PUMPING LEMMA FOR REGULAR SETS

Dr. A. Beulah

AP/CSE

INTRODUCTION

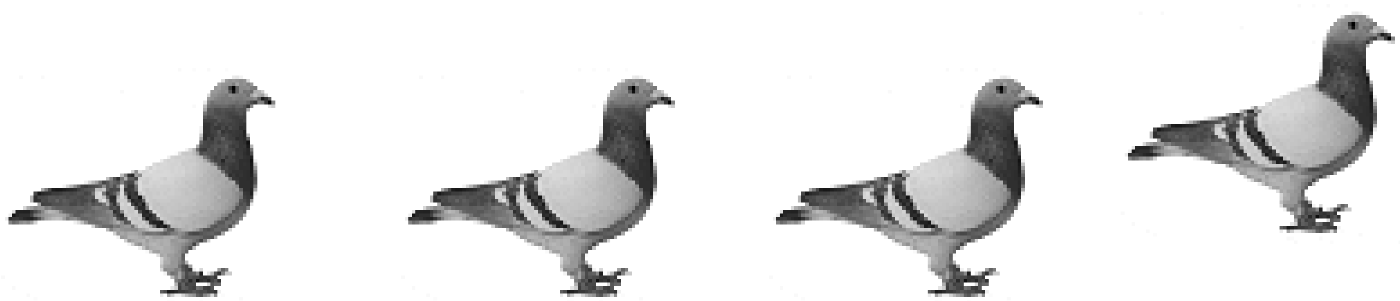
- A Regular language is a formal language that can be expressed using a regular expression
- A regular language satisfies the following equivalent properties:
  - it is the language accepted by a nondeterministic finite automaton
  - it is the language accepted by a deterministic finite automaton
  - it can be generated by a regular grammar
  - it can be generated by a prefix grammar
  - it can be accepted by a read-only Turing machine
- Regular set is a set of strings of a Regular Language
- For every regular language there is a FA that accepts the language

THE PIGEONHOLE PRINCIPLE

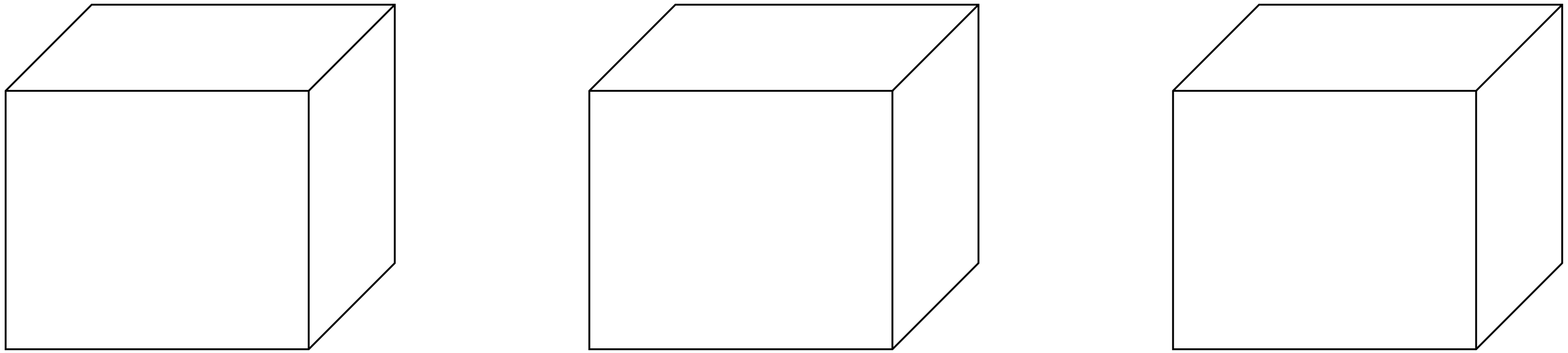
- If you put *n* pigeons into *m* pigeonholes, and $n > m > 0$, then at least at least two pigeons are in the same pigeonhole.

THE PIGEONHOLE PRINCIPLE

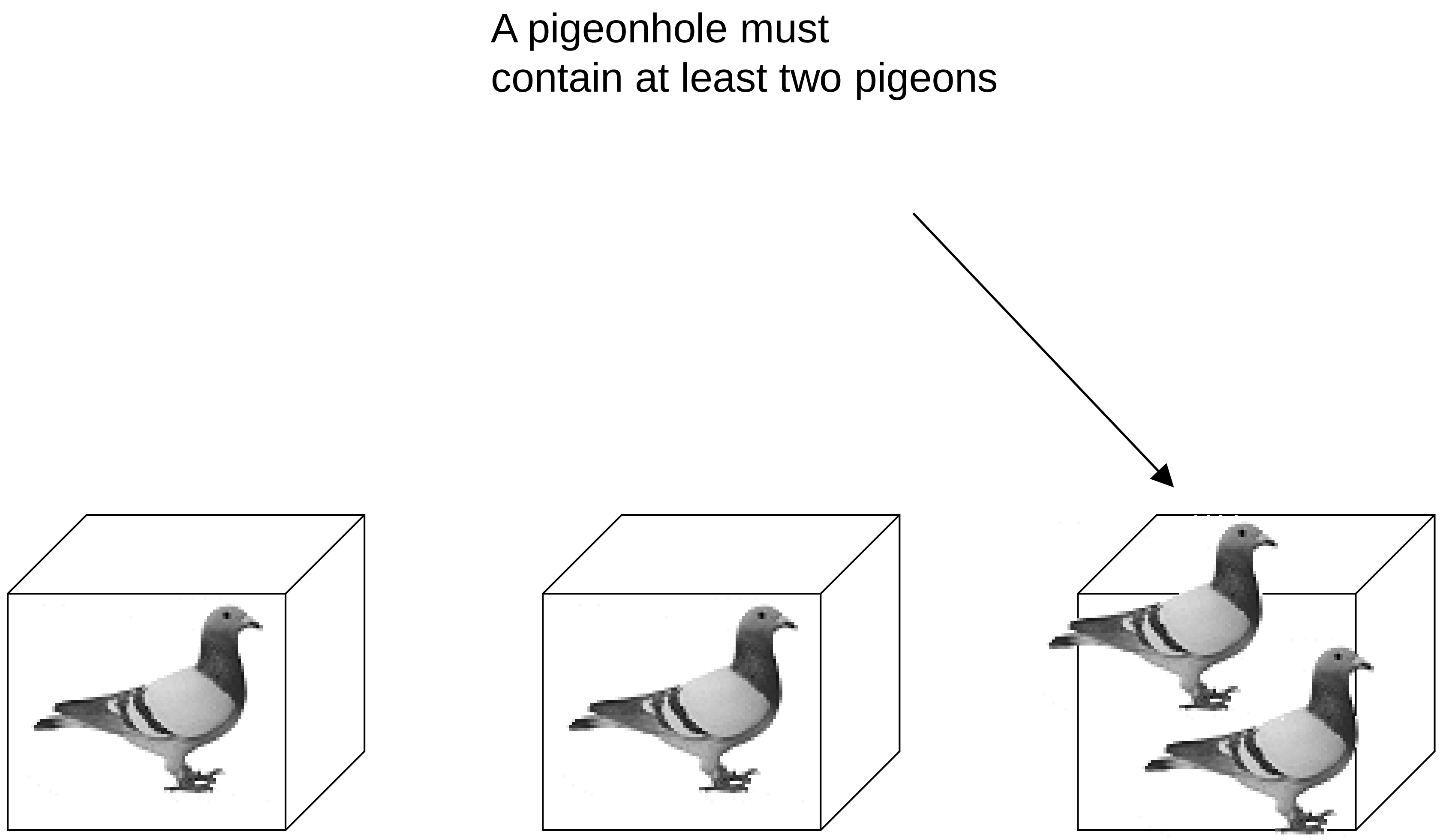
4 pigeons



3 pigeonholes

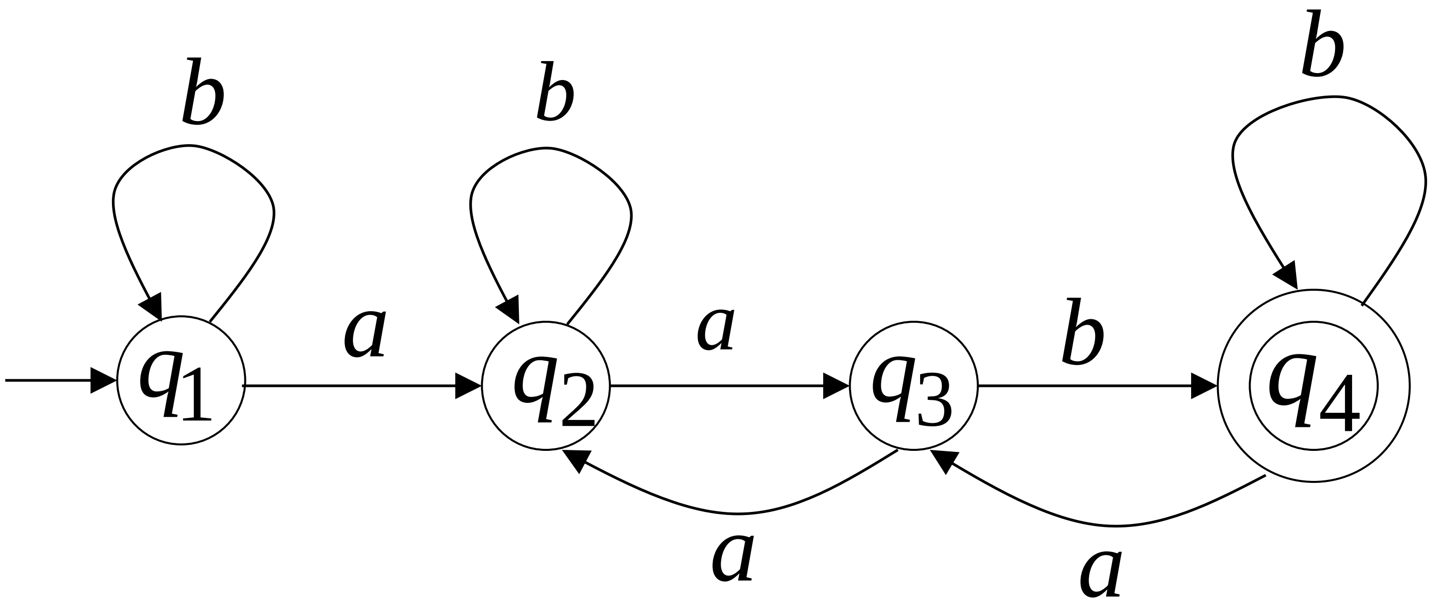


THE PIGEONHOLE PRINCIPLE



PIGEONHOLE PRINCIPLE & DFAS

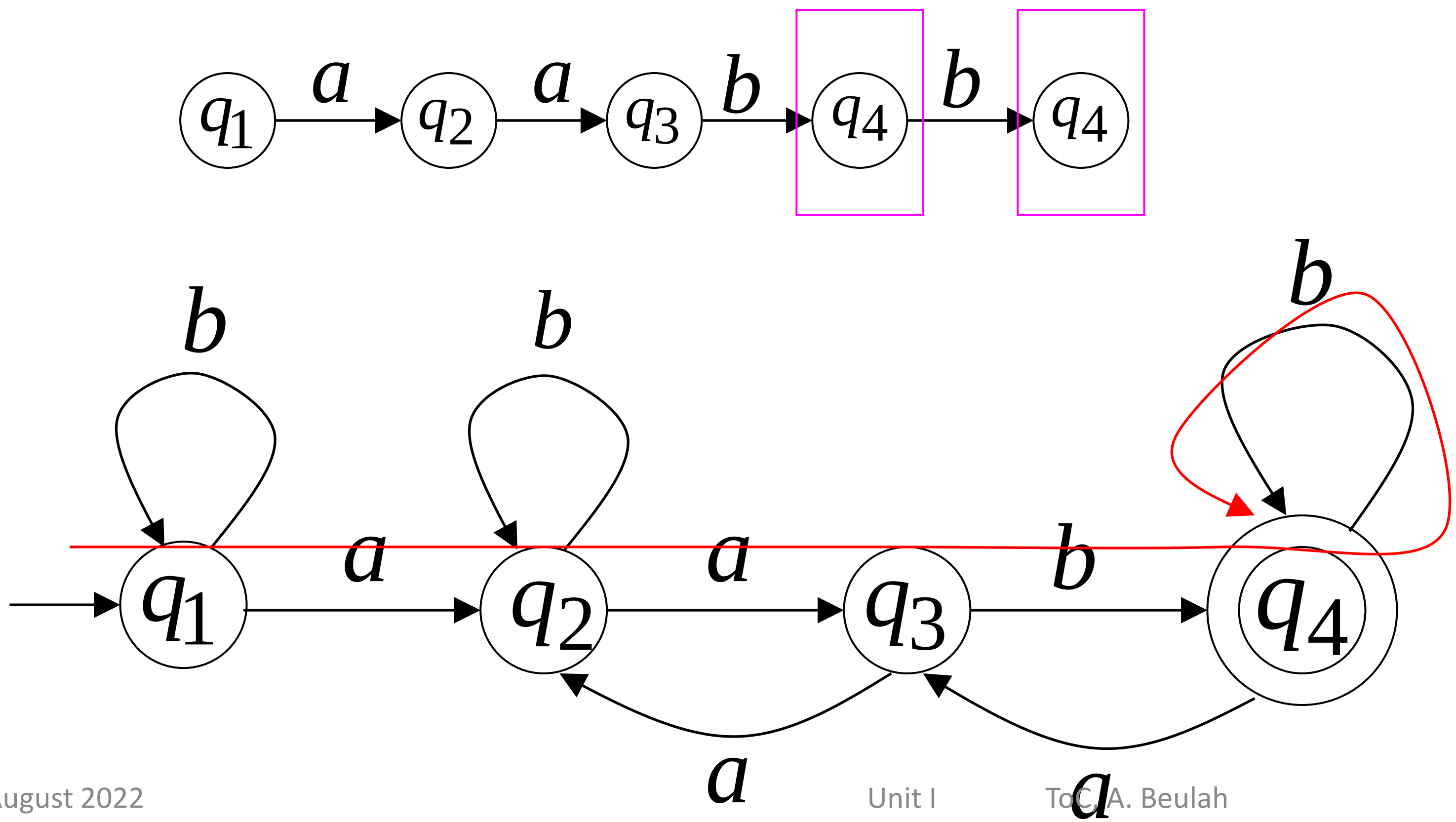
Consider a DFA with 4 states



PIGEONHOLE PRINCIPLE & DFAS

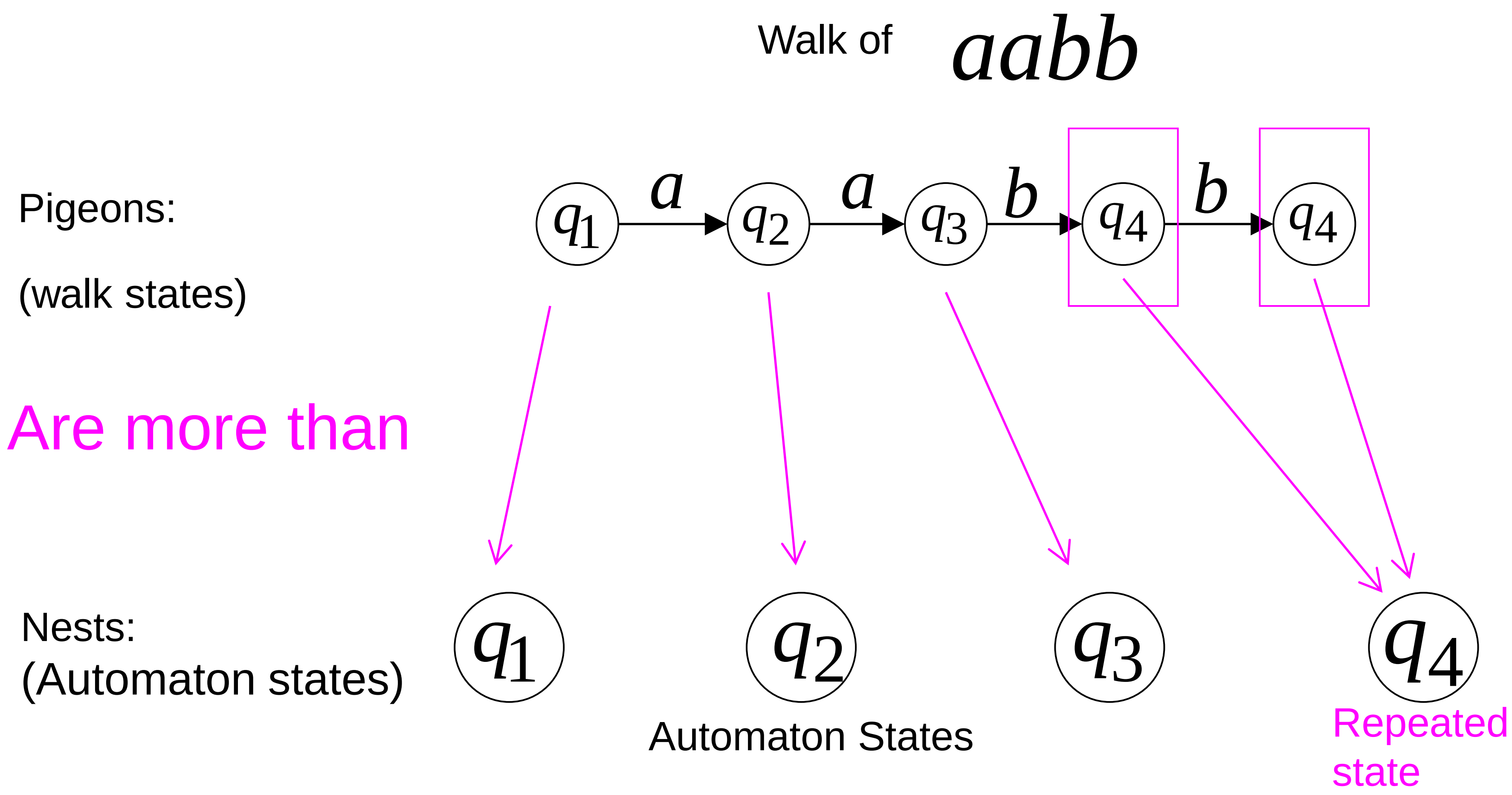
- Consider the walk of a "long'' string: aabb (>=4)

A state is repeated in the walk of $aabb$



PIGEONHOLE PRINCIPLE & DFAS

- The state is repeated as a result of the pigeonhole principle



PUMPING LEMMA

- Describes an essential property of all regular languages
- For a particular language, any sufficiently long string in the language contains a section, or sections, that can be removed, or repeated any number of times(pumping), with the resulting string remaining in that language
- The pumping lemma is used to prove that a particular language is non-regular

PUMPING LEMMA

- Let L be a regular language. Then there is a constant n (which depends on L/ number of states in FA) such that for every string w in L such that $|w| \geq n$, we can break w into three strings, $w = xyz$, such that $y \neq \varepsilon$ ie $|y| > 0$, $|xy| \leq n$, and for all $i \geq 0$, xy^iz is also in L.
- Proof
  - Let n be $|Q|$.
  - If $w \in L$ and $|w| \geq n$. Let $w = a_1 a_2 \ldots a_m$, where $m \geq n$.
  - $\delta(q_0, a_1 a_2 \ldots a_i) = q_i$, $i = 1, 2, \ldots, m$.

$L = \{ w \mid \ldots \}$

$|w| \geq n$

$w = xyz$

(i) $|xy| \leq n$

(ii) $|y| > 0$

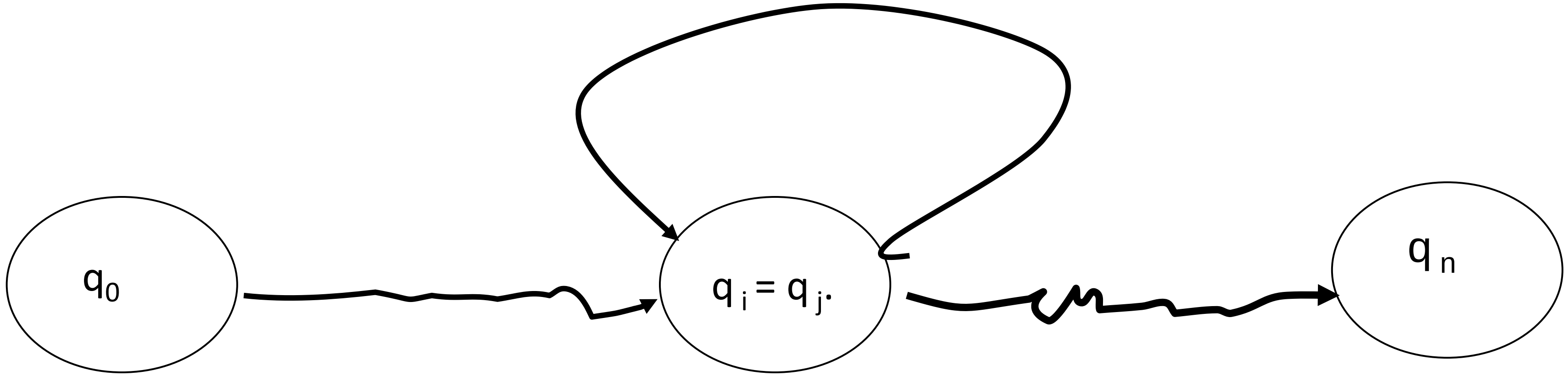
i.e. $y \neq \varepsilon$

xy^iz $\forall i$ $i \geq 0$

$\in L$

PUMPING LEMMA

– Since there are only n states in Q and $m \geq n$, by the pigeon hole theorem there are two states of $q_0, q_1, q_2, \ldots,$ and q_n are same, say $0 \leq i < j \leq n$ and $q_i = q_j$.

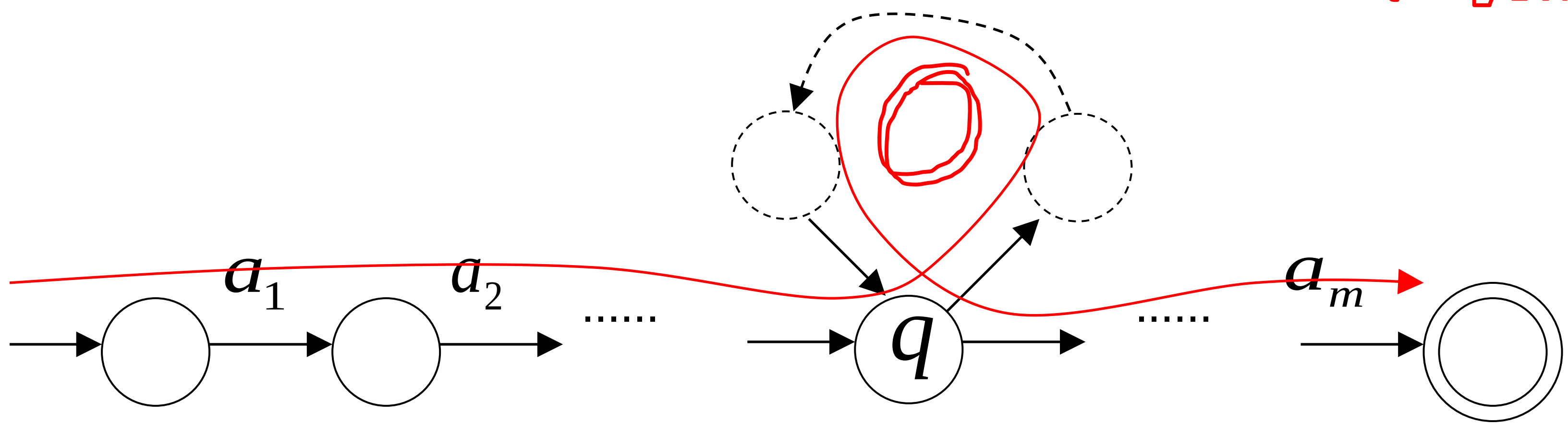


PUMPING LEMMA

- If w ∈L and |w| ≥ n, then, at least one state is repeated in the walk of w

Walk in DFA of $w = a_1 a_2 \cdots a_m$

xyz y^i 0 - - - ∞



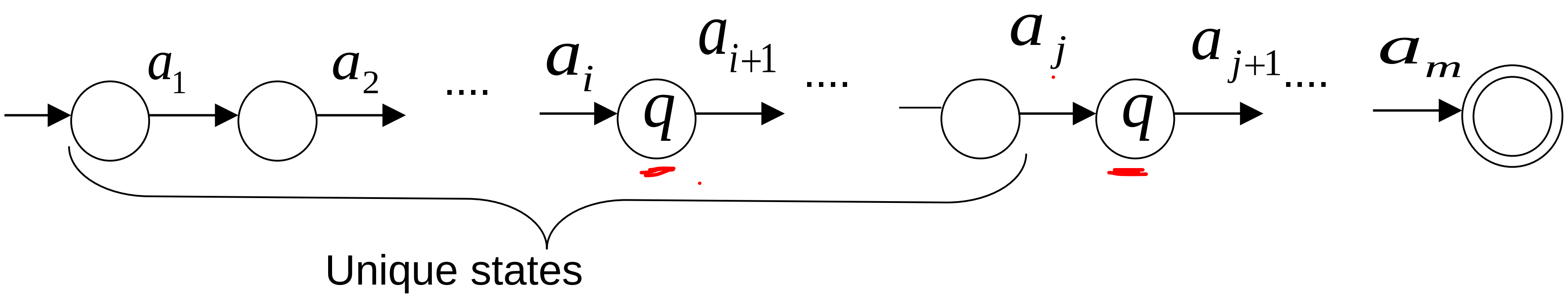
Repeated state in DFA

PUMPING LEMMA

- There could be many states repeated
- Take q to be the first state repeated

$w = a_1 a_2 \cdots a_i a_{i+1} \cdots a_j a_{j+1} \cdots a_m$

$i = 1$

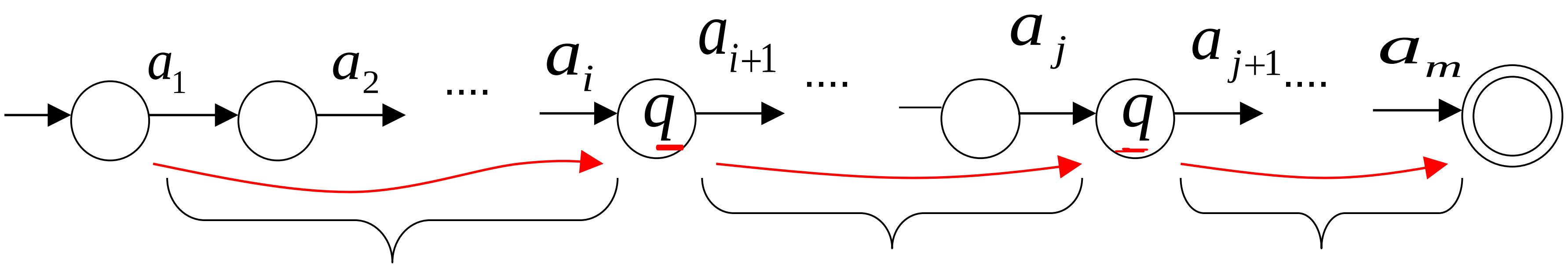


PUMPING LEMMA

- We can write w = xyz

$|x| \le n$

$y \ne \epsilon$

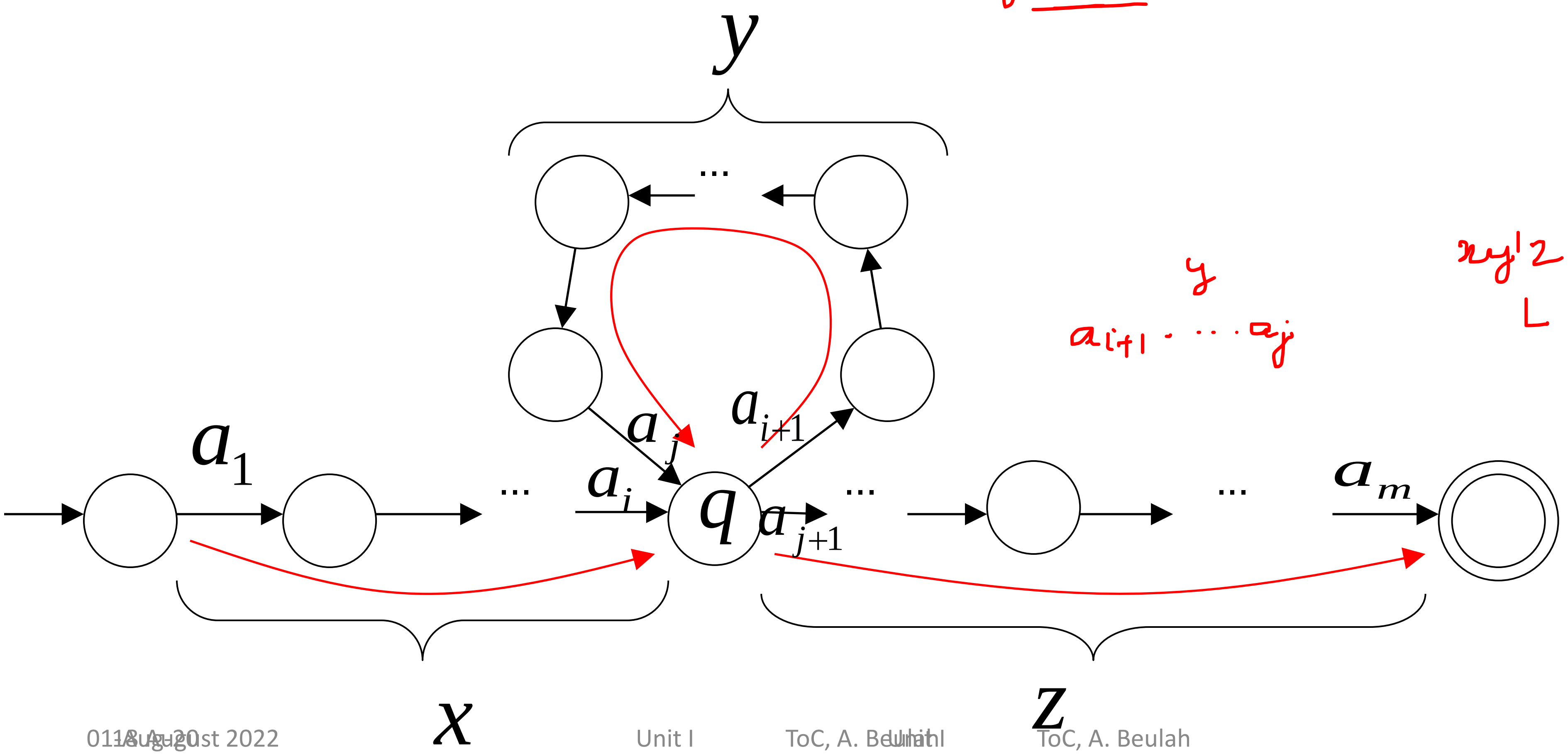


$x = a_1 \cdots a_i$ $\quad$ $y = a_{i+1} \cdots a_j$ $\quad$ $z = a_{j+1} \cdots a_m$

PUMPING LEMMA

- In DFA: write w = xyz
- 1 occurrence of y (xyz)

y^i $\underline{i=1}$



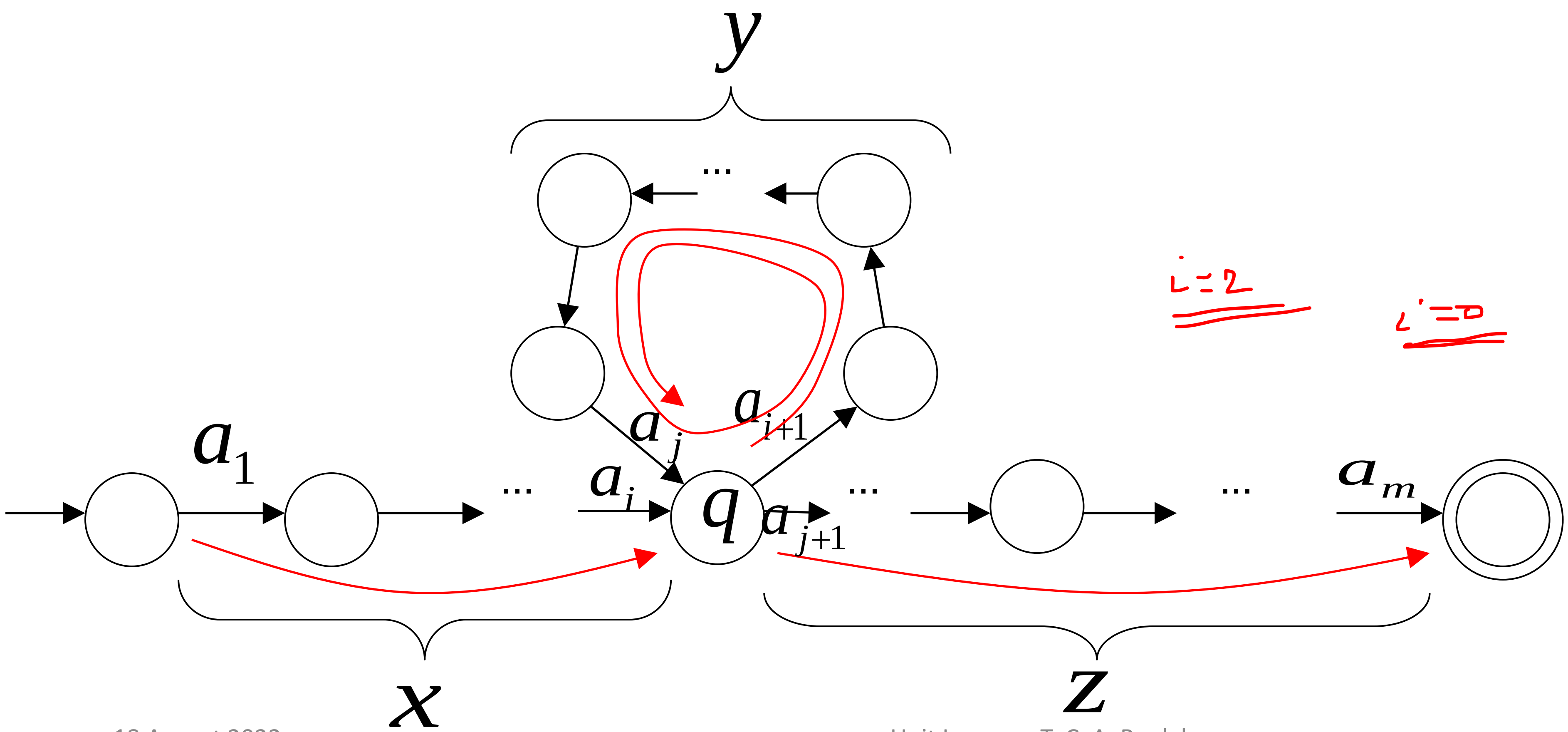
y
$a_{i+1} \cdot \cdots \cdot a_j$

xy^1z
L

18-Aug-20 Unit I ToC, A. Beulah

PUMPING LEMMA

- 2 occurrences of y (xyyz)



$i=2$

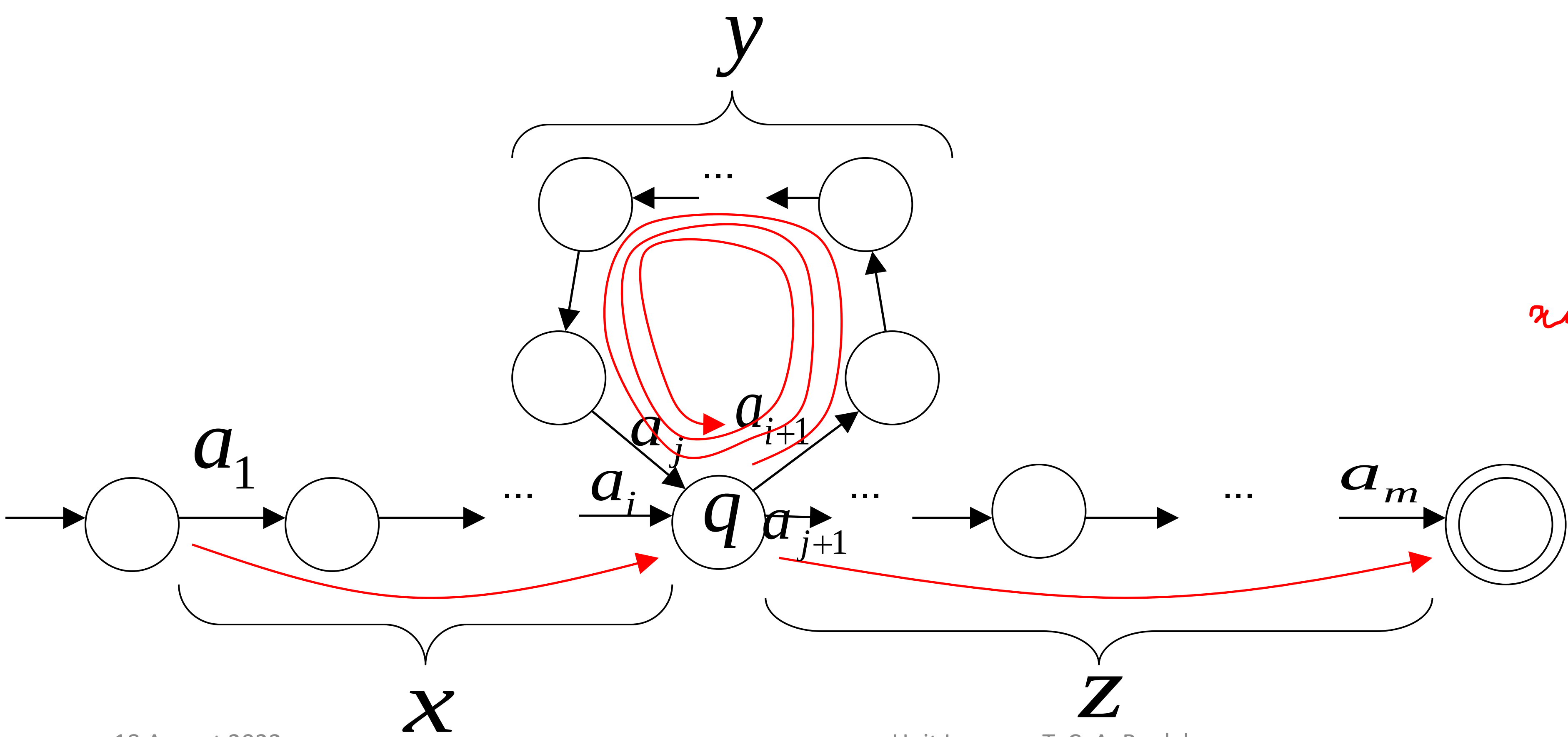
$i=0$

xy^2z

xz

PUMPING LEMMA

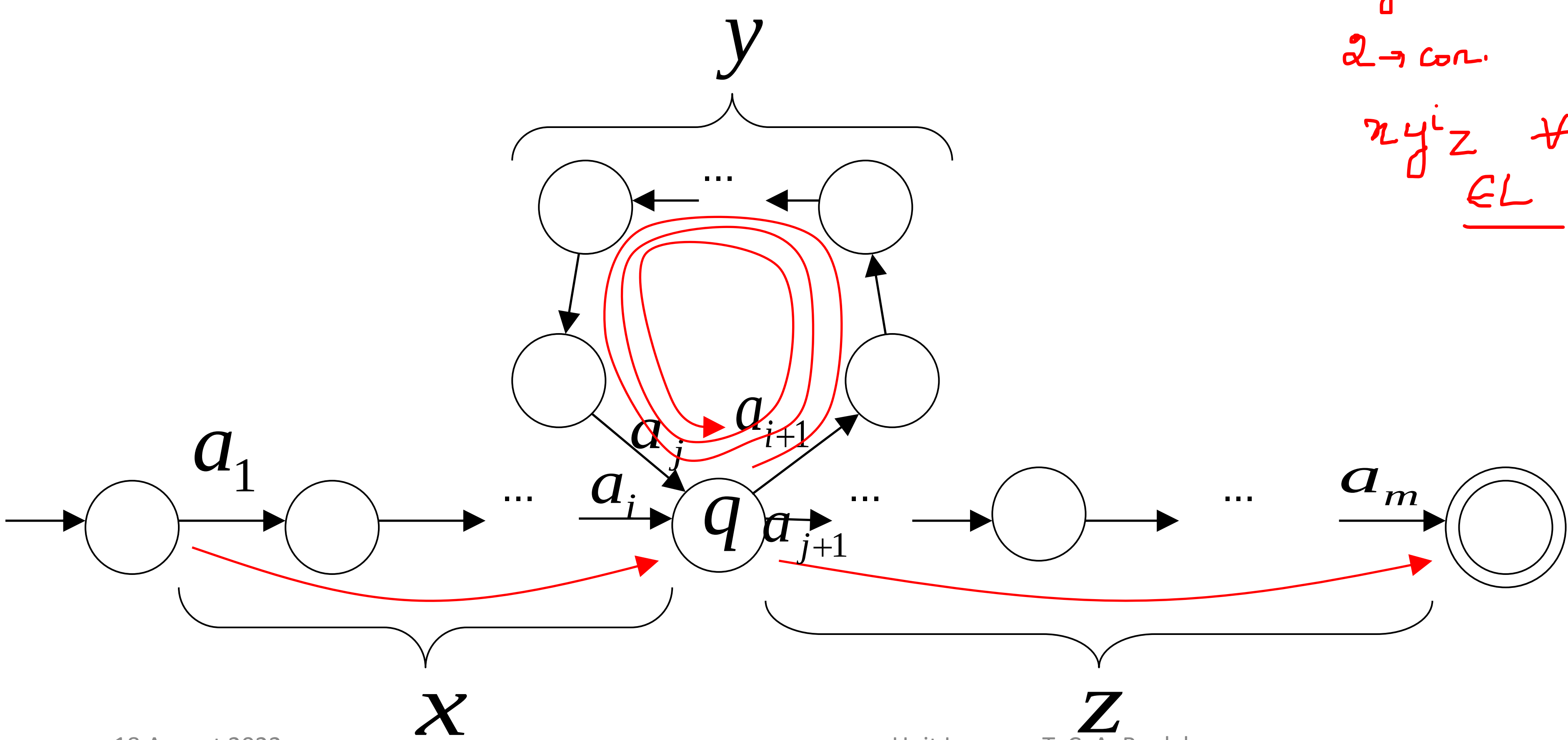
- 3 occurrences of y (xyyyz)



xy^3z
$\in L$ →

PUMPING LEMMA

- Many occurrences of y (xy^iz)
- xy^iz is also in L



$w = xyz$

$z \to$ con.

$xy^iz \;\; \forall i \geq 0$

$\in L$

PUMPING LEMMA

- $\delta(q_0, a_1 a_2 \ldots a_i) = q_i = q_j$
- $\delta(q_i, a_{i+1} .. a_j) = q_i$, and
- $\delta(q_i, a_{j+1} \ldots a_m) = q_n$

It is obvious that $\delta(q_i, y^i) = q_i$ for $i \geq 0$.

So, if the FA accepts $w = xyz$, it also accepts xy^iz.

APPLICATION

- Useful to prove a language L is not a regular set
- Method
  - Select an arbitrary 'n'
  - Choose a string w in L where $|w| \geq n$
  - For any partition of w = xyz such that
    - $|xy| \leq n$ and $|y| \geq 1$, show a contradiction;
    - i.e. show that there is a string xy^kz not in L;
    - k will depend on n, x, y, and z

EXAMPLE

- $L=\{a^m b^m / m>=1\}$

$w = xyz \qquad |w| \geq n$

$|xy| \leq n$

$|y| > 0$

$y \neq \epsilon$

$xy^iz \quad \forall i \geq 0 \quad \in L$

$a^m b^m / m \geq n \qquad n \rightarrow$

$w = a^p b^p \qquad |a^p b^p| \geq n$

$|w| \geq n$

xyz

$xy = a^r \qquad |a^r| \leq n$

$y = a^q \qquad y \neq \epsilon$

$z = a^{p-r} b^p$

$xy^iz = xy\, y^{i-1} z$

$xy^iz = a^r . (a^q)^{i-1} . a^{p-r} b^p$

$i=0$

$a^r (a^q)^{-1} a^{p-r} b^p$

$= \cancel{a^r} a^{-q} a^{\cancel{p-r}} b^p$

$= a^{p-q} b^p \notin L$

$i=1$

$a^r (a^q)^0 a^{p-r} b^p$

$= a^p b^p \in L$

$i=2 \quad a^r (a^q)^1 a^{p-r} b^p$

$= a^{p+q} b^p \notin L$

$\therefore$ The language is not regular.

EXAMPLE

- L={a^m/m is a prime}

$w = a^p$ assume $|a^p| \geq n$

$|w| \geq n$

$xy = a^r$ where $|a^r| \leq n$

assume $y = a^q$ $y \neq \varepsilon$

$z = a^{p-r}$

$\Rightarrow xy^iz$

$x\,y\,y^{i-1}z$

$a^r (a^q)^{i-1} a^{p-r}$

$i = 0$

$a^r a^{-q} a^{p-r}$ $a^{p-q} \notin L$ p-q might not be prime

$i = 1$

$a^r (a^q)^0 a^{p-r}$ $a^p \in L$

p is a prime no accepted

$i=2$

$a^r a^q a^{p-r}$ $a^{p+q} \notin L$ p+q might not be prime

EXAMPLE

- $L=\{a^{i^2} / i >= 1\}$
- a^i^2

SUMMARY

- Definition of Pumping lemma – Regular Language
- Application of pumping lemma

TEST YOUR KNOWLEDGE

- If we select a string w such that w∈L, and w=xyz. Which of the following portions cannot be an empty string?
  a) x
  b) y
  c) z
  d) all of the mentioned
- Which of the following one can relate to the given statement:
  Statement: If n items are put into m containers, with n>m, then atleast one container must contain more than one item.
  a) Pumping lemma
  b) Pigeon Hole principle
  c) Count principle
  d) None of the mentioned

REFERENCE

- Hopcroft J.E., Motwani R. and Ullman J.D, “Introduction to Automata Theory, Languages and Computations”, Second Edition, Pearson Education, 2008