Non-regular languages

Non-regular languages

$$\{a^n b^n : \quad n \geq 0\}$$

$$\{vv^R : \quad v \in \{a,b\}^*\}$$

Regular languages

$$a^*b$$

$$b^*c+a$$

$$b+c(a+b)^*$$

$$etc...$$

How can we prove that a language L is not regular?

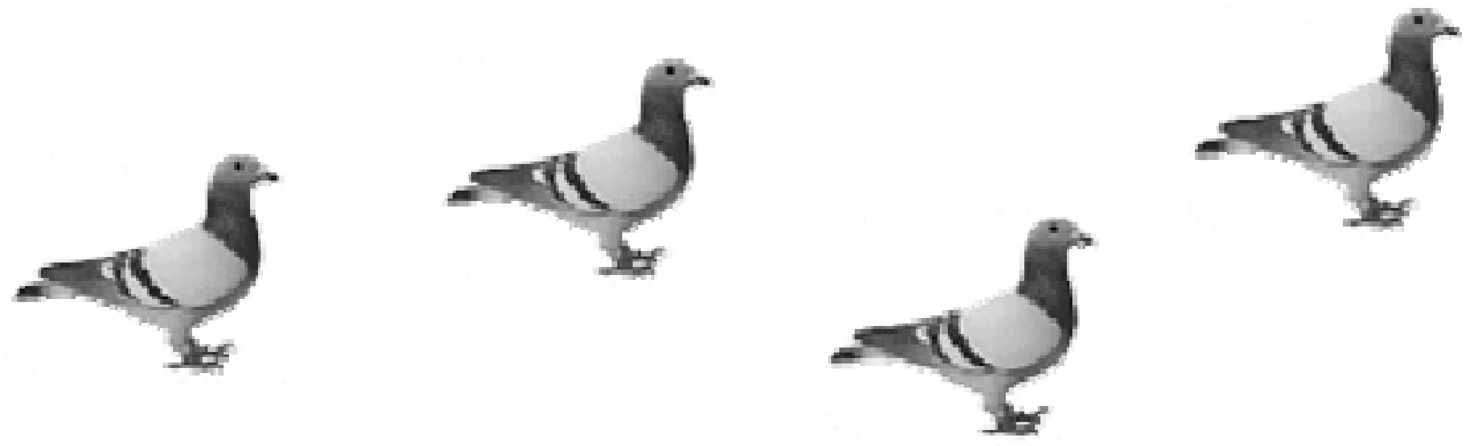
Prove that there is no DFA that accepts L

Problem: this is not easy to prove

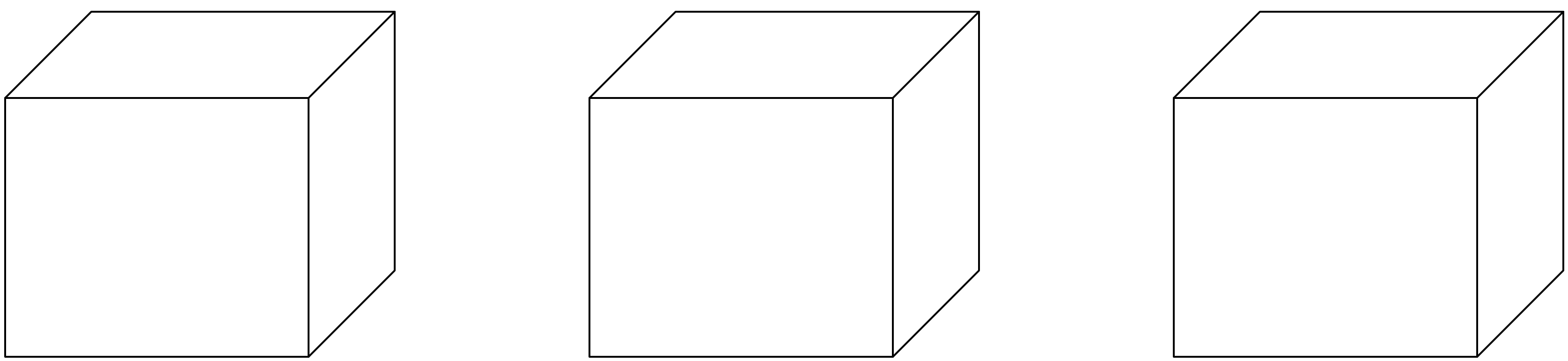
Solution: the Pumping Lemma !!!

The Pigeonhole Principle

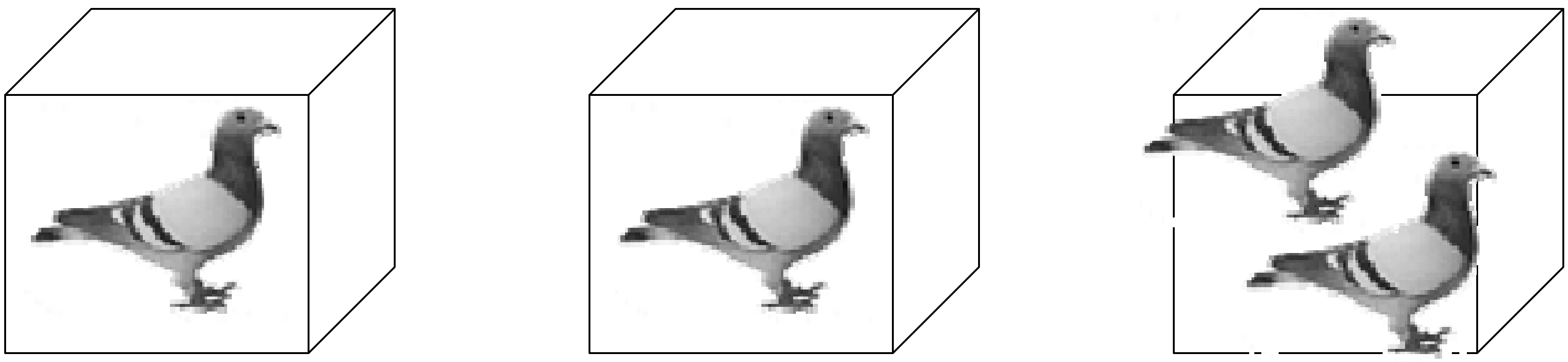
4 pigeons



3 pigeonholes

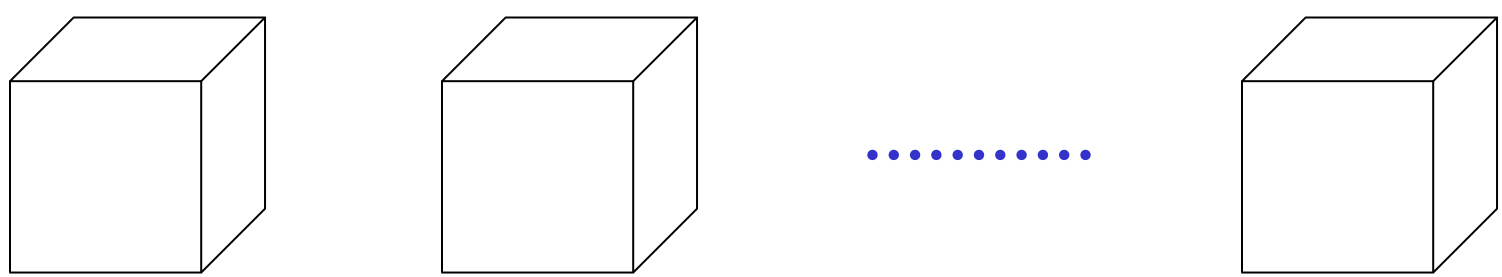


A pigeonhole must contain at least two pigeons



n pigeons

m pigeonholes $n > m$



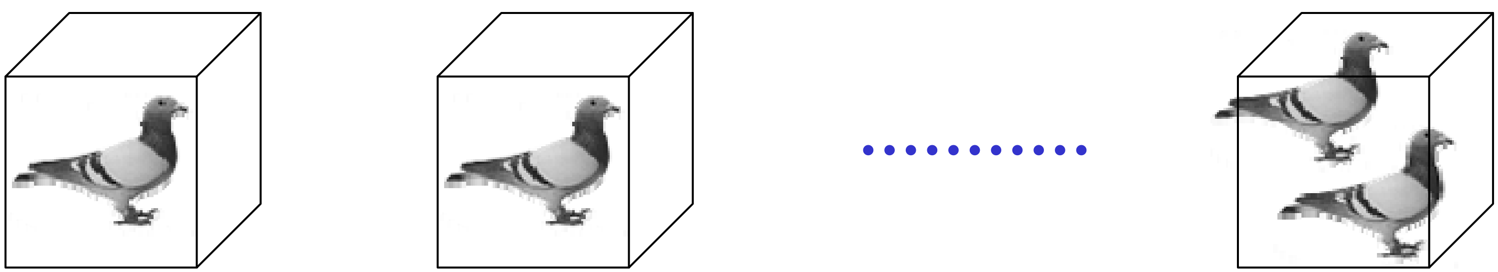
The Pigeonhole Principle

n pigeons

m pigeonholes

$$n > m$$

There is a pigeonhole with at least 2 pigeons

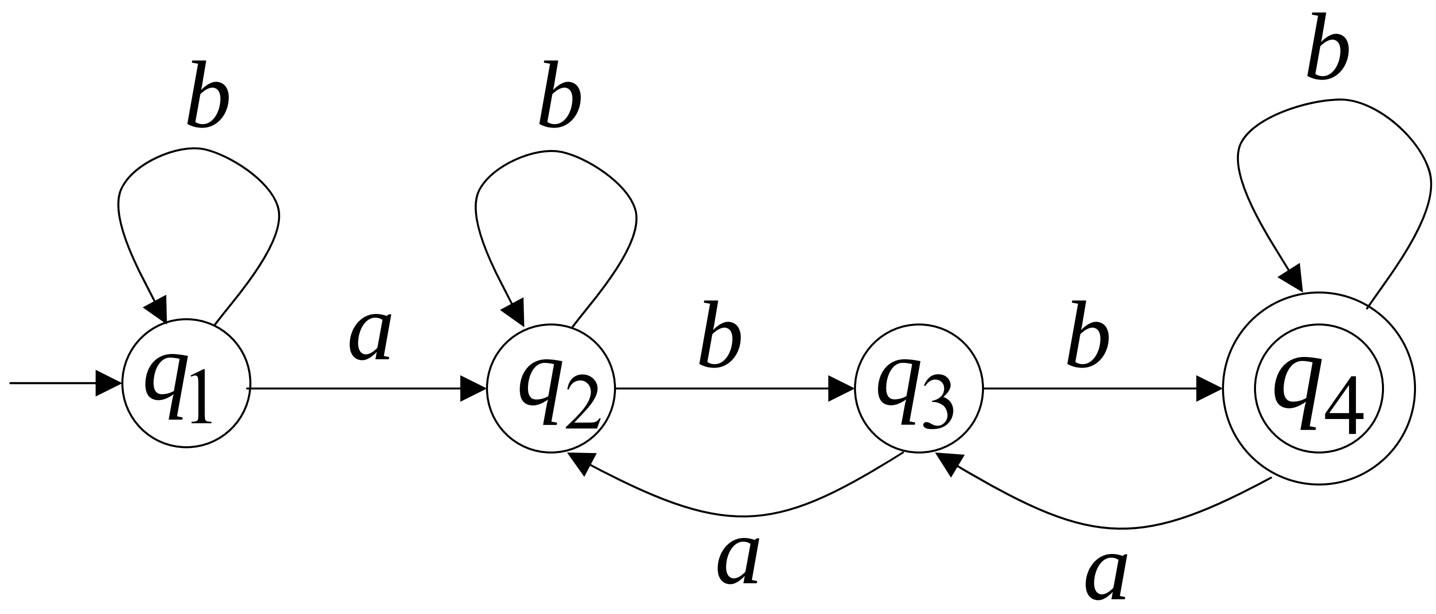


The Pigeonhole Principle

and

DFAs

DFA with 4 states

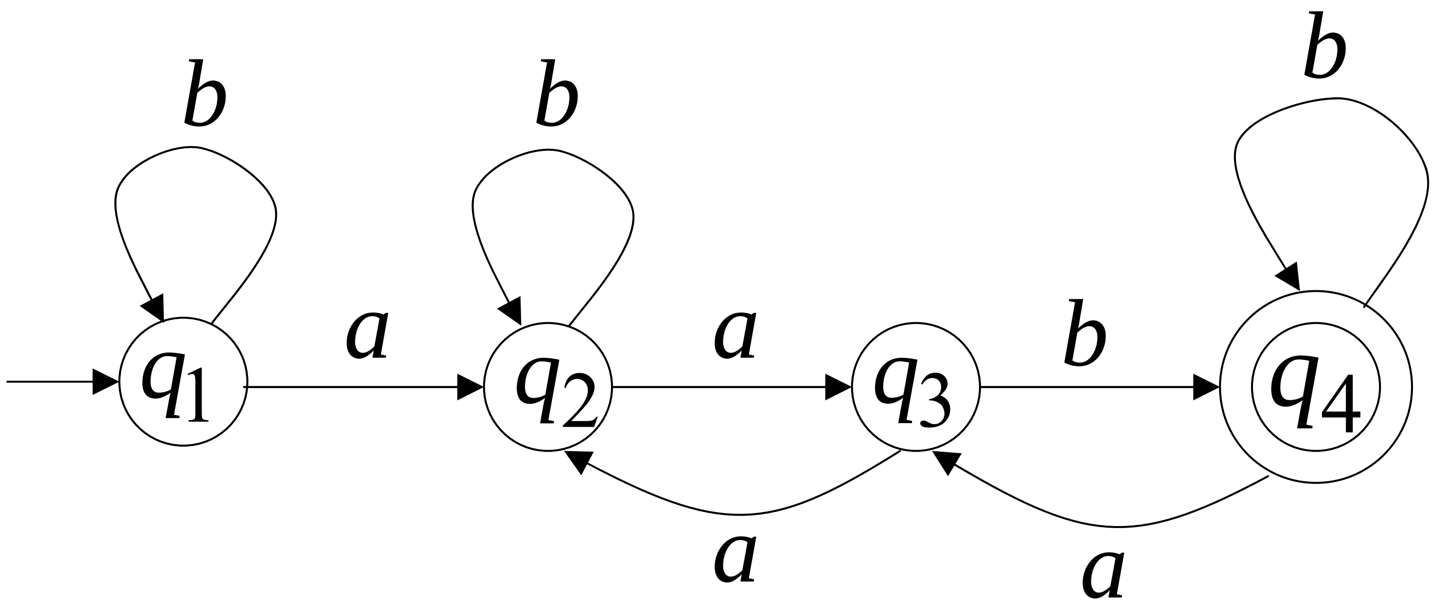


In walks of strings: a

aa

aab

no state is repeated



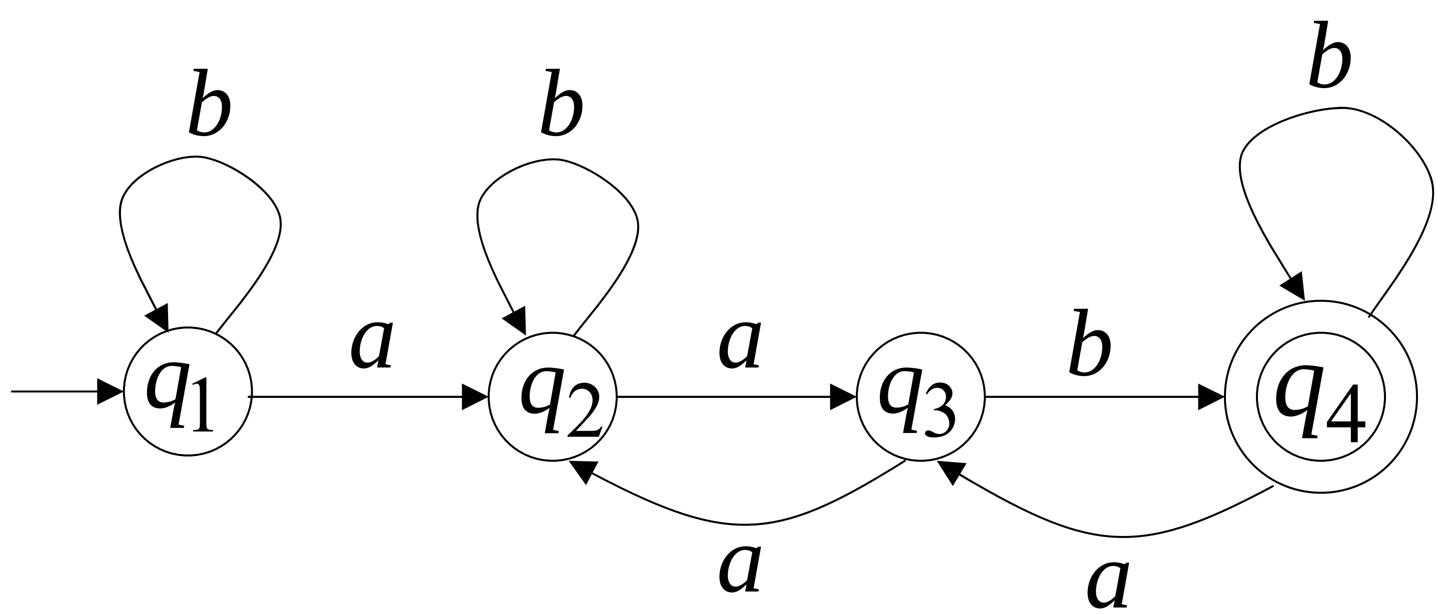
In walks of strings: $aabb$

$bbaa$

$abbabb$

$abbbabbabb\ldots$

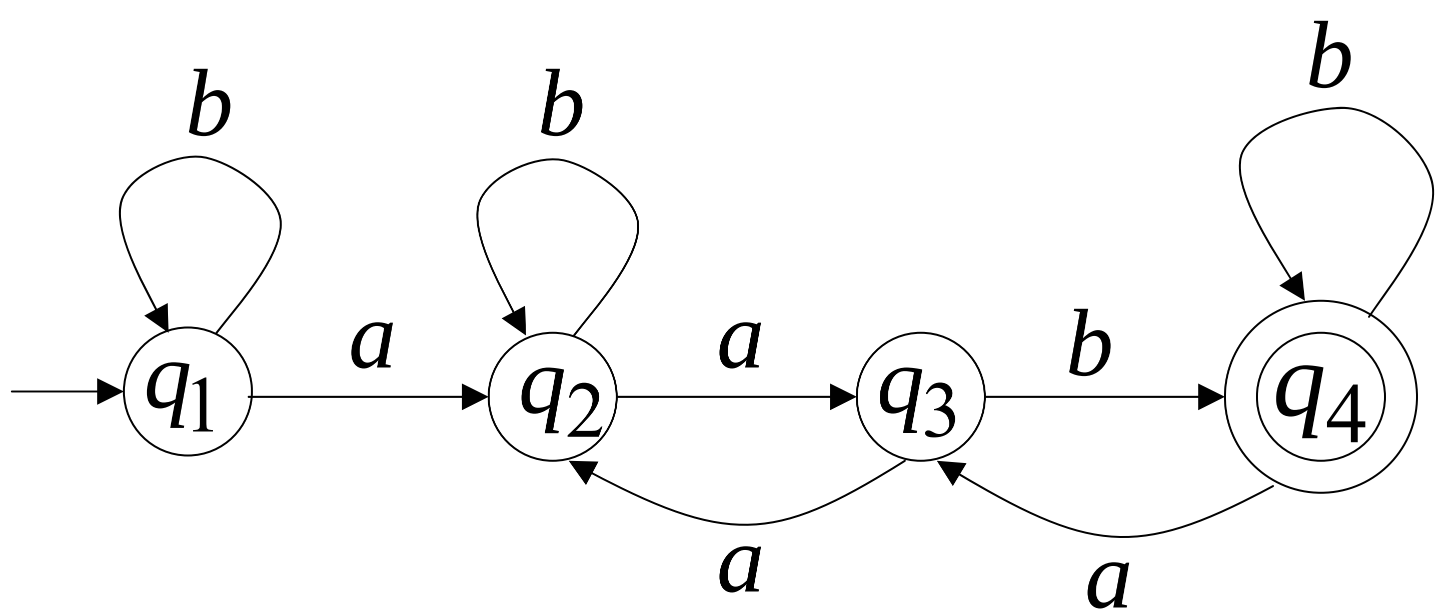
a state
is repeated



If string w has length $|w| \geq 4$:

Then the transitions of string w are more than the states of the DFA

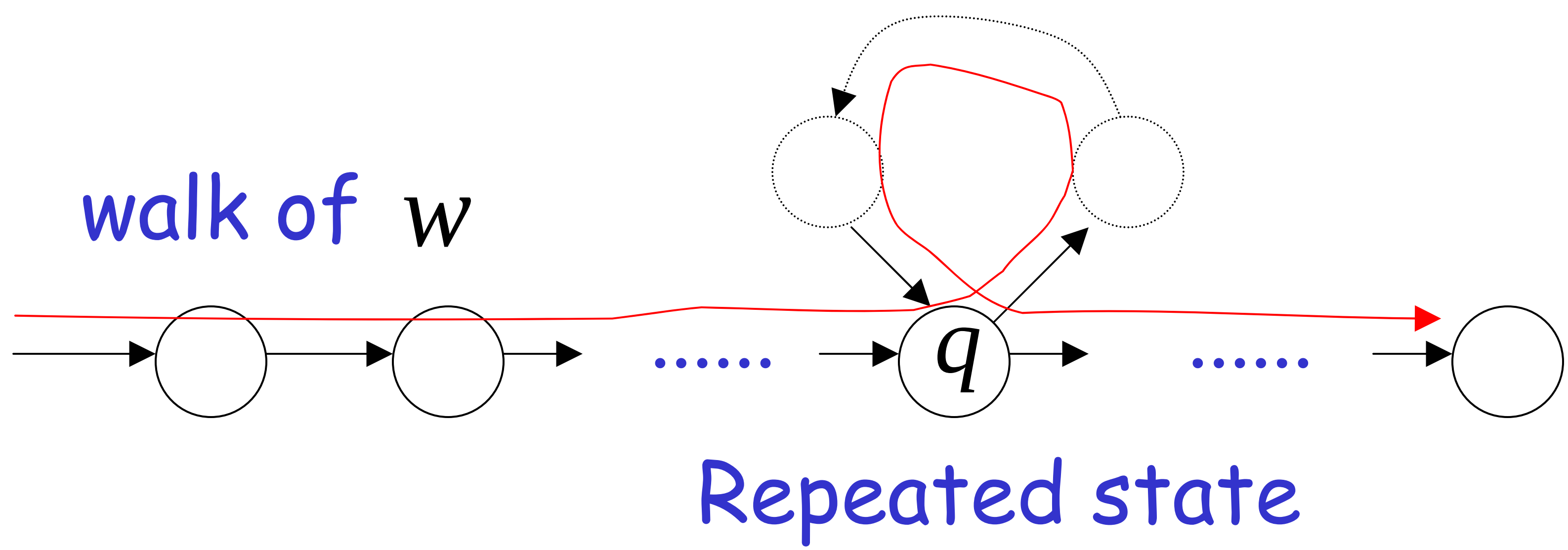
Thus, a state must be repeated



In general, for any DFA:

String w has length $\geq$ number of states

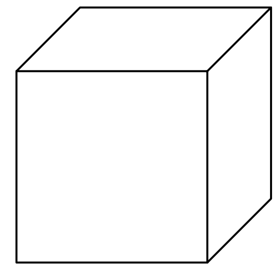
A state q must be repeated in the walk of w



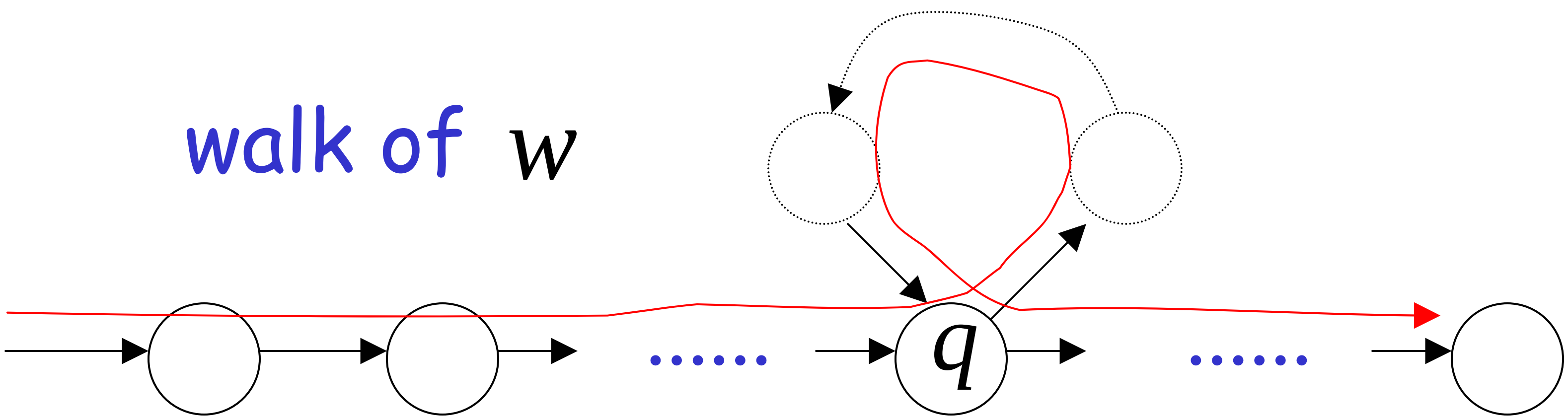
In other words for a string w:

$\xrightarrow{a}$ transitions are pigeons

q states are pigeonholes



walk of w

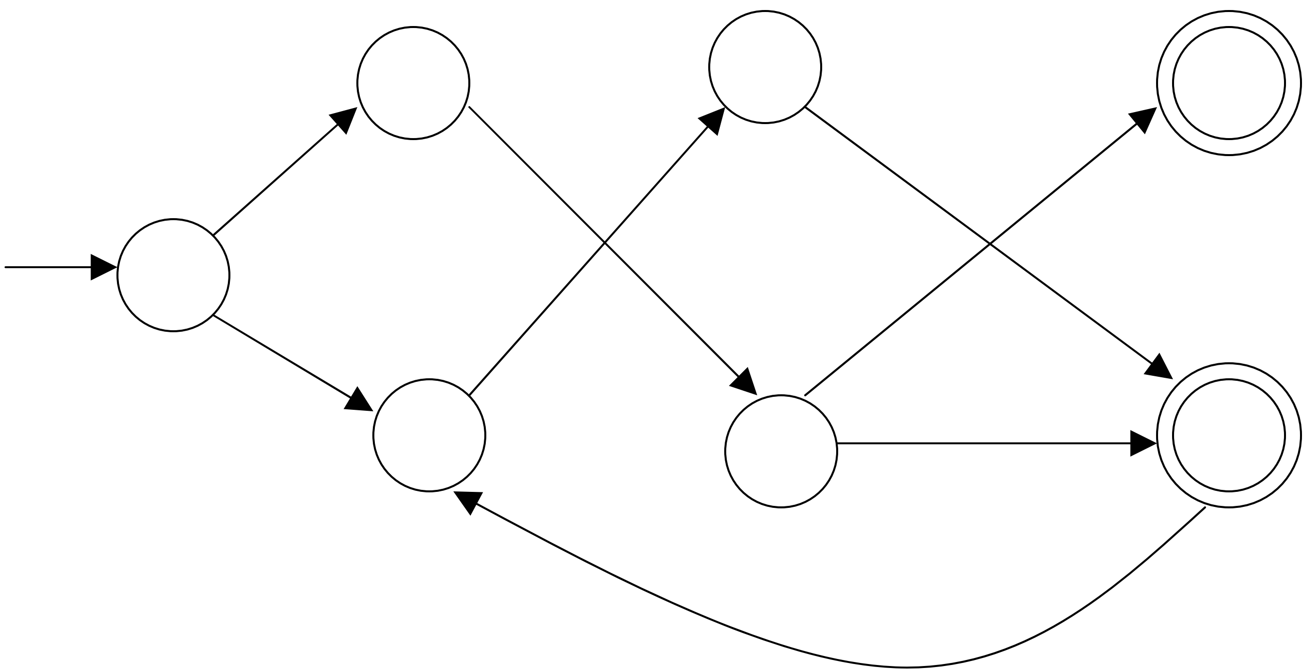


Repeated state

The Pumping Lemma

Take an infinite regular language L

There exists a DFA that accepts L



m
states

Take string w with $w \in L$

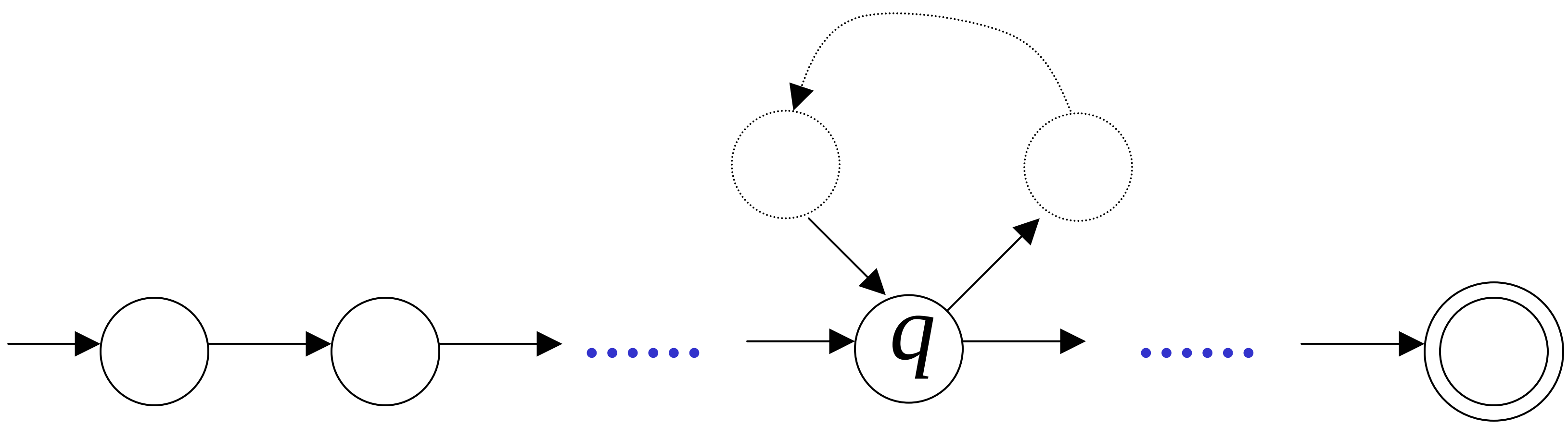
There is a walk with label w :

walk w

If string w has length $|w| \geq m$ (number of states of DFA)

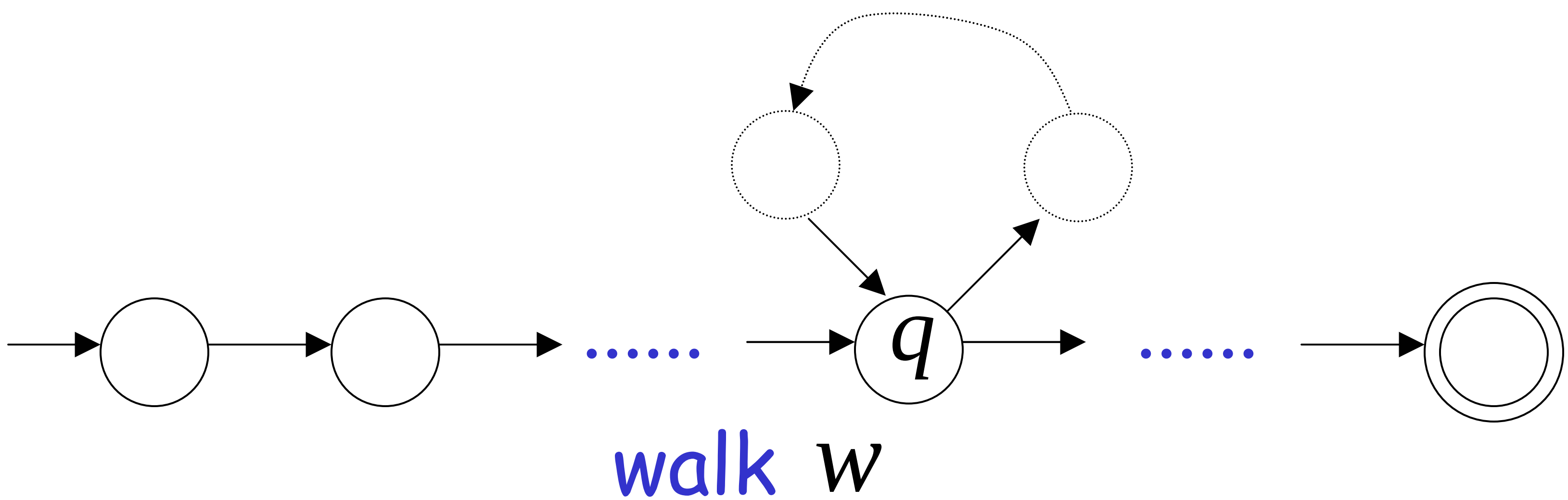
then, from the pigeonhole principle:

a state is repeated in the walk w

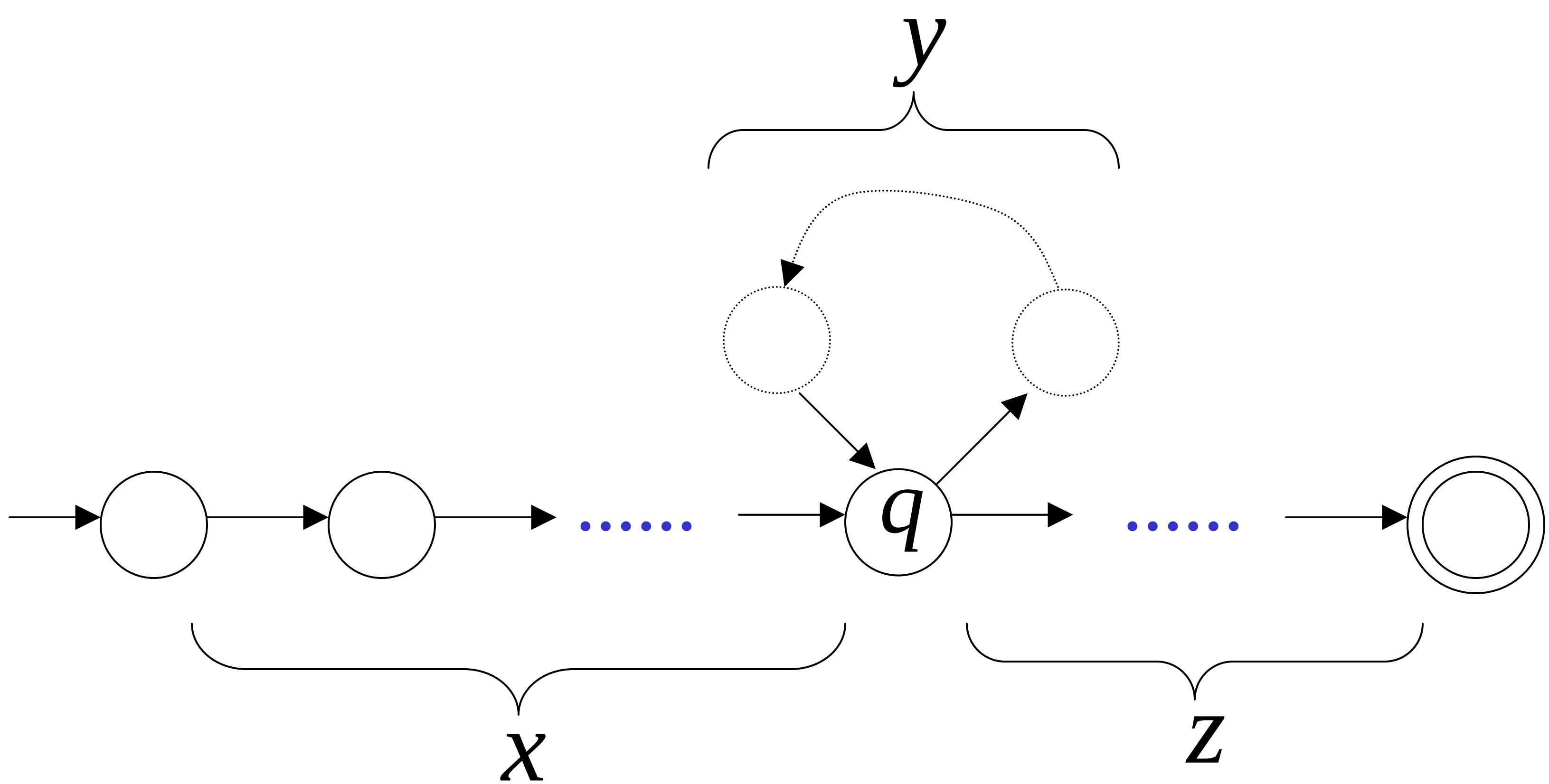


walk w

Let q be the first state repeated in the walk of w

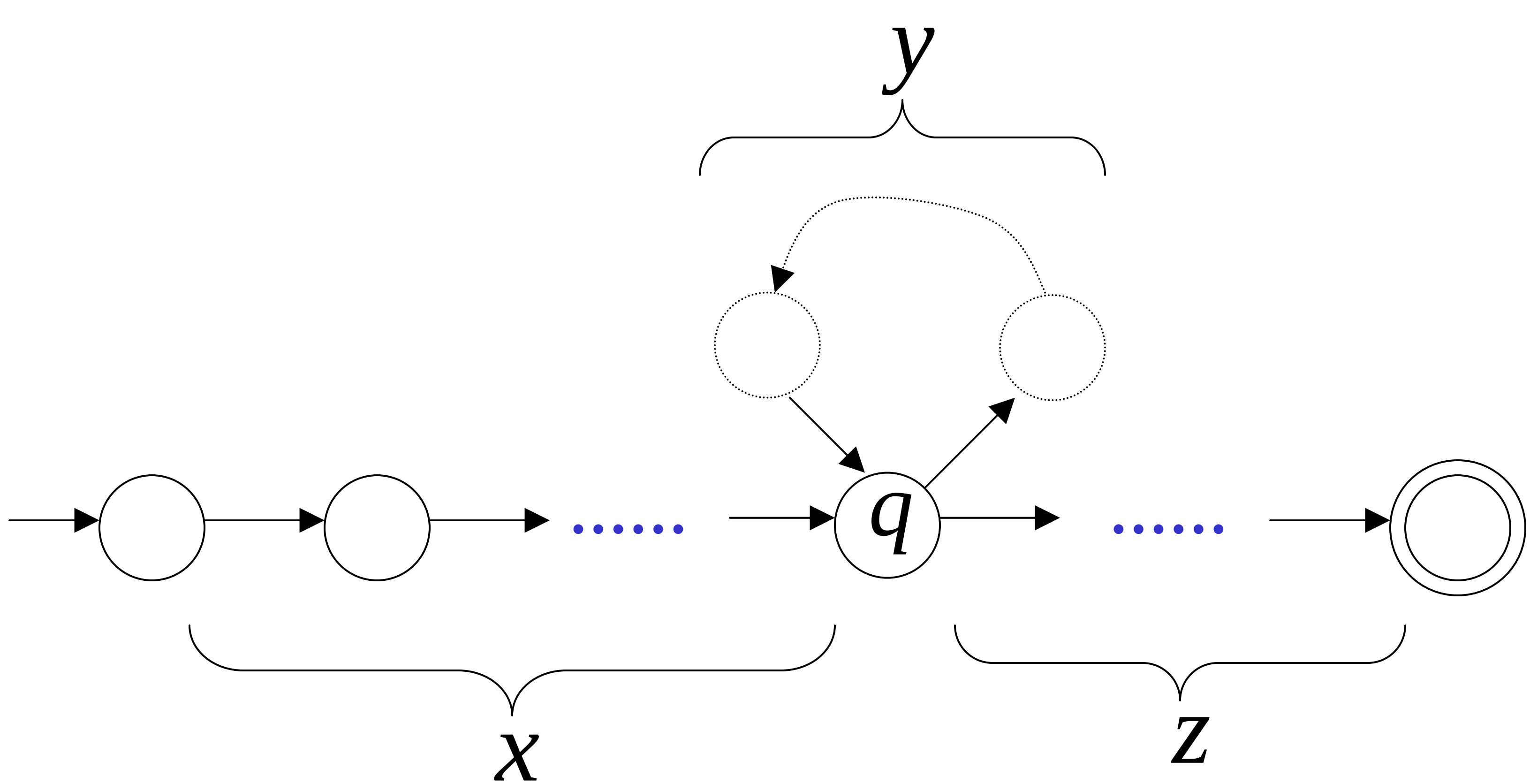


Write $w = x\ y\ z$

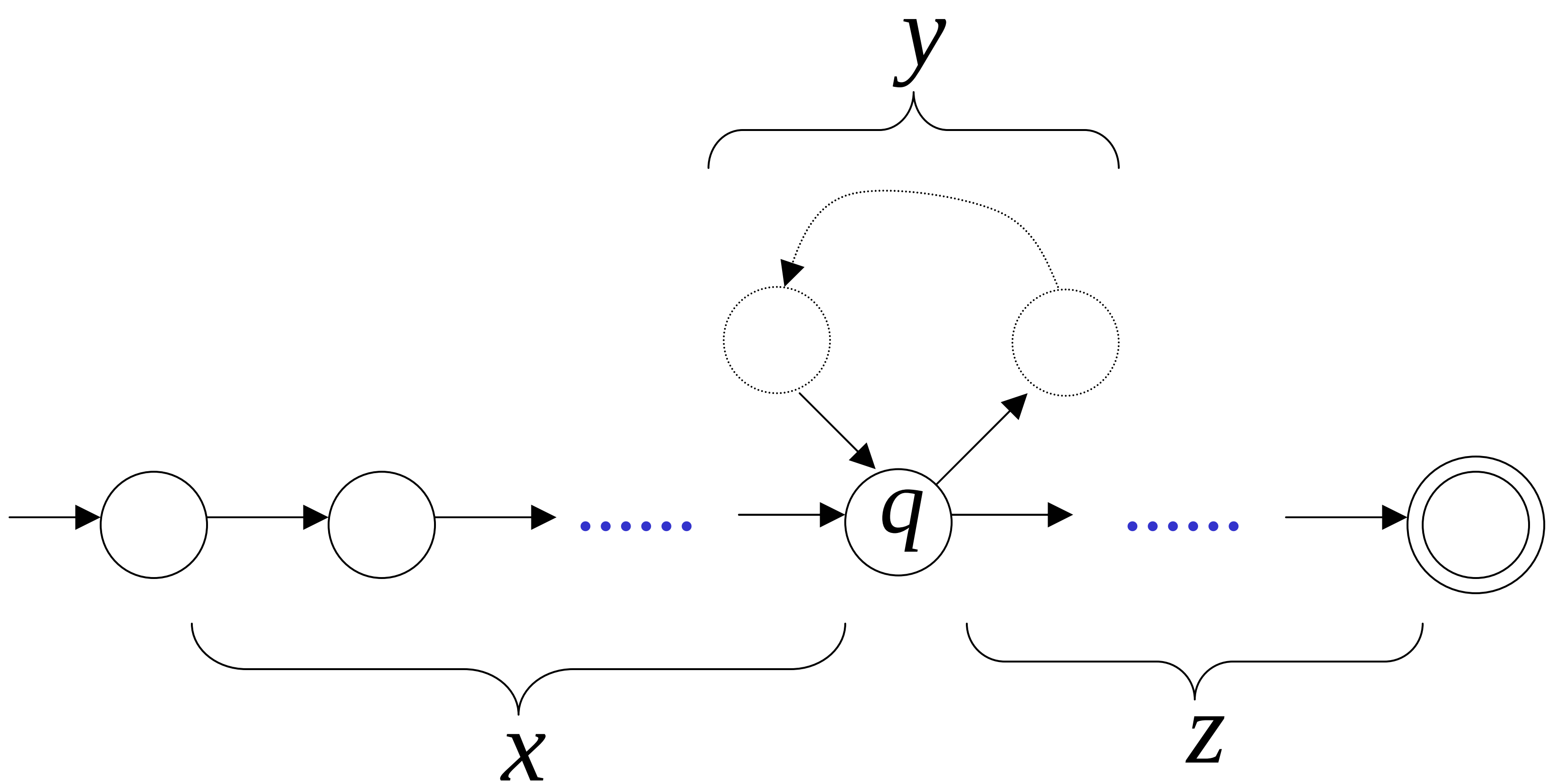


Observations: length $|x\,y| \leq m$ number of states of DFA

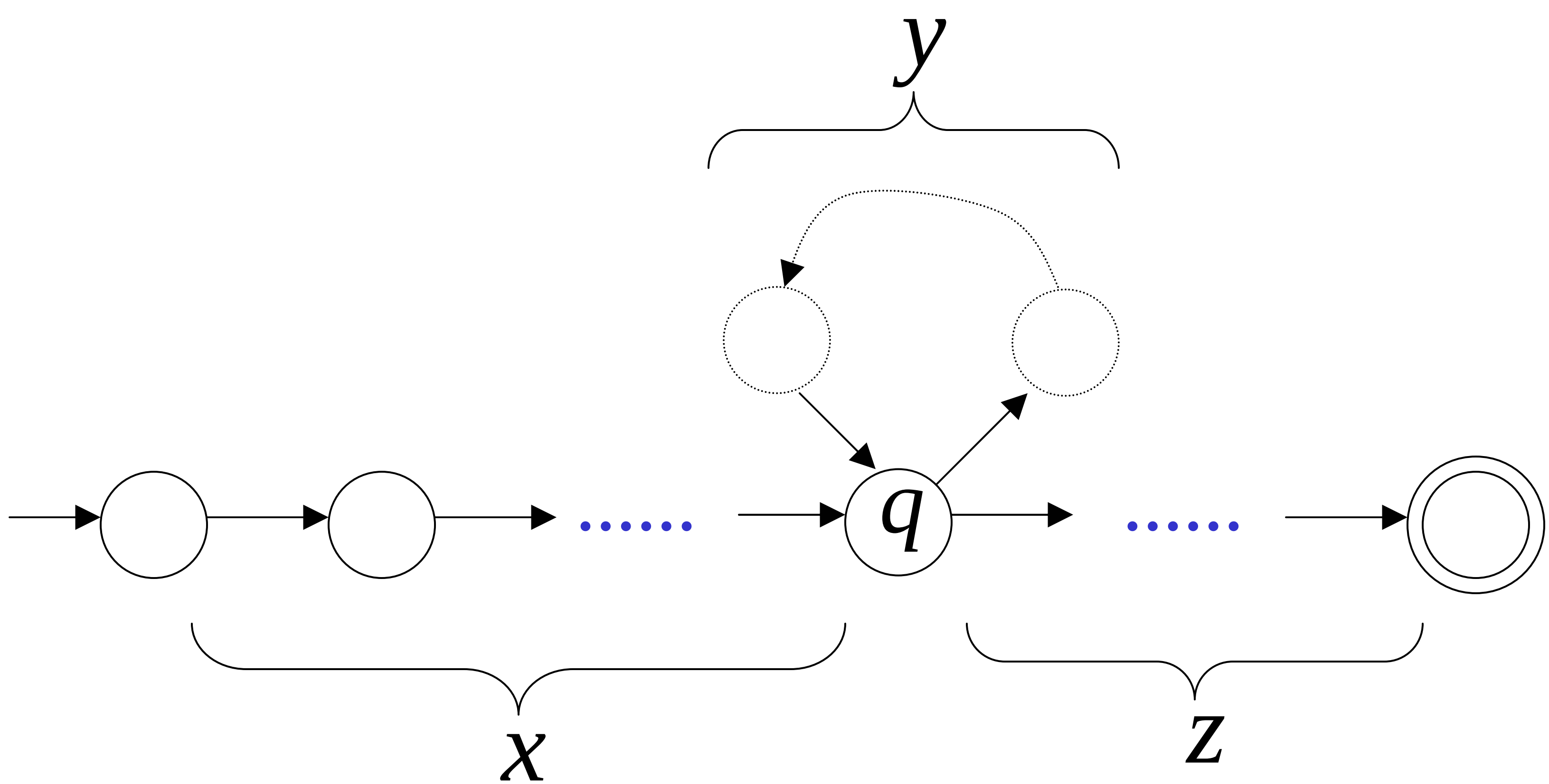
length $|y| \geq 1$



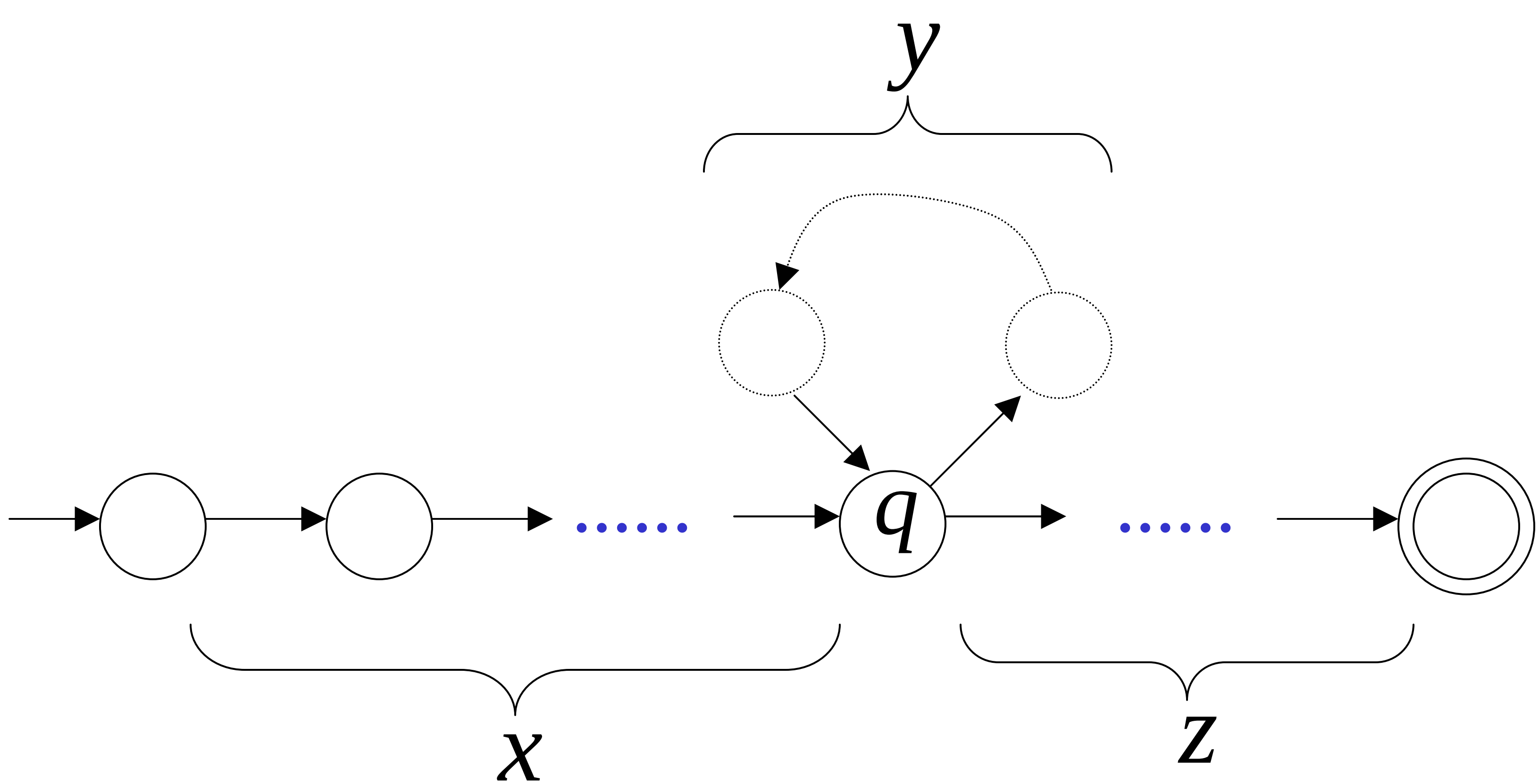
Observation: The string $x\,z$ is accepted



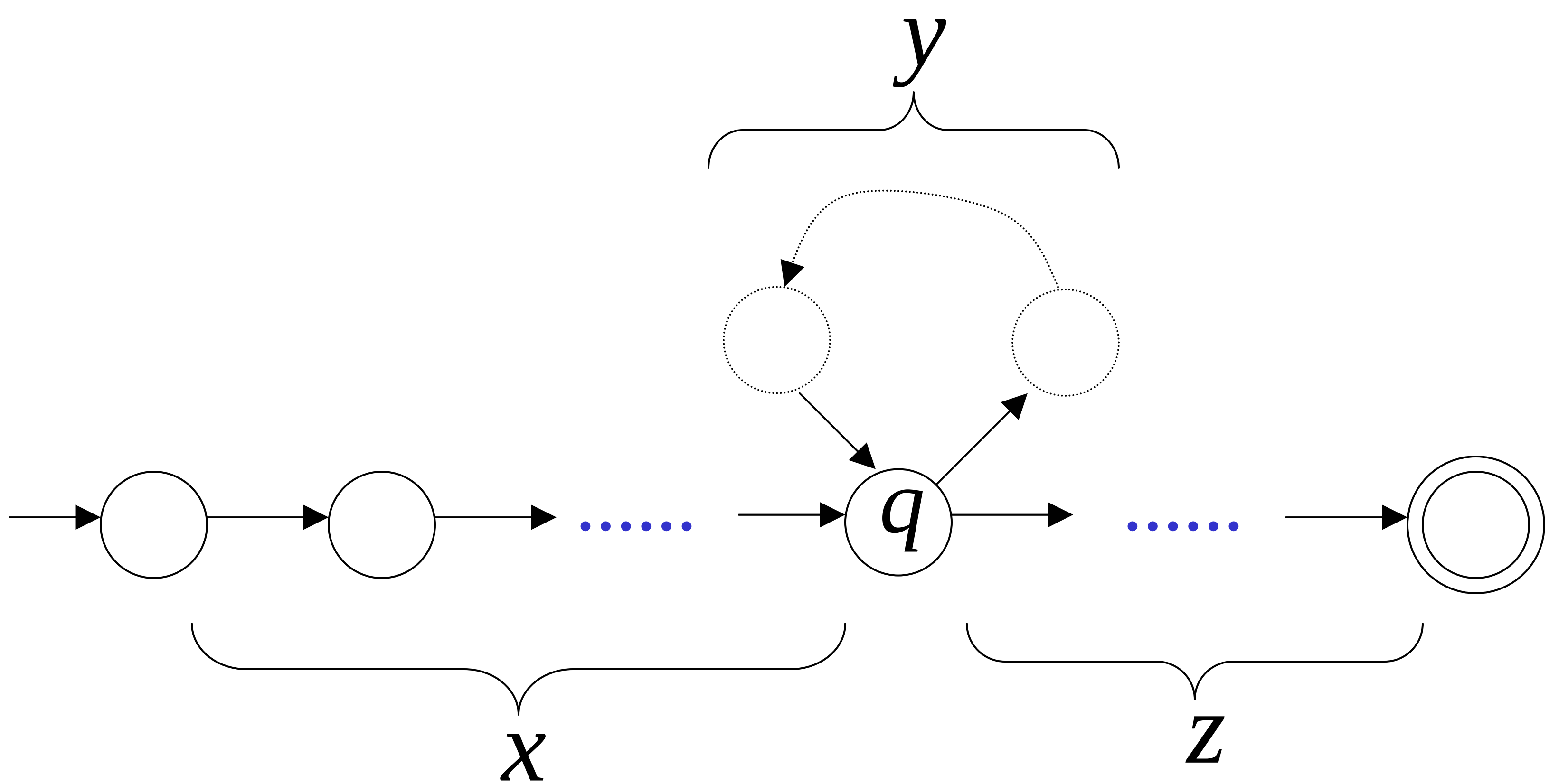
Observation: The string $x\ y\ y\ z$ is accepted



Observation: The string $x\,y\,y\,y\,z$ is accepted

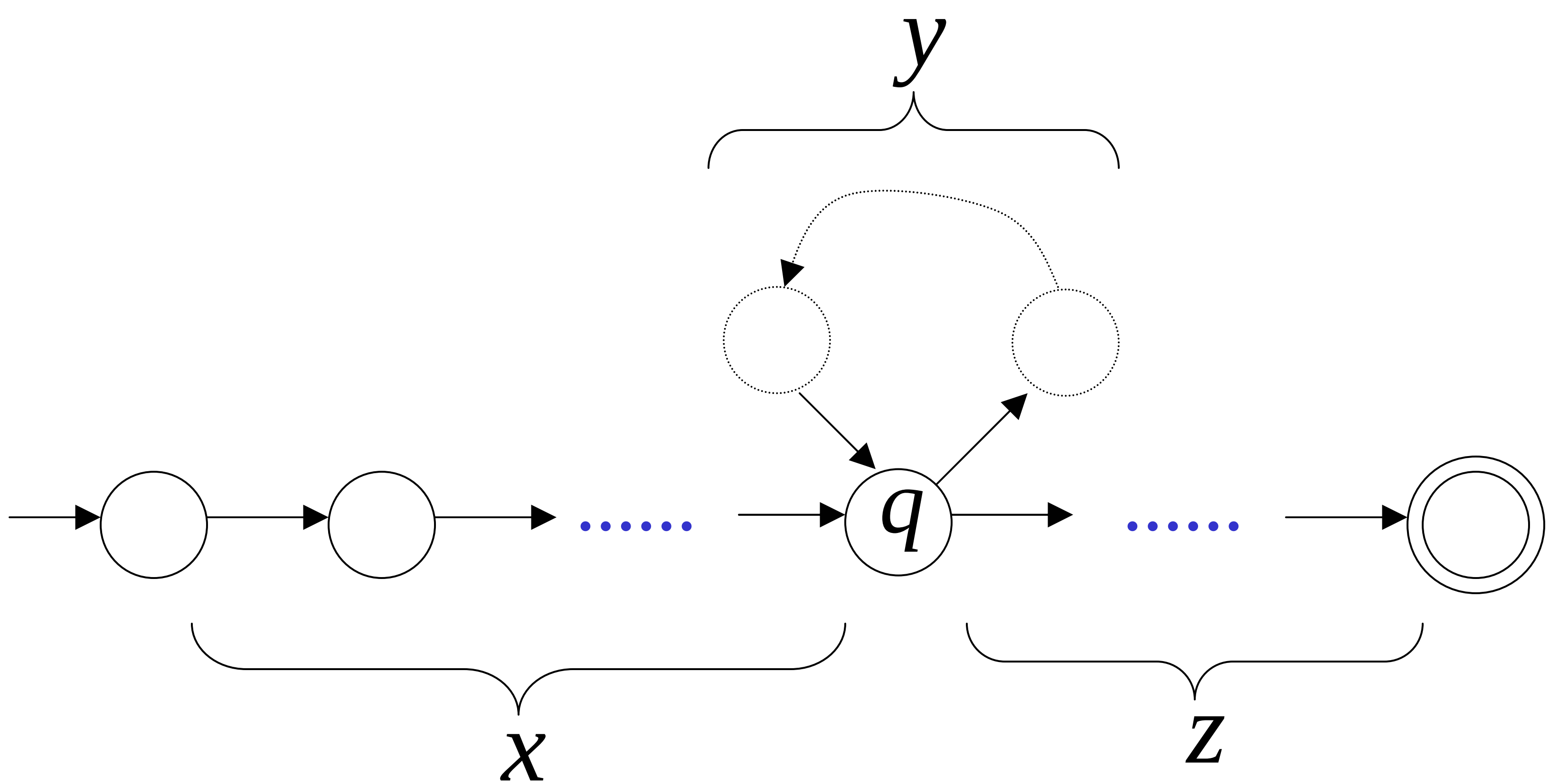


In General: The string $x\,y^i\,z$ is accepted $i = 0, 1, 2, \ldots$



In General: $x\,y^i\,z \in L$ $\qquad i = 0, 1, 2, \ldots$

Language accepted by the DFA



In other words, we described:

The Pumping Lemma !!!

The Pumping Lemma:

- Given a infinite regular language L
- there exists an integer m
- for any string $w \in L$ with length $|w| \geq m$
- we can write $w = x\,y\,z$
- with $|x\,y| \leq m$ and $|y| \geq 1$
- such that: $x\,y^i\,z \in L \qquad i = 0, 1, 2, \ldots$

Nonregular languages

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

The pumping lemma for regular languages

- A special property that all regular languages have

There exists certain languages that can not be recognized by any finite automaton

- $B = \{0^n 1^n \mid n \geq 0\}$
- $C = \{w \mid w$ has an equal number of 0s and 1s$\}$.

Nonregular languages do not have the special property

The Pumping Lemma

If A is a regular language, then there is a number p (the pumping length) where, if s is any string in A of length at least p, then s may be divided into three pieces, s = xyz, satisfying conditions:

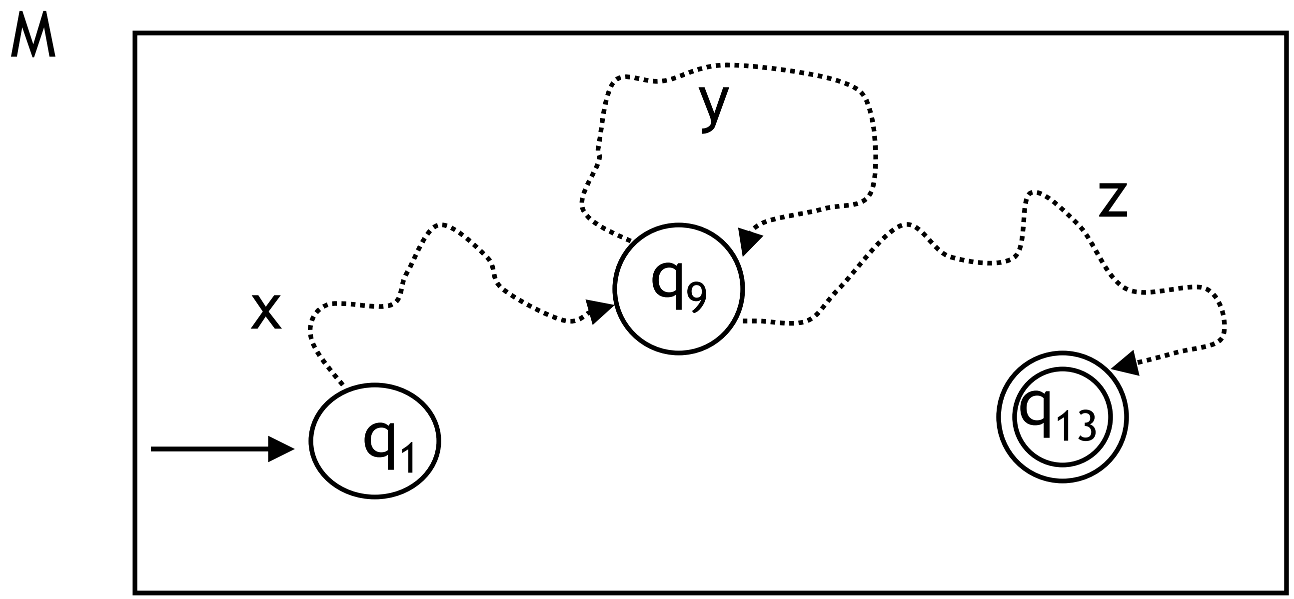
1. for each $i \geq =0$, $xy^iz \in A$,
2. $|y|>0$, and
3. $|xy| \leq p$.

Examples:

0001*, {0}

Proof of Pumping Lemma

Now we divide s into 3 pieces xyz



- Condition 1: $xy^iz \in A$
- Condition 2: $|y| > 0$
- Condition 3: $|xy| \leq k$, (q_9 is the first repetition in the sequence)

Proof of Pumping Lemma

Proof: let $M = \{Q, \Sigma, \delta, q_1, F\}$ be a DFA recognizing A and k be the number of states of M.

let $s = s_1 s_2 \ldots s_n$ be a string in A of length n, where $n \geq k$.

```````````````let $s = r_1 r_2 \ldots r_{n+1}$ be the sequence of states that M enters while processing s, so $r_{i+1} = \delta(r_i, s_i)$, for $1 \leq i \leq n$.

The sequence has length n+1, which is at least k+1 states. Along the first k+1 elements, two must be the same state.

We call the first $r_j$, and the second $r_l$. The $j \leq k$, $l \leq k+1$.

Now let $x = s_1 \ldots s_{j-1}$, $y = s_j \ldots s_{l-1}$, $z = s_l \ldots s_n$

As x takes M from $r_1$ to $r_j$, y takes M from $r_j$ to $r_l$, and z takes M from $r_l$ to $r_{n+1}$ which is an accept state, M must accept $xy^iz$ for $i \geq 0$.

We know that $j \neq l$. so $|y|>0$, and $l \leq p+1$, so $|xy| \leq k$. Thus we have satisfied all three conditions of the pumping lemma.
```````````````

The use of Pumping Lemma

- Use the pumping lemma to prove a language A is not regular
- First assume that language A is regular in order to obtain a contradiction
  - The use of pumping lemma guarantees the existence of a pumping length such that all strings in A of length at least k can be pumped
- Next find a string s in A of length at least k , but can’t be pumped
- Finally, demonstrate that s can not be pumped by considering all ways of dividing s into x, y and z

Example 1

Let A = $\{0^n1^n \mid n \geq 0\}$, prove that A is not regular

Proof:

Assume that A is regular, let K be the pumping length given by the pumping lemma. Chose s = 0^K1^k. Because s is a member of B and has length more than k, the pumping lemma guarantees that s can be broken into xyz, where for any $i \geq =0$, $xy^iz \in A$. We consider three cases to show that this result is impossible

(1).The string y only consists of only 0s. In this case xyyz will have more 0s than 1s. So xyyz is not in A

(2).The string y only consists of 1 s, then xyyz will have more 1s than 0s

(3).The string y consists of both 0s and 1s. In this case, xyyz may have same number of 0s and 1s, but they will be out of order. So xyyz is not in A

Thus a contradiction is unavoidable if we make the assumption that A is regular, so A is not regular

Example 2

Let the language $B = \{0^i 1^j \mid I > j\}$, prove that B is not regular.

Proof:

Assume that B is regular, let p be the pumping length given by the pumping lemma. Chose

$$s = 0^{p+1} 1^p,$$

Because s is a member of B and has length more than p, the pumping lemma guarantees that s can be broken into xyz, where for any $i \geq =0$, $xy^iz \in B$. We show that this result is impossible.

Because $|xy| \leq p$, y must consists of only 0s. Consider the string $xy^0z = xz$. Removing y decreases the number of 0s. xyz has only 1 more 0s than 1s. So xz is not in B.

Thus a contradiction is unavoidable if we make the assumption that B is regular, so B is not regular.