Assignment 5

ADITEYA BARAL [ab12057]

Q1 (a) Units $= 2(n-m) - [\phi_n - \phi_m]$

$= 2(99-0) - (4-0)$ $\quad [\phi_n = 4$ since $99 = (1100011)_2]$

$\Rightarrow \boxed{\text{Work units} = 194}$

(b) Units $= 2(n-0) - [\phi_n - 0]$

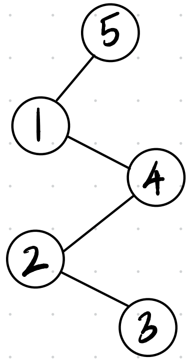
$\Rightarrow \boxed{\text{Units} = 2n - \phi_n}$ $\quad [\phi_n \rightarrow$ number of 1's in binary representation of n$]$

(c) $\boxed{\text{Units} = 2(n-m) - [\phi_n - \phi_m]}$

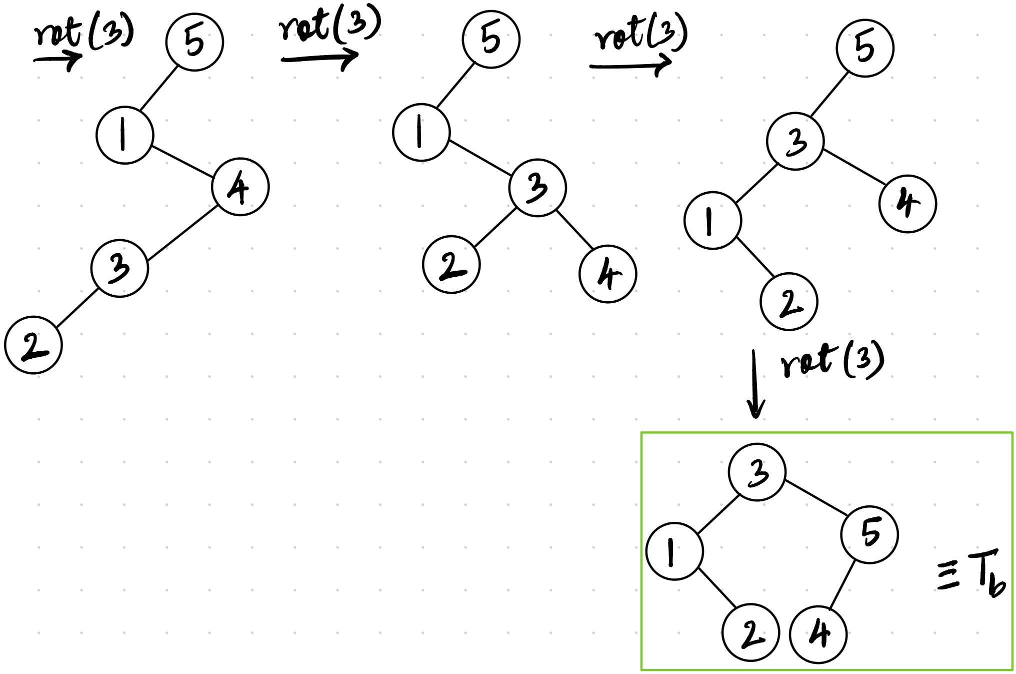
(d) Cost $= 2(200-100) - [3-3]$

$\Rightarrow \boxed{\text{Cost} = 200}$

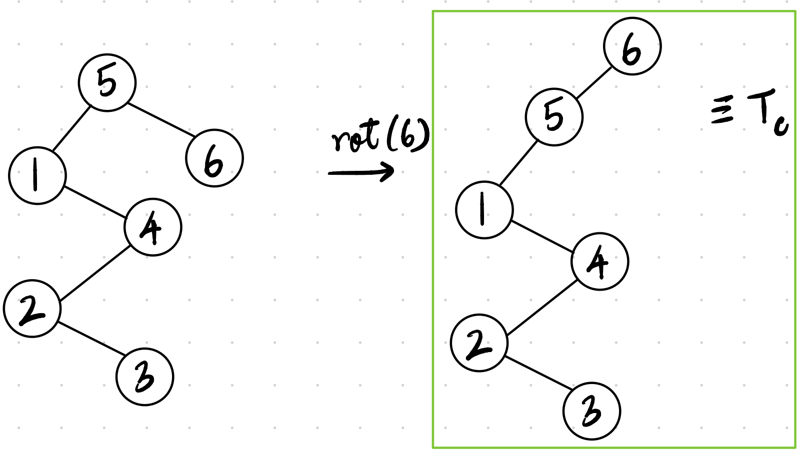
Q2 (a) $T_a \equiv$



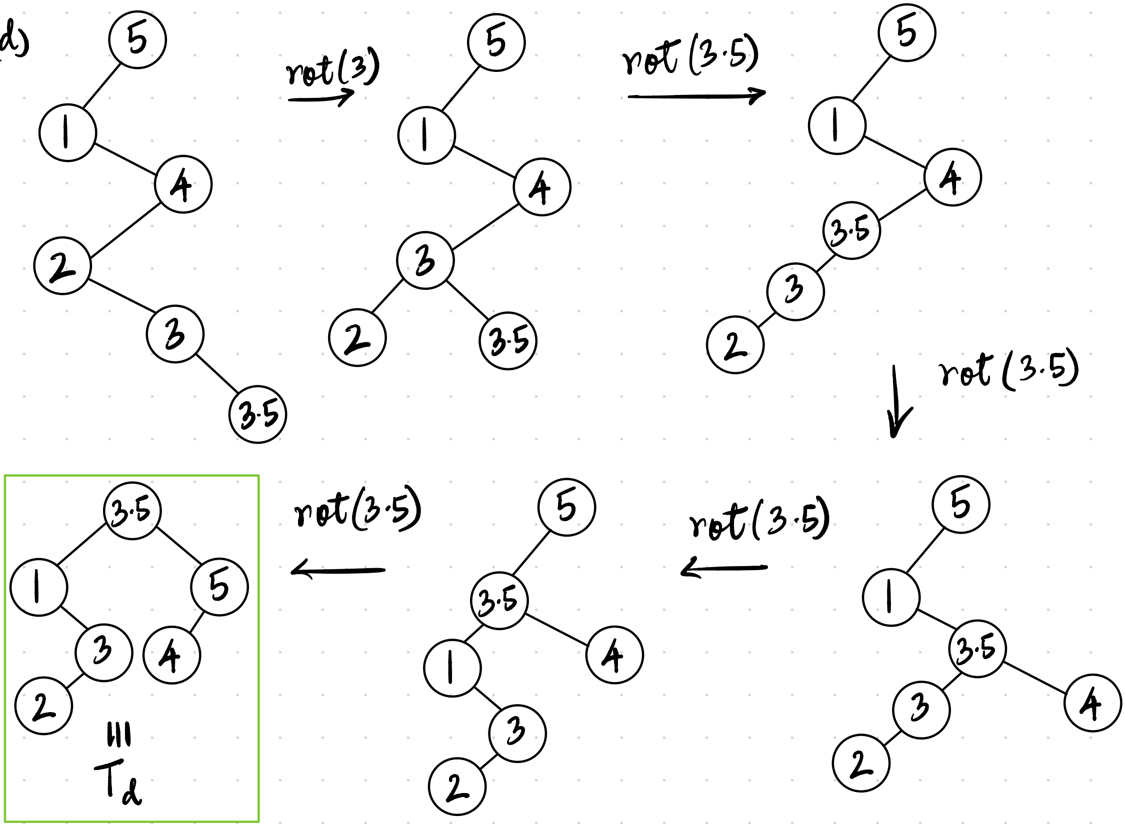
(b) T_a



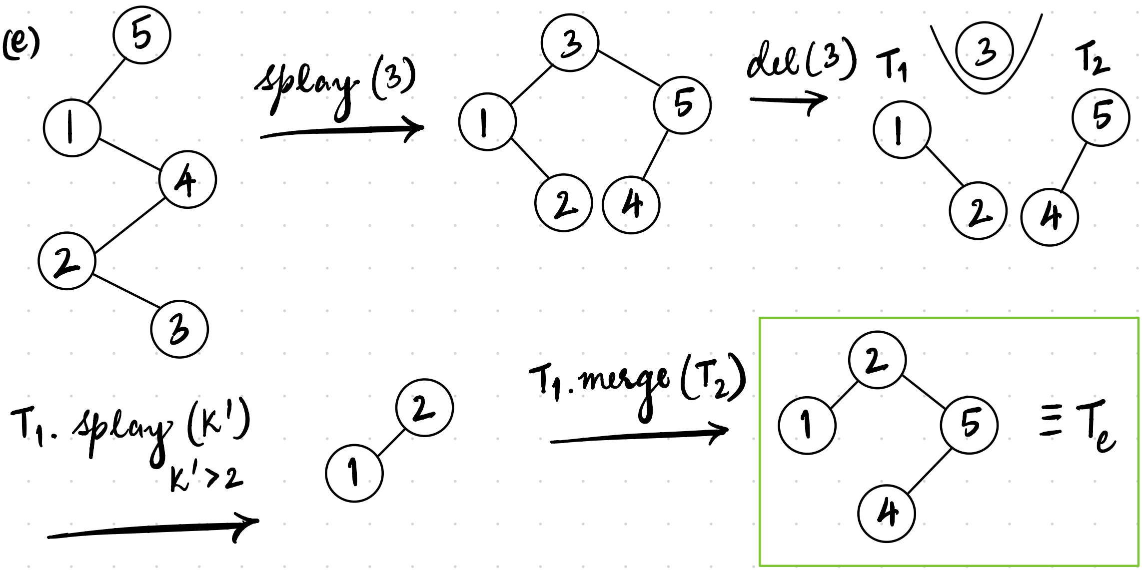
(c)



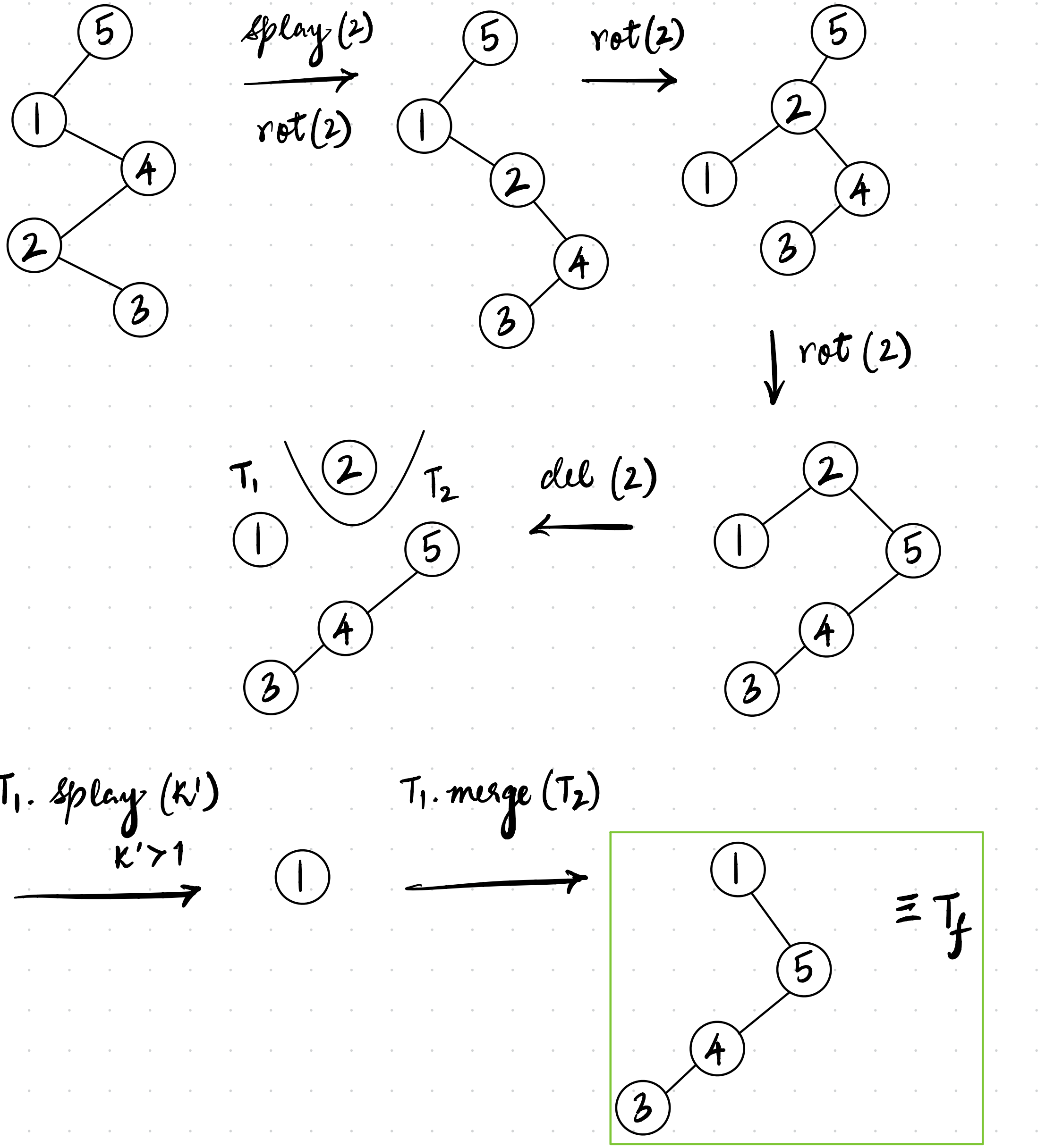
(d)


rot(3)
rot(3.5)
rot(3.5)
rot(3.5)
rot(3.5)
$\equiv T_d$

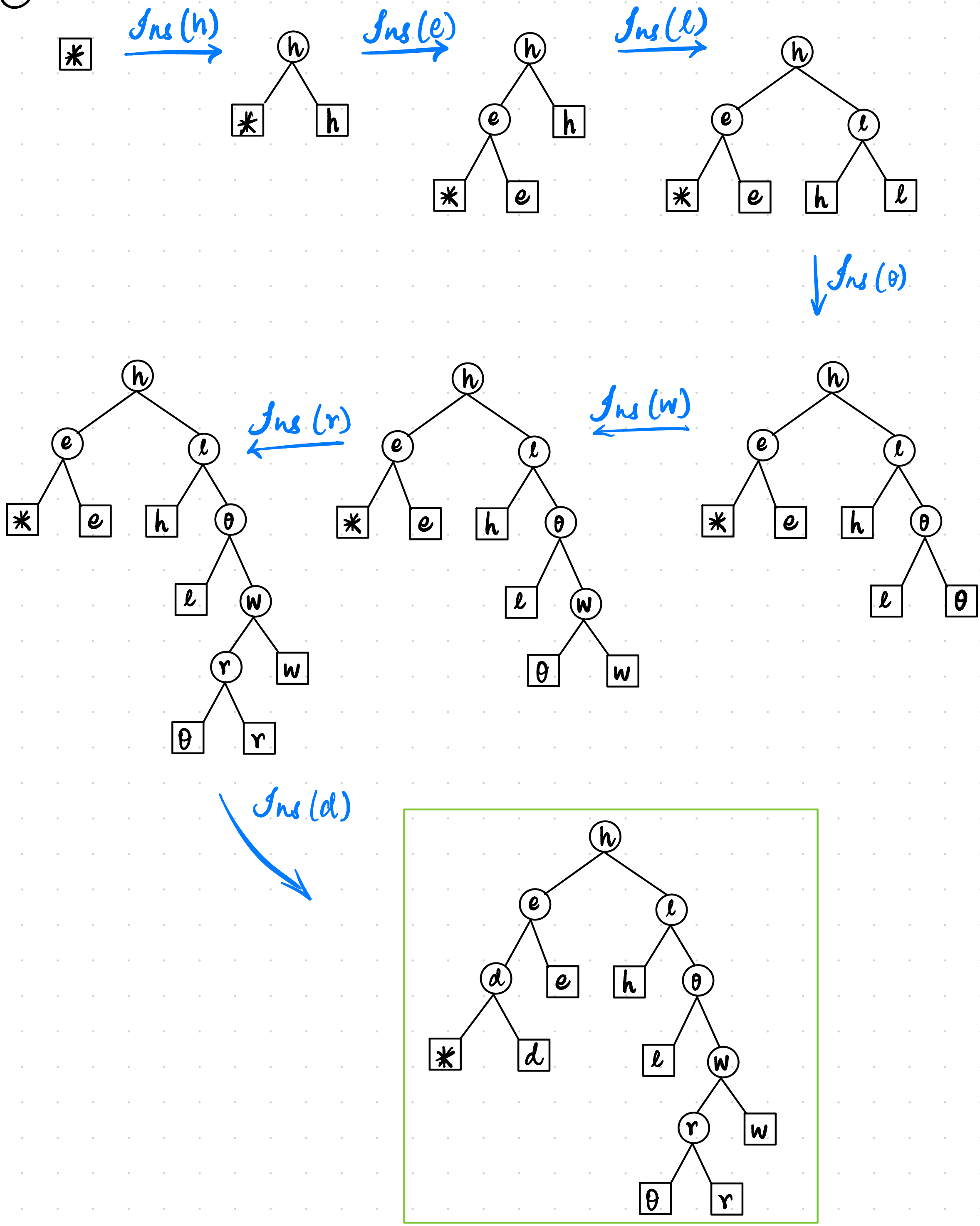
(e)


splay(3)
del(3)
T_1
T_2
T_1.splay(K')
K' > 2
T_1.merge(T_2)
$\equiv T_e$

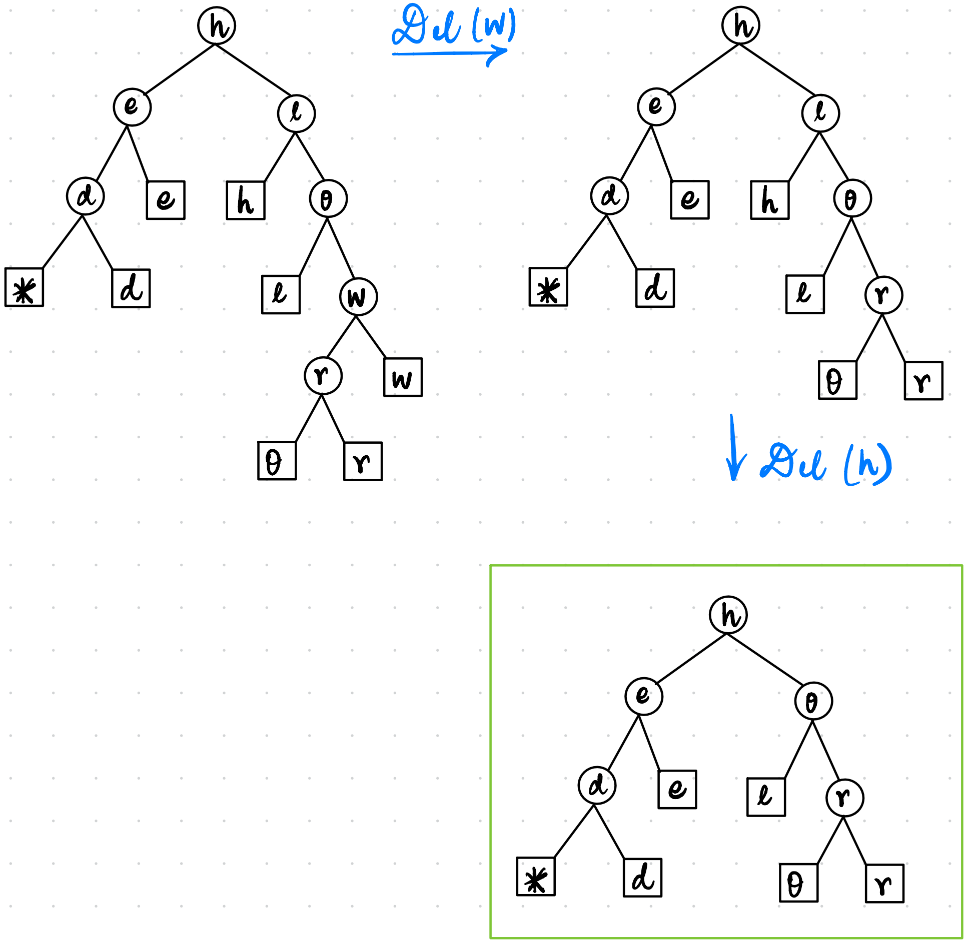
(f)



Q3 (a)



(b)



Q4

(a)

i=0	i=1 [a]	i=2 [b]	i=3 [c]	i=4 [b]

i=0

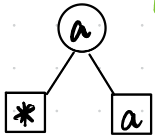
$\boxed{*}$

i=1 [a]

b → $\boxed{*}$

q → -

emit → ASC (a) [8]



p → -

8 bits

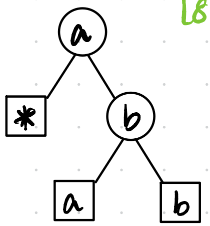
i=2 [b]

b → $\boxed{a}$

emit → $C_{i-1}(*) = 0$ [1]

q → a

emit → ASC(b) [8]



p → a

splay (p) → no effect

1+8 bits

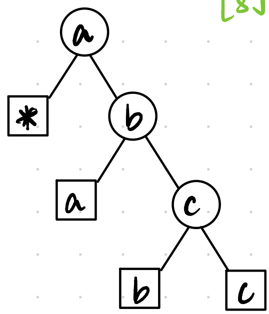
i=3 [c]

b → $\boxed{b}$

emit → $C_{i-1}(*) = 0$ [1]

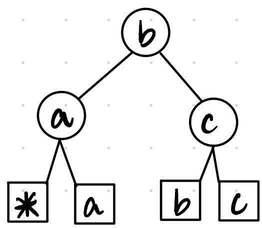
q → a

emit → ASC (c) [8]

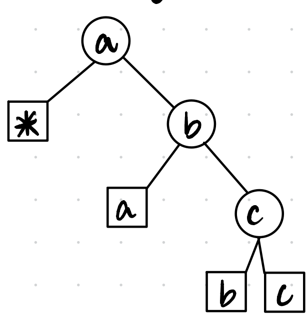


p → b

splay (p)



splay (q)



1+8 bits

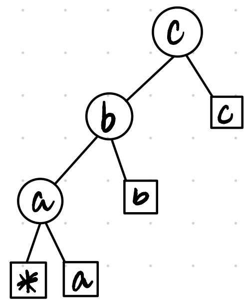
i=4 [b]

b → $\boxed{b}$

β → 110 [3]

p → c

splay (p)



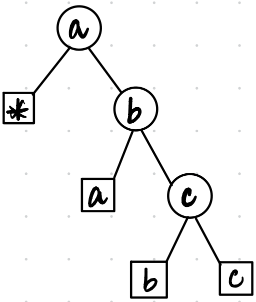
3 bits

Total bits = $8 + (1+8) + (1+8) + 3 + 3$
$= 32$ bits

$i=5$ [a]

$b \to \boxed{a}$
$\beta \to 001$ [3]
$p \to$ (a)

splay(p)

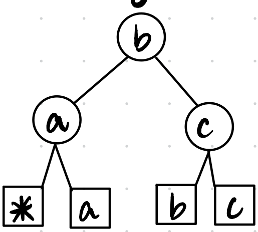


3 bits

(b)

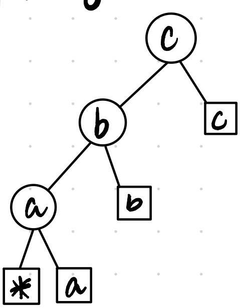
$i=6$ [a]

$b \to \boxed{a}$
$p \to$ (b)
emit → 10
splay(p)



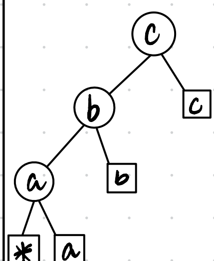
$i=7$ [b]

$b \to \boxed{b}$
$p \to$ (c)
emit → 10
splay(p)



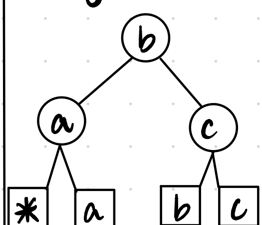
$i=8$ [c]

$b \to \boxed{c}$
$p \to$ (c)
emit → 1
splay(p)



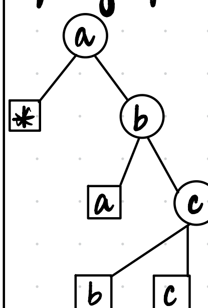
$i=9$ [b]

$b \to \boxed{b}$
$p \to$ (b)
emit → 01
splay(p)



$i=10$ [a]

$b \to \boxed{a}$
$p \to$ (a)
emit → 01
splay(p)



For $m \geq 2$, we see that a total of $2+2+1+2+2=9$ bits are emitted for each m. $\Rightarrow$ 32 + 9 bits for each of $(m-1)$ repetitions

$$\Rightarrow \boxed{\text{Total bits} = 32 + 9(m-1)}$$

(c) Total bits without splay compression

$= 5$ characters $\times$ 8 bits per character $\times$ m repeats
$= 40m$

$$\text{Bits saved} = 40m - (32 + 9(m-1))$$

$$= 40m - 32 - 9m + 9$$

$$\boxed{\Rightarrow \text{Bits saved} = 31m - 23}$$

Q5 (a)

X ↓ \ Y →		T	T	C	A	C	G	C	A
		0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	1	1	1
A	0	0	0	0	1	1	1	1	2
C	0	0	0	1	1	2	2	2	2
T	0	1	1	1	1	2	2	2	2
C	0	1	1	2	2	2	2	3	3
G	0	1	1	2	2	2	3	3	3
A	0	1	1	2	3	3	3	3	4
A	0	1	1	2	3	3	3	3	4

$\boxed{L(X,Y) = 4}$

(b) Shown in (a).

○ : $X[i] = Y[j]$

→ : Reachable nodes from (m,n)

→ : Extra edges of $G^+(X,Y)$

→ : Non-reachable nodes from (m,n)

$G_1(X,Y)$ has 3 kinds of edges:

Diagonal : $(i,j) - (i-1,j-1) \rightarrow X[i] = Y[j]$

Horizontal : $(i,j) - (i-1,j) \rightarrow M[i,j] = M[i-1,j]$

Vertical : $(i,j) - (i,j-1) \rightarrow M[i,j] = M[i,j-1]$

For a unique path $p = (u_1, u_2, \ldots, u_\ell)$ in $G_1(X,Y)$, we define a string representation of p if there are k diagonal edges on this path and any edge $u_{i_j} - u_{i_{j+1}}$ for $j = 1 \ldots k$ is a diagonal edge for $1 \leq i_1 < i_2 < \ell$ and the string(p) is given by $a_1 a_2 \ldots a_k$ where a_i is a matched letter $X[i] = Y[j]$.

If u_ℓ is a sink, then we also have a diagonal edge $u_\ell - u'$ where $u' \notin V_1(X,Y)$. We can add u' to V_1 and $u_\ell - u'$ to $G_1(X,Y)$ we obtain augmented $G_1^+(X,Y)$ and the matched letter $u - u'$ can be appended to string(p). to form string$^+(p)$. Reversing this gives us an element of the set $LCS(X,Y)$

(c)

$$\boxed{LCS(X,Y) = \{TCAA, TCGA, CCGA, ACGA, ACCA\}}$$

Q6

		Y	A	H	O	O
	0	3	6	9	12	15
G	3	1	4	7	10	13
O	6	4	2	5	7	10
O	9	7	5	4	5	7
G	12	10	8	6	6	7
L	15	13	11	9	8	8
E	18	16	14	12	10	9

Total paths = 2

Alignment cost $\Delta = 1+1+3+1+2+1 = 9$

Optimal Alignments:

G	O	O	G	L	E
Y	–	A	H	O	O

(or)

G	O	O	G	L	E
Y	A	–	H	O	O

Q7 $W = a_i^2 + a_j^2 + a_k^2$

(a) $a = \{4, 1, 3, 2, 2, 1\}$

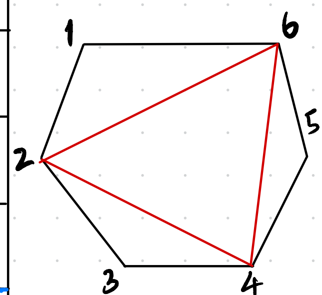
	1	2	3	4	5	6
1	16	0	26	35	44	47
2		1	0	14	23	29
3			9	0	17	23
4				4	0	9
5					4	0
6						1

$\boxed{\Rightarrow \text{Optimal Cost} = 47}$

(b)

	1	2	3	4	5	6
1	1			2	2	2
2		2			4	4
3			3			4
4				4		
5					5	
6						6

$K[1,6] = 2$
$K[2,6] = 4$



Q8 (a) Assume there exists two distinct intervals such that $I < J_1$ and $I < J_2$.

$\Rightarrow S(J_1) < f(I) < f(J_1)$ and $S(J_2) < f(I) < f(J_2)$

Assume a point $x = f(I) - \varepsilon \Rightarrow$ a point in I just before I ends. Since $x > s(J_1)$ and $x > s(J_2)$, $\Rightarrow$ intervals I, J_1 and J_2 all contain x. Thus, cover#$(A, x) = 3$, which contradicts the original assumption. Therefore, for cover#A ≤ 2, there can only be one of either J_1 or J_2 such that $I < J_1$ or $I < J_2$, but not both. $\Rightarrow$

(b) Assume that $I \cap J \neq \phi \Rightarrow$ I and J overlap and $\exists x$ in $I \cap J$ intervals. We know that $I < K \Rightarrow S(K) < f(I) < f(K)$ and $J < K \Rightarrow S(K) < f(J) < f(K)$. If $S(I) > S(K)$ or $s(J) > s(K)$, then we will have a point x that exists in all three intervals I, J and K since $I \cap J \neq \phi$ and $S(K) < f(I)$ and $S(K) < f(J)$. Thus, at some point all three intervals overlap making cover#A $= 3$. This violates the initial condition and establishes that if cover #A ≤ 2, then $I \cap J = \phi$ for $I < K$ and $J < K$. This also occurs if $s(I) \leq s(K)$ and $s(J) \leq S(K)$ at $x = S(K)$. Since all three intervals overlap, the same violation occurs.

(c) If three distinct intervals exist such that $I \cap J \neq \phi \Rightarrow$ there is a point x in the overlap of I and J. Since $I < K$ and $J < K$, we also know that $s(K) < f(I) < f(K)$ and $s(k) < f(J) < f(K)$. Since cover#A ≥ 3, we know that there is another point p such that all three intervals overlap at p. $\Rightarrow p \in I, p \in J, p \in K$. Assume $p = f(I) - \varepsilon$. Then the condition $I < K$ holds. Similarly if $p = f(J) - \varepsilon$, then the condition $J < K$ also holds. Thus if cover #A ≥ 3, the conditions of the statement are satisfied. Thus if I, J, K are the three distinct intervals in A and all three overlap at x, then cover $\#A \geq 3$.

If $I \prec K$, then $s(K) < f(I) < f(K) \Rightarrow K$ overlaps with I. Similarly, if $J \prec K$ then $s(K) < f(J) < f(K) \Rightarrow K$ overlaps with J. Additionally since $I \cap J \neq \phi$, there is a point where all three intervals overlap. $\Rightarrow$ cover $\#A$ is 3 at that point. Thus, since $s(I) \leq x < f(I)$, $s(J) \leq x < f(J)$ and $s(K) \leq x < f(K)$ assuming $f(I) < f(J) < f(K)$, then we also obtain the following, given that cover $\#A \geq 3$:

$$s(K) \leq x < f(I) < f(K) \Rightarrow s(K) < f(I) < f(K)$$
$$\Rightarrow I \prec K$$

$$s(K) \leq x < f(J) < f(J) \Rightarrow s(K) < f(J) < f(K)$$
$$\Rightarrow J \prec K$$

$$I \cap J \neq \phi \quad (\text{since } x \in I, J)$$

Thus, cover $\#A \geq 3$ if and only if there are three distinct intervals $I, J, K \in A$ such that $I \cap J \neq \phi$ and $I \prec K$ and $J \prec K$.

(d) Initial : $Bin(1) = \phi$
$Bin(2) = \phi$
$fTime(1) = 0$
$fTime(2) = 0$
$idx = 1$

Iteration 1 : $I_1 = [0, 3)$
$s(I_1) = 0 \not> fTime(1) = 0$
and
$s(I_1) = 0 \not> fTime(3-1) = 0$

$\Rightarrow$ discard I_1

Iteration 2 : $I_2 = [1,4)$

$S(I_2) = 1 > fTime(1) = 0$

$\Rightarrow Bin(1) = \{I_2\}$

$fTime(1) = 4$

Iteration 3 : $I_3 = [4,5)$

$S(I_3) = 4 \not> fTime(1) = 4$

$S(I_3) = 4 \not> fTime(2) = 0$

$\Rightarrow idx = 2,\ Bin(2) = \{I_3\}$

$fTime(2) = 5$

Iteration 4 : $I_4 = [2,6)$

$S(I_4) = 2 \not> fTime(2) = 5$

$S(I_4) = 2 \not> fTime(1) = 4$

$\Rightarrow$ discard I_4

Iteration 5 : $I_5 = [5,7)$

$S(I_5) = 5 \not> fTime(2) = 5$

$S(I_5) = 5 \not> fTime(1) = 4$

$\Rightarrow idx = 1,\ Bin(1) = \{I_2, I_5\}$

$fTime(1) = 7$

Iteration 6 : $I_6 = [4,8)$

$S(I_6) = 4 \not> fTime(1) = 7$

$S(I_6) = 4 \not> fTime(2) = 5$

$\Rightarrow$ discard I_6

Iteration 7 : $I_7 = [7,9)$

$S(I_7) = 7 \not> fTime(1) = 7$

$S(I_7) = 7 > fTime(2) = 5$

$\Rightarrow idx = 2,\ Bin(2) = \{I_3, I_7\}$

$fTime(2) = 4$

$\Rightarrow Bin(1) = \{I_2, I_5\},\ Bin(2) = \{I_3, I_7\}$

$\Rightarrow \boxed{B = Bin(1) \cup Bin(2) = \{I_2, I_3, I_5, I_7\}}$

(e) To prove correctness, we must show the following,

(i) Output set B has cover # <=2 ⇒ No three intervals overlap

(ii) B is maximum size 2-feasible subset possible

We know that Bin (1) ∩ Bin (2) = ϕ and I_i into a Bin only if $s(I_i) \geq$ fTime(k) of the Bin(k). ⇒ $s(I_i)$ start time of I_i is after all intervals in Bin (k) have finished ⇒ $s(I_i) \geq$ max fTime (I, $I \in$ Bin (K)). Thus, there is no overlap (ϕ) and each bin has no conflicts. ⇒ All bins are compatible sets cover # Bin (1) ≤ 1 and cover # Bin (2) ≤ 1.

Since we have two bins which are disjoint, any point p can be at most in one interval per bin. Since B = Bin (1) ∪ Bin (2), the total cover is at most 2 ⇒ cover # B ≤ 2. [i]

An interval I_i is discarded if $s[I_i] \not\geq$ fTime (1) and $s[I_i] \not\geq$ fTime (2). This implies that there is an overlap with the last interval in both bins Bin (1) and Bin (2). If it had been added to either Bin, we would have three overlapping intervals, contradicting the condition cover # A ≤ 2. Since this would lead to a cover # ≥ 3, the discarding is essential to obtain an optimal solution.