WU

Compiler Design

What is a Compiler ?

A program or group of programs that translates one language into another

$$\boxed{\text{Source Program}} \longrightarrow \boxed{\text{Compiler}} \longrightarrow \boxed{\text{Target Program}}$$

→ Understand Regular Language with Examples

→ A string (word) over a set Σ is a finite sequence of elements from Σ.

→ ε = A string with no elements is null string (ε = Espilon)

→ Σ is the alphabet

→ Σ^* is the set of all finite strings over Σ.

→ A language over Σ is a subset of Σ^*

→ $u, v \in \Sigma^*$, uv is concatenation of u and v

u^R is reversal of u

Examples:

1. $u = abab$, $v = bbb$, $uv = ababbbb$
2. $u^R = baba$

$(uv)^R = v^R u^R = bbbbaba$

→ Let X, Y be languages

$X = \{a, ab\}$, $Y = \{cc\}$

$XY = \{acc, abcc\}$

$X^0 = \{\varepsilon\}$, this is not $\emptyset$, $0^0 = 1$ $(\text{Anything})^0 = 1$

$X^n = X \dots X = XX^{n-1} = X^{n-1}X$

$X^* = \bigcup_{i=0}^{\infty} X^i = X^0 \cup X^1 \cup X^2 \cup \dots \cup X^i$

$X^+ = \bigcup_{i=1}^{\infty} X^i = X^1 \cup X^2 \cup \dots \cup X^i$

Regular Expression

X^* means zero or more occurance of X

X^+ means one or more occurrance of X

$X^* = \{\varepsilon, X, XX, XXX, \ldots\}$

$X^+ = \{X, XX, XXX, \ldots\}$

Different Notations

(a|b) - a OR b

(a|b)* - Zero or more occurances of a OR b

(a|b)+ - one or more occurrences of a OR b

[a-z] - all alphabets from a to z

[A-Z] - all alphabets from A to Z

Example: Languages $X = \{a, ab\}$ $Y = \{cc\}$

$X^0 = \{\varepsilon\}$

$X^1 = X = \{a, ab\}$

$X^2 = XX = \{aa, aab, aba, abab\}$

$X^3 = XX^2$

$Y^* = \{\varepsilon\} \cup \{cc\} \cup \{cccc\} \cup \{cccccc\} \ldots \{(cc)^i\}$

How to describe the language Y^* using proper English?

The set of strings over {c} that consists of even number of c's.

Definition (Regular set)

Let Σ be an alphabet

1. $\emptyset$, $\{\varepsilon\}$, $\{a\}$ for every $a \in \Sigma$ are regular sets over Σ.
2. If X and Y are regular sets over Σ, then $X \cup Y$, XY, X^* are regular sets over Σ.

Definition (Regular Expression)

Let Σ be an alphabet

1. $\emptyset$, ε, a for every $a \in \Sigma$ are regular expression over
2. If u, v are regular expression over Σ, then $(u|v)$, uv, u^* are regular expressions over Σ.

Ex 1: Concatenation $ab = \{a\}\{b\} = \{ab\}$

$(a|b)c = \{a,b\}\{c\} = \{ac, bc\}$

Ex 2: Kleene Closure* – repetition

$(a|bb)^* = \{\varepsilon, a, bb, aa, abb, bba, bbbb, abbbb, abbaaabbbba, \ldots\}$

Not Included: ab, bbab, aabbb, aabbabab

How to describe the language using proper English?

The set of strings over $\{a,b\}$ where every maximal substring of b's has even length.

Ex 3. $(a|b)^* = \{a,b\}^*$

$= \{\varepsilon\} \cup \{a,b\} \cup \{a,b\}^2 \cup \{a,b\}^3 \cup \ldots$

What's the language ?

$(a|b)^* = \{\varepsilon, a, b, ab, aab, ba, bb, \ldots\}$

The set of all finite strings over $\{a,b\}$ including ε.

Ex 4: $(a|c)^* b (a|c)^*$

$S_1 b S_2$, S_1 and S_2 are any finite strings of a, c including ε.

For Example: b, ab, abc, aabc, acacbacac, ...

Not Included: acaaccacca, aabba, ababc

// Above b is not included, also b included twice

The set of strings over $\{a,b,c\}$ that contains exactly one b.

Ex 5: Write a regular expression for the following

1. The set of strings over $\{a, b, c\}$ that contain exactly 3 letters that are not c.

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

// Here any no of $c^*c^*c^*$ is always c^*

$c^* = c^*c^* = \{\varepsilon, c, cc, ccc, \ldots\}\{\varepsilon, c, cc, ccc, \ldots\}$

2. The set of strings over $\{a, b, c\}$ that contain at most 3 letters that are not c

$c^*(a|b|\varepsilon)c^*(a|b|\varepsilon)c^*(a|b|\varepsilon)c^*$

3. The set of strings over $\{a, b, c\}$ that contain atleast 3 letters that are not c.

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

Ex 6 Write a regular expression for the set of strings over $\{a, b\}$ that contain even number of b's.

Even number $2K$ (K can be 0)

2 b's : $a^*ba^*ba^*$

Even no of b's : $((a^*ba^*ba^*)^* | a^*)$

$(a|c)^* \neq a^*c^*$

$a^*(ba^*ba^*)^*$, $(a^*ba^*b)^*a^*$, $(a^*ba^*ba^*)^*a^*$, $a^*(ba^*ba^*)^*a^*$

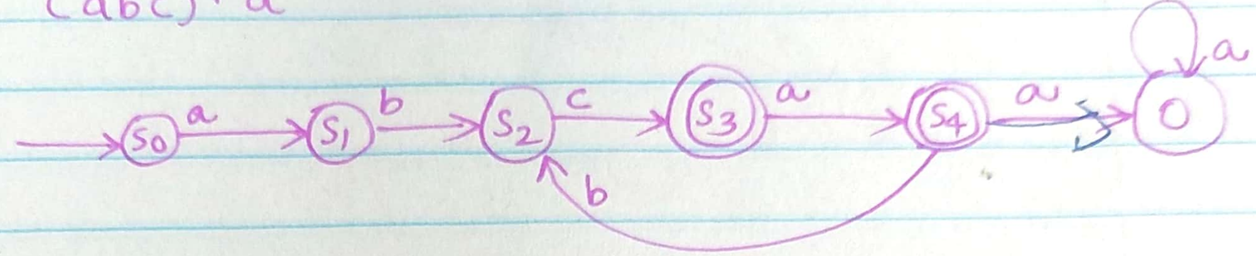
But Not :

$a^*(ba^*b)^*a^*$, $(a^*ba^*ba^*)^*$

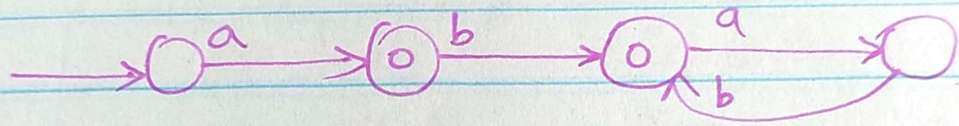
Even no of b's : $(a|c)^*b(a|c)^*b(a|c)^*$ for $\{a, b, c\}$

DFA

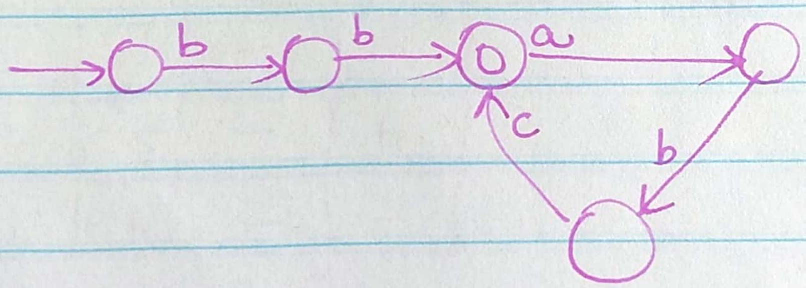
4. $(abc)^+a^*$



3. $(ab)^+ \mid a$

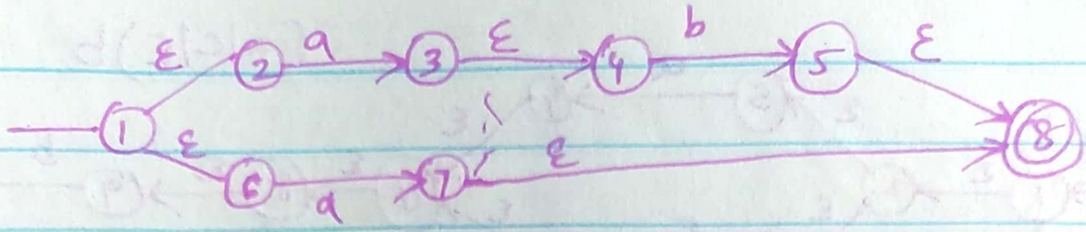


1. $-bb(abc)^*$

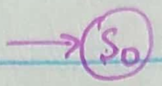


NFA to DFA

a: 2,6
b: 4



i) Start state $S_0 = \overline{1} = \{\overline{1}\} = \{1,2,6\}$



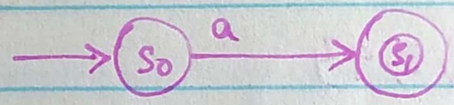
ii) $S_0 = \{1,2,6\}$

$(S_0)_a = \{1,2,6\}_a = \{3,7\}$

$\overline{(S_0)_a} = \{\overline{3,7}\} = \{3,4,7,8\} = S_1$

$S_1 = \{3,4,7,8\}$

$(S_0)_b = \{1,2,6\}_b = \emptyset$

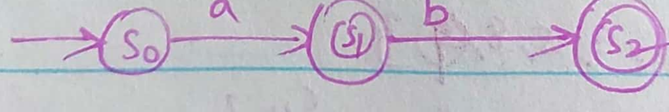


iii) $S_1 = \{3,4,7,8\}$

$(S_1)_a = \{3,4,7,8\}_a = \emptyset$

$(S_1)_b = \{3,4,7,8\}_b = \{5\}$

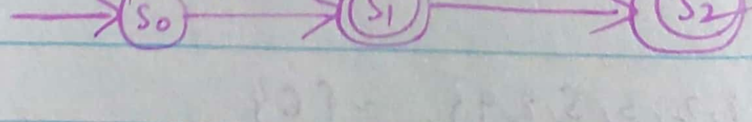
$\overline{(S_1)_b} = \{\overline{5}\} = \{5,8\} = S_2$



iv) $S_2 = \{5,8\}$

$(S_2)_a = \{5,8\}_a = \emptyset$

$(S_2)_b = \{5,8\}_b = \emptyset$

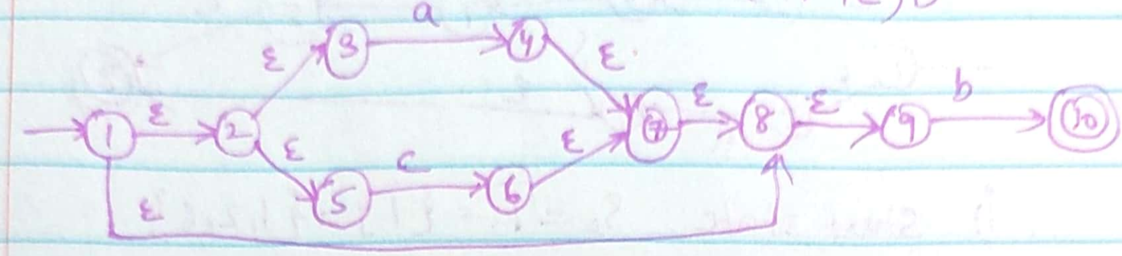


Convert NFA to DFA

$((a|c)^* \epsilon\ ((a|c)^* b)$

$(a|c|\epsilon)b$

a: 3
b: 9
c: 5



$\bar{1} = \{1, 2, 3, 5, 8, 9\}$

$\bar{2} = \{2, 3, 5\}$

$\bar{3} = \{3\}$

$\bar{4} = \{4, 7, 8, 9\}$

$\bar{5} = \{5\}$

$\bar{6} = \{6, 7, 8, 9\}$

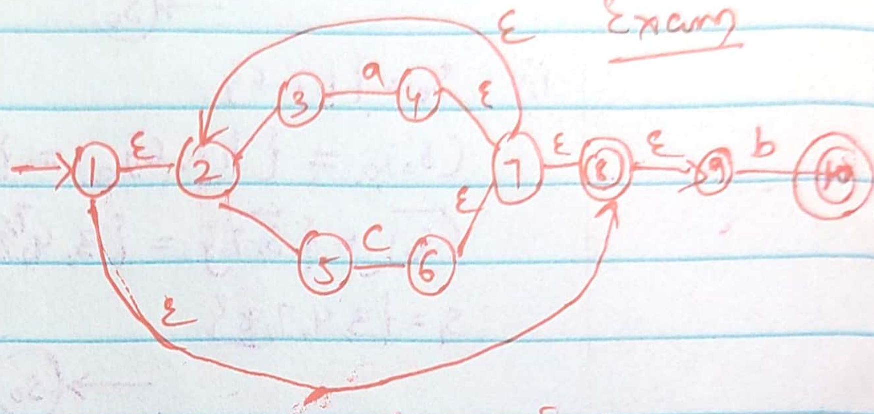
$\bar{7} = \{7, 8, 9\}$

$\bar{8} = \{8, 9\}$

$\bar{9} = \{9\}$

$\bar{10} = \{10\}$

Exam



8, 10

$(a|c)^* b$

$S_0 = \bar{1} = \{1, 2, 3, 5, 8, 9\}$

$(S_0)_a = \{1, 2, 3, 5, 8, 9\}_a = \{4\}$

$\overline{(S_0)_a} = \{\bar{4}\} = \{4, 7, 8, 9\} = S_1$

$(S_0)_b = \{1, 2, 3, 5, 8, 9\}_b = \{10\}$

$\overline{(S_0)_b} = \{\bar{10}\} = \{10\} = S_2$

$(S_0)_c = \{1, 2, 3, 5, 8, 9\}_c = \{6\}$

$\overline{(S_0)_c} = \{\bar{6}\} = \{6, 7, 8, 9\} = S_3$

$S_1 = \{4,7,8,9\}$

$(S_1)_a = \{4,7,8,9\}_a = \emptyset$

$(S_1)_b = \{4,7,8,9\}_b = \{\overline{10}\}$

$\overline{(S_1)_b} = \{\overline{10}\} = \{10\} = S_2$

$(S_1)_c = \{4,7,8,9\}_c = \emptyset$

8, 10

$S_2 = \{10\}$

$(S_2)_a = \{10\}_a = \emptyset$

$(S_2)_b = \{10\}_b = \emptyset$

$(S_2)_c = \{10\}_c = \emptyset$

S0
S1 — 10
S2
S3 — 10
S4

$S_3 = \{6,7,8,9\}$

$(S_3)_a = \{6,7,8,9\}_a = \emptyset$

$(S_3)_b = \{6,7,8,9\}_b = \{10\}$

$\overline{(S_3)_b} = \{\overline{10}\} = \{\overline{10}\} = S_2$

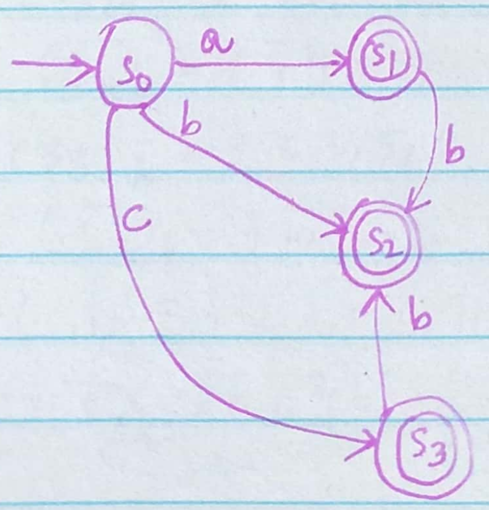
$(S_3)_c = \{6,7,8,9\}_c = \emptyset$

Kishor

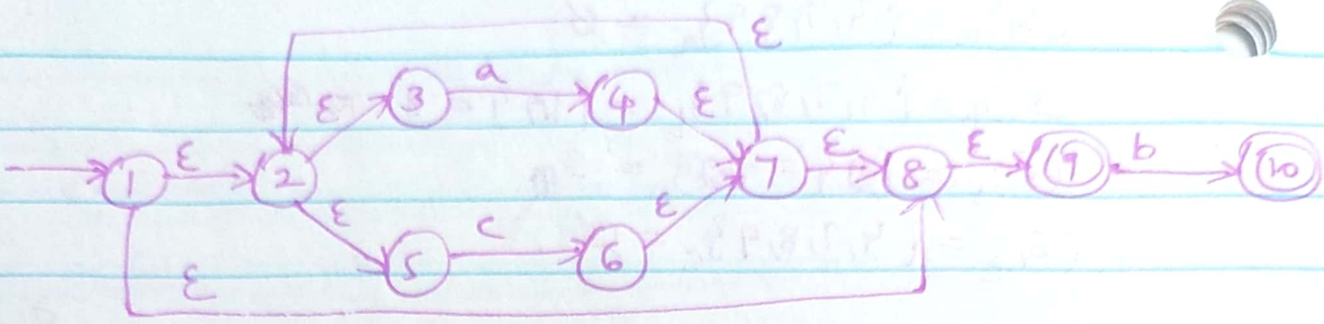
~~$S_4 = \{\ldots\}$~~

$\{8, 10, 1, \ldots\}$

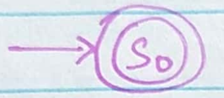
$\{8, 1, 2\}$



NFA to DFA



1. $S_0 = \overline{1} = \{1,2,3,5,8,9\}$, 9 is final state



2. $(S_0)_a = \{1,2,3,5,8,9\}_a = \{4\}$

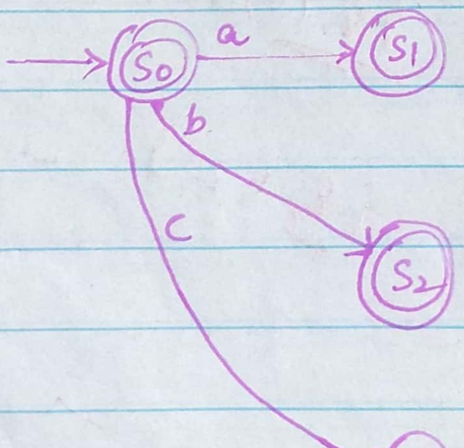
$\overline{(S_0)_a} = \{\overline{4}\} = \{\overset{2,3,5}{4,7,8,9}\} = S_1$

$(S_0)_b = \{1,2,3,5,8,9\}_b = \{10\}$

$\overline{(S_0)_b} = \{\overline{10}\} = \{10\} = S_2$

$(S_0)_c = \{1,2,3,5,8,9\}_c = \{6\}$

$\overline{(S_0)_c} = \{\overline{6}\} = \{2,3,5,6,7,8,9\} = S_3$



$S_1 = \{2,3,4,5,7,8,9\}$

$(S_1)_a = \{2,3,4,5,7,8,9\}_a = \{4\}$

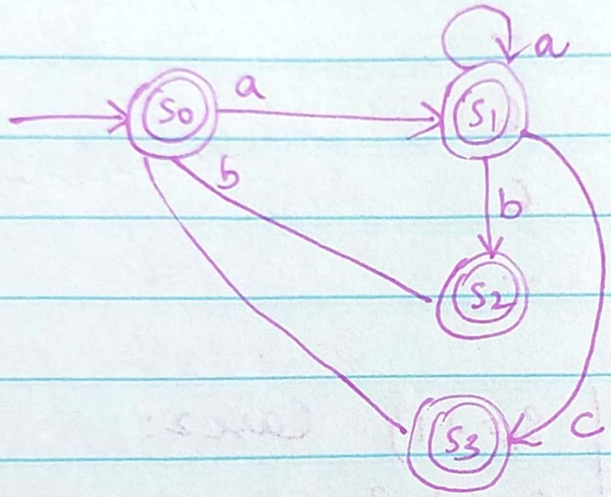
$\overline{(S_1)_a} = \{\overline{4}\} = S_1$

$(S_1)_b = \{2,3,4,5,7,8,9\}_b = \{10\}$

$\overline{(S_1)_b} = \{\overline{10}\} = \{10\} = S_2$

$(S_1)_c = \{2,3,4,5,7,8,9\}_c = \{6\}$

$\overline{(S_1)_c} = \{\overline{6}\} = \{2,3,5,6,7,8,9\} = S_3$



$S_2 = \{10\}$

$(S_2)_a = \{10\}_a = \emptyset$

$(S_2)_b = \{10\}_b = \emptyset$

$(S_2)_c = \{10\}_c = \emptyset$

$S_3 = \{2,3,5,6,7,8,9\}$

$(S_3)_a = \{2,3,5,6,7,8,9\}_a = \{4\}$

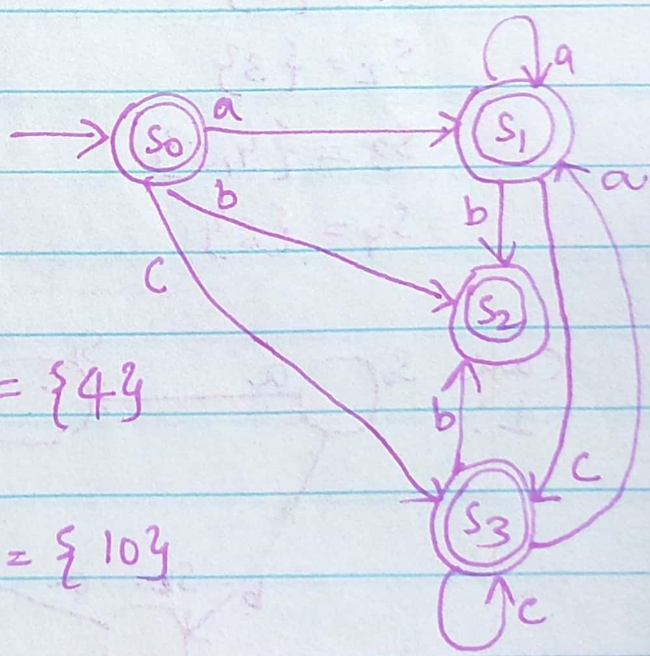
$\overline{(S_3)_a} = \{\overline{4}\} = S_1$

$(S_3)_b = \{2,3,5,6,7,8,9\}_b = \{10\}$

$\overline{(S_3)_b} = \{\overline{10}\} = S_2$

$(S_3)_c = \{2,3,5,6,7,8,9\}_c = \{6\}$

$\overline{(S_3)_c} = \{\overline{6}\} = S_3$

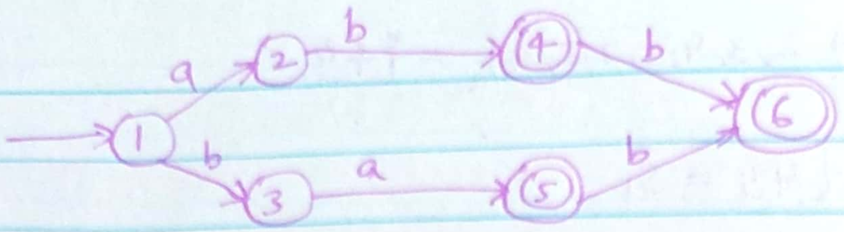


DFA

Minimization Algorithm

Initial state 1, 2, 3

Final state 4, 5, 6



	States	a	b
S_0	1	2	3
S_1	2	–	4
S_2	3	5	–
S_3	4	–	6
	5	–	6
S_4	6	–	–

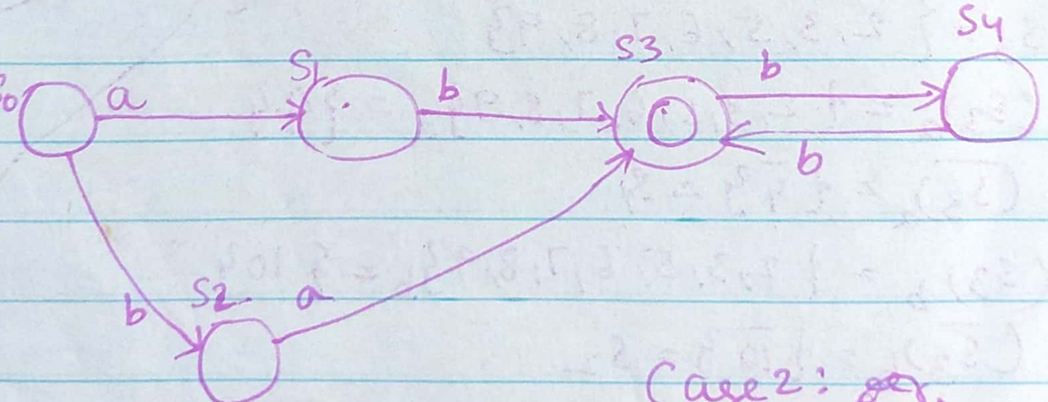
Case 1:

$S_0 = \{1\}$
$S_1 = \{2\}$
$S_2 = \{3\}$
$S_3 = \{4, 5\}$
$S_4 = \{6\}$

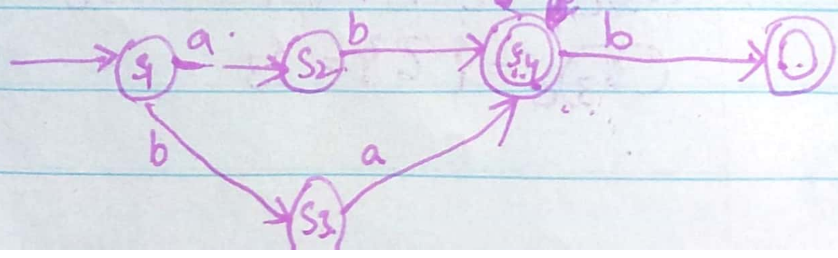
Case 2:

$S_1 = \{1\}$
$S_2 = \{2\}$
$S_3 = \{3\}$
$S_4 = \{4, 5\}$
$S_5 = \{6\}$

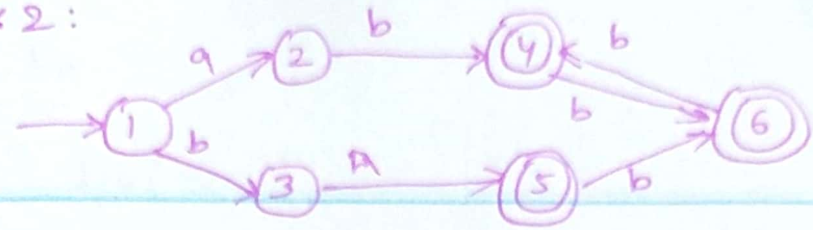
Case 1:



Case 2:

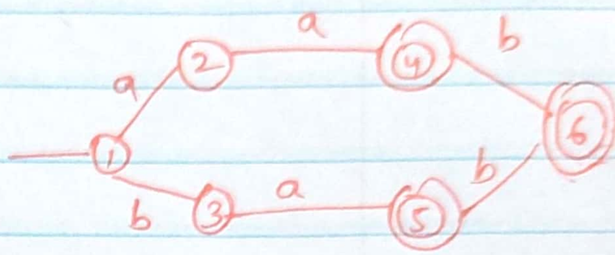


Ex 2:



Initial states: 1,2,3

Final states: 4,5,6



	States	a	b
S_1	1	2	3
S_2	2	–	4
S_3	3	5	–
S_4	4	–	6
	5	–	6
S_5	6	–	4

– 6

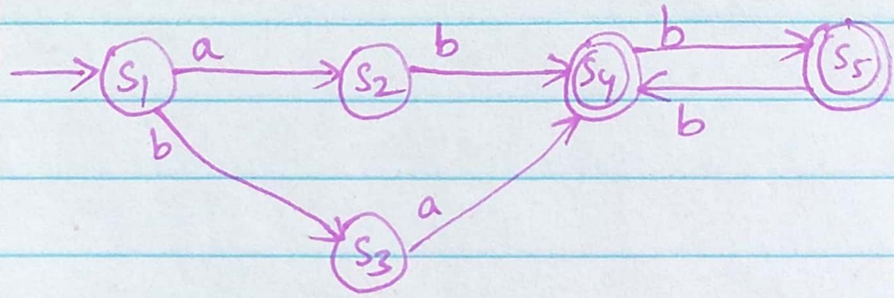
$S_1 = \{1\}$

$S_2 = \{2\}$

$S_3 = \{3\}$

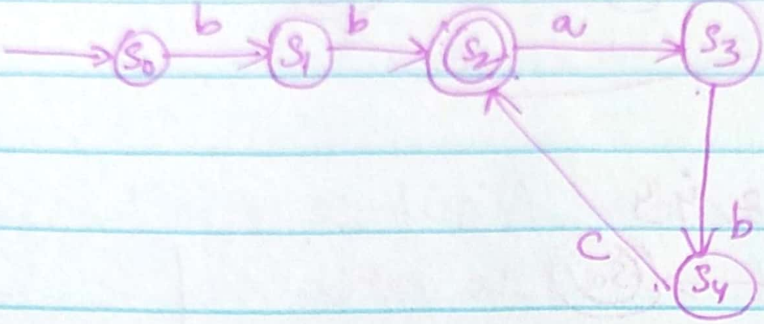
$S_4 = \{4,5\}$

$S_5 = \{6\}$



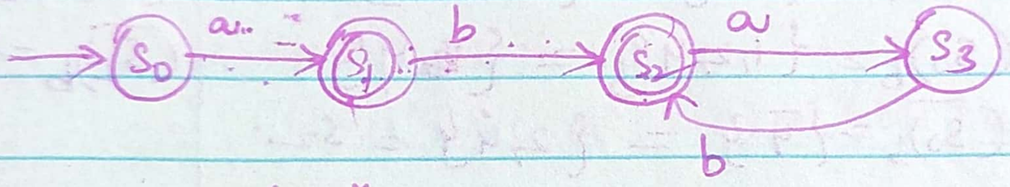
DFA

1. $bb(abc)^*$



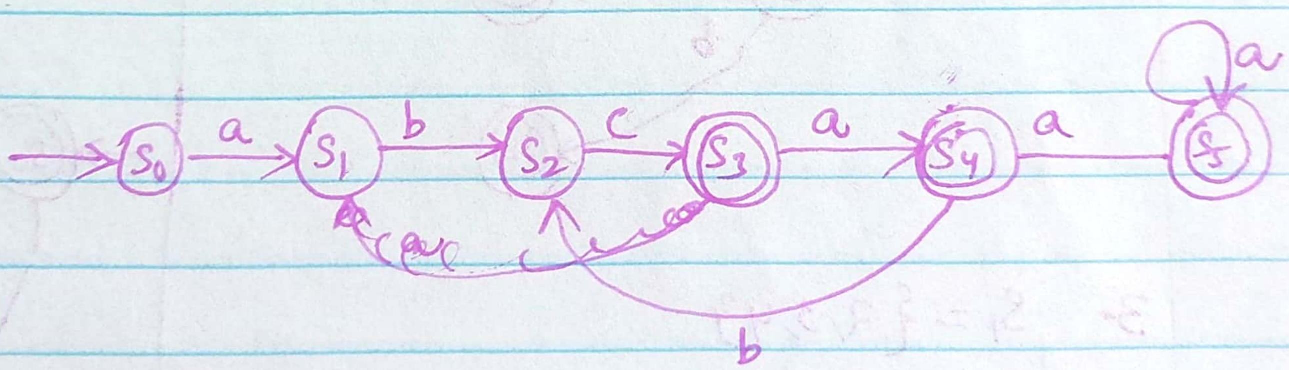
$ab^+ = ab, abab,$
$ababab.$

2. $(ab)^+ \mid a$ $\{a, \underline{ab}, abab\}$ ababab

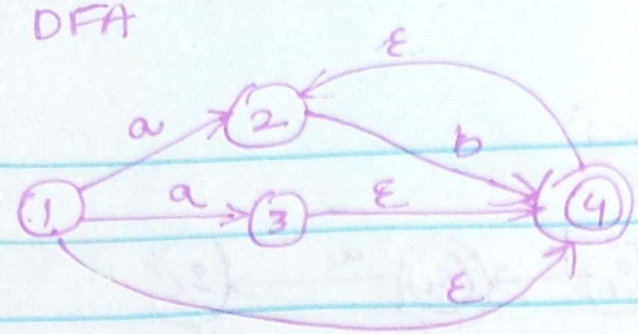


3. $(abc)^+ a^*$

abca



NFA to DFA



a: 1

b: 2

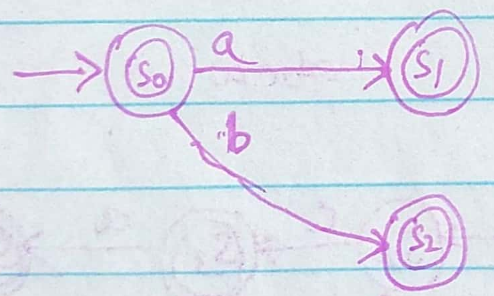
$S_0 = \overline{1} = \{1,2,4\}$

2. $(S_0)_a = \{1,2,4\}_a = \{2,3\}$

$\overline{(S_0)_a} = \{\overline{2,3}\} = \{2,3,4\} = S_1$

$(S_0)_b = \{1,2,4\}_b = \{4\}$

$\overline{(S_0)_b} = \{\overline{4}\} = \{2,4\} = S_2$

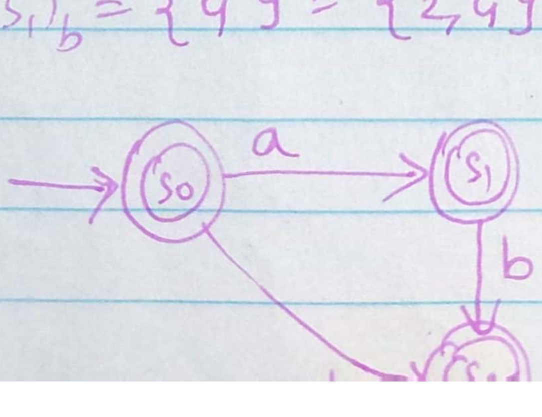


3. $S_1 = \{2,3,4\}$

$(S_1)_a = \{2,3,4\}_a = \emptyset$

$(S_1)_b = \{2,3,4\}_b = \{4\}$

$\overline{(S_1)_b} = \{\overline{4}\} = \{2,4\} = S_2$



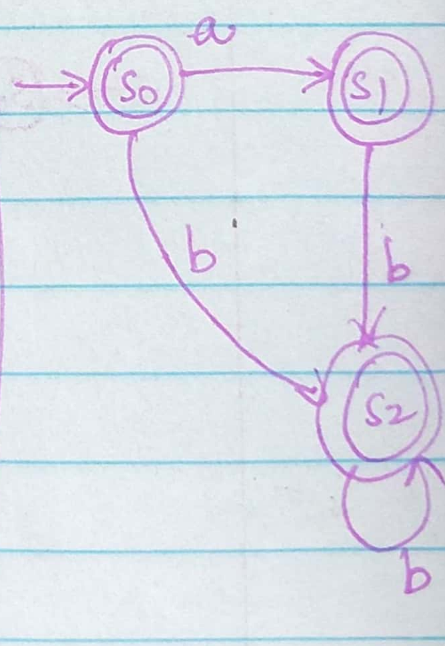
$S_2 = \{2,4\}$

$(S_2)_a = \{2,4\} = \emptyset$

$(S_2)_b = \{2,4\}_b = \{4\}$

$\overline{(S_2)_b} = \{\overline{4}\} = \{2,4\}$

$= S_2$



RE

Test C

All strings over {a,b,c} that contain no aa's

$$(b|c)^* \mid (b|c)^*(a(b|c)^+)^* a (b|c)^*)$$

3, after 7,8,9

All strings of digits such that all the 2's and 3's occur after all 8's and 9's

$$(0|1|4|5|6|7|8|9)^+ (0|1|2|3|4|5|6|7)^*$$

$*$

All strings over {a,b,c} in which numbers of b's plus numbers of c's is 4

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

$$(0|1|2|4|5|6|7|8|9|)^+(0|1|2|3|4|5|\cancel{6})^*$$

NFA to DFA

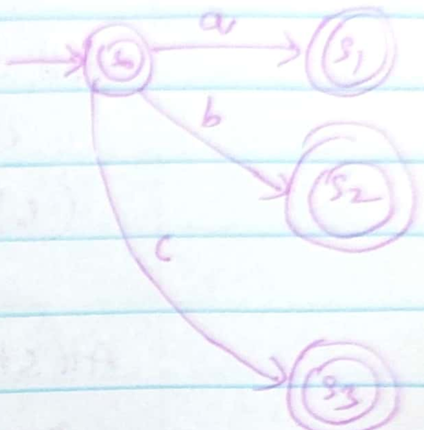
$a: 3$
$b: 9$
$c: 5$

9,10 ⓪

1. $S_0 = \{\bar{1}\} = \{1,2,3,5,8,9\}$ ⟶ ①

$(S_0)_a = \{1,2,3,5,8,9\}_a = \{4\}$

$\overline{(S_0)_a} = \{\bar{4}\} = \{2,3,5,4,7,8,9\} = S_1$



$(S_0)_b = \{1,2,3,5,8,9\}_b = \{10\}$

$\overline{(S_0)_b} = \{\overline{10}\} = \{10\} = S_2$

$(S_0)_c = \{1,2,3,5,8,9\}_c = \{6\}$

$\overline{(S_0)_c} = \{\bar{6}\} = \{2,3,5,7,8,9\} = S_3$

2. $S_1 = \{2,3,4,5,7,8,9\}$

$(S_1)_a = \{2,3,4,5,7,8,9\}_a = \{4\}$

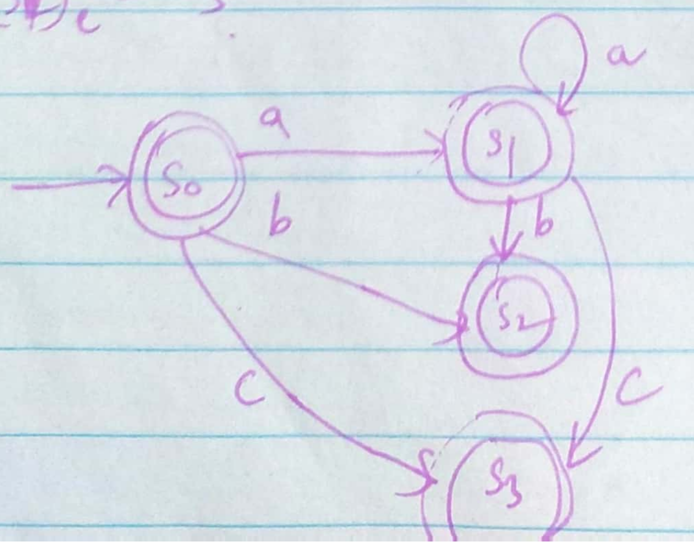
$\overline{(S_1)_a} = \{\bar{4}\} = S_1$

$(S_1)_b = \{2,3,4,5,7,8,9\}_b = \{10\}$

$\overline{(S_1)_b} = S_2$

$(S_1)_c = \{2,3,4,5,7,8,9\}_c = \{6\}$

$\overline{(S_1)_c} = S_3$



$S_2 = \{10\}$

$(S_2)_a = \{10\}_a = \emptyset$

$(S_2)_b = \{10\}_b = \emptyset$

$(S_2)_c = \{10\}_c = \emptyset$

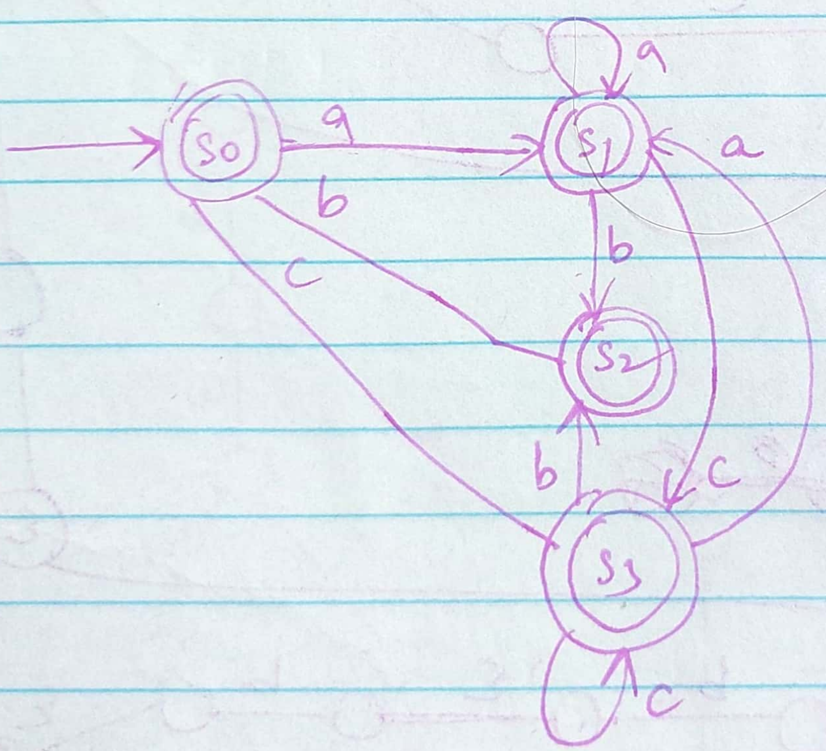
$S_3 = \{2, 3, 5, 7, 8, 9\}$

$(S_3)_a = \{2, 3, 5, 7, 8, 9\}_a = \{4\}$

$(\overline{S_3})_a = S_1$.

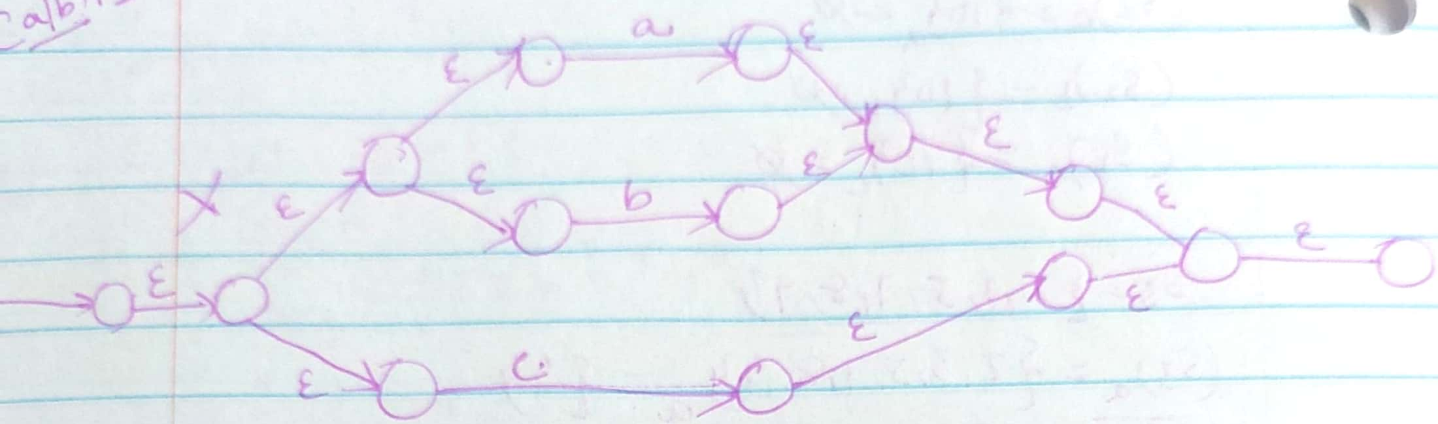
$(S_3)_b = \{10\}$, $(\overline{S_3})_b = \{10\} = S_2$

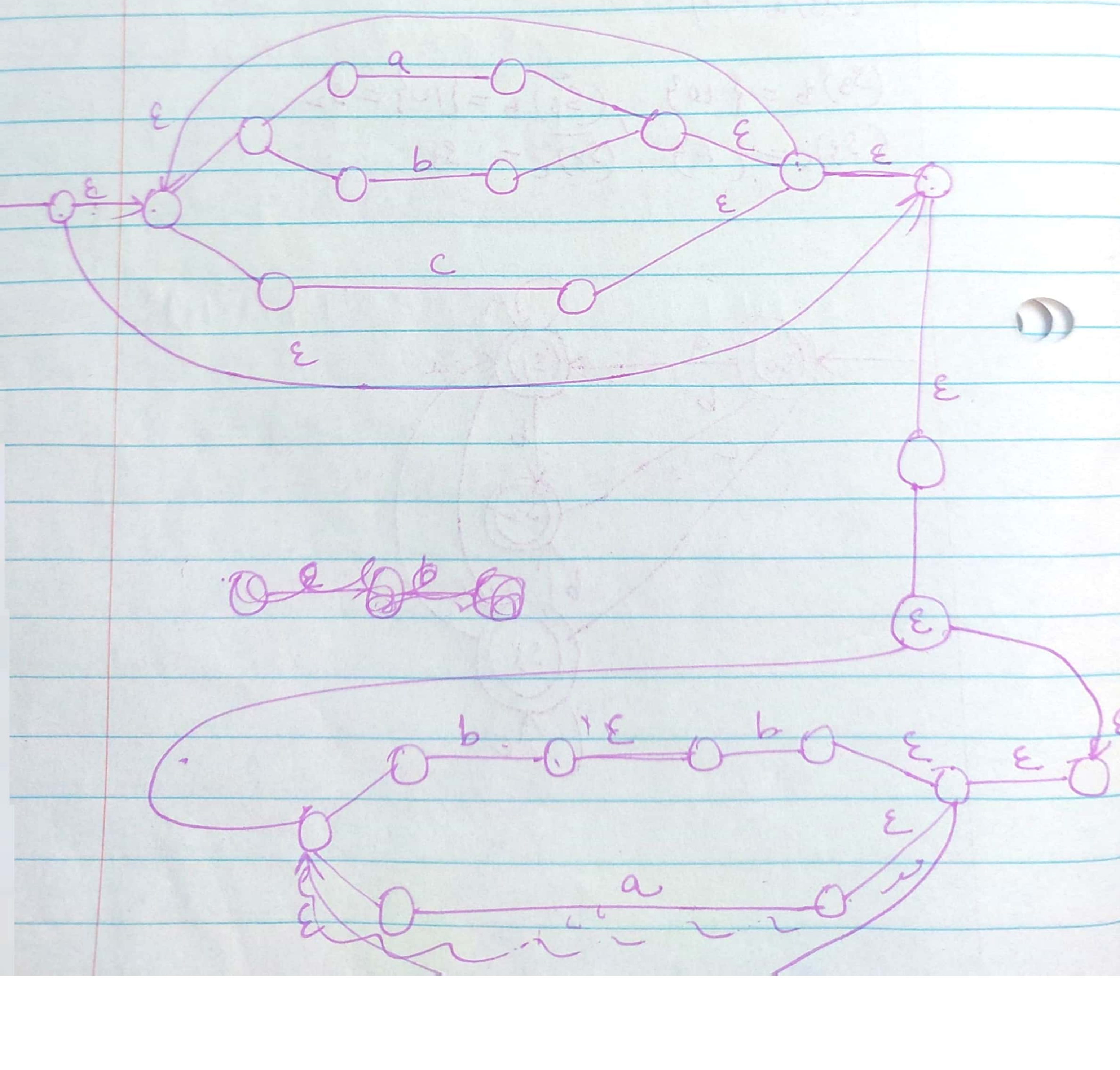
$(S_3)_c = \{8\}$, $(\overline{S_3})_c = \underline{S_3}$



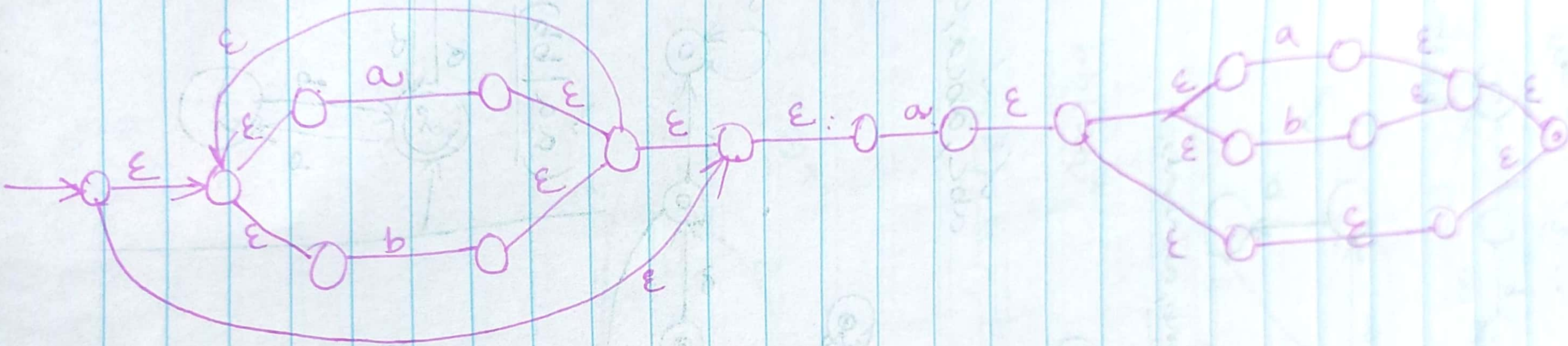
(a|b)|c

Thomson $(a|b|c)^*(bb|a)$

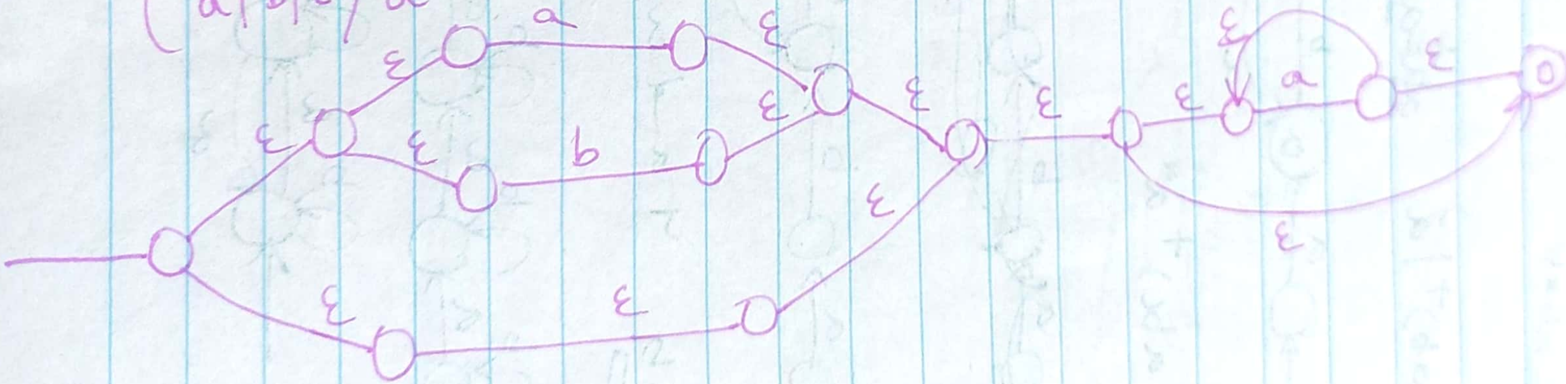




$(a|b)^* a (a|b|\varepsilon)$



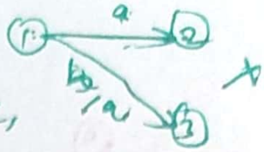
$(a|b|\varepsilon)\, a^*$

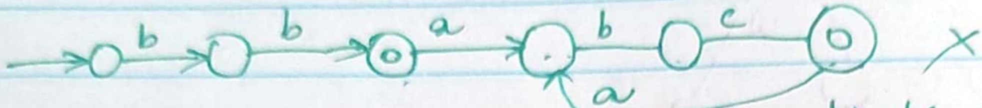


২০

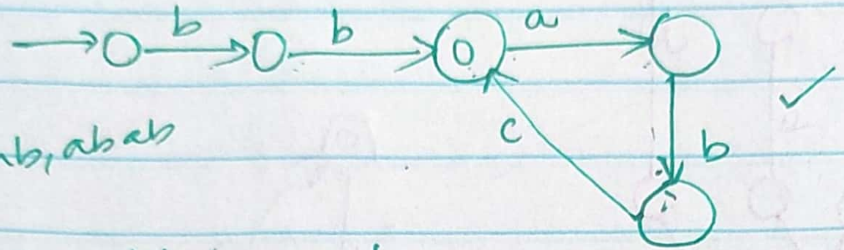
$bb(abc)^*$

$(abc)^* = \{\varepsilon, abc, abcabc, \ldots\}$



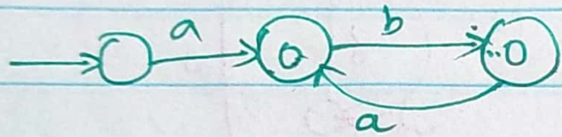


bbabc

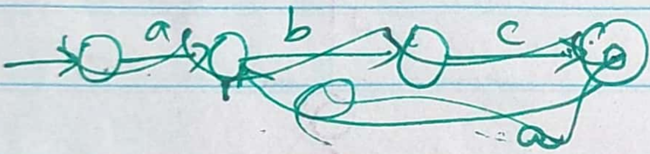


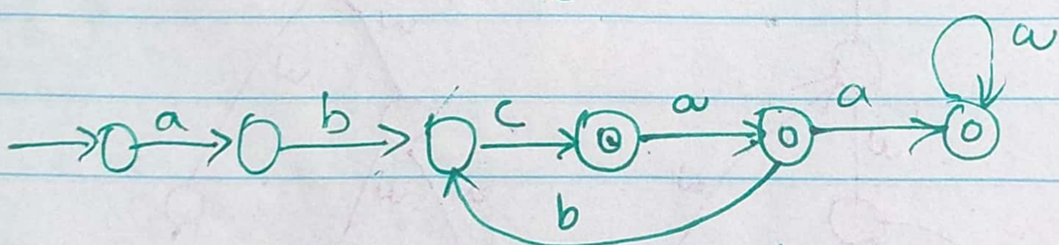
$ab^* = 0, ab, abab$

$(ab)^+ | a$ $ab^+ = ab, abab, ababab, \ldots$

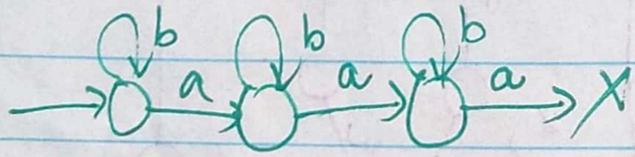


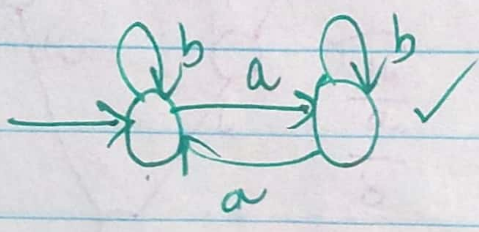
$(abc)^+ a^*$ abc, abca, abcabcaa,



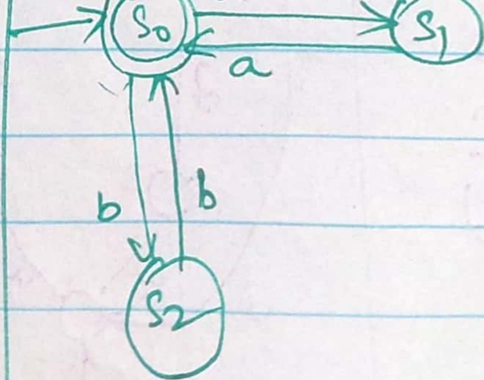


a, b 4 and even

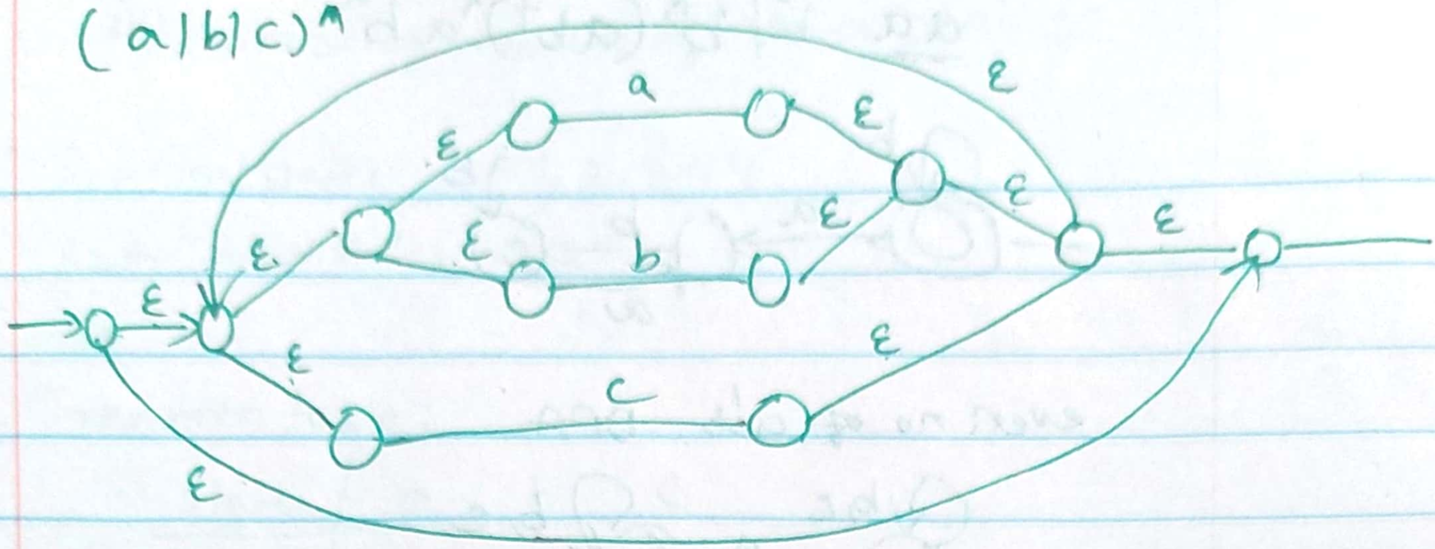




$(aa|bb)^*$ DFA

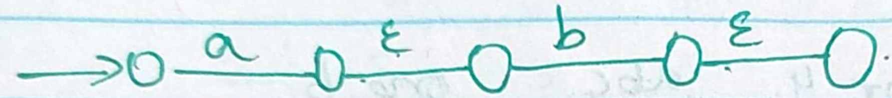


$(a|b|c)^*$

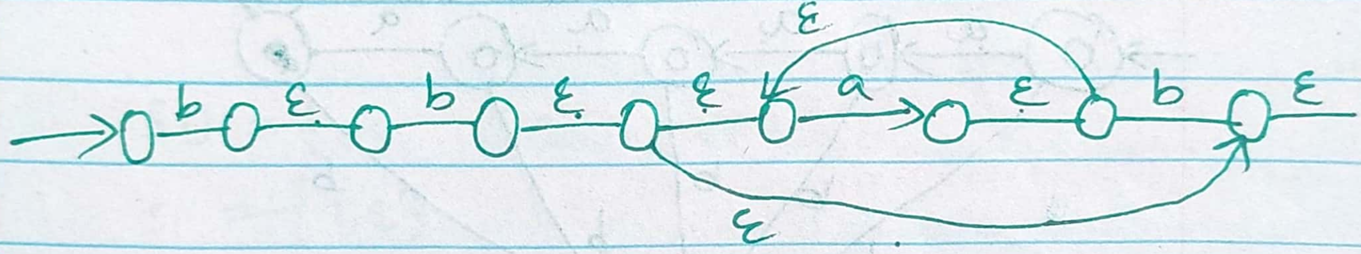


ε

abc

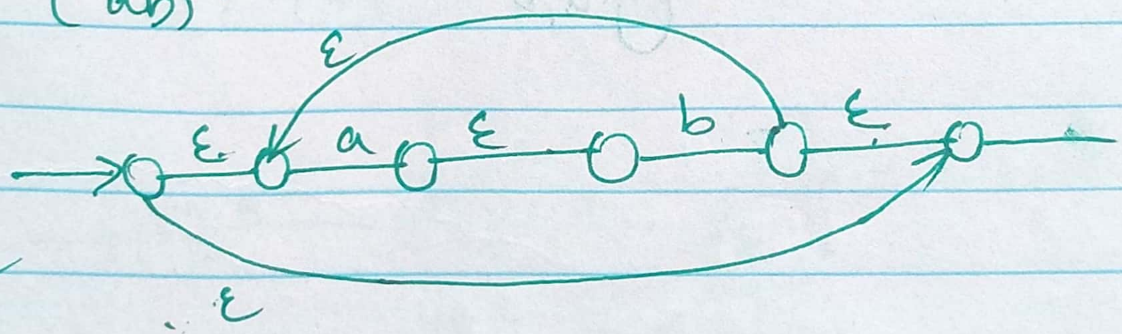


$bb(ab)^*$



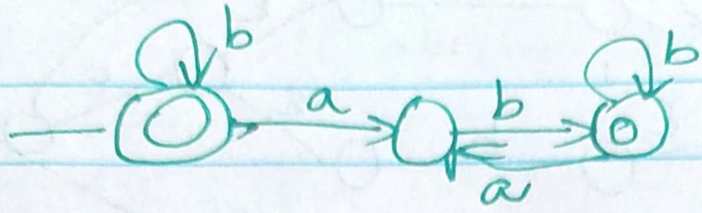
7ε

$(ab)^*$

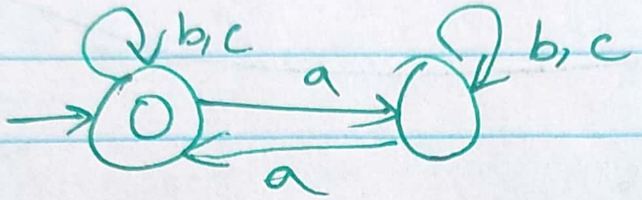


5ε

aa $b^* | b^*(ab^+)^* a b^*$

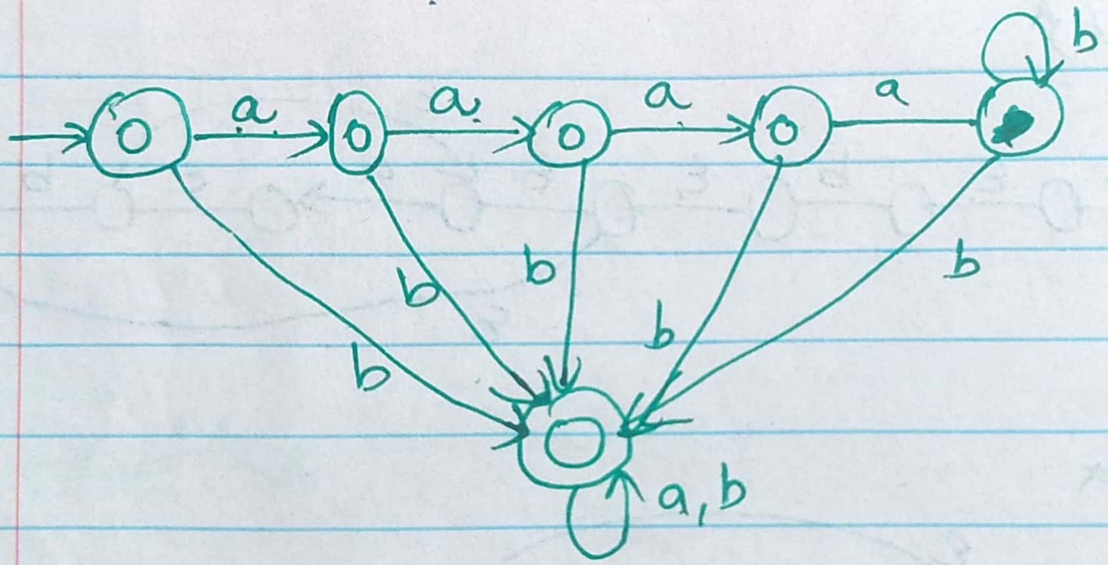


even no of a's DFA



greater or 4. abc. one b

a, b > 3 one b



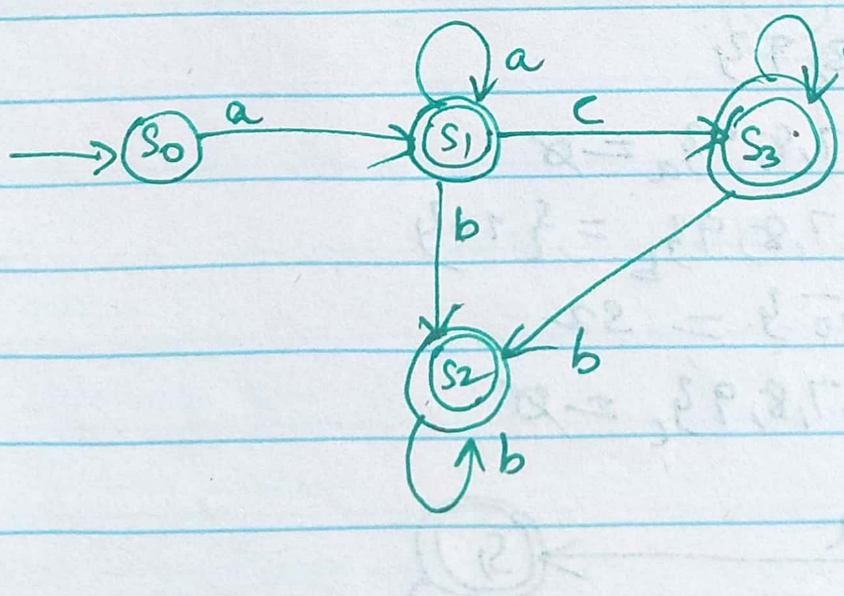
Minimizing the no of states DFA

1. S_1 = final states = {1, 2, 3, 4}
   S_2 = Initial states = {0}

2. Transition table:

states	a	b	c
0	1	–	–
1	2	3	4
2	2	3	4
3	–	3	–
4	–	3	4

$S0 = \{0\}$ // Cant split
$S_1 = \{1, 2\}$
$S2 = \{3\}$
$S3 = \{4\}$



NFA to DFA (20)

a: 3
b: 9
c: 5

$S_0 = \{\bar{7}\} = \{1, 2, 3, 5, 8, 9\}$

7, 10 (FS)

destinct state

$(S_0)_a = \{1,2,3,5,8,9\}_a = \{4\}$

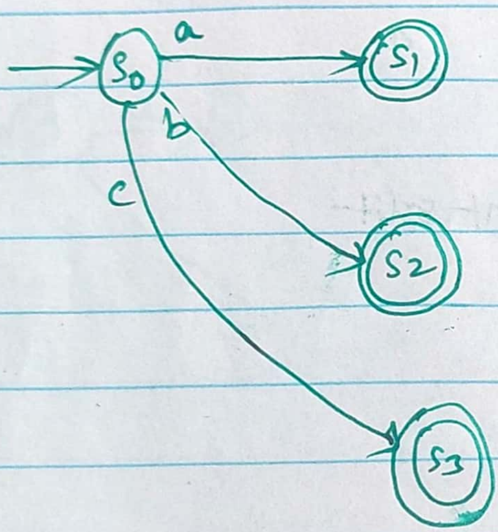
$\overline{(S_0)_a} = \{\bar{4}\} = \{4, 7, 8, 9\} = S_1$

$(S_0)_b = \{1,2,3,5,8,9\}_b = \{10\}$

$\overline{(S_0)_b} = \{\overline{10}\} = \{10\} = S_2$

$(S_0)_c = \{1,2,3,5,8,9\}_c = \{6\}$

$\overline{(S_0)_c} = \{\bar{6}\} = \{6, 7, 8, 9\} = S_3$



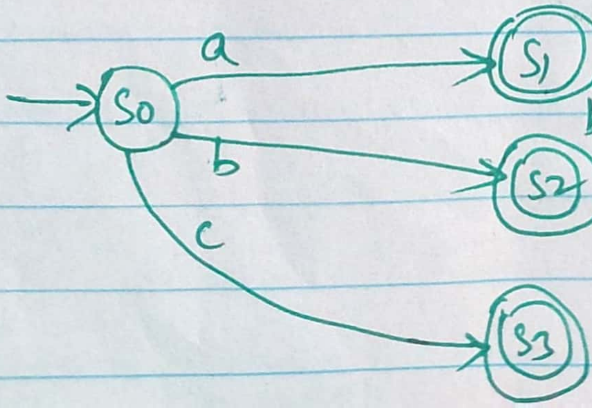
$S_1 = \{4, 7, 8, 9\}$

$(S_1)_a = \{4,7,8,9\}_a = \emptyset$

$(S_1)_b = \{4,7,8,9\}_b = \{10\}$

$\overline{(S_1)_b} = \{\overline{10}\} = S_2$

$(S_1)_c = \{4,7,8,9\}_c = \emptyset$



$S_2 = \{10\}$

$(S_2)_a = \emptyset$

$(S_2)_b = \emptyset$

$(S_2)_c = \emptyset$

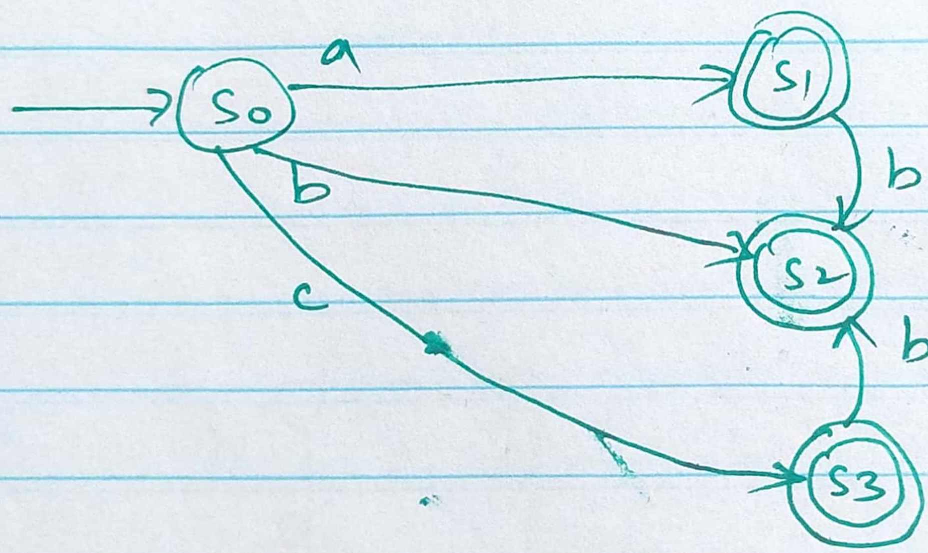
$S_3 = \{6, 7, 8, 9\}$

$(S_3)_a = \{6, 7, 8, 9\}_a = \emptyset$

$(S_3)_b = \{6, 7, 8, 9\}_b = \{10\}$

$(S_3)_b = \{10\} = S_2$

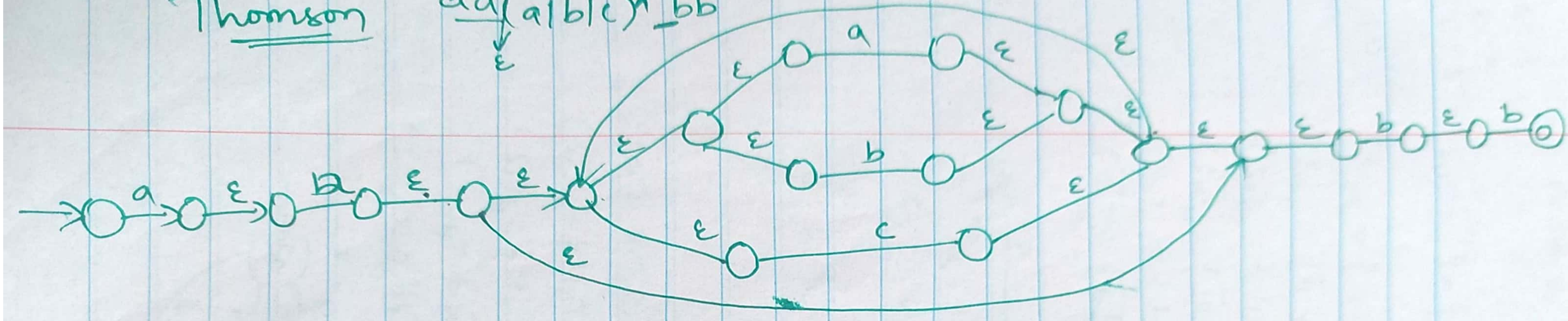
$(S_3)_c = \{6, 7, 8, 9\}_c = \emptyset$



Thomson

$aa(a|b|c)^{*}bb$

ε



Thom (15)

$(bb|a|c)(a|b)^{*}$

$(bb|a|c)(a|b)^*$