Regular expressions over an alphabet ∑

- Algebraic expressions defined recursively
- Each regular expression R denotes a language L(R)

Regular expressions are :

1) ε, φ and a for any a Є ∑ are regular expressions. These are Basic or Elementary Regular expressions. Moreover we have regular expressions defined recursively.
2) If R is a R.E. then (R) is also a R.E. denoting L((R)) = L(R).
3) If R is a R.E. R* is a R.E. denoting (L(R))*.
4) If R1, R2 are R.E. then R1 + R2 is also a R.E. denoting L(R1) U L(R2).
5) If R1, R2 are R.E. then R1 . R2 or R1 R2 is a R.E. denoting L(R1) L(R2).

Any expression is a R.E. if it is obtained from elementary R.E.'s by a finite number of applications of *,. and +. Prrecedence is *, . followed by + and can be altered by ().

Theorem : Every regular expression denotes a regular language.

Proof : Every elementary R.E. clearly denotes a regular language by simple construction of the corresponding DFA's. By closure properties of regular languages, the theorem follows.

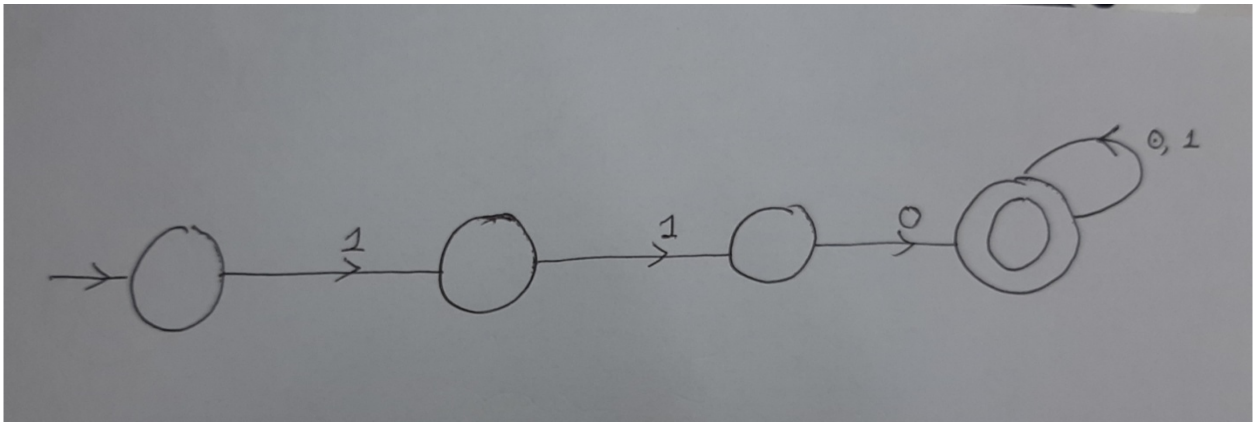
Conversion of R.E. R to ε-NFA

The elementary R.E.'s present in R are first converted and then are combined using prescriptions of combining the respective ε-NFA's.

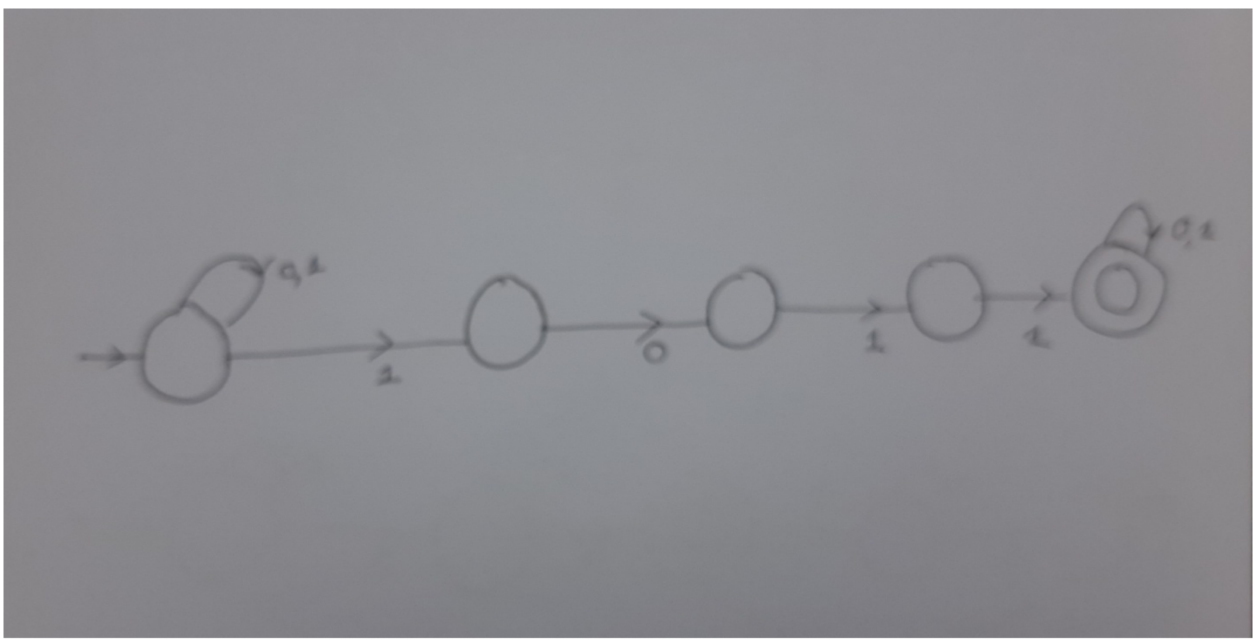
Design R.E. and convert to ε-NFA

1) Set of all strings that begin with 110

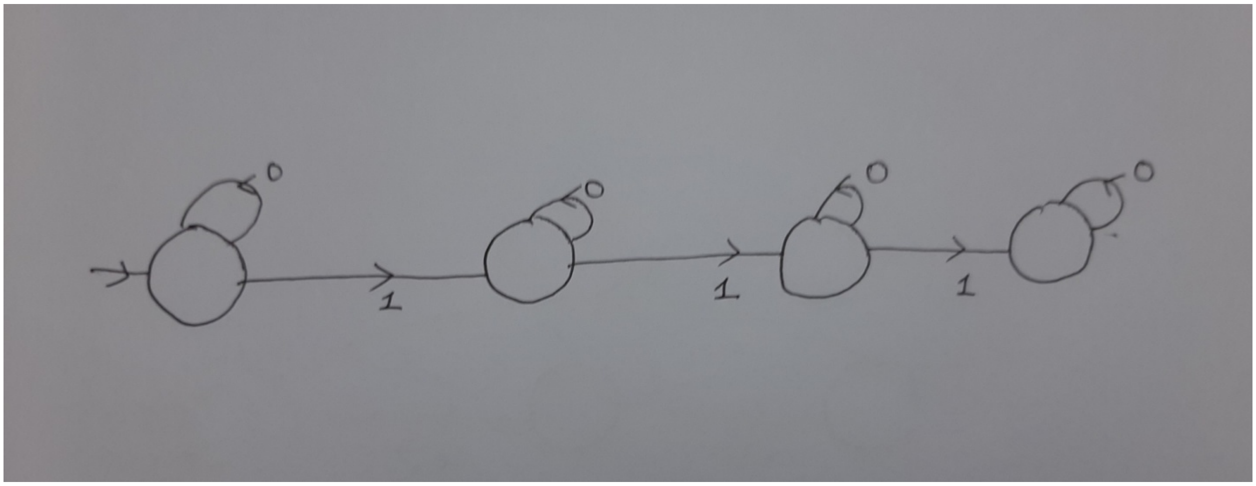
110(0+1)*



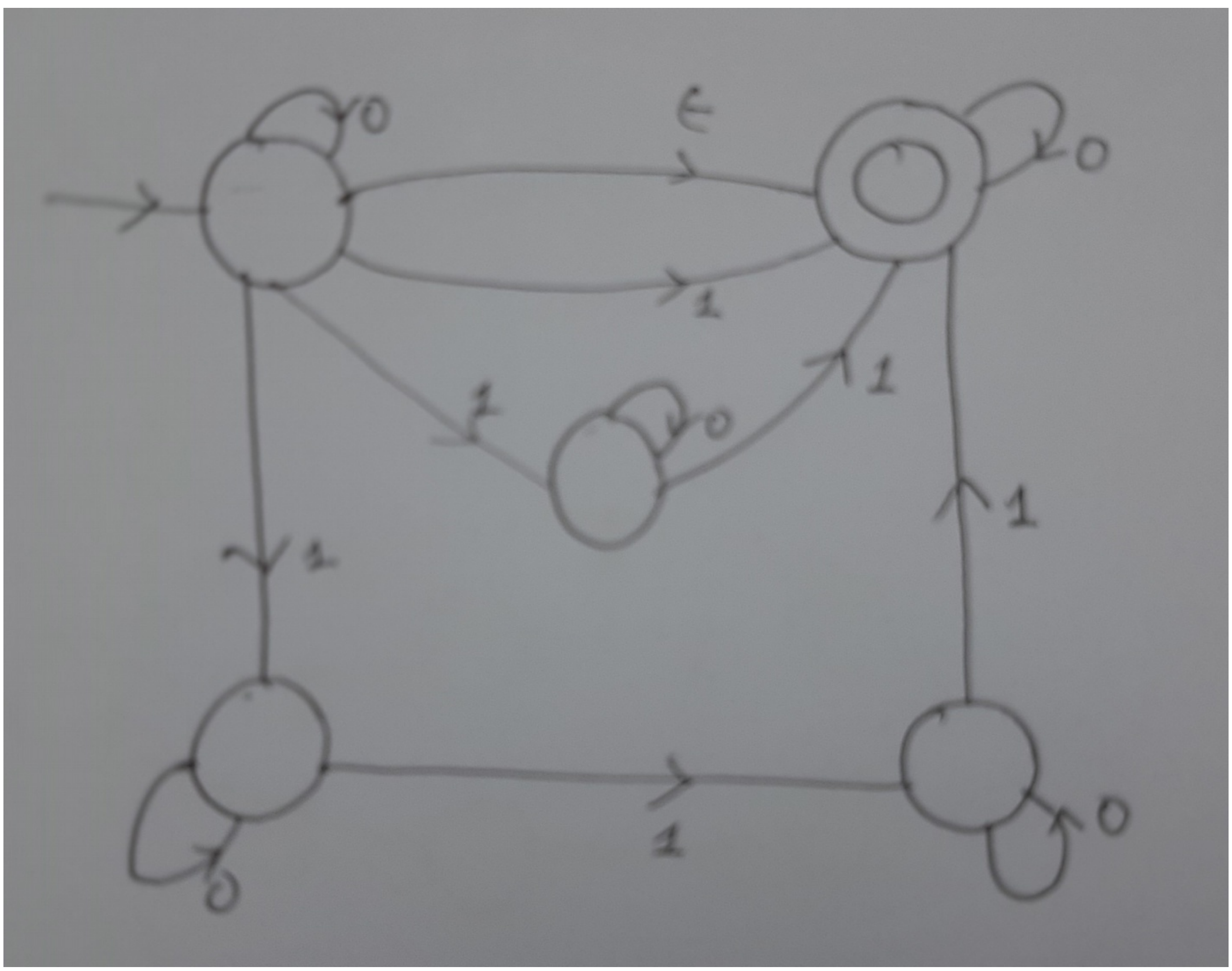
2) Substring 1011 (0+1)*1011(0+1)*



3) Exactly 3 1’s 0*10*10*10*

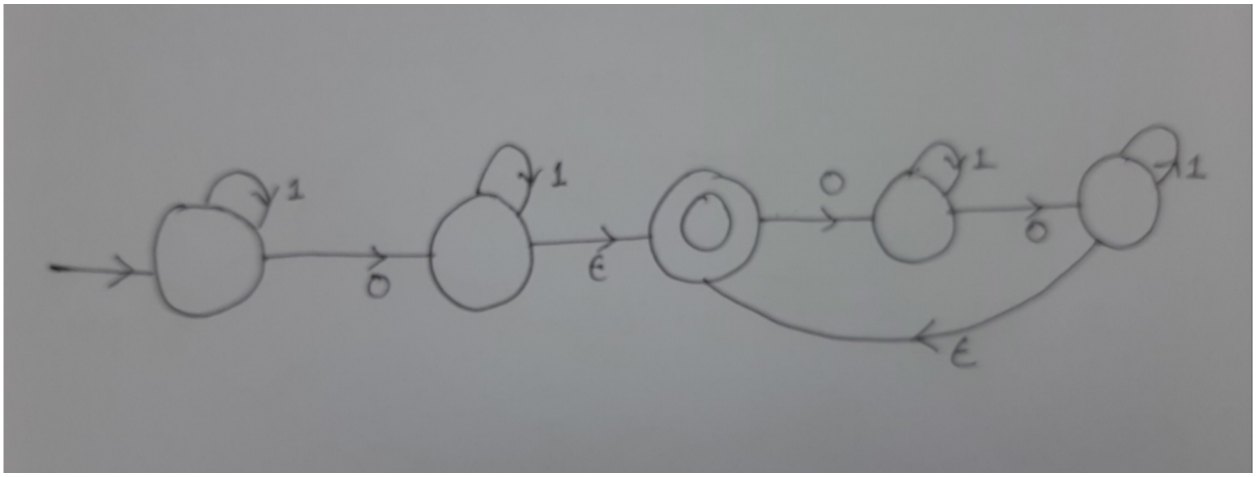


4) At most 3 1's $0^* + 0^*10^*+0^*10^*10^*+$
$0^*10^*10^*10^* = 0^*(\varepsilon+1+10^*1+10^*10^*1)0^*$



5) At least 3 1's HW

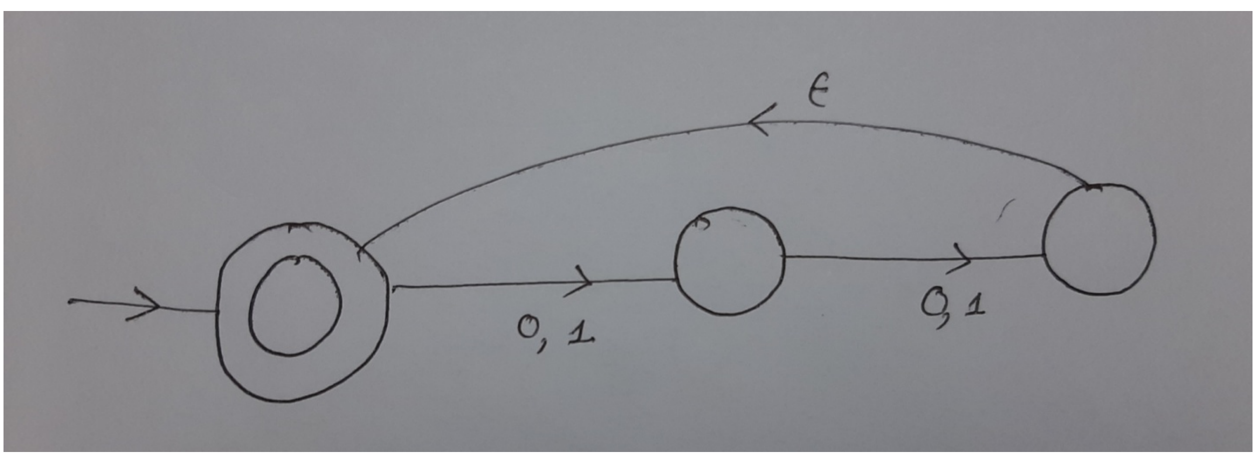
6) Number of 0's odd 1*01*(01*01*)*



7) Number of 0's 1 mod 3 HW

8) Start and end with the same symbol
HW

9) Even length ((0+1)(0+1))*



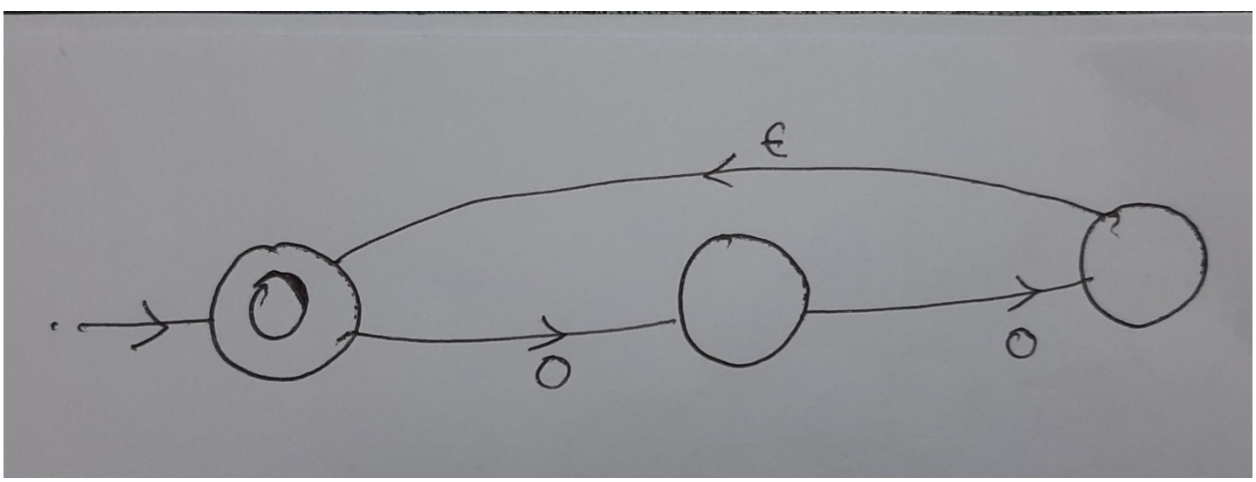
10) Odd length HW

Describe the language :

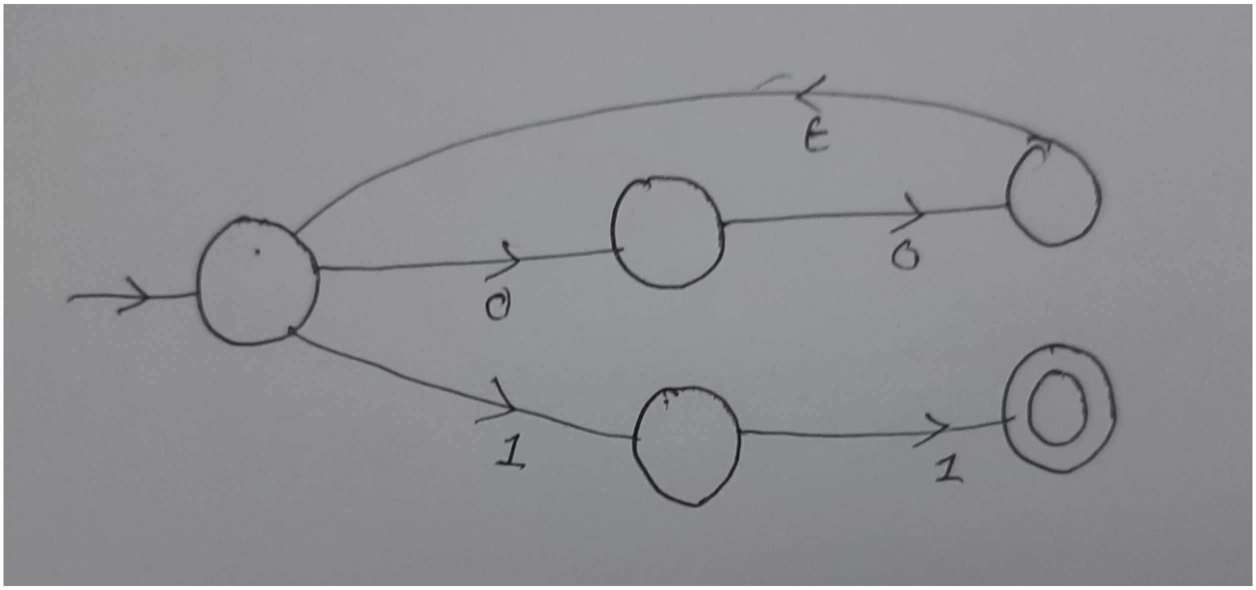
1) 1*(011*)*   every 0 followed by at least one 1
2) 1*(011*)*(0+ε)   HW
3) (0+1)(0+1)((0+1)(0+1)(0+1))*   HW
4) (0+10+11)(0+1)* start with 0 or start with 1 and length at least 2.

Convert to ε-NFA

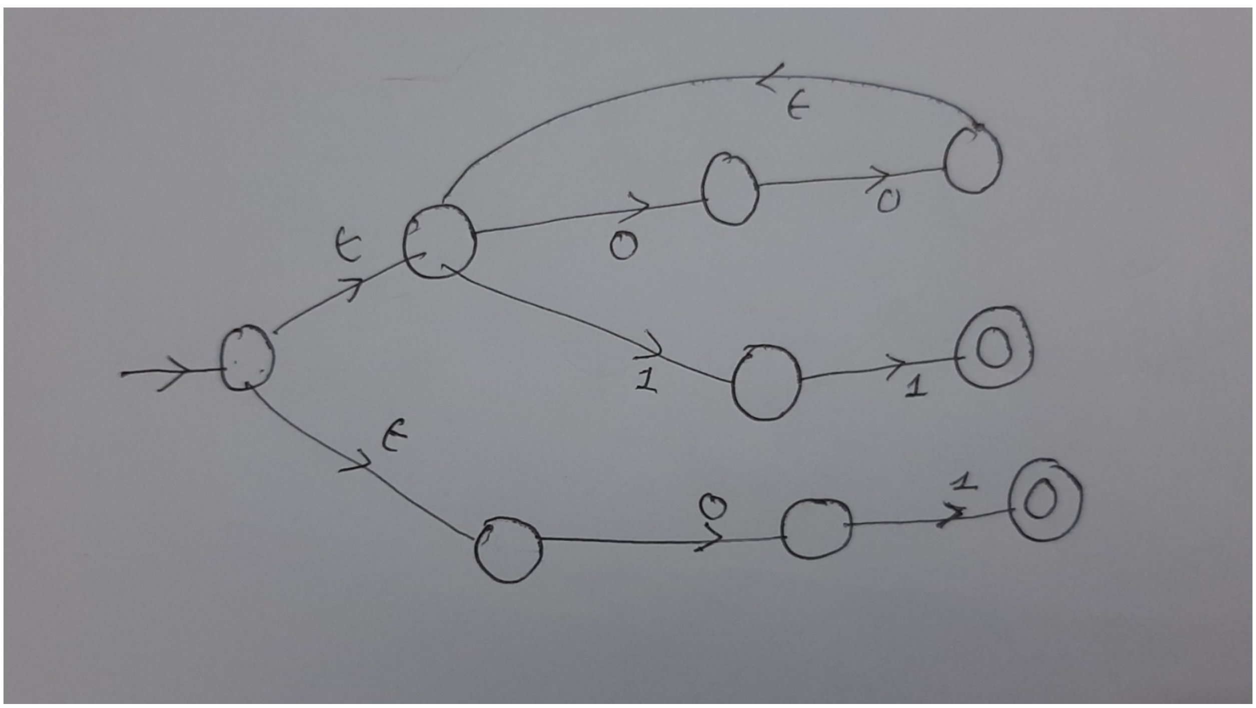
1) ((00)*11+01)*



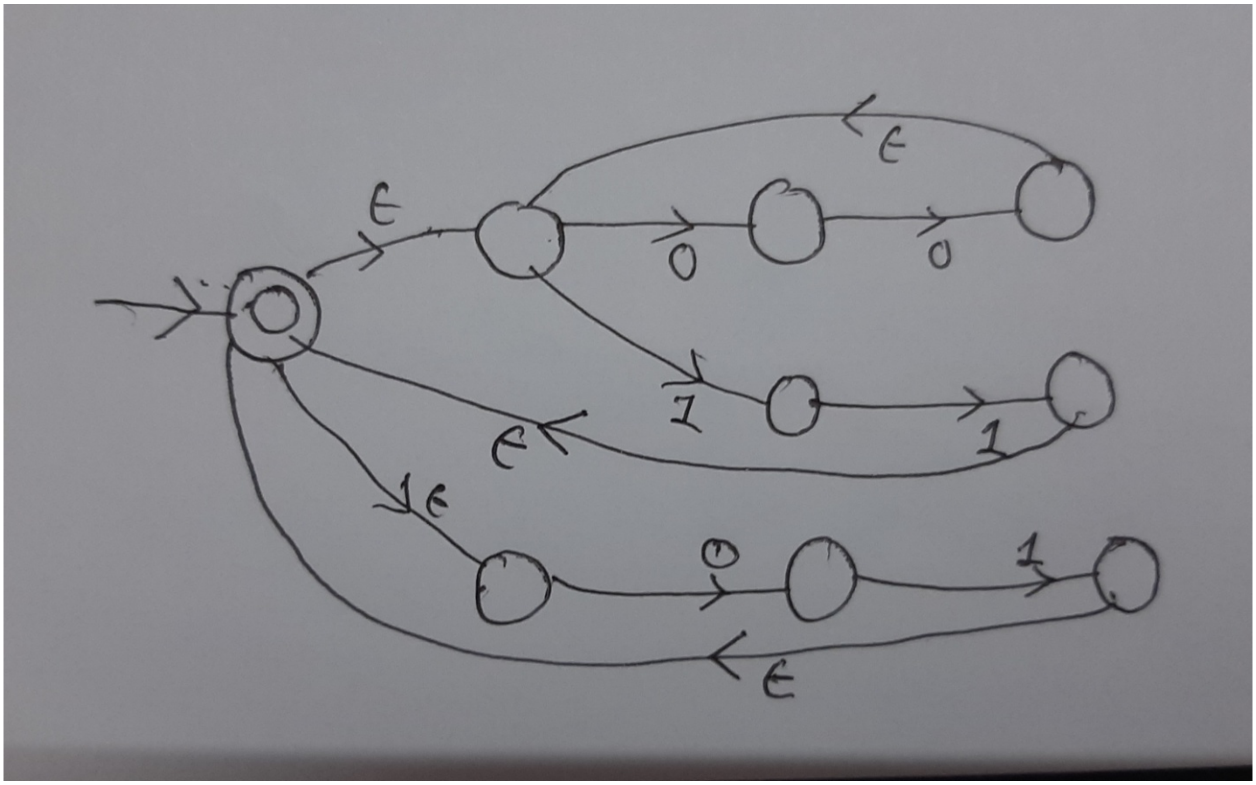
(00)*



(00)*11

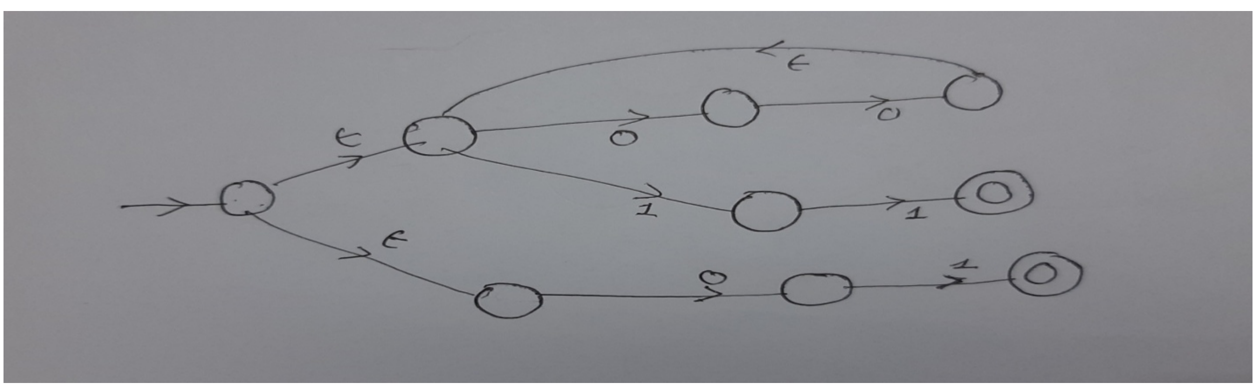


(00)*11+01

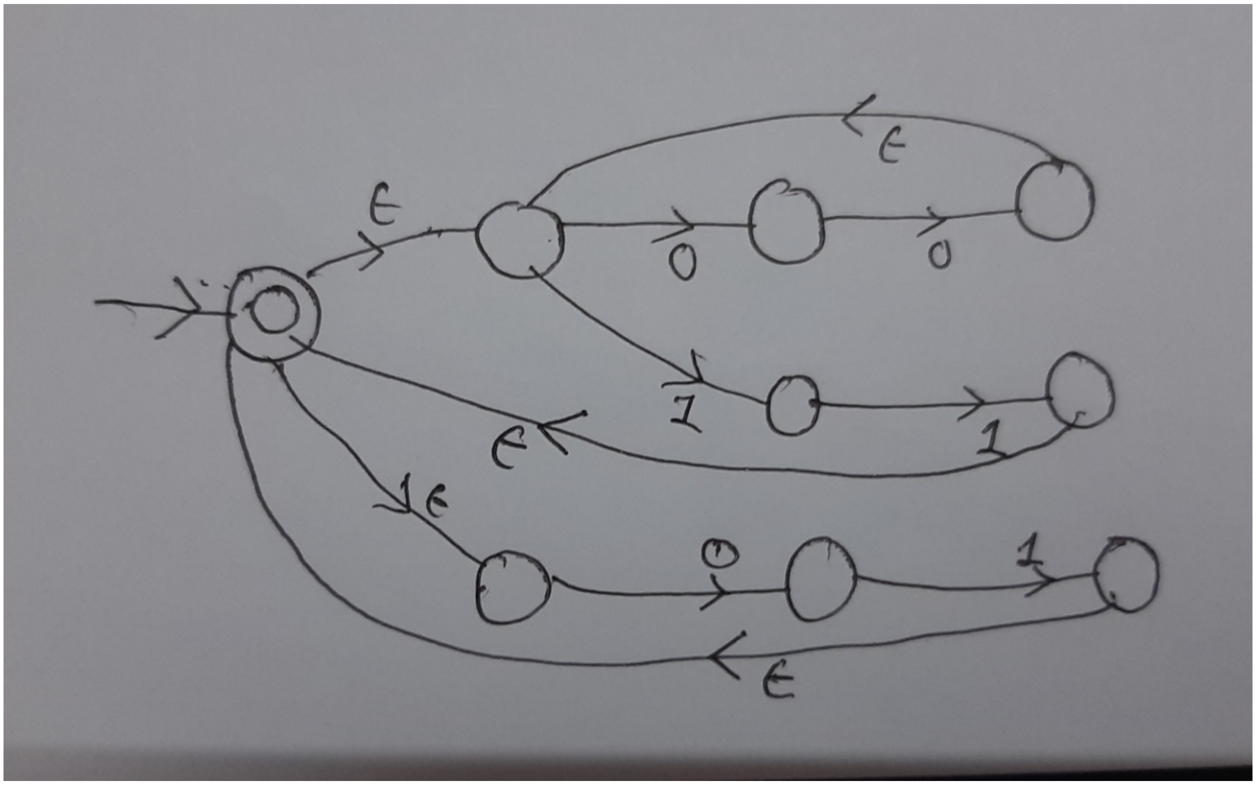


((00)*11+01)*

One can cut down the steps. Thus

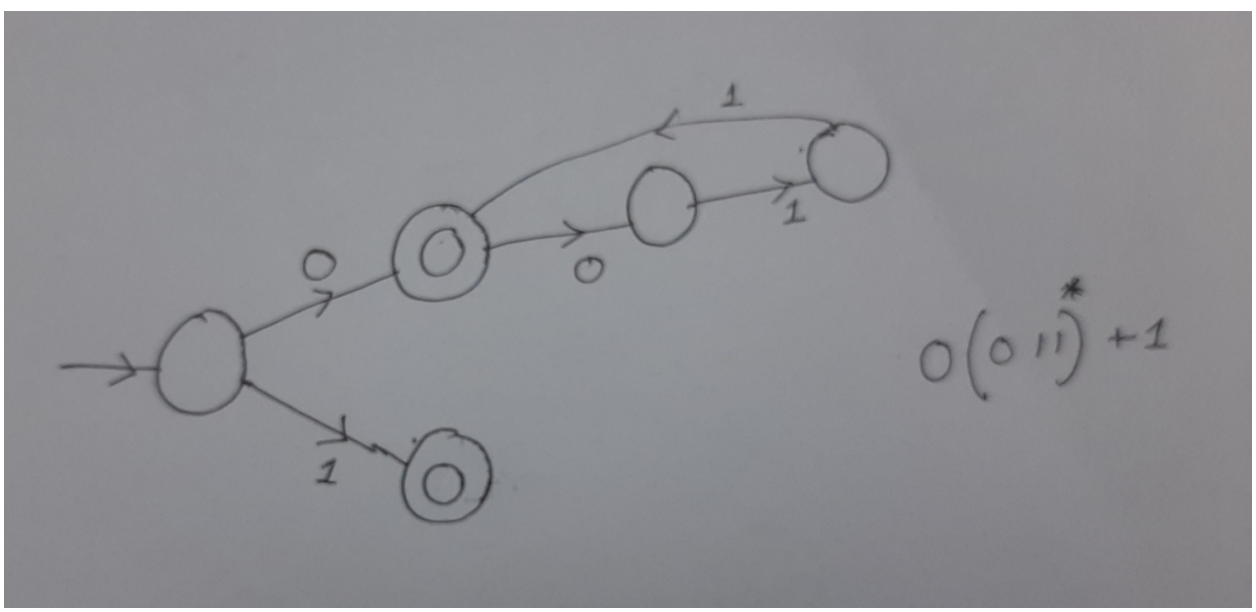


(00)*11+01

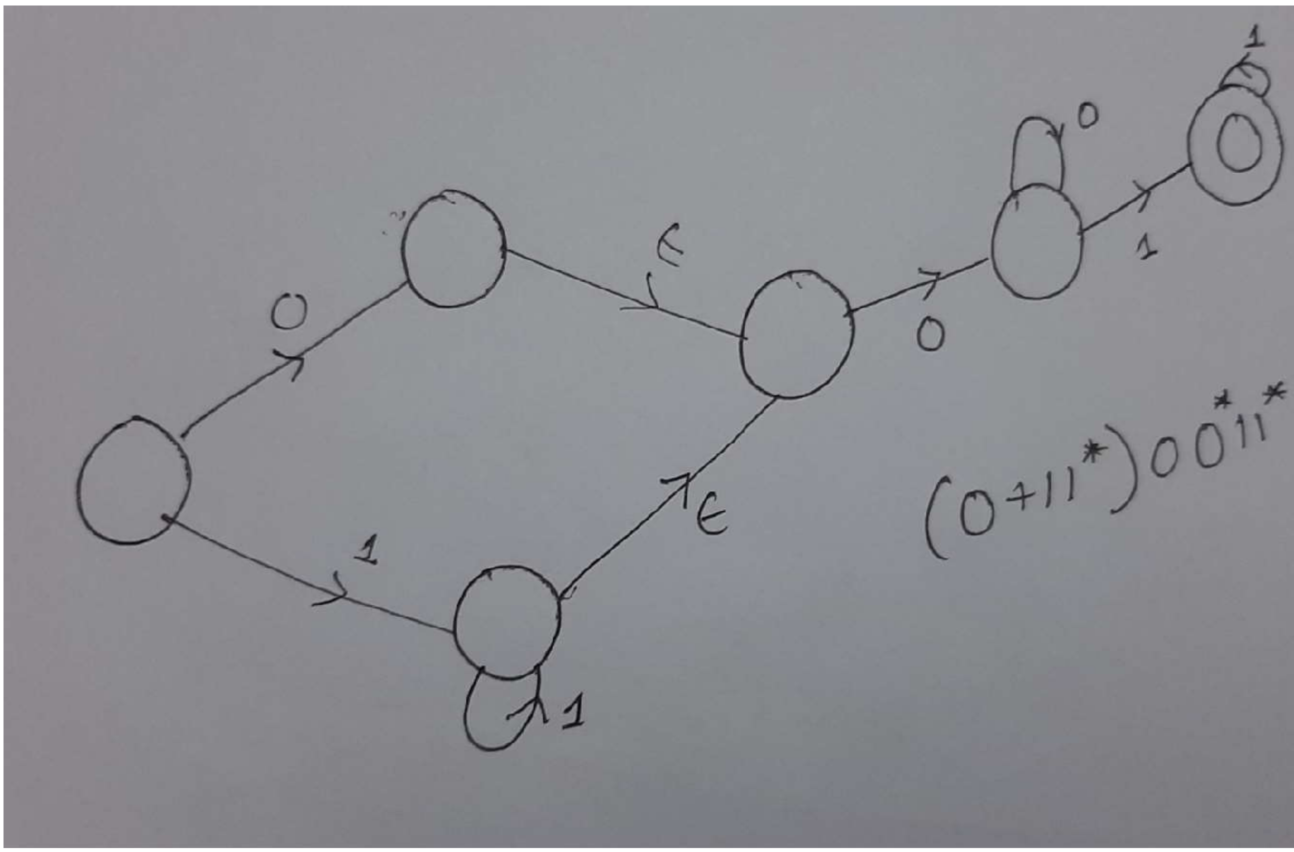


((00)*11+01)*

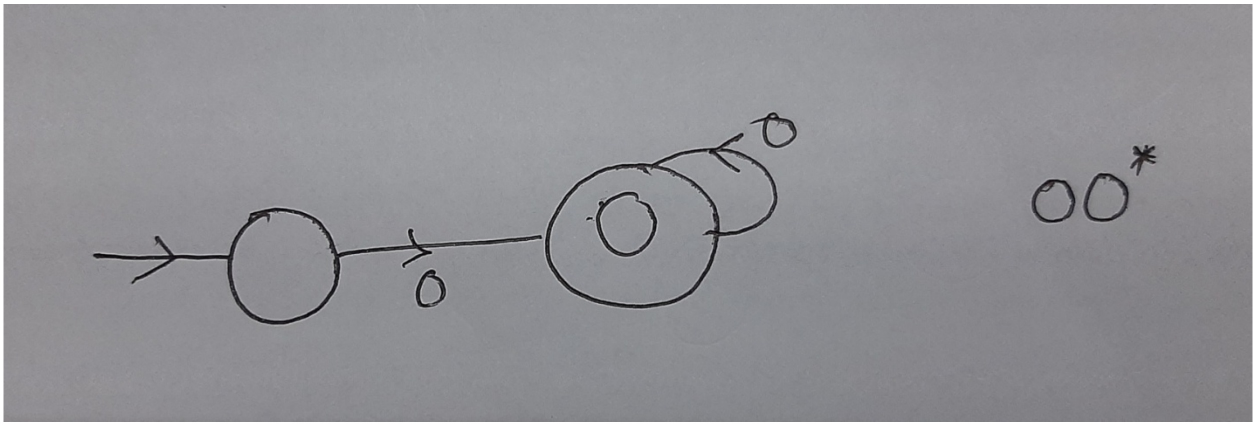
1a) 0(011)* + 1

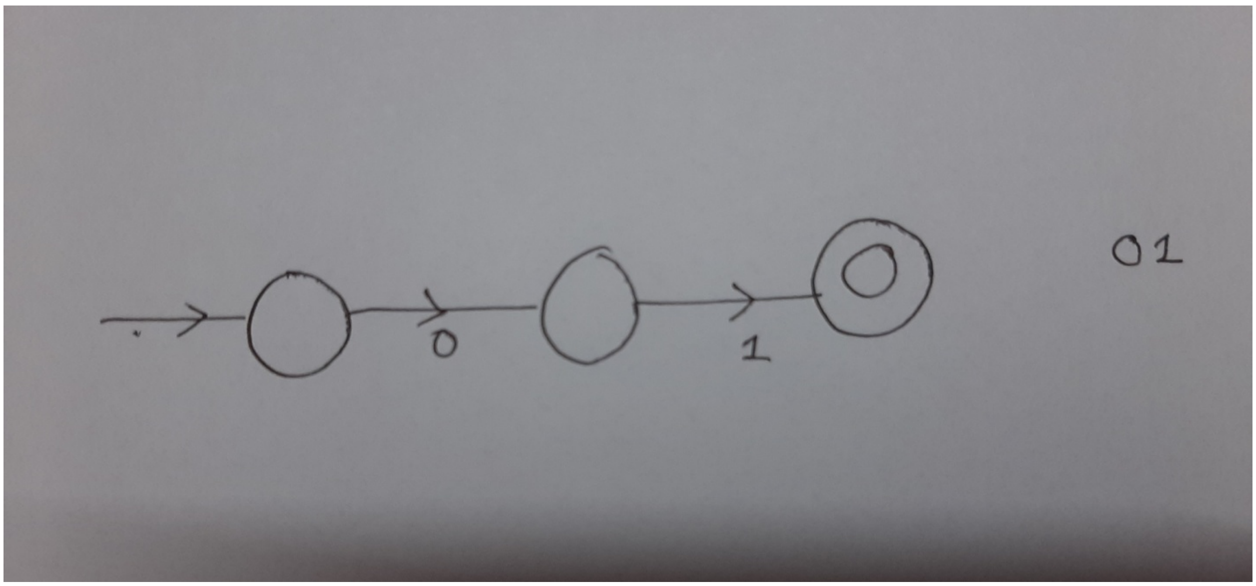


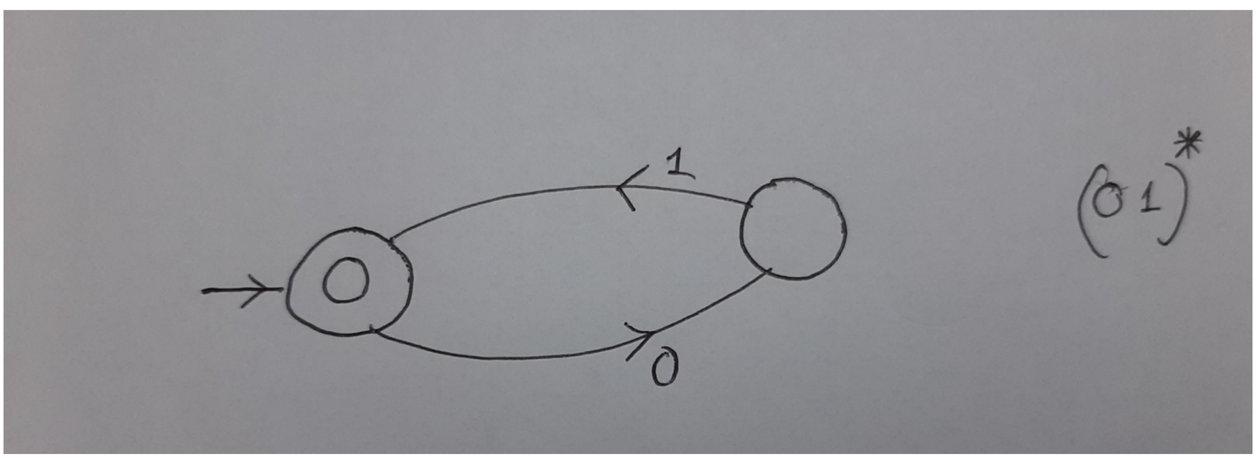
2) (0+11*)00*11*

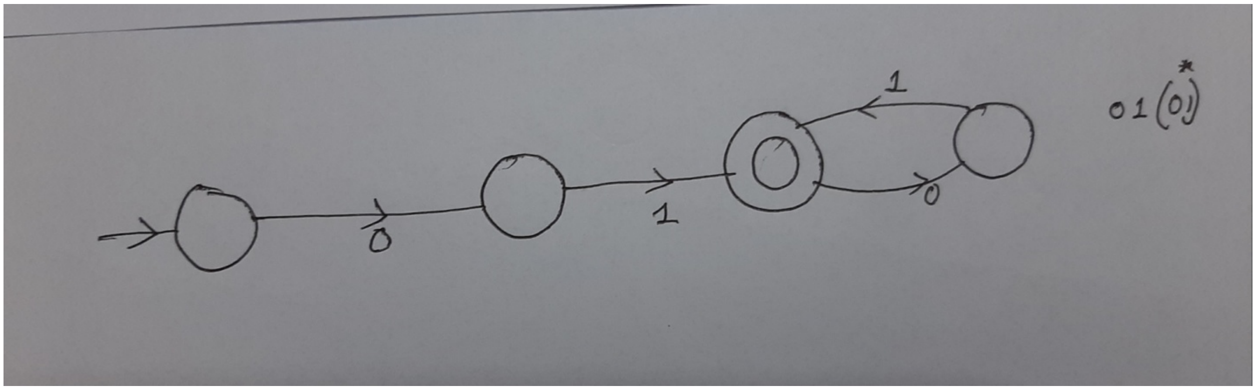


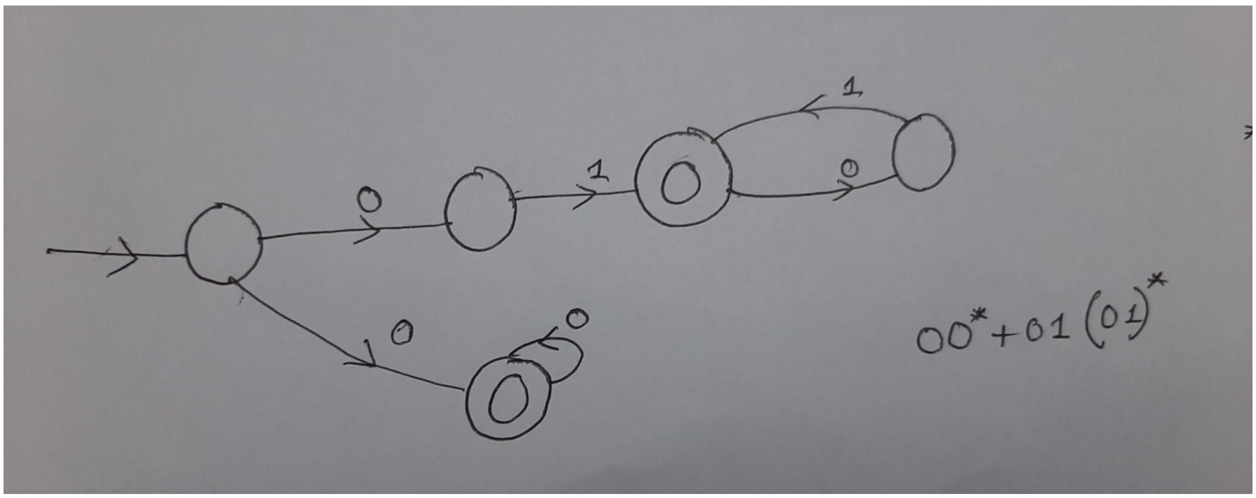
3) 00*+01(01)*











Regular expression identities :

1) $\varphi + R = R + \varphi = R$

2) $\varphi R = R\varphi = \varphi$

3) $\varepsilon R = R\varepsilon = R$

4) $\varepsilon^* = \varepsilon$, $\varphi^* = \varepsilon$

5) $R + R = R$, $R^*R^* = R^*$, $RR^* = R^*R$, $L(RR^*) = R^+$
6) $\varepsilon + RR^* = R^* = \varepsilon + R^*R$
7) $(PQ)^*P = P(QP)^*$
8) $(P+Q)^* = (P^*Q^*)^* = (P^* + Q^*)^*$
9) $P+Q=Q+P$, $(P+Q)R=PR+QR$, $R(P+Q)=RP+RQ$

Arden's Theorem : The equation $R = Q + RP$ has a solution $R = QP^*$ where P,Q, and R are regular expressions. The solution is unique provided ε is not in P.

For the proof of the first part we put $R=QP^*$ in the equation.

$LHS = QP^*$

$RHS = Q + QP^*P = Q(\varepsilon + P^*P) = QP^*$

Hence $R=QP^*$ is a solution.

The proof of the second part is advanced.

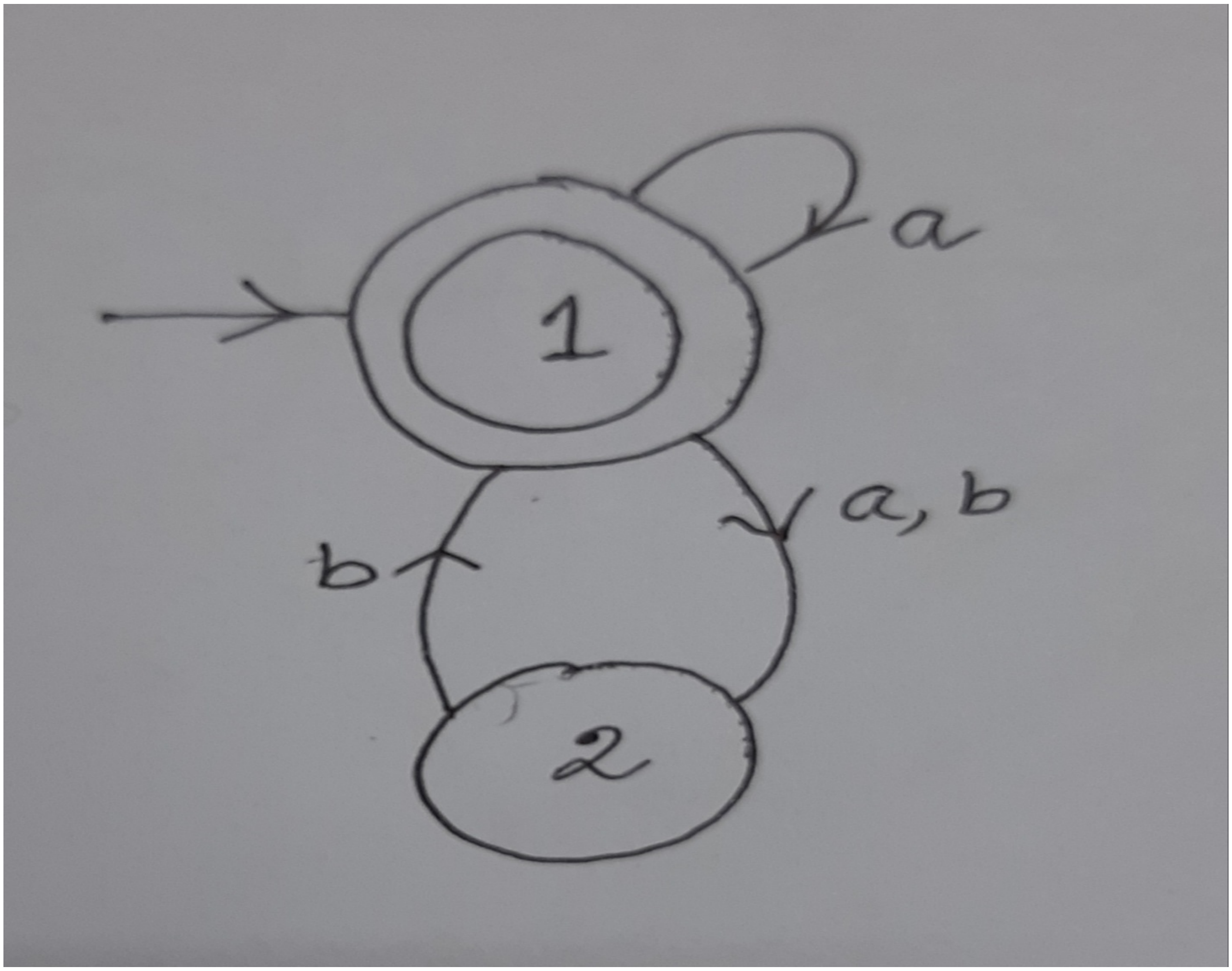
In conversion of an ε-NFA to a regular expression the second part is not needed. So we don't have to check whether ε does not belong to P.

Method of conversion of an ε-NFA to a R.E. :

Let R_i be the R.E. for the set of strings that take the automaton from the start state to the state i. From the transition diagram/table we write down equations for the R_i's and solve them by repeated applications of Arden's Theorem. Finally the required R.E. is obtained as

$$R = \sum_{i \in F} R_i$$

Example 1



$R_1 = \varepsilon + R_1\, a + R_2\, b$ (1)

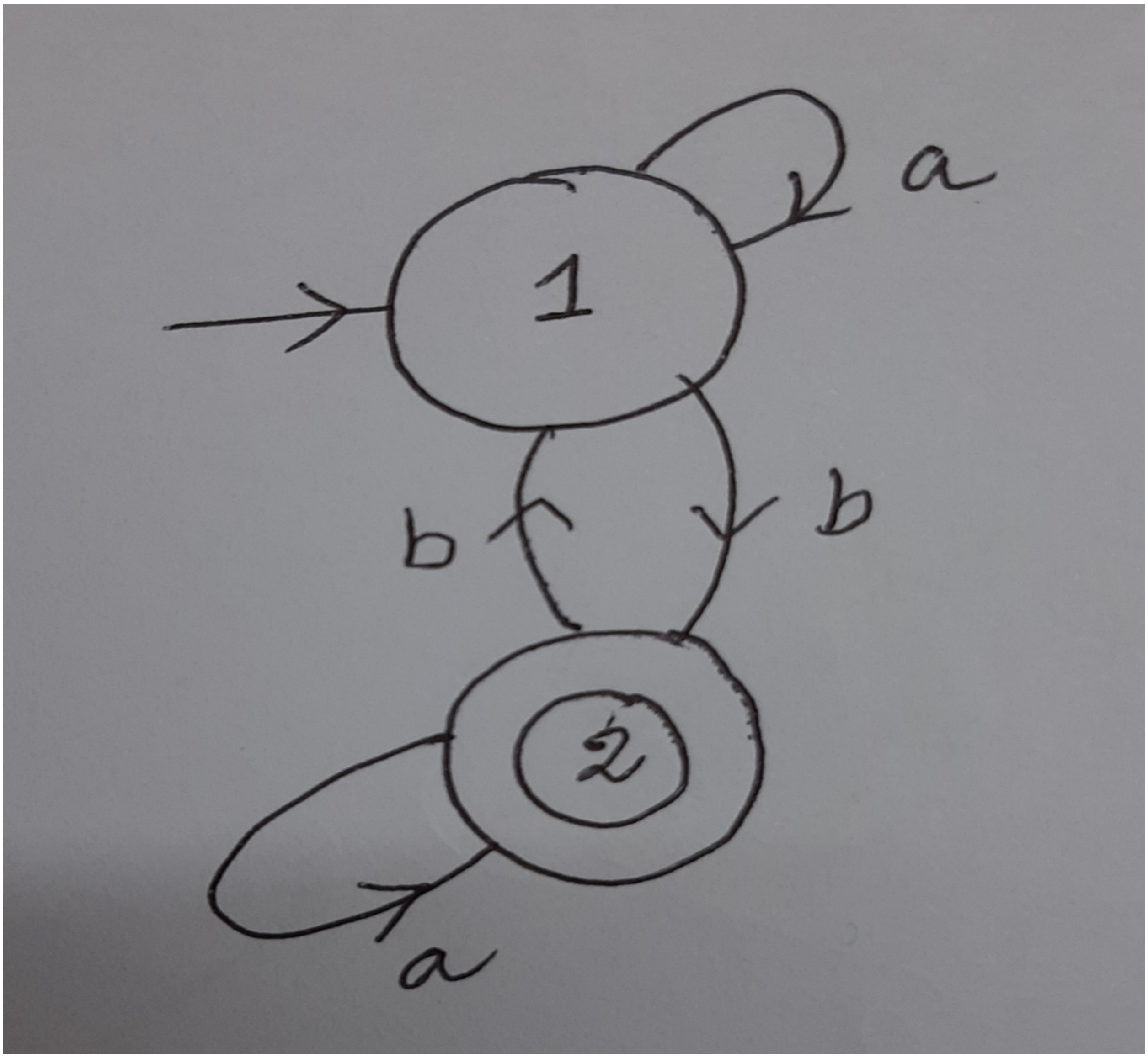
$R_2 = R_1\, (a + b)$ (2)

Putting (2) in (1)

$R_1 = \varepsilon + R_1\, a + R_1\, (a + b)\, b = \varepsilon + R_1(a+(a+b)b)$

Using Arden's Theorem $R_1=(a+(a+b)b)^*$ which is the required regular expression.

Example 2



$R_1 = \varepsilon + R_1\,a + R_2\,b$ (1)

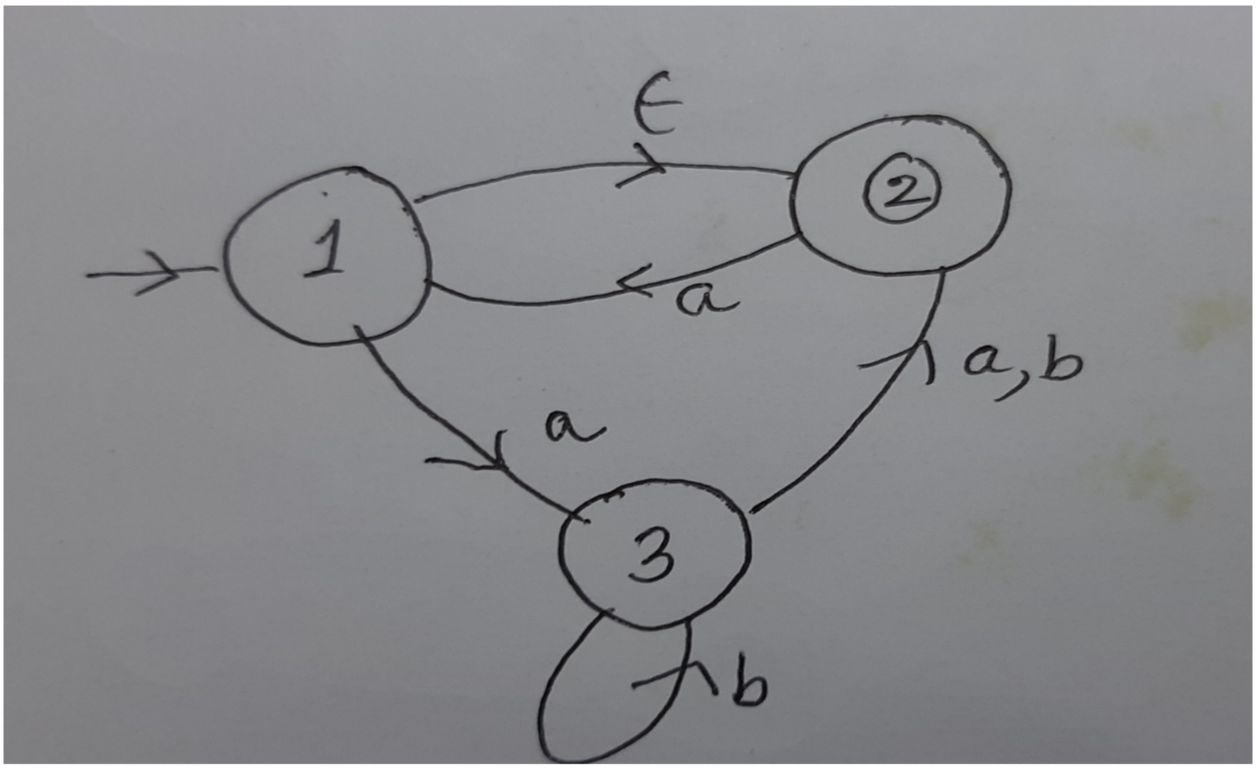
$R_2 = R_1\,b + R_2\,a$ (2)

Using Arden's Theorem in (1) $R_1=(\varepsilon+R_2b)a^*$ (3)

Putting this in (2) $R_2=(\varepsilon+R_2b)a^*b+R_2a= a^*b+R_2(ba^*b+a)$. Using Arden's Theorem

$R_2=a^*b+(ba^*b+a)^*$ which is the required R.E.

Example 3



$R_1 = \varepsilon + R_2\, a$ (1)

$R_2 = R_1 + R_3\, (a + b)$ (2)

$R_3 = R_1\, a + R_3\, b$ (3)

From (3) using Arden's Theorem and using (1)

$R_3 = R_1 \, a \, b^* = (\varepsilon + R_2 \, a) \, a \, b^*$ (4)

Putting (4) in (2) and using (1)

$R_2=\varepsilon+R_2a+(\varepsilon+R_2a)ab^*(a+b)=\varepsilon+ab^*(a+b)+R_2(a+aab^*(a+b))$. Using Arden's Theorem we get the required R.E. $R_2=(\varepsilon+ab^*(a+b))(a(\varepsilon+ab^*(a+b)))^*$

HW