#### FINITE STATE AUTOMATA

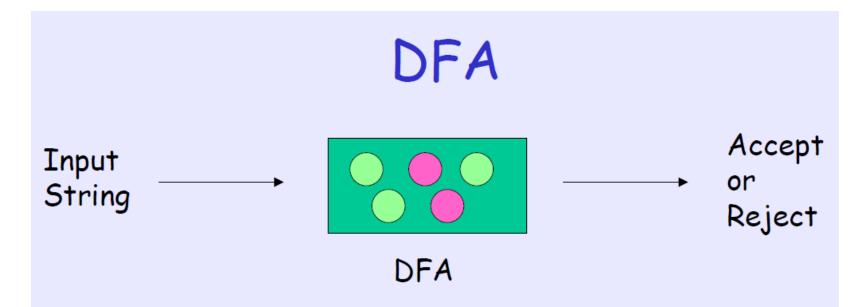
- Finite State Automata (FSA) are the simplest automata.
- Only the finite memory in the control unit is available.
- The memory can be in one of finite states at a given time – hence the name.
  - One can remember only a (fixed) finite number of properties of the past input.
  - Since input strings can be of arbitrary length, it is not possible to remember unbounded portions of the input string.
- It comes in Deterministic and Nondeterministic flavors.

Example: Switch

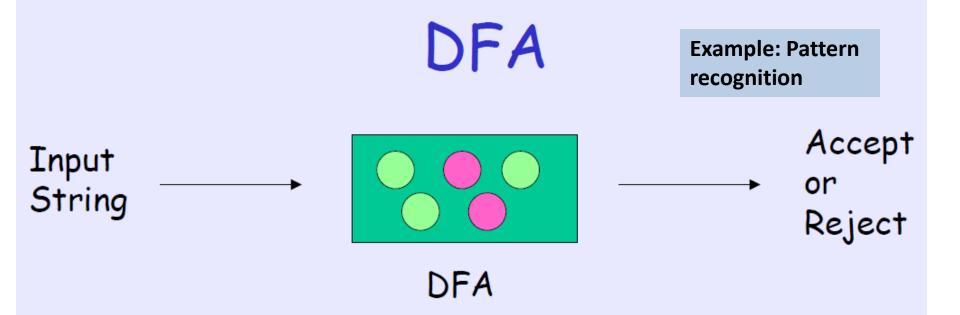
# DETERMINISTIC FINITE STATE AUTOMATA (DFA)

- A DFA starts in a start state and is presented with an input string.
- It moves from state to state, reading the input string one symbol at a time.
- What state the DFA moves next depends on
  - the current state,
  - current input symbol
- When the last input symbol is read, the DFA decides whether it should accept the input string

### Finite State Machines

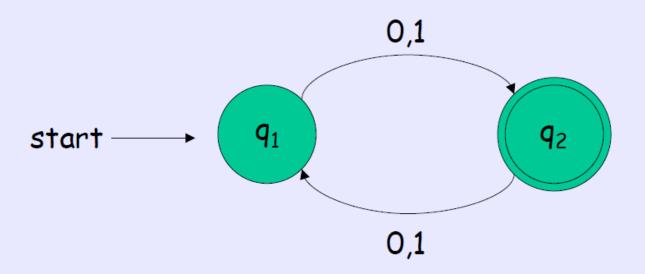


- A machine with finite number of states, some states are accepting states, others are rejecting states
- · At any time, it is in one of the states
- It reads an input string, one character at a time



- After reading each character, it moves to another state depending on what is read and what is the current state
- If reading all characters, the DFA is in an accepting state, the input string is accepted.
- Otherwise, the input string is rejected.

# Example of DFA



- The circles indicates the states
- If accepting state is marked with double circle
- The arrows pointing from a state q indicates how to move on reading a character when current state is q

#### DFA - FORMAL DEFINITION

- A Deterministic Finite State Acceptor (DFA) is defined as the 5-tuple  $M = (Q, \Sigma, \delta, q_0, F)$  where
  - Q is a finite set of states
  - Σ is a finite set of symbols the alphabet
  - $\delta: Q \times \Sigma \to Q$  is the next-state function
  - $q_0 \in Q$  is the (label of the) start state
  - F ⊆ Q is the set of final (accepting) states

#### DFA - FORMAL DEFINITION

- A Deterministic Finite State Acceptor (DFA) is defined as the 5-tuple  $M = (Q, \Sigma, \delta, q_0, F)$  where
  - Q is a finite set of states
  - Σ is a finite set of symbols the alphabet
  - $\delta: Q \times \Sigma \to Q$  is the next-state function
  - $q_0 \in Q$  is the (label of the) start state
  - $F \subseteq Q$  is the set of final (accepting) states

Note, there must be exactly one start state. Final states can be many or even empty!

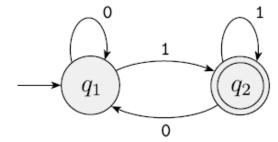
# Some Terminology

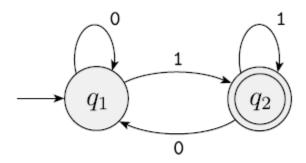
#### Let M be a DFA

- Among all possible strings, M will accept some of them, and M will reject the remaining
- The set of strings which M accepts is called the language recognized by M
- That is, M recognizes A if
   A = { w | M accepts w }

### L(M)

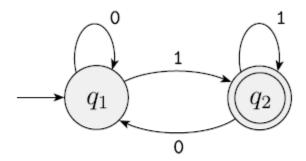
If A is the set of all strings that machine M accepts, we say that A is the language of machine M and write L(M) = A. We say that M recognizes A or that M accepts A.





In the formal description,  $M_2$  is  $(\{q_1, q_2\}, \{0,1\}, \delta, q_1, \{q_2\})$ . The transition function  $\delta$  is

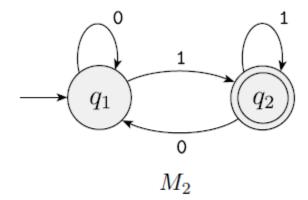
$$\begin{array}{c|cccc} & 0 & 1 \\ \hline q_1 & q_1 & q_2 \\ q_2 & q_1 & q_2. \end{array}$$



In the formal description,  $M_2$  is  $(\{q_1, q_2\}, \{0,1\}, \delta, q_1, \{q_2\})$ . The transition function  $\delta$  is

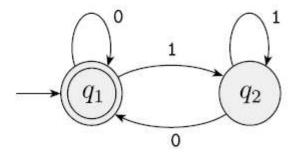
$$egin{array}{c|cccc} & 0 & 1 \\ \hline q_1 & q_1 & q_2 \\ q_2 & q_1 & q_2. \end{array}$$

Remember that the state diagram of  $M_2$  and the formal description of  $M_2$  contain the same information, only in different forms. You can always go from one to the other if necessary.



$$L(M_2) = \{ \stackrel{\cdot}{w} | \stackrel{\cdot}{w} \text{ ends in a 1} \}.$$

Consider the finite automaton  $M_3$ .



#### FIGURE 1.10

State diagram of the two-state finite automaton  $M_3$ 

Can you describe this in the 5 tuple form? In particular, can you write down the transition table?

#### Consider the finite automaton $M_3$ .

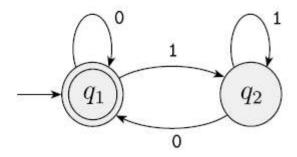
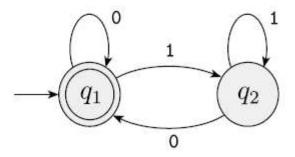


FIGURE 1.10

State diagram of the two-state finite automaton  $M_3$ 

What language  $M_3$  recognizes?

Consider the finite automaton  $M_3$ .



#### FIGURE 1.10

State diagram of the two-state finite automaton  $M_3$ 

What language  $M_3$  recognizes?

 $L(M_3) = \{w | w \text{ is the empty string } \varepsilon \text{ or ends in a 0} \}.$ 

#### EXAMPLE 1.11 .....

The following figure shows a five-state machine  $M_4$ .

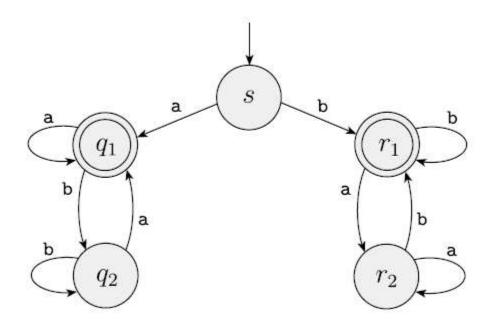


FIGURE 1.12 Finite automaton  $M_4$ 

#### EXAMPLE 1.11 .....

The following figure shows a five-state machine  $M_4$ .

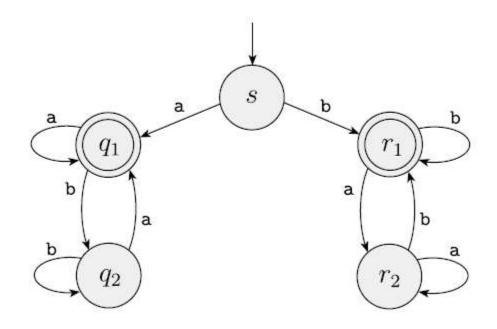


FIGURE 1.12 Finite automaton  $M_4$ 

 $L(M_4)$  = all strings that begin and end with the same character.

## DFA for complement of a language

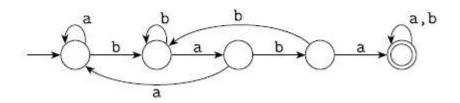
Flip final and non-final states.

- 1.5 Each of the following languages is the complement of a simpler language. In each part, construct a DFA for the simpler language, then use it to give the state diagram of a DFA for the language given. In all parts, Σ = {a, b}.
  - Aa.  $\{w \mid w \text{ does not contain the substring ab}\}$
  - Ab.  $\{w \mid w \text{ does not contain the substring baba}\}$

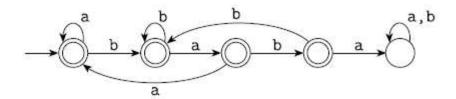
1.5 (a) The left-hand DFA recognizes  $\{w | w \text{ contains ab}\}$ . The right-hand DFA recognizes its complement,  $\{w | w \text{ doesn't contain ab}\}$ .



(b) This DFA recognizes  $\{w | w \text{ contains baba}\}.$ 



This DFA recognizes  $\{w | w \text{ does not contain baba}\}.$ 



# Designing a DFA (Quick Quiz)

 How to design a DFA that accepts all binary strings representing a multiple of 5? (E.g., 101, 1111, 11001, ...)

# Formally

Let  $M = (Q, \Sigma, \delta, q_0, F)$  be a finite automaton and let  $w = w_1 w_2 \cdots w_n$  be a string where each  $w_i$  is a member of the alphabet  $\Sigma$ . Then M accepts w if a sequence of states  $r_0, r_1, \ldots, r_n$  in Q exists with three conditions:

- 1.  $r_0 = q_0$ ,
- 2.  $\delta(r_i, w_{i+1}) = r_{i+1}$ , for i = 0, ..., n-1, and
- 3.  $r_n \in F$ .

## Regular language [Ref: Sipser Book]

#### DEFINITION 1.16

A language is called a *regular language* if some finite automaton recognizes it.

### The regular operations

#### DEFINITION 1.23

Let A and B be languages. We define the regular operations union, concatenation, and star as follows:

- Union:  $A \cup B = \{x | x \in A \text{ or } x \in B\}.$
- Concatenation:  $A \circ B = \{xy | x \in A \text{ and } y \in B\}.$
- Star:  $A^* = \{x_1 x_2 \dots x_k | k \ge 0 \text{ and each } x_i \in A\}.$

- These are similar to arithmetic operations.
- Note, \* is a unary operator.

The class of regular languages is closed under the union operation.

In other words, if  $A_1$  and  $A_2$  are regular languages, so is  $A_1 \cup A_2$ .

- The proof is by construction.
- We build a DFA for the union from the individual DFAs.

- The idea is simple: While reading the input simultaneously follow both machines.
  - Put a finger on current state. You need two fingers.
     You can move these two fingers as per the respective transition function.

#### PROOF

Let  $M_1$  recognize  $A_1$ , where  $M_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ , and  $M_2$  recognize  $A_2$ , where  $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$ .

Construct M to recognize  $A_1 \cup A_2$ , where  $M = (Q, \Sigma, \delta, q_0, F)$ .

Q = {(r<sub>1</sub>, r<sub>2</sub>)| r<sub>1</sub> ∈ Q<sub>1</sub> and r<sub>2</sub> ∈ Q<sub>2</sub>}.
 This set is the *Cartesian product* of sets Q<sub>1</sub> and Q<sub>2</sub> and is written Q<sub>1</sub> × Q<sub>2</sub>.
 It is the set of all pairs of states, the first from Q<sub>1</sub> and the second from Q<sub>2</sub>.

2. Σ, the alphabet, is the same as in M₁ and M₂. In this theorem and in all subsequent similar theorems, we assume for simplicity that both M₁ and M₂ have the same input alphabet Σ. The theorem remains true if they have different alphabets, Σ₁ and Σ₂. We would then modify the proof to let Σ = Σ₁ ∪ Σ₂.

3.  $\delta$ , the transition function, is defined as follows. For each  $(r_1, r_2) \in Q$  and each  $a \in \Sigma$ , let

$$\delta((r_1, r_2), a) = (\delta_1(r_1, a), \delta_2(r_2, a)).$$

Hence  $\delta$  gets a state of M (which actually is a pair of states from  $M_1$  and  $M_2$ ), together with an input symbol, and returns M's next state.

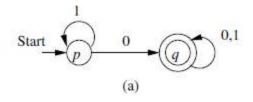
**4.**  $q_0$  is the pair  $(q_1, q_2)$ .

5. F is the set of pairs in which either member is an accept state of  $M_1$  or  $M_2$ . We can write it as

$$F = \{(r_1, r_2) | r_1 \in F_1 \text{ or } r_2 \in F_2\}.$$

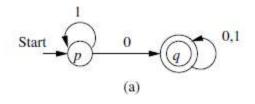
This expression is the same as  $F = (F_1 \times Q_2) \cup (Q_1 \times F_2)$ . (Note that it is *not* the same as  $F = F_1 \times F_2$ . What would that give us instead?<sup>3</sup>)

## Union Example

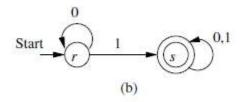


What is the language recognized by this DFA?

## Union Example



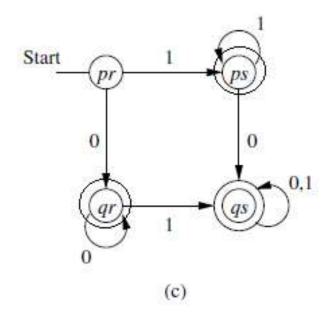
What is the language recognized by this DFA?



What is the language recognized by this DFA?

### Find DFA for the union

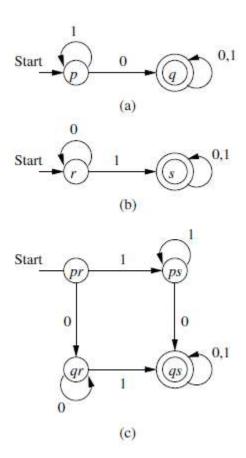
### Find DFA for the union



### What about intersection?

- Intersection of two regular languages is also regular.
- Proof: by construction. Similar. Only final states will change.

### Intersection



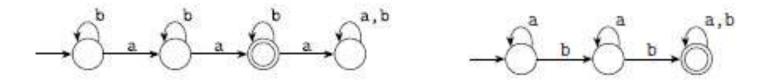
# What else we can do with product principle?

- Set difference.
  - How?

$$A - B = A \cap \overline{B}$$

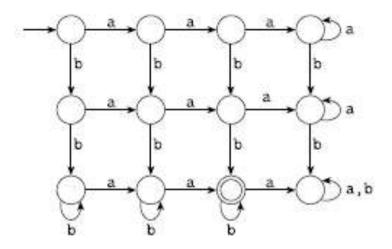
- 1.4 Each of the following languages is the intersection of two simpler languages. In each part, construct DFAs for the simpler languages, then combine them using the construction discussed in footnote 3 (page 46) to give the state diagram of a DFA for the language given. In all parts, Σ = {a, b}.
  - a.  $\{w \mid w \text{ has at least three a's and at least two b's}\}$
  - Ab.  $\{w \mid w \text{ has exactly two a's and at least two b's}\}$ 
    - c.  $\{w \mid w \text{ has an even number of a's and one or two b's}\}$
  - Ad.  $\{w \mid w \text{ has an even number of a's and each a is followed by at least one b}\}$ 
    - e.  $\{w | w \text{ starts with an a and has at most one b} \}$
    - f.  $\{w \mid w \text{ has an odd number of a's and ends with a b}\}\$
    - g.  $\{w \mid w \text{ has even length and an odd number of a's}\}$

1.4 (b) The following are DFAs for the two languages {w | w has exactly two a's} and {w | w has at least two b's}.



Now find product machine.

Combining them using the intersection construction gives the following DFA.



• This can be minimized. {Some states are redundant}.

### NONDETERMINISM

- Useful concept, has great impact on ToC/algorithms.
- DFA is deterministic: every step of a computation follows in a unique way from the preceding step.
  - When the machine is in a given state, and upon reading the next input symbol, we know deterministically what would be the next state.
  - Only one next state.
  - No choice !!

### NONDETERMINISM

- In a nondeterministic machine, several choices may exist for the next state at any point.
- Nondeterminism is a generalization of determinism.

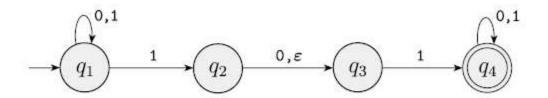


FIGURE 1.27

The nondeterministic finite automaton  $N_1$ 

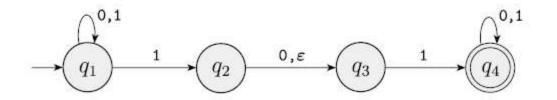
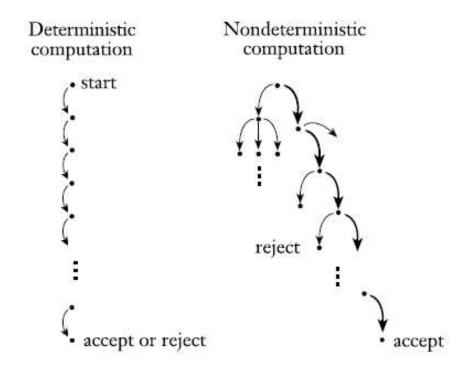


FIGURE 1.27
The nondeterministic finite automaton  $N_1$ 

- More than one arrow from from  $q_1$  on symbol 1.
- No arrow at all from  $q_3$  on 0.
- There is & over an arrow!

# How does an NFA compute?



Deterministic and nondeterministic computations with an accepting branch

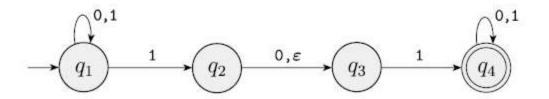


FIGURE 1.27

The nondeterministic finite automaton  $N_1$ 

On input **010110** 

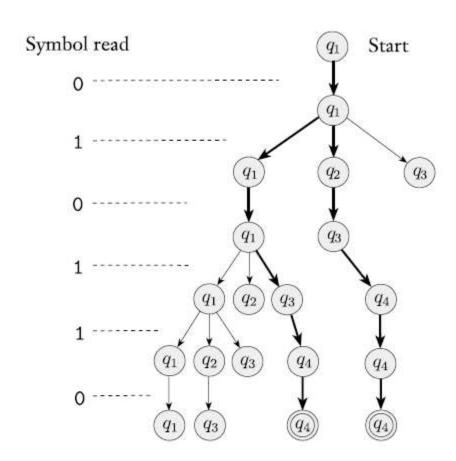
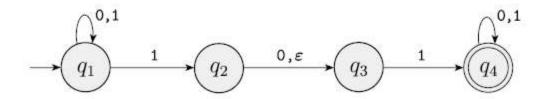


FIGURE 1.29 The computation of  $N_1$  on input 010110



### FIGURE 1.27 The nondeterministic finite automaton $N_1$

What is the language accepted by this NFA?

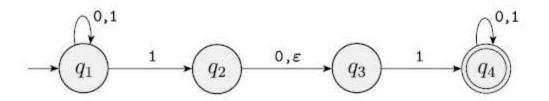


FIGURE 1.27
The nondeterministic finite automaton  $N_1$ 

 It accepts all strings that contain either 101 or 11 as a substring.

- Constructing NFAs is sometimes easier than constructing DFAs.
  - Later we see that every NFA can be converted into an equivalent DFA.

Let A be the language consisting of all strings over  $\{0,1\}$  containing a 1 in the third position from the end (e.g., 000100 is in A but 0011 is not). The following four-state NFA  $N_2$  recognizes A.

- Building DFA for this is possible, but difficult.
- Try this.

# But NFA is easy to build.

EXAMPLE 1.30

Let A be the language consisting of all strings over  $\{0,1\}$  containing a 1 in the third position from the end (e.g., 000100 is in A but 0011 is not). The following four-state NFA  $N_2$  recognizes A.

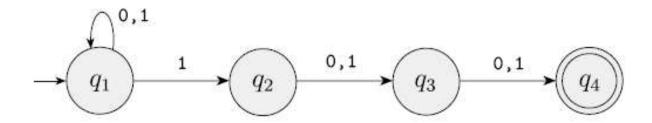
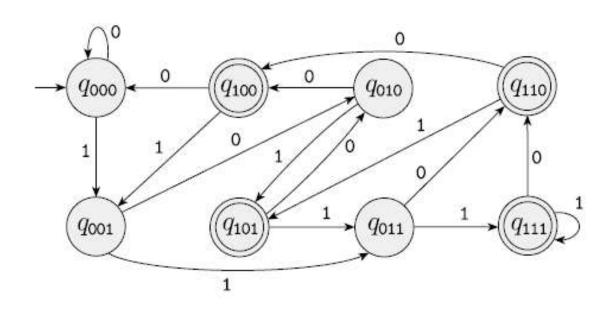


FIGURE 1.31
The NFA  $N_2$  recognizing A

### DFA for A



A DFA recognizing A

See number of states and complexity!

### Formal definition of NFA

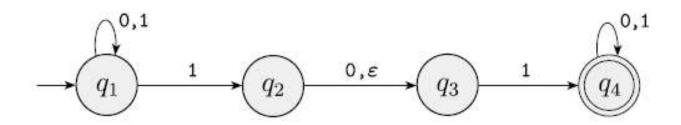
We use  $\Sigma_{\varepsilon}$  to mean  $\Sigma \cup \{\varepsilon\}$ 

#### DEFINITION 1.37

A nondeterministic finite automaton is a 5-tuple  $(Q, \Sigma, \delta, q_0, F)$ , where

- 1. Q is a finite set of states,
- 2.  $\Sigma$  is a finite alphabet,
- 3.  $\delta: Q \times \Sigma_{\varepsilon} \longrightarrow \mathcal{P}(Q)$  is the transition function,
- 4.  $q_0 \in Q$  is the start state, and
- 5.  $F \subseteq Q$  is the set of accept states.

Recall the NFA  $N_1$ :



The formal description of  $N_1$  is  $(Q, \Sigma, \delta, q_1, F)$ , where

1. 
$$Q = \{q_1, q_2, q_3, q_4\},\$$

2. 
$$\Sigma = \{0,1\},\$$

3. 
$$\delta$$
 is given as

	0	1	8
$q_1$	$\{q_1\}$	$\{q_1,q_2\}$	Ø
$q_2$	$\{q_1\}$ $\{q_3\}$	Ø	$\{q_3\}$
$q_3$	Ø	$\{q_4\}$	Ø
$q_4$	$\{q_4\}$	$\{q_4\}$	Ø,

4.  $q_1$  is the start state, and

5. 
$$F = \{q_4\}.$$

The formal definition of computation for an NFA is similar to that for a DFA. Let  $N = (Q, \Sigma, \delta, q_0, F)$  be an NFA and w a string over the alphabet  $\Sigma$ . Then we say that N accepts w if we can write w as  $w = y_1 y_2 \cdots y_m$ , where each  $y_i$  is a member of  $\Sigma_{\varepsilon}$  and a sequence of states  $r_0, r_1, \ldots, r_m$  exists in Q with three conditions:

- 1.  $r_0 = q_0$ ,
- 2.  $r_{i+1} \in \delta(r_i, y_{i+1})$ , for i = 0, ..., m-1, and
- 3.  $r_m \in F$ .

# Equivalence of NFAs and DFAs

 We say two machines are equivalent if they recognize the same language.

THEOREM 1.39 ------

Every nondeterministic finite automaton has an equivalent deterministic finite automaton.

### Proof

- Proof by construction.
  - We build a equal DFA for the given NFA

Let  $N=(Q,\Sigma,\delta,q_0,F)$  be the NFA recognizing some language A. We construct a DFA  $M=(Q',\Sigma,\delta',q_0',F')$  recognizing A.

• First, for understanding purpose, we assume that there are no edges with £ transitions.

Let  $N=(Q,\Sigma,\delta,q_0,F)$  be the NFA recognizing some language A. We construct a DFA  $M=(Q',\Sigma,\delta',q_0',F')$  recognizing A.

- Q' = P(Q).
   Every state of M is a set of states of N. Recall that P(Q) is the set of subsets of Q.
- **2.** For  $R \in Q'$  and  $a \in \Sigma$ , let  $\delta'(R, a) = \{q \in Q | q \in \delta(r, a) \text{ for some } r \in R\}$ .

$$\delta'(R, a) = \bigcup_{r \in R} \delta(r, a).$$

- 3.  $q_0' = \{q_0\}$ .

  M starts in the state corresponding to the collection containing just the start state of N.
- 4.  $F' = \{R \in Q' | R \text{ contains an accept state of } N\}$ . The machine M accepts if one of the possible states that N could be in at this point is an accept state.

# Subset construction: Example

# 

### NFA to DFA construction: Example

**DFA**:

{q₁}

\*{q<sub>2</sub>}

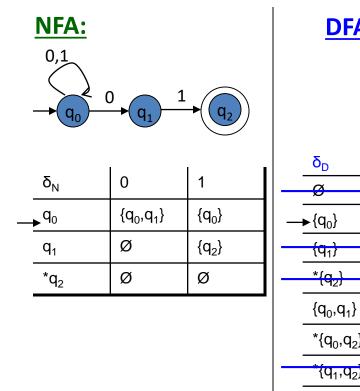
 $*{q_0,q_2}$ 

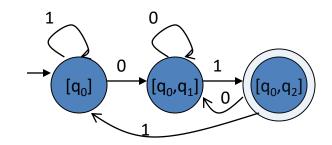
 $\{q_1,q_2\}$ 

 $\{q_0, q_1, q_2\}$ 

 $L = \{w \mid w \text{ ends in } 01\}$ 

Idea: To avoid enumerating all of power set, do "lazy creation of states"



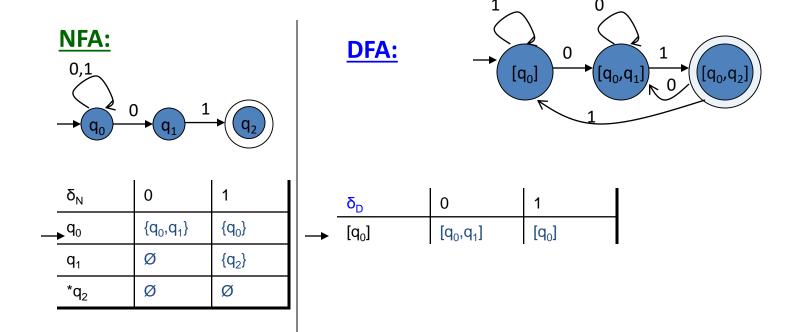


	$\delta_{\text{D}}$	0	1
	▶[q <sub>0</sub> ]	$[q_0, q_1]$	[q <sub>0</sub> ]
$\Longrightarrow$	$[q_0,q_1]$	[q <sub>0</sub> ,q <sub>1</sub> ]	$[q_0, q_2]$
	*[q <sub>0</sub> ,q <sub>2</sub> ]	[q <sub>0</sub> ,q <sub>1</sub> ]	[q <sub>0</sub> ]

- Enumerate all possible subsets 0.
- 1. **Determine transitions**
- Retain only those states reachable from {q<sub>0</sub>}

# NFA to DFA: Repeating the example using *LAZY CREATION*

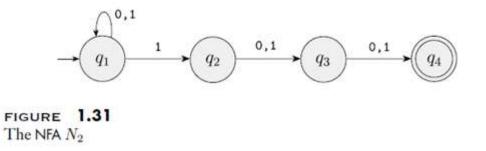
• L = {w | w ends in 01}



#### Main Idea:

Introduce states as you go (on a need basis)

# Can you convert the following



What is the language accepted by this?

# Now, considering & arrows

 For this purpose, we define E-CLOSURE of a set of states R as E(R) or ECLOSE(R)

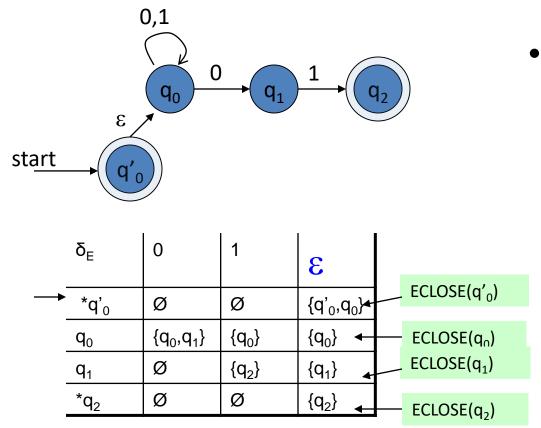
Formally, for  $R \subseteq Q$  let

 $E(R) = \{q \mid q \text{ can be reached from } R \text{ by traveling along } 0 \text{ or more } \varepsilon \text{ arrows} \}.$ 

• E(R) is  $\epsilon$ -CLOSURE of R.

# Example of an $\varepsilon$ -NFA

L = {w | w is empty, or if non-empty will end in 01}



 $\epsilon$ -closure of a state q, ECLOSE(q), is the set of all states (including itself) that can be *reached* from q by repeatedly making an arbitrary number of  $\epsilon$ -transitions.

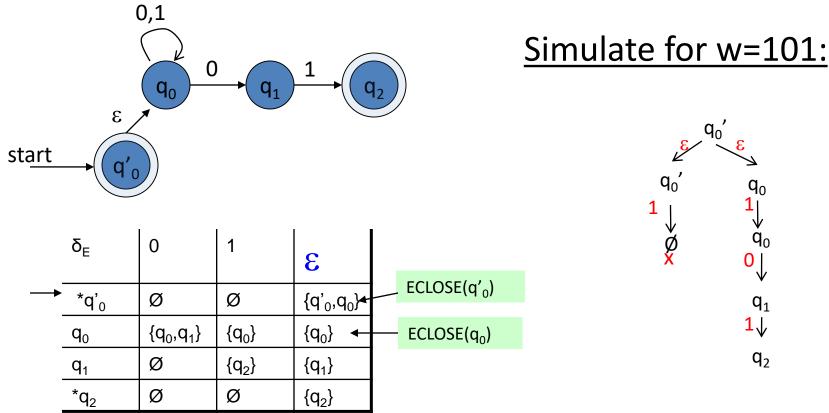
To simulate any transition:

Step 1) Go to all immediate destination states.

Step 2) From there go to all their  $\epsilon$ -closure states as well.

# Example of an $\varepsilon$ -NFA

L = {w | w is empty, or if non-empty will end in 01}

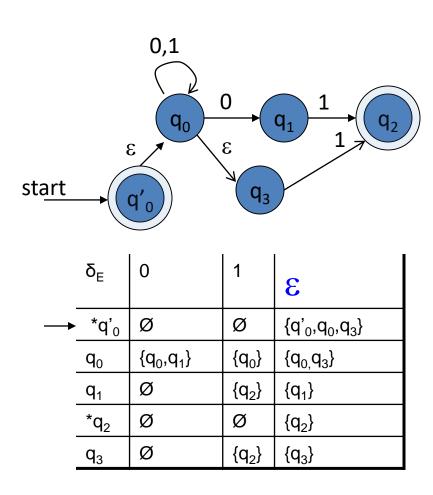


#### To simulate any transition:

Step 1) Go to all immediate destination states.

Step 2) From there go to all their  $\varepsilon$ -closure states as well.

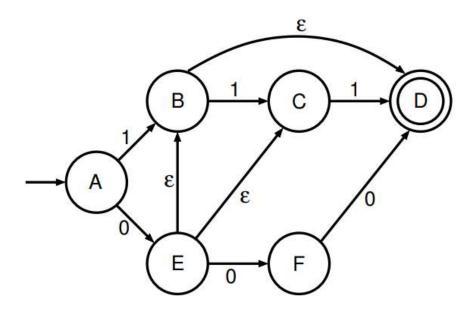
# Example of another $\varepsilon$ -NFA



### Simulate for w=101:

65

### Find ECLOSE for all states



# Epsilon-NFA to DFA

### Formal definition

- $-E = (Q_E, \Sigma, \delta_E, q_0, F_E)$  be a E-NFA
- We define DFA,  $D = (Q_D, \Sigma, \delta_D, q_D, F_D)$ 
  - $Q_D = 2^{QE}$
  - $q_D = ECLOSE(q_0)$
  - F<sub>D</sub> = sets containing at least one state from F<sub>E</sub>

# Epsilon-NFA to DFA

### Computing $\delta_D$

- $-\delta_{D}(S, a)$  for  $S \in Q_{D}, a \in \Sigma$ 
  - Let  $S = \{ p_1, p_2, ..., p_n \}$
  - Compute the set of all states reachable from states in S on input a using transitions from E.

$$\{r_1, r_2, \dots, r_m\} = \bigcup_{i=1}^n \delta_E(p_i, a)$$

 δ<sub>D</sub>(S, a) will be the union of the ε closures of the elements of {r<sub>1</sub>, ..., r<sub>m</sub>}

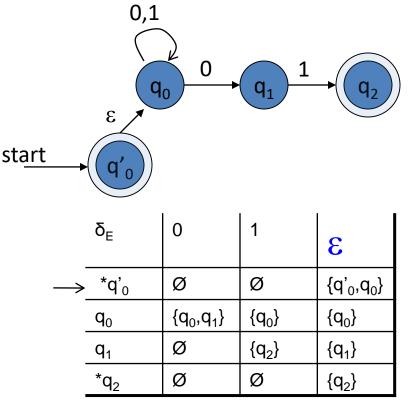
$$\delta_{D}(S,a) = \bigcup_{j=1}^{m} ECLOSE(r_{j})$$

### Epsilon-NFA to DFA

- Step 1: We will take the  $\varepsilon$ -closure for the starting state of NFA as a starting state of DFA.
- **Step 2:** Find the states for each input symbol that can be traversed from the present. That means the union of transition value and their closures for each state of NFA present in the current state of DFA.
- Step 3: If we found a new state, take it as current state and repeat step
  2.
- **Step 4:** Repeat Step 2 and Step 3 until there is no new state present in the transition table of DFA.
- Step 5: Mark the states of DFA as a final state which contains the final state of NFA.

# Example: $\varepsilon$ -NFA $\rightarrow$ DFA

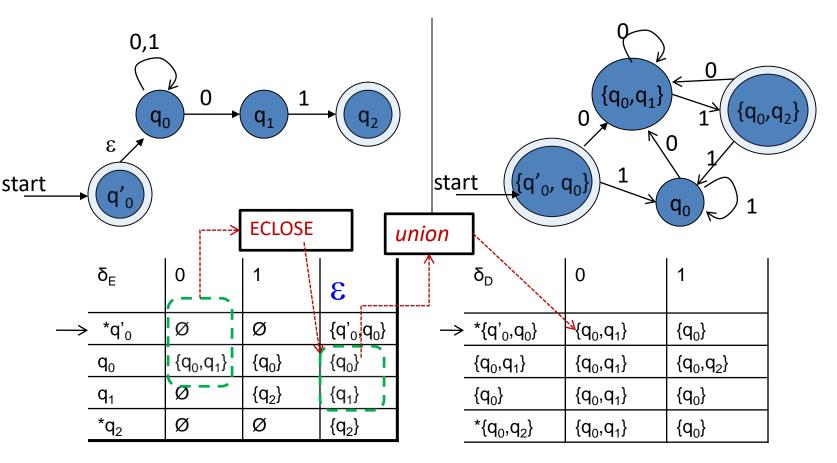
L = {w | w is empty, or if non-empty will end in 01}



	$\delta_{D}$	0	1
$\rightarrow$	*{q' <sub>0</sub> ,q <sub>0</sub> }		

# Example: $\varepsilon$ -NFA $\rightarrow$ DFA

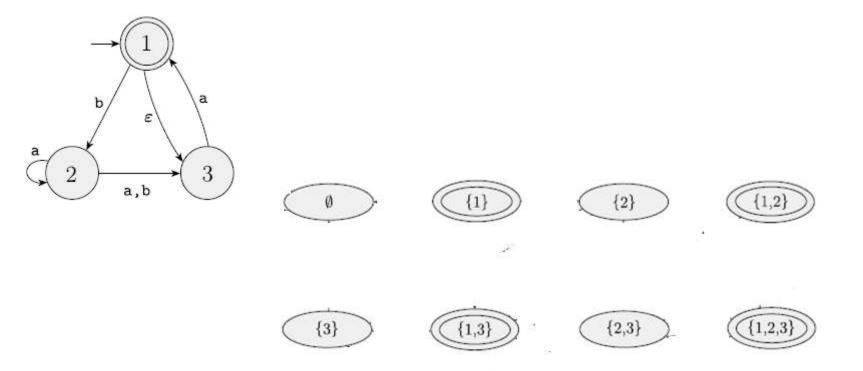
L = {w | w is empty, or if non-empty will end in 01}



# Example

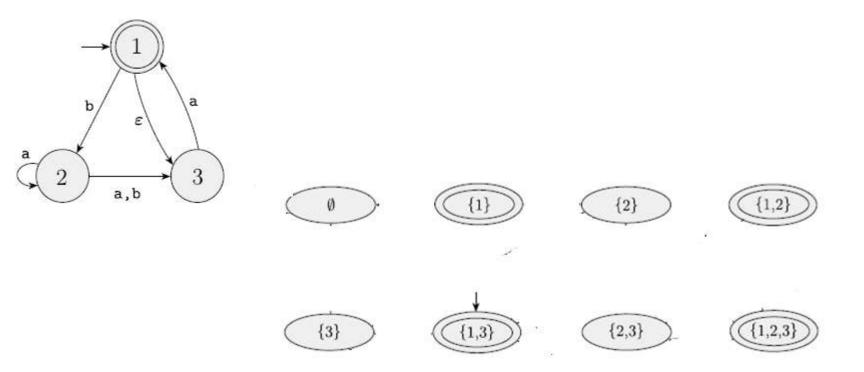
 $\begin{array}{c} \longrightarrow \\ 1 \\ \downarrow \\ 2 \\ \downarrow \\ a,b \end{array}$ 

FIGURE 1.42 The NFA  $N_4$ 

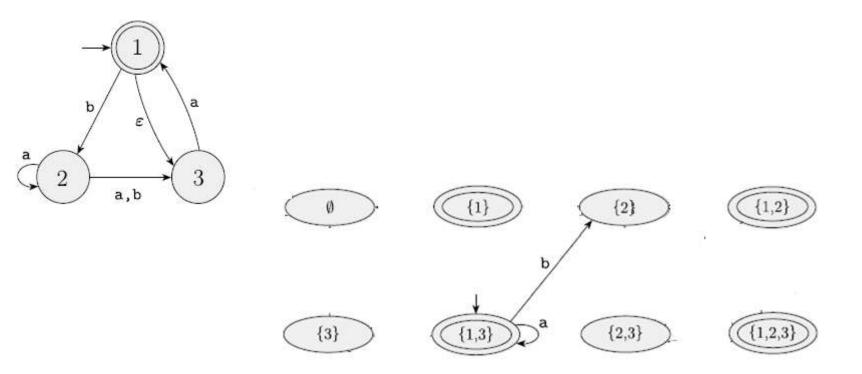


All possible states of the DFA. (to be constructed; Final states are shown)

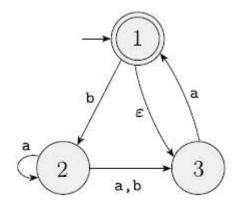
- Now we need to add edges, and
- identify the initial state.



- Identify the initial state.
  - Note, it is not {1}



• Adding edges, ...



# After all edges ...

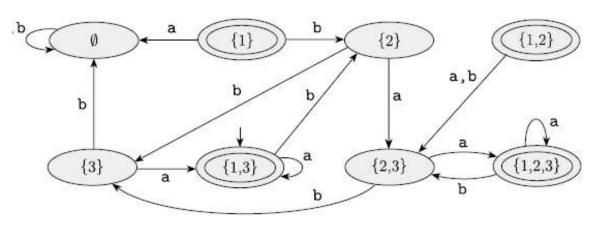


FIGURE 1.43 A DFA D that is equivalent to the NFA  $N_4$ 

- But, some states are not reachable!
- Simplification can remove this.

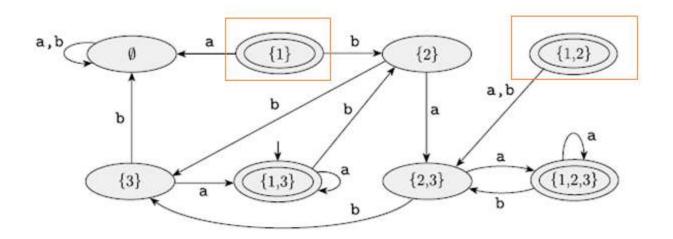
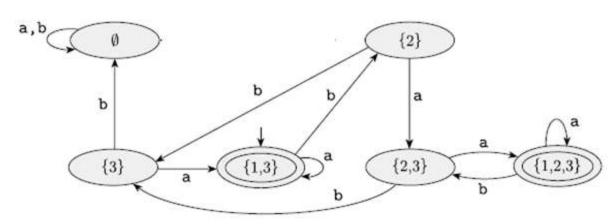
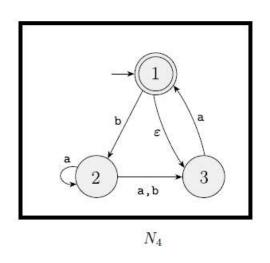


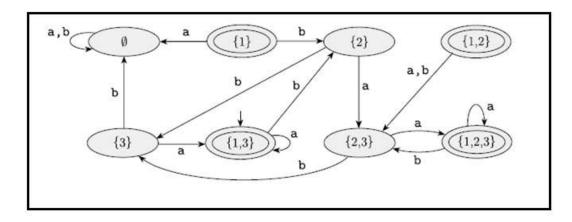
FIGURE 1.43 A DFA D that is equivalent to the NFA  $N_4$ 



DFA D' which is equivalent to D.

Note: D' and D are different machines; but, they are equivalent.



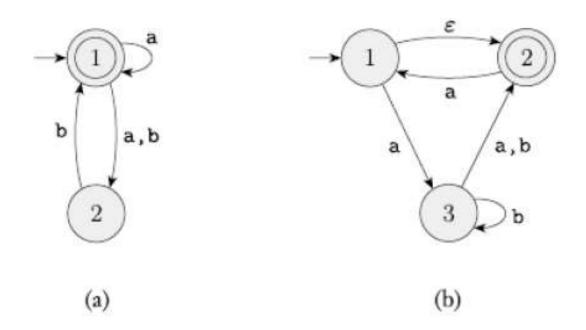


A DFA D that is equivalent to the NFA  $N_4$ 

- Being in state 1 of  $N_4$  upon reading input a the machine  $N_4$  can be in state 1.
- Convince yourself that in the DFA D there are no mistakes.
  - From state  $\{1\}$  with input a the DFA D goes to state  $\phi$ .

### Exercise

Convert the following NFAs to equivalent DFAs.

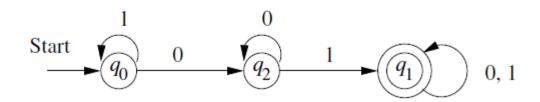


(Problem Source: Sipser's book exercise problem 1.16)

### Exercise

- 1.7 Give state diagrams of NFAs with the specified number of states recognizing each of the following languages. In all parts, the alphabet is {0,1}.
  - <sup>A</sup>a. The language  $\{w | w \text{ ends with 00}\}$  with three states
    - d. The language {0} with two states
  - g. The language  $\{\varepsilon\}$  with one state
  - h. The language 0\* with one state
- Can you convert each of above NFAs into a corresponding DFA.

# Some Notation adapted



The transition diagram for the DFA accepting all strings with a substring 01

	0	1
$ \begin{array}{c}                                     $	$q_2$	$q_0$
$*q_1$	$egin{array}{c} q_2 \\ q_1 \\ q_2 \end{array}$	$q_1$
$q_2$	$q_2$	$q_1$

This also has all 5 components. This table is complete description of the DFA as the diagram.

### Equivalency of DFA, NFA, ε-NFA

### What have we shown

- For every DFA, there is an NFA that accepts the same language and visa versa
- For every DFA, there is a E-NFA that accepts the same language, and visa versa
- Thus, for every NFA there is a E-NFA that accepts the same language, and visa versa
- DFAs, NFAs, and E-NFA s are equivalent!