A✓ , B = \{1,2,3,4,6,7,8\} , C = \{1,2,4,5,6,7\}\}$

$\delta'(A,a) = B$

$\delta'(A,b) = C$

Pick an unmarked set from S_1', B, mark it, and $\forall x \in \Sigma$ let $U = \varepsilon\text{-closure}(move(B,x))$, if $U \notin S'$, add unmarked U to S' and add transition: $\delta'(B,x) = U$

$S_2' = \{ A✓ , B✓ , C , D = \{1,2,4,5,6,7,9\}\}$

$\delta'(B,a) = B$

$\delta'(B,b) = D$

Pick an unmarked set from S_2', C, mark it, and $\forall x \in \Sigma$ let $U = \varepsilon\text{-closure}(move(C,x))$, if $U \notin S'$, add unmarked U to S' and add transition: $\delta'(C,x) = U$

$S_3' = \{ A✓ , B✓ , C✓ , D \}$

$\delta'(C,a) = B$

$\delta'(C,b) = C$

NFA -> DFA algorithm

(set of states construction - page 153 of text)

Pick an unmarked set from S_3', D, mark it, and $\forall$ x $\in \Sigma$ let U = ε-closure(move(D,x)), if U $\notin$ S', add unmarked U to S' and add transition: $\delta'(D,x) = U$

$S_4' = \{ A✓, B✓, C✓, D✓, E = \{1,2,4,5,6,7,10\} \}$

$\delta'(D,a) = B$

$\delta'(D,b) = E$

Pick an unmarked set from S_4', E, mark it, and $\forall$ a $\in \Sigma$ let U = ε-closure(move(E,a)), if U $\notin$ S', add unmarked U to S' and add transition: $\delta'(E,a) = U$

$S_5' = \{ A✓, B✓, C✓, D✓, E✓ \}$

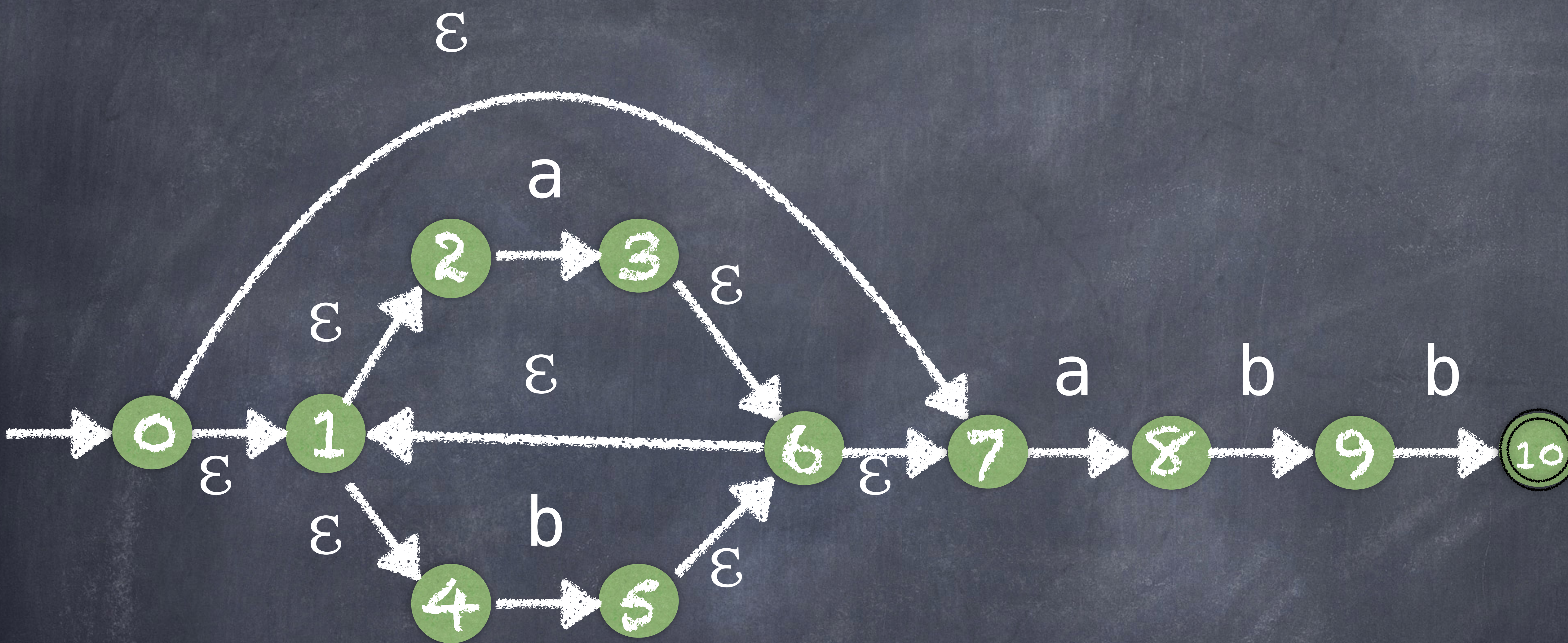
$\delta'(E,a) = B$

$\delta'(E,b) = C$

Since there are no unmarked sets in S_5' the algorithm has reached a fixed point. STOP.

F' is the subset of S' all of whose members contain a state in F: {E}

The original NFA



The resulting DFA

DFA = ({A, B, C, D, E}, {a, b}, A, δ', {E}), where

δ'(A,a) = B

δ'(A,b) = C

δ'(B,a) = B

δ'(B,b) = D

δ'(C,a) = B

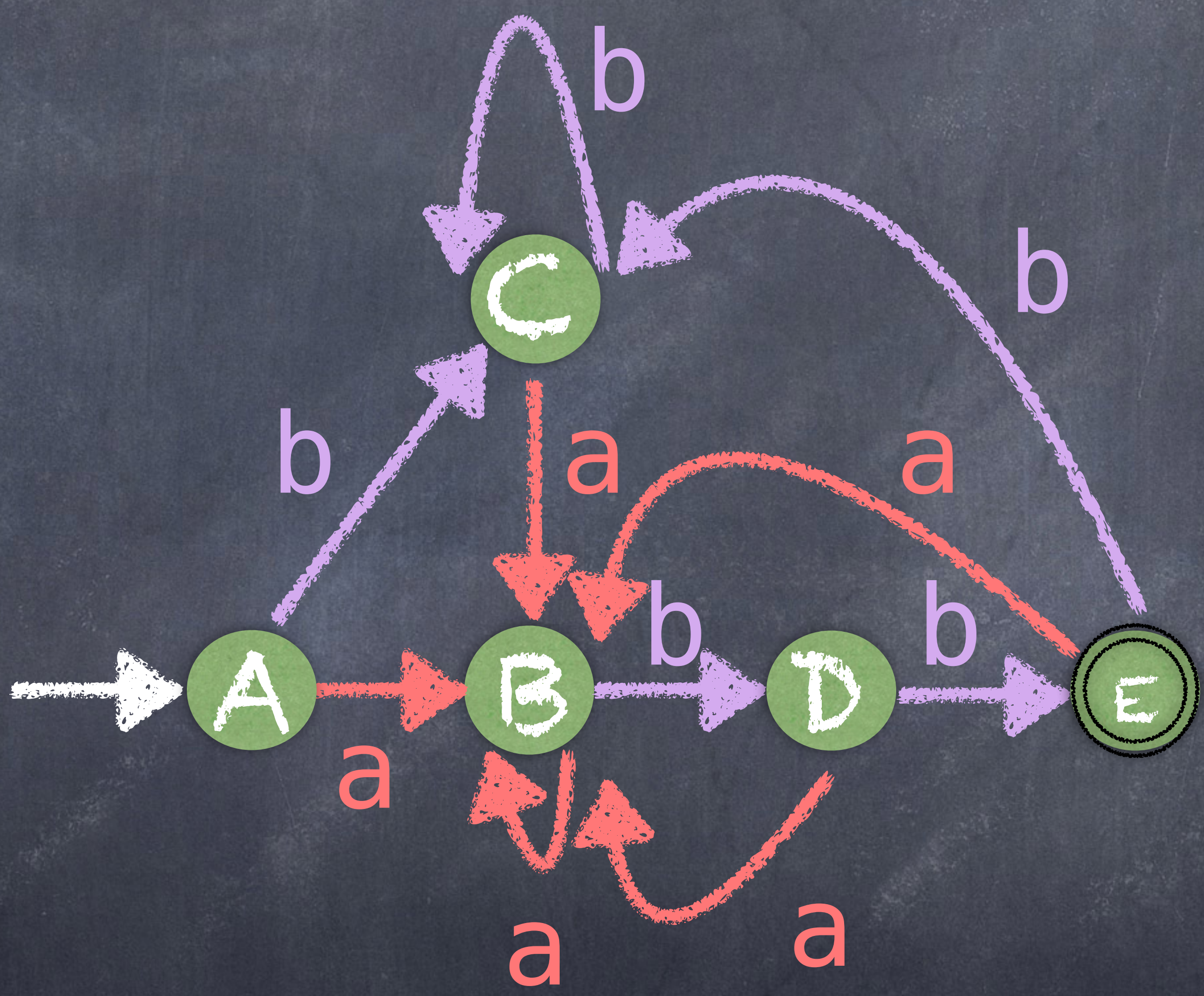
δ'(C,b) = C

δ'(D,a) = B

δ'(D,b) = E

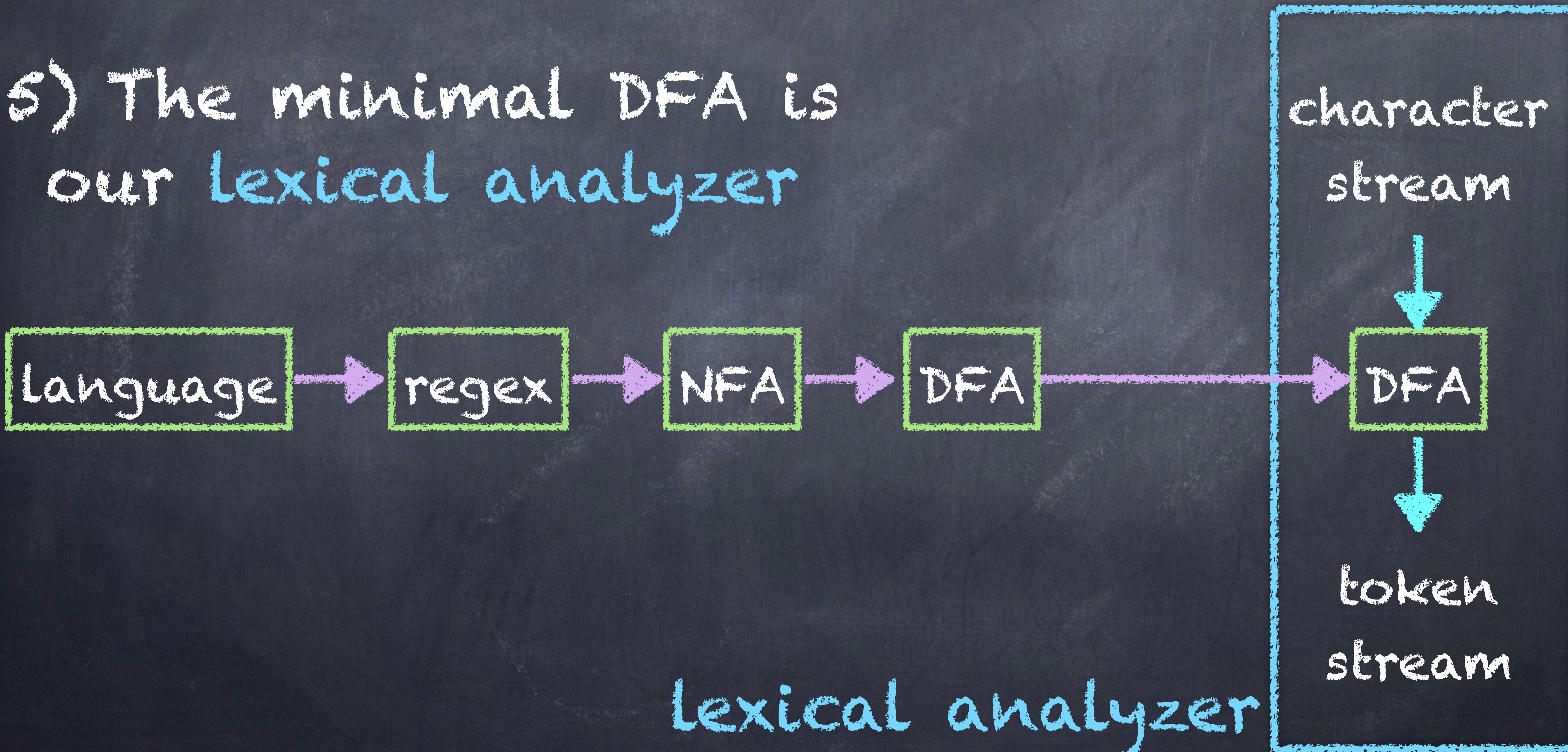
δ'(E,a) = B

δ'(E,b) = C



Process of building lexical analyzer

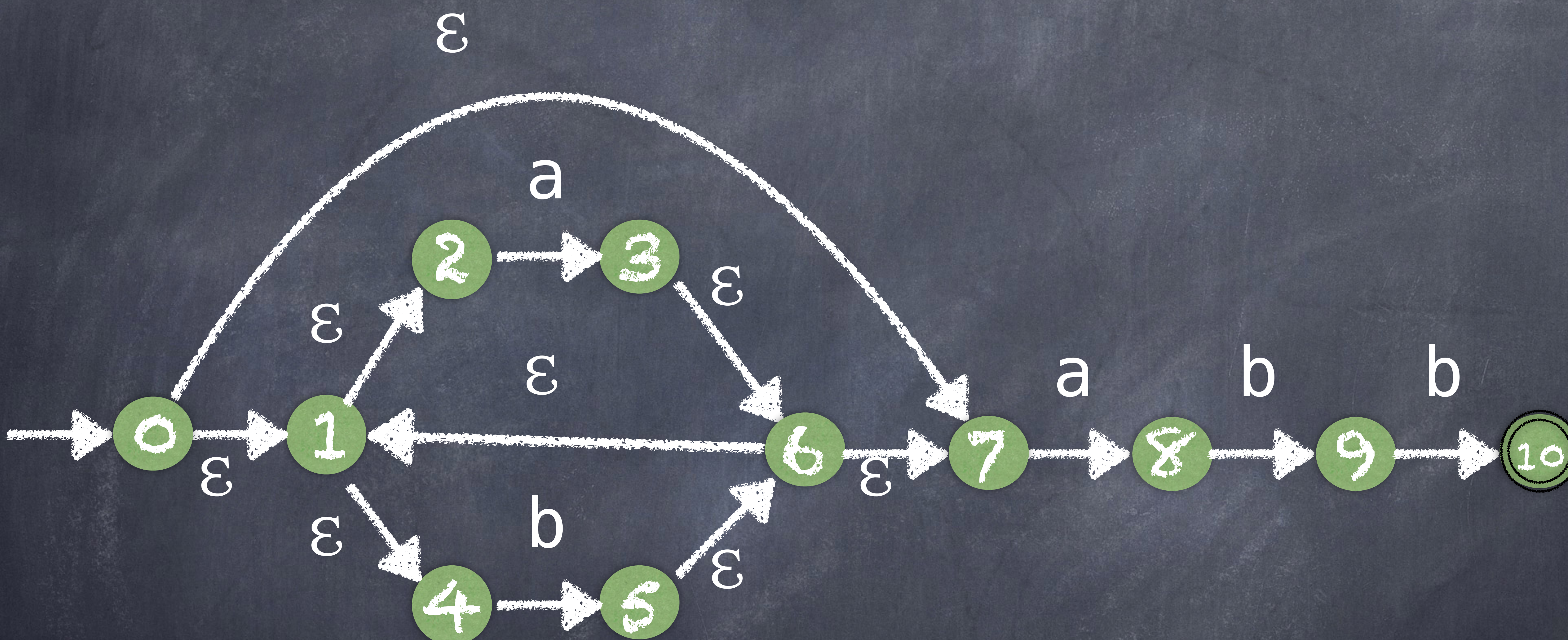
5) The minimal DFA is our lexical analyzer



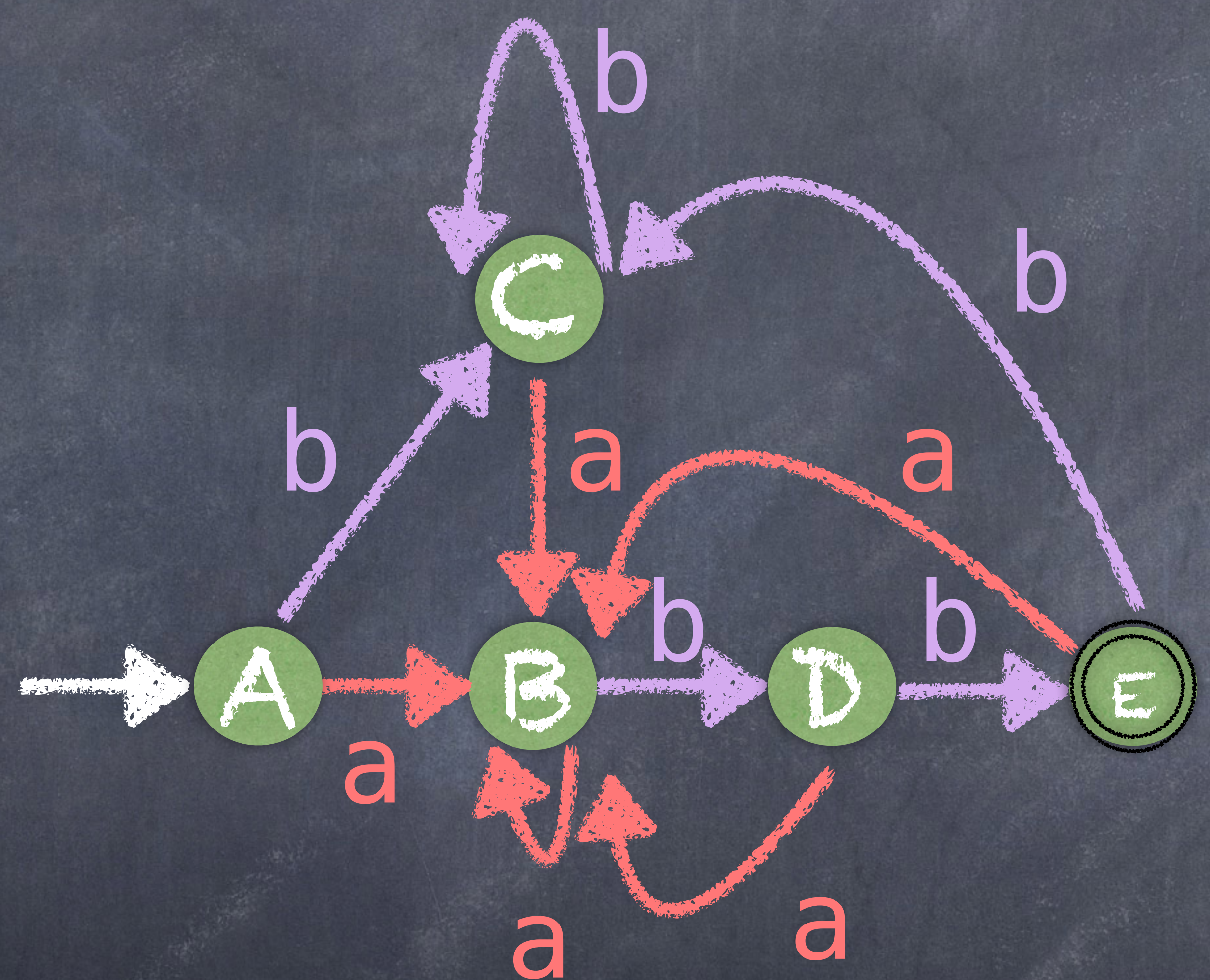
focus above:
NFA to DFA conversion

next step: DFA minimization

NFA for (a|b)*abb



DFA for (a|b)*abb



Minimization Algorithm

DFA -> minimal DFA algorithm

- INPUT: An DFA $D = (S, \Sigma, \delta, s_0, F)$
- OUTPUT: A DFA $D' = (S', \Sigma, \delta', s_0', F')$ such that
  - S' is as small as possible, and
  - $\mathcal{L}(D)=\mathcal{L}(D')$
- ALGORITHM:

1. Let $\pi = \{ F, S-F \}$
2. Let $\pi' = \pi$. For every group G of π:

   partition G into subgroups such that two states s and t are in the same subgroup iff for all input symbols a, states s and t have transitions on a to states in the same group of π

   Replace G in π' by the set of all subgrops formed
3. if $\pi'=\pi$ let $\pi''=\pi$, otherwise set $\pi=\pi'$ and repeat 2.
4. Choose one state in each group of π'' as a representative for that group.

   a) The start state of D' is the representative of the group containing the start state of D

   b) The accepting states of D' are the representatives of those groups that contain an accepting state of D

   c) Adjust transitions from representatives to representatives.

ORIGINAL DFA

$D = (S, \Sigma, s_0, \delta, F)$

$S = \{A, B, C, D, E\}$

$\Sigma = \{a, b\}$

$s_0 = A$

$\delta = \{(A,a) \to B, (A,b) \to C,$
$(B,a) \to B, (B,b) \to D,$
$(C,a) \to B, (C,b) \to C,$
$(D,a) \to B, (D,b) \to E,$
$(E,a) \to B, (E,b) \to C\}$

$F = \{E\}$

Finding the minimal set of distinct sets of states

π_0 = { F, S-F } = { {E}, {A,B,C,D} }

Pick a non-singleton set X = {A,B,C,D} from π_0 and

check behavior of states on all transitions on symbols in Σ (are they to states in X or to other groups in the partition?)

(A,a)->B, (B,a)->B, (C,a)->B, (D,a)->B
(A,b)->C, (B,b)->D, (C,b)->C, (D,b)->E

D behaves differently, so put it in its own partition.

Finding the minimal set of distinct sets of states

π_1 = { {E}, {A, B, C}, {D} }

Pick a non-singleton set X = {A,B,C} from π_1 and

check behavior of states on all transitions on symbols in Σ (are they to states in X or to other groups in the partition?)

(A,a)->B, (B,a)->B, (C,a)->B
(A,b)->C, (B,b)->D, (C,b)->C

B behaves differently, so put it in its own partition.

Finding the minimal set of distinct sets of states

π_2 = { {E}, {A, C}, {B}, {D} }

Pick a non-singleton set X = {A,C} from π_2 and check

behavior of states on all transitions on symbols in Σ (are they to states in X or to other groups in the partition?)

(A,a)->B, (C,a)->B
(A,b)->C, (C,b)->C

A and C both transition outside the group on symbol a, to the same group (the one containing B). Therefore A and C are indistinguishable in their behaviors, so do not split this group.

Finding the minimal set of distinct sets of states

$\pi_3 = \{ \{E\}, \{A, C\}, \{B\}, \{D\} \} = \pi_2$

We have reached a fixed point! STOP

Pick a representative from each group

π_{FINAL} = { {E}, {A, C}, {B}, {D} }

MINIMAL DFA

$D' = (S', \Sigma, s'_0, \delta', F')$

$S' = \{B, C, D, E\}$ -> the representatives

$\Sigma = \{a, b\}$ -> no change

$s'_0 = C$ -> the representative of the group that contained D's starting state, A

δ = (on next slide)

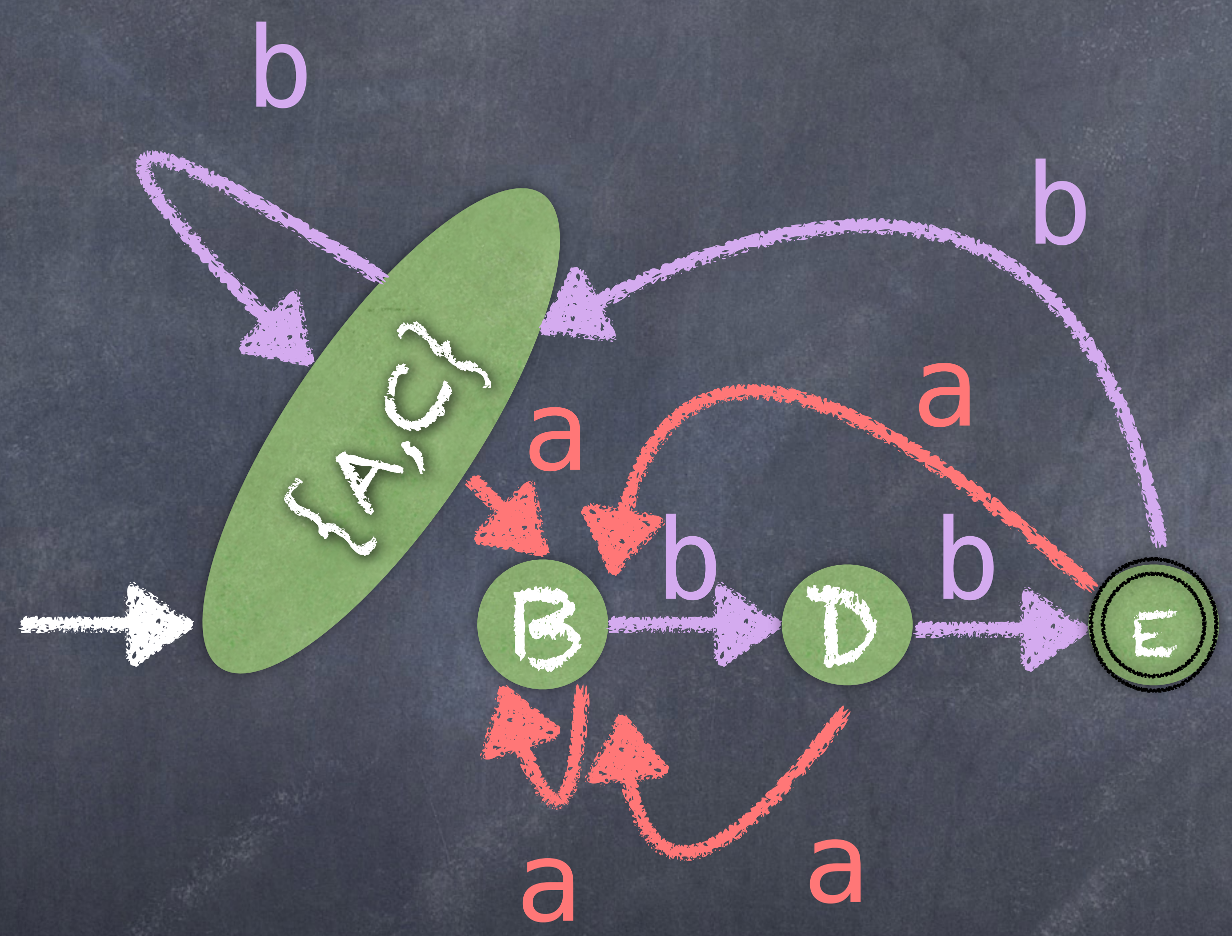
$F = \{E\}$ -> the representatives of all the groups that contained any of D's final states (which, in this case, was just $\{E\}$)

The new transition function δ'

- For each state $s \in S'$, consider its transitions in D, on each $a \in \Sigma$.
- if $\delta(s,a) = t$, then $\delta'(s,a) = r$, where r is the representative of the group containing t.

$$\delta = \{ (B,a) \to B,\ (B,b) \to D,\ (C,a) \to B,\ (C,b) \to C,\ (D,a) \to B,\ (D,b) \to E,\ (E,a) \to B,\ (E,b) \to C \}$$

Minimal DFA for (a|b)*abb



DFA for (a|b)*abb

Non-minimized