



AHEAD Online

Design and Analysis of Algorithms

Dr. Swaminathan J

Department of Computer Science

Amrita Vishwa Vidyapeetham

Objectives

- To provide a warm-up on algorithms.
- To appreciate the need to design good algorithms.
- Algorithm vs. Program – To know how they differ?
- To measure the running time of iterative algorithm and recursive algorithms by counting operations.

Place your
Webcam Video here
Size 100%

Hello World

Place your
Webcam Video here
Size 38%

- Consider the two programs given below and answer the following questions.

1

```
print("Hello world")
```

2

```
print("Hello ")  
print("world")
```

Note down
your answers

- What is the **time taken** by the first program to run? Can't say
- Will the program take same time on **every machine** it is run? No
- Will the program take same time on **every run** on same machine? No
- Will the first program **run faster** than the second one? Not always
- Do the **differences** in running times matter at all? No

Sum of two Integers

Place your
Webcam Video here
Size 38%

- Consider the programs given below.

1
sum(a,b) :
return a+b

2
sum(a,b) :
c = a+b
return c

3
a = input()
b = input()
print(a+b)

Again, note
down your
answers

1. Will the first program **run faster** than the second one?
2. How about the third program? It takes **input from user**.
3. Should user input delays be added to the running time?
4. Which program uses **more memory** among the three?
5. Do the **differences** in memory usage matter at all?

Not always

Yes

Not fair

Program 2?

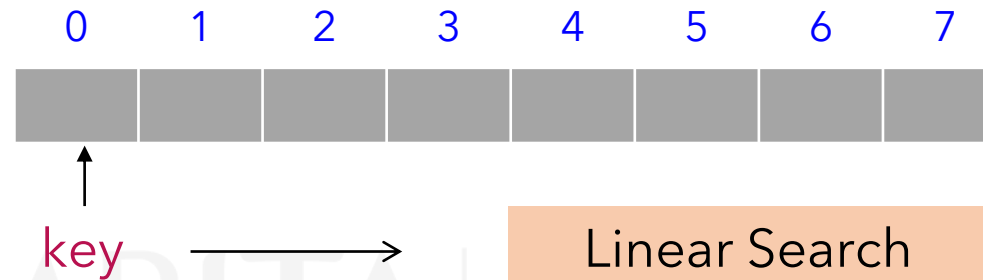
No

Searching an array

Place your
Webcam Video here
Size 38%

Given an array A of size n , check if a key is in A .

```
for (i=0; i<n; i++)  
    if ( key == A[i] )  
        return true  
return false
```



- Running time depends on the number of comparisons made.
(the number of comparisons depends on array size)

- Best case 1 comparison (key = A[0])
- Worst case n comparisons (key = A[n] or key not in A)
- Average case $\frac{n+1}{2}$ comparisons (key appears in every position with equal probability)

$$\frac{1+2+ \dots +n}{n} = \frac{n(n+1)/2}{n} = \frac{n+1}{2}$$

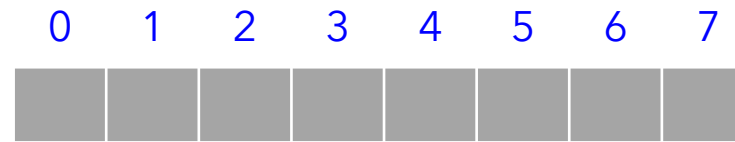
Memory used
 $n+1$

Searching a sorted array

Place your
Webcam Video here
Size 38%

Given a sorted **array A** of **size n**, check if a **key** is in **A**.

```
if ( key == A[ $\frac{n}{2}$ ] )  
    return true  
if ( key < A[mid] )  
    Search the left half  
else  
    Search the right half
```



Binary
Search

Algorithm
design
matters!

- After each step, search space reduces by half.

- $n \rightarrow \frac{n}{2} \rightarrow \frac{n}{4} \rightarrow \dots \rightarrow 1$ ($\log_2 n$ comparisons)

- What difference does it make? Say $n = 10^9$.

- Linear search: 10^9 comparisons (worst case)

• Binary search: $\log_2 10^9$

$$= \log_2 (10^3)^3$$

$$= 3 \log_2 10^3$$

$$\approx 3 \log_2 2^{10}$$

since $10^3 \approx 2^{10}$

$$= 30 \log_2 2$$

$$= 30 \text{ comparisons (worst case)}$$

Algorithm vs. Program

Place your
Webcam Video here
Size 38%

- An algorithm is a **finite sequence of computational steps** that transforms an input into an output.

Algorithm	Program
Written in pseudo code	Written in a programming language
Targeted for analysis and reasoning by humans	Targeted for execution on a real machine
Well structured	Need not be well-structured
Includes only what is necessary for analysis	Must be complete and syntactically correct.
Input output instructions are not included	Input output instructions are included

Algorithm: LinearSearch

Input: A[1..n], key

Output: true/false

for i \leftarrow 1 to n **do**

if key = A[i] **then**

return true

return false

Running time of Iterative algorithm

Place your
Webcam Video here
Size 38%

- Lets measure the running time of linear search by counting the number of operations.

Algorithm: Search

Input: A[1..n], key

Output: true/false

```
1. i ← 1
2. while i ≤ n do
3.     if key = A[i] then
4.         return true
5.     i ← i + 1
6. return false
```

Running Time

Worst case	Best case
1	1
n+1	1
2n	2
0 / 1	1
2n	0
1 / 0	0
5n+3	5

Closed form equation

$$T(n) = 5n + 3$$

n is the input size

op (assignment)
ops (comparison)
ops (comparison, indexing)
op (return)
op (increment, assignment)
op (return)
ops

It is the **worst case** that is used to characterize any algorithm.

- Best case never.
- Average case occasionally.

All types of **operations** are assumed to take the **same time**. In reality, it is not so.

Running time of Recursive algorithm

Place your
Webcam Video here
Size 38%

- Lets measure the running time of linear search by counting the number of operations.

Recurrence equation

$$T(n) = \begin{cases} 4 & \text{if } n = 0 \\ T(n-1) + 5 & \text{otherwise} \end{cases}$$

How to
solve this
recurrence
equation?

Algorithm: Search

Input: A[1..n], key

Output: true/false

1. if $n = 0$ then
2. return false
3. if key = A[n] then
4. return true
5. else
6. return Search(A[1..n-1], key)

Running Time		
Worst case	Best case	
1	1	op (comparison)
1	0	op (return)
2	2	ops (comparison, indexing)
1	1	op (return)
$T(n-1) + 1$	0	op (recursion + return)
$T(n-1) + 5$	4	ops

Solving Recurrences

Place your
Webcam Video here
Size 38%

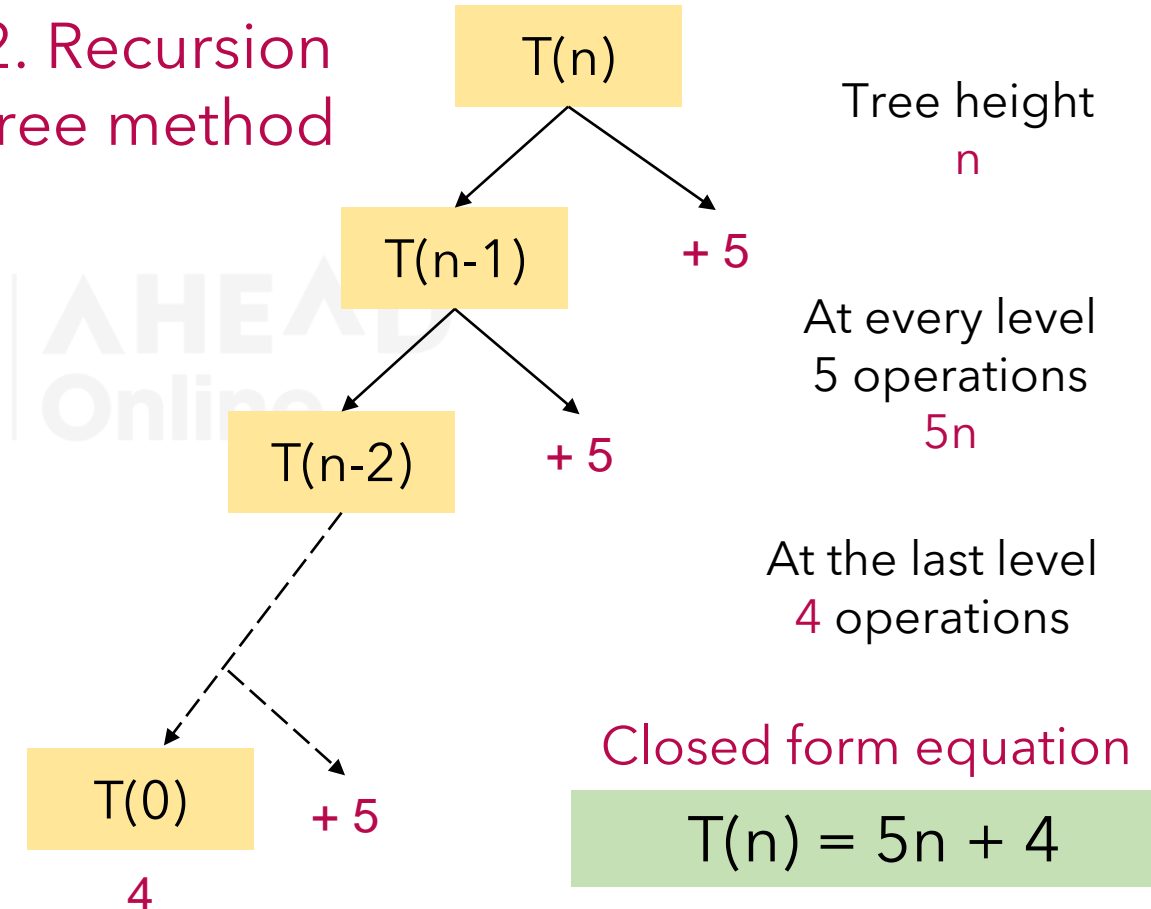
Two methods to solve recurrence equations are given.

$$T(n) = \begin{cases} 4 & \text{if } n = 0 \\ T(n-1) + 5 & \text{otherwise} \end{cases}$$

1. Substitution method

$$\begin{aligned} T(n) &= T(n-1) + 5 &&= T(n-1) + 1*5 \\ &= [T(n-2) + 5] + 5 &&= T(n-2) + 2*5 \\ &= [T(n-3) + 5] + 2*5 &&= T(n-3) + 3*5 \\ &= \\ &= [T(0) + 5] + (n-1)*5 &&= T(0) + n*5 \\ &= 4 + 5n &&= 5n + 4 \end{aligned}$$

2. Recursion tree method



Summary

- We showed why design of good algorithms are important.
- We demonstrated how to measure the running time of iterative algorithm and recursive algorithms by counting operations.

Place your
Webcam Video here
Size 100%



AHEAD Online

Design and Analysis of Algorithms

Dr. Swaminathan J

Department of Computer Science

Amrita Vishwa Vidyapeetham

Objectives

- To introduce time and space complexity.
- To understand orders of growth in functions.
- To group functions based on their orders of growth.
- To introduce the idea of bounds.
- To introduce asymptotic notations.

Place your
Webcam Video here
Size 100%

Time & Space Complexity

Place your
Webcam Video here
Size 38%

Time complexity refers to the **running time** of an algorithm.

- It is measured by the **number of primitive operations** performed by the algorithm as a function of **input size**.
 - If n is the input size, $f(n)$ denotes the number of operations, Time complexity $T(n) = f(n)$.
- What is its role in algorithm analysis?
 - It allows us to **compare two algorithms** and choose which is more time efficient.
 - It allows us to **classify real-world problems** in terms of complexity classes.

For linear search
we noted that
 $T(n) = 5n + 2$

Similar to Easy,
Medium, Hard.

Space complexity refers to the **number of variables** required by the algorithm as a function of input size.

Space implies
memory

Orders of Growth

Place your
Webcam Video here
Size 38%

Given below are the pairs of functions. Determine which among the two **grows at a faster** rate?

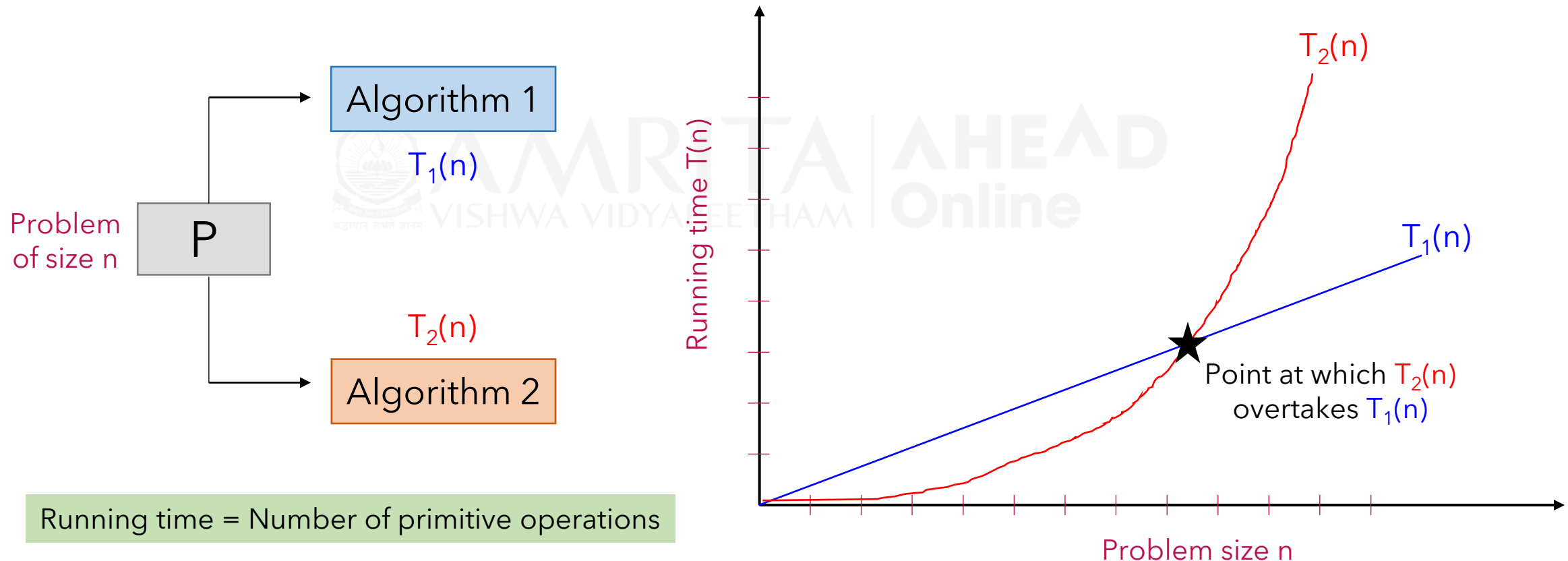
$T_1(n)$	$T_2(n)$	
n	n^2	n^2 grows faster
$1000n$	n^2	n^2 overtakes $1000n$ as n increases
$n(n+1)$	$10n^2$	$n(n+1) = n^2 + n$... Think about it
$100n^2$	$0.1n^3$	$0.1n^3$ overtakes $100n^2$ as n increases
$\log n$	$10^{-6}n$	$10^{-6}n$ overtakes $\log n$ as n increases
$\log^2 n$	$\log n^2$	since $\log^2 n = (\log n)^2$ and $\log n^2 = 2\log n$
$\log n$	\sqrt{n}	\sqrt{n} grows faster than $\log n$
2^n	n^2	2^n grows faster
10^{18}	n	10^{18} is a big constant..... Think about it

Note down
your answers

Orders of growth in Algorithms

Place your
Webcam Video here
Size 38%

- To be able to compare the running times of different algorithms and choose the better.

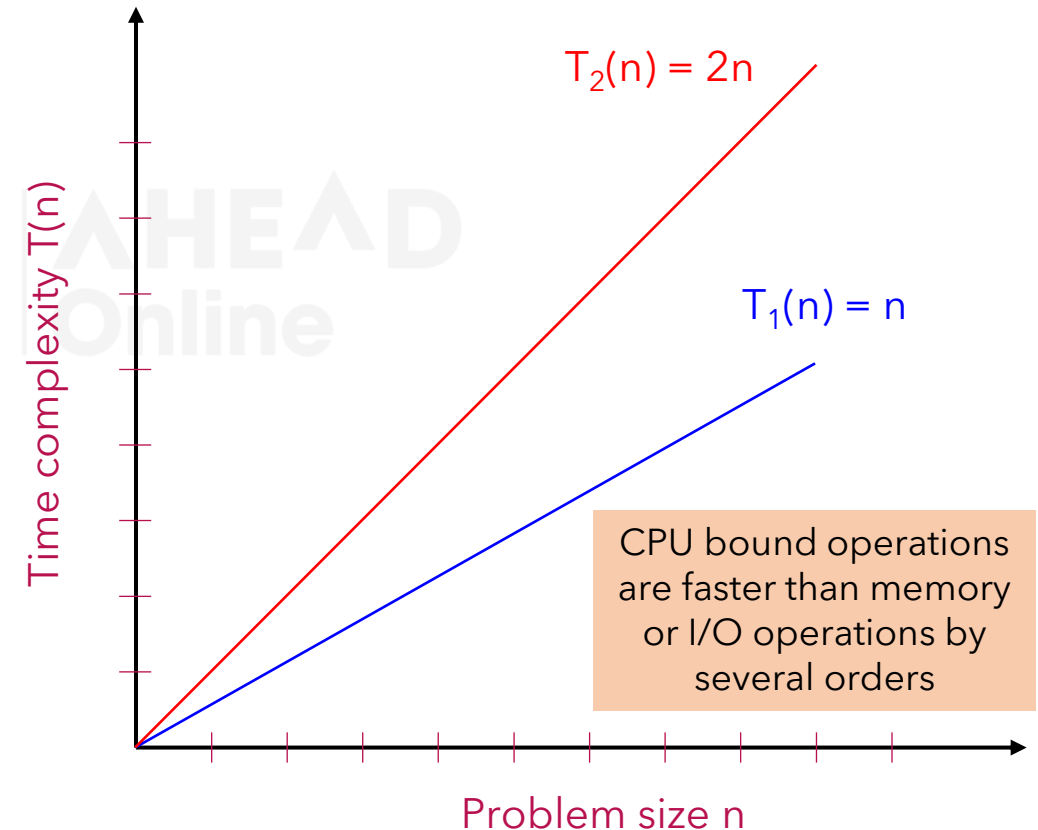
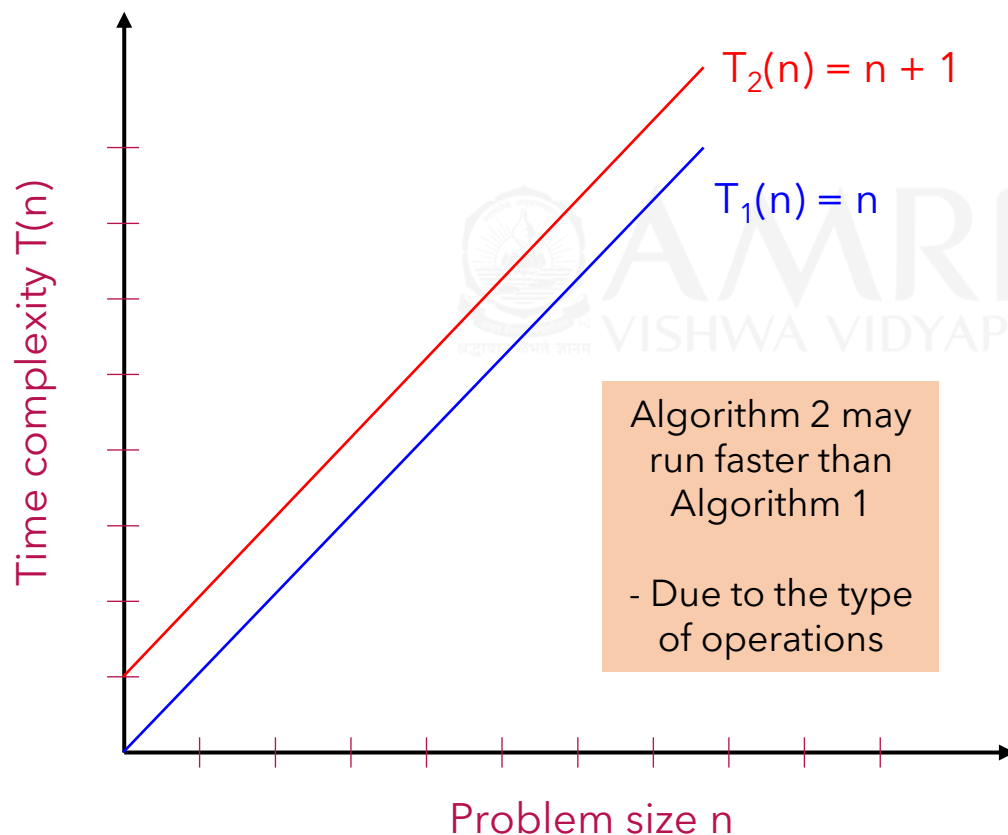


Functions with similar growth

Place your
Webcam Video here
Size 38%

Which among the two algorithms is efficient?

- Discussed using **linear** functions



Grouping functions based on growth

Place your
Webcam Video here
Size 38%

- Constants don't matter
- Lower order terms don't matter

But certainly

$$\cancel{2n + 3} \text{ vs. } \cancel{5n - 7}$$

Both are linear

$$\cancel{4n^2 - n + 3} \text{ vs. } \cancel{2n^2 - 6}$$

Both are quadratic

$$2^n \text{ grows faster than } n^2$$

How about these pairs?

$$3^n > 2^n$$

✓

$$2^{n+1} > 2^n$$

~~$2 \cdot 2^n$~~ ✗

$$(n+1)! > n!$$

~~$(n+1)n!$~~ ✓

$$n^n > \dots > n! \dots > 3^n > 2^n$$

Exponential

$$n^k > \dots > n^3 > n^2 > n > \sqrt{n}$$

Polynomial

$$(\log n)^2 > \log n > \log \log n > k$$

Logarithmic

The idea of bounds

Place your
Webcam Video here
Size 38%

Lets say there is a medical shop nearby to your house.

- And it is **roughly 200 metres** away from your home.

If someone knocks at your door and asks you,

- **How far** is the nearest medical shop from here?

Now, your answer can be one of the following.

- At least 200 m
- At least 180 m
- At least 150 m
- At least 10 m
- At least 1 m

Lower bounds

- About 200 m
- About 180 m
- About 220 m

Tight bounds

- At most 200 m
- At most 220 m
- At most 250 m
- At most 1 km
- At most 5 km

Upper bounds

Which ones
are correct?

Walkable

Let's connect it to Algorithms

Place your
Webcam Video here
Size 38%

We are interested in determining **how long** would it take for an algorithm to run to compute a solution?

In the example, we would **classify** distances to several destinations based on whether they are reachable

- ✓ By walk
- ✓ By car
- ✓ By flight
- ✓ By rocket
- **Not reachable**

For example, reaching a grocery shop, theatre or a medical store belong to **walkable class** of distances since both are walkable distance away.

Given algorithms to solve several problems, we want to **classify** them based on their time complexities.

- ✓ Constant
- ✓ Logarithmic
- ✓ Polynomial
- ✓ Linear, quadratic, cubic, ...
- **Exponential**

Finding max element of an array or searching a key in an array belong to **linear class** of algorithms since their running times are of form **$an + b$** .

Asymptotic notations

Place your
Webcam Video here
Size 38%

Let $f(n)$ be the running time (number of operations) of an algorithm where n is the input size.

- We want to characterize the algorithm based on its order of growth.
- Three notations, known as asymptotic notations, are introduced towards this.

Big Omega / $\Omega(.)$

$f(n) \in \Omega(g(n))$ if there exists

- a real constant $c > 0$
- an integer $n_0 \geq 0$

such that

- $f(n) \geq c.g(n)$ for $n \geq n_0$

Big Theta / $\theta(.)$

$f(n) \in \theta(g(n))$ if there exists

- real constants $c_1, c_2 > 0$
- an integer $n_0 \geq 0$

such that

- $c_1.g(n) \leq f(n) \leq c_2.g(n)$ for $n \geq n_0$

Big O / $O(.)$

$f(n) \in O(g(n))$ if there exists

- a real constant $c > 0$
- an integer $n_0 \geq 0$

such that

- $f(n) \leq c.g(n)$ for $n \geq n_0$

Is $f(n) \in O(g(n))$?

Place your
Webcam Video here
Size 38%

Let's work out some examples with $f(n) = 100n^2$.

1. Is $f(n) \in O(n^2)$?

- $f(n) = 100n^2$
- $g(n) = n^2$

Choose $c = 101$.

$100n^2 \leq 101n^2$ is
true for all $n \geq 0$.

i.e. $c = 101, n_0 = 0$.

In fact any real value
>100 will serve as c .

2. Is $f(n) \in O(n^3)$?

- $f(n) = 100n^2$
- $g(n) = n^3$

$c = 101, n_0 = 0$ will serve
our purpose.

We can also choose c to
be 99. $100n^2 \leq 101n^3$
will be true for all $n \geq 2$.
i.e. $c = 99, n_0 = 2$.

$f(n) \in O(g(n))$ if there exists

- a real constant $c > 0$
 - an integer $n_0 \geq 0$
- such that
- $f(n) \leq c.g(n)$ for $n \geq n_0$

3. Is $f(n) \in O(n)$?

- $f(n) = 100n^2$
- $g(n) = n$

For any c , however
high, $100n^2 \not\leq c.n$.
Hence, $f(n) \notin O(n)$.

Understanding $f(n) \in O(g(n))$

Place your
Webcam Video here
Size 38%

- $f(n)$ belongs to the class of functions upper bounded by $g(n)$.
- Since, $g(n)$ grows at least as fast as $f(n)$, $c.g(n)$ overtakes $f(n)$.

$f(n) = 100n$ and $g(n) = n^2$. Is $f(n) \in O(n^2)$?

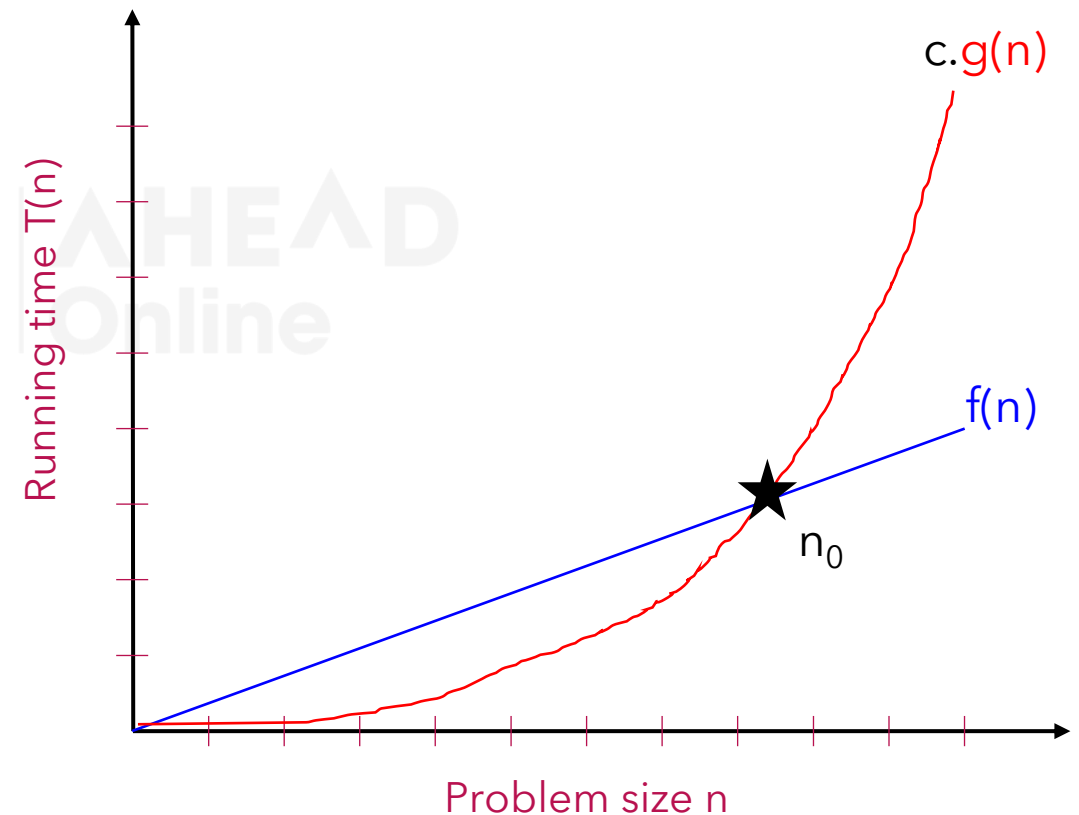
To find c & n_0 such that $f(n) \leq c.g(n)$ for $n \geq n_0$

Let's take $c = 15$. At $n = 7$, $15n^2$ overtakes $100n$.

n	1	2	3	4	5	6	$n_0=7$
$100n$	100	200	300	400	500	600	700
$15n^2$	15	60	135	240	375	540	735

Therefore, n^2 upper bounds $100n$

Note $g(n)$ denotes a class and has no constants or lower order terms.



Understanding $f(n) \in \Omega(g(n))$

Place your
Webcam Video here
Size 38%

- $f(n)$ belongs to the class of functions lower bounded by $g(n)$.
- Since, $g(n)$ grows slower than $f(n)$, $c.g(n)$ cannot keep pace with $f(n)$. i.e. $f(n)$ will overtake $c.g(n)$ whatever be c .

$f(n) = 100n$ and $g(n) = \log n$
Is $f(n) \in \Omega(n^2)$?

To find c and n_0 such that

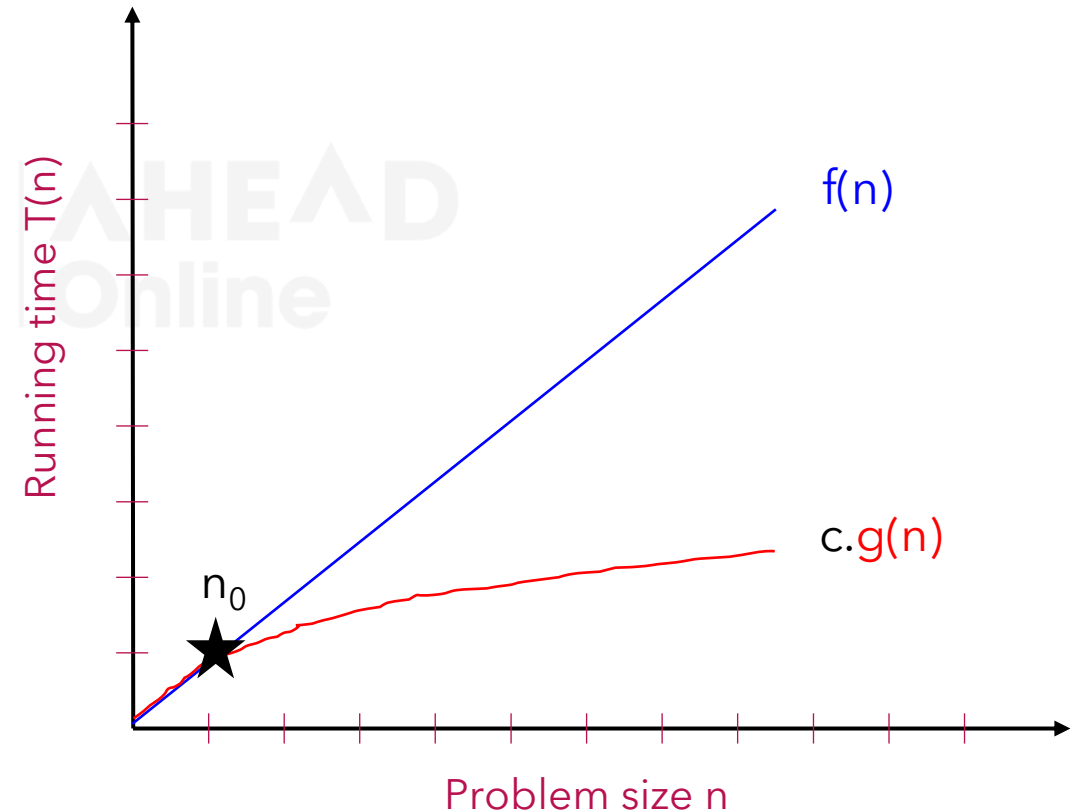
- $f(n) \geq c.g(n)$ for $n \geq n_0$

$$\begin{aligned} f(n) &= 100n \\ &\geq 100 \log n \quad \text{for } n \geq 1 \end{aligned}$$

$$\text{So, } c = 100, \quad n_0 = 1$$

$c = 200, n_0 = 2$ is a correct too

Hence, $\log n$ lower bounds $100n$



Understanding $f(n) \in \Theta(g(n))$

Place your
Webcam Video here
Size 38%

- $f(n)$ belongs to the class of functions that is both upper and lower bounded by $g(n)$.
- Just by changing constants, $g(n)$ can serve as both bounds.

$f(n) = 100n$ and $g(n) = n$
Is $f(n) \in \Theta(n)$?

To find c_1 , c_2 and n_0 such that

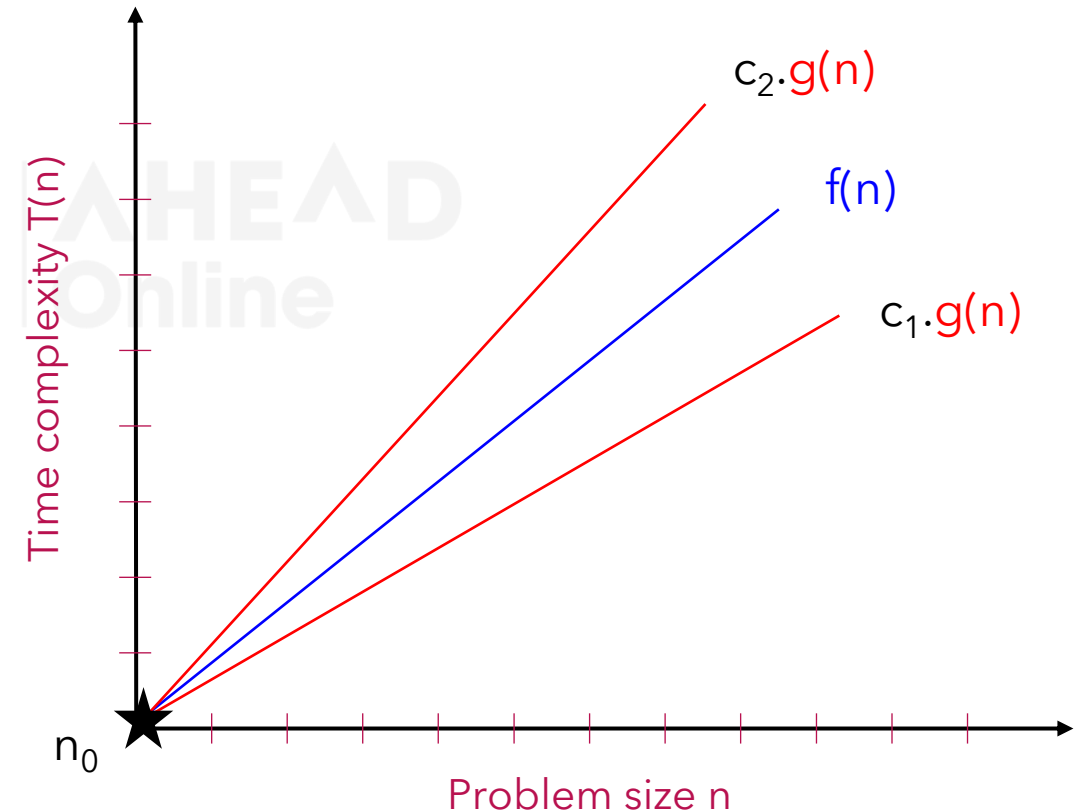
$$\blacksquare \quad c_1 \cdot g(n) \leq f(n) \leq c_2 \cdot g(n) \text{ for } n \geq n_0$$

$$f(n) = 100n \geq 99n \quad \text{for } n \geq 0$$

$$= 100n \leq 101n \quad \text{for } n \geq 0$$

$$\text{So, } c_1 = 99, c_2 = 101, n_0 = 0$$

In fact, any value < 100 fits c_1 , > 100 fits c_2 .

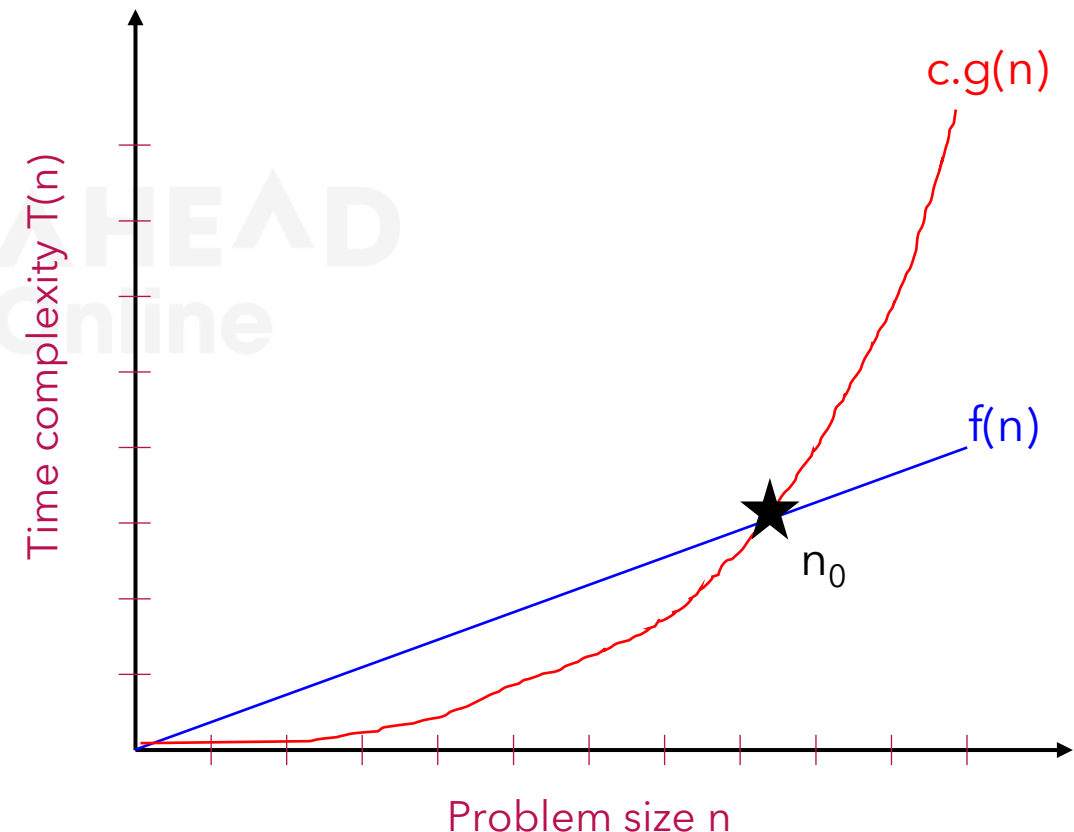


Understanding asymptotic notations

Place your
Webcam Video here
Size 38%

Let $f(n)$ refer to running time of an algorithm. $g(n)$ refers to a function without constants or lower order terms.

$f(n) \in O(g(n))$	<ul style="list-style-type: none">$g(n)$ upper bounds $f(n)$ for all $n \geq n_0$$g(n)$ grows at least as fast as $f(n)$$c \cdot g(n)$ overtakes $f(n)$ for some c
$f(n) \in \Omega(g(n))$	<ul style="list-style-type: none">$g(n)$ lower bounds $f(n)$ for all $n \geq n_0$$g(n)$ grows at most as fast as $f(n)$$f(n)$ overtakes $c \cdot g(n)$ for some c
$f(n) \in \theta(g(n))$	<ul style="list-style-type: none">$g(n)$ tight bounds $f(n)$ for all $n \geq n_0$$g(n)$ grows about as fast as $f(n)$By choosing appropriate constants, $f(n)$ can be shown to overtake or overtaken by $g(n)$



Time complexity & Asymptotic notations

Place your
Webcam Video here
Size 38%

For a given algorithm, we know

- n is the input size.
- $f(n)$ is the time complexity of the algorithm.

$\Omega(g(n))$ refers to the class of all algorithms whose running time $f(n)$ is lower bounded by $g(n)$.

$\Theta(g(n))$ refers to the class of all algorithms whose running time $f(n)$ is tightly bounded by $g(n)$.

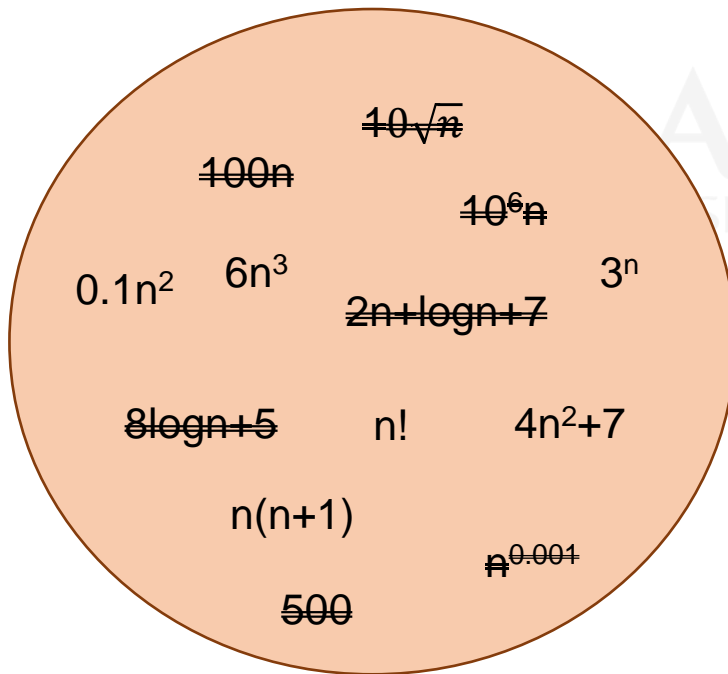
$O(g(n))$ refers to the class of all algorithms whose running time $f(n)$ is upper bounded by $g(n)$.

Examples

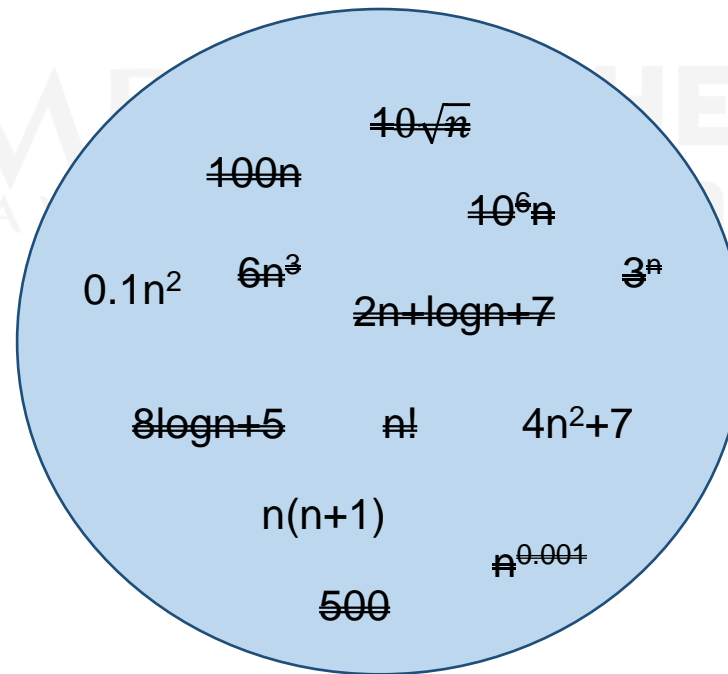
Place your
Webcam Video here
Size 38%

Let's take $g(n) = n^2$. Consider the three classes below.

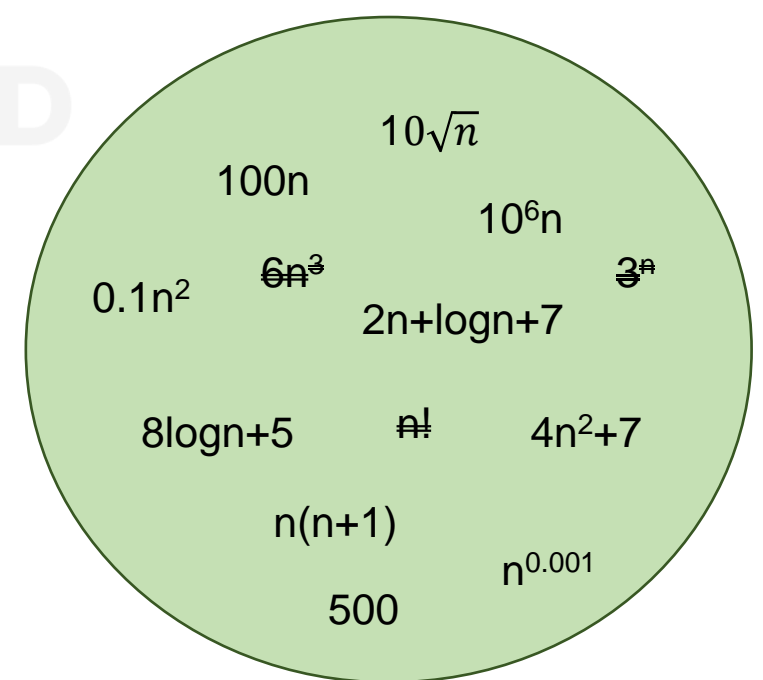
$\Omega(n^2)$



$\Theta(n^2)$



$O(n^2)$



Properties of Asymptotic functions

Place your
Webcam Video here
Size 38%

1. If $f(n) = a_k n^k + a_{k-1} n^{k-1} + \dots + a_1 n + a_0$, then
 - $f(n) \in O(n^k)$
 - $f(n) \in \Omega(n^k)$
 - $f(n) \in \Theta(n^k)$
2. If $f(n) \in O(n^k)$, then $f(n) \in O(n^{k+1})$
3. If $f(n) \in \Omega(n^k)$, then $f(n) \in \Omega(n^{k-1})$
4. If $f(n) \in O(n^k)$ and $f(n) \in \Omega(n^k)$, then $f(n) \in \Theta(n^k)$
5. If $f(n) \in O(g(n))$ and $g(n) \in O(h(n))$, then $f(n) \in O(h(n))$
6. If $f_1(n) \in \Theta(g_1(n))$ and $f_2(n) \in \Theta(g_2(n))$, then
 - $f_1(n) + f_2(n) \in \Theta(\max\{g_1(n), g_2(n)\})$
 - $f_1(n) * f_2(n) \in \Theta(g_1(n) * g_2(n))$

Summary

- We discussed time complexity and asymptotic notations.



Place your
Webcam Video here
Size 100%





AHEAD Online

Design and Analysis of Algorithms

Dr. Swaminathan J

Department of Computer Science

Amrita Vishwa Vidyapeetham

Objectives

- To introduce the **divide and conquer** strategy.
- To demonstrate how the strategy can be applied for practical problems such as sorting, searching and selection.

Place your
Webcam Video here
Size 100%

Problems on algorithmic strategies

Place your
Webcam Video here
Size 38%

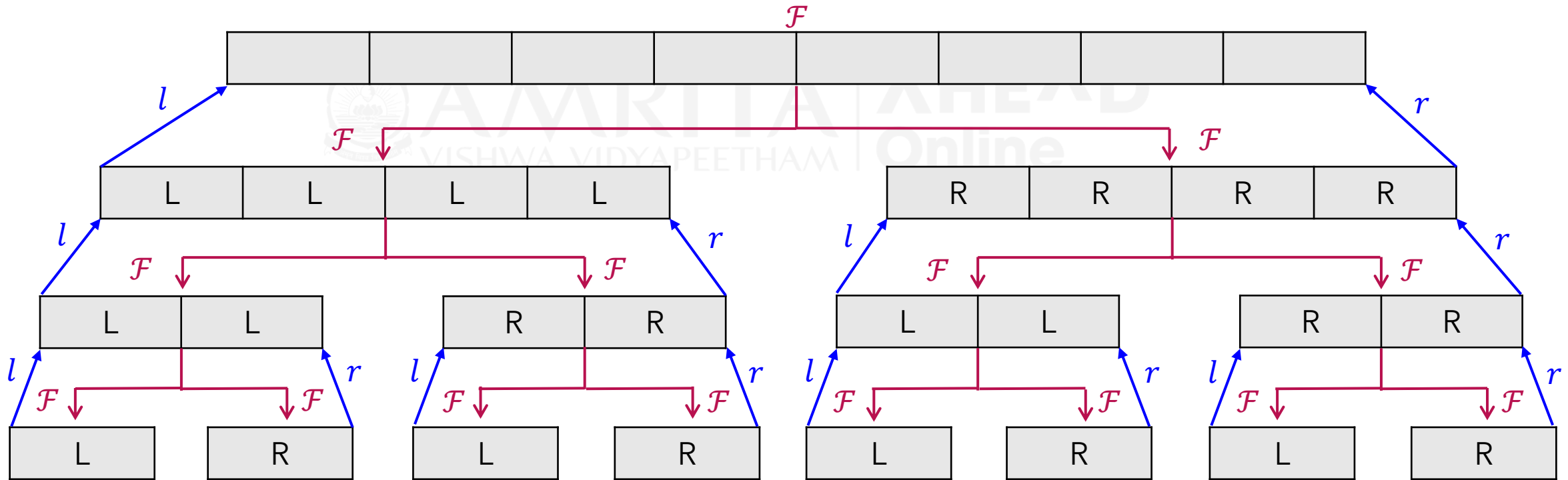
- **Divide and Conquer** : Binary Search – Merge sort – Quick sort – Multiplication of Large Integers.
- **Dynamic programming**: Principle of optimality – Coin changing problem, Computing a Binomial Coefficient – Floyd's algorithm – Multi stage graph – Optimal Binary Search Trees – Knapsack Problem and Memory functions.
- **Greedy Technique**: Container loading – Huffman Trees – Task scheduling – Fractional Knapsack.
- **Iterative methods**: The Simplex Method – The Maximum-Flow Problem – Maximum Matching in Bipartite Graphs, Stable marriage Problem.
- **Backtracking**: N-Queens problem – Hamiltonian Circuit Problem – Subset Sum Problem.
- **Branch and Bound**: – LIFO and FIFO search – Assignment problem – Knapsack Problem – Travelling Salesman Problem
- **Approximation Algorithms** for NP-Hard Problems – Travelling Salesman problem – Knapsack problem revisited.

Divide-and-Conquer (Data view)

Place your
Webcam Video here
Size 38%

A recursive approach to solving a problem

- **Divide** the problem into subproblems until small enough
- **Solve** the subproblems (**recurse**) and return the solution
- **Combine** the solutions of the subproblems



Place your
Webcam Video here
Size 38%

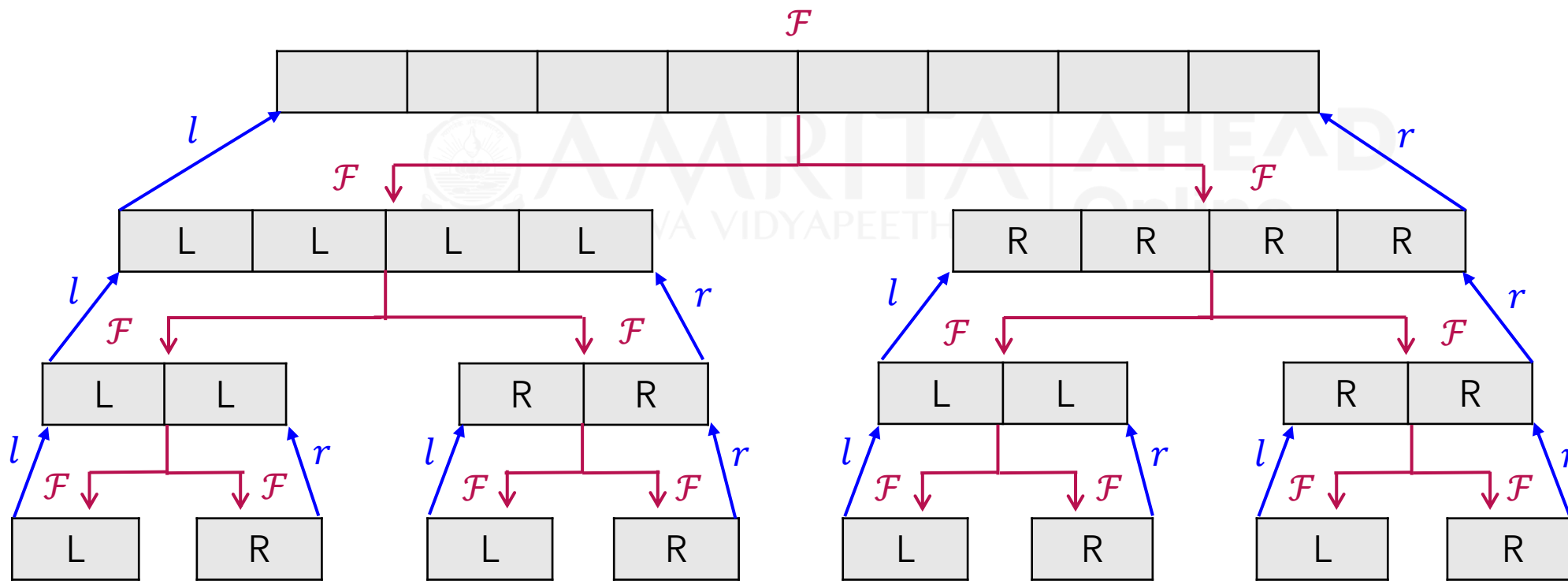
F ← Recursive step
C ← Combine step

Divide-and-Conquer (Algorithm)

Place your
Webcam Video here
Size 38%

Provides data and control views together.

- **Data view** in the form of diagram
- **Control view** in the form of the code



Array $a[1..n]$

```
F(a, s, e) {  
  if (s == e)  
    Base case  
    return result
```

Divide step
 $m = (s + e)/2$

```
l = F(a, s, m)  
r = F(a, m+1, e)
```

Combine step
return result

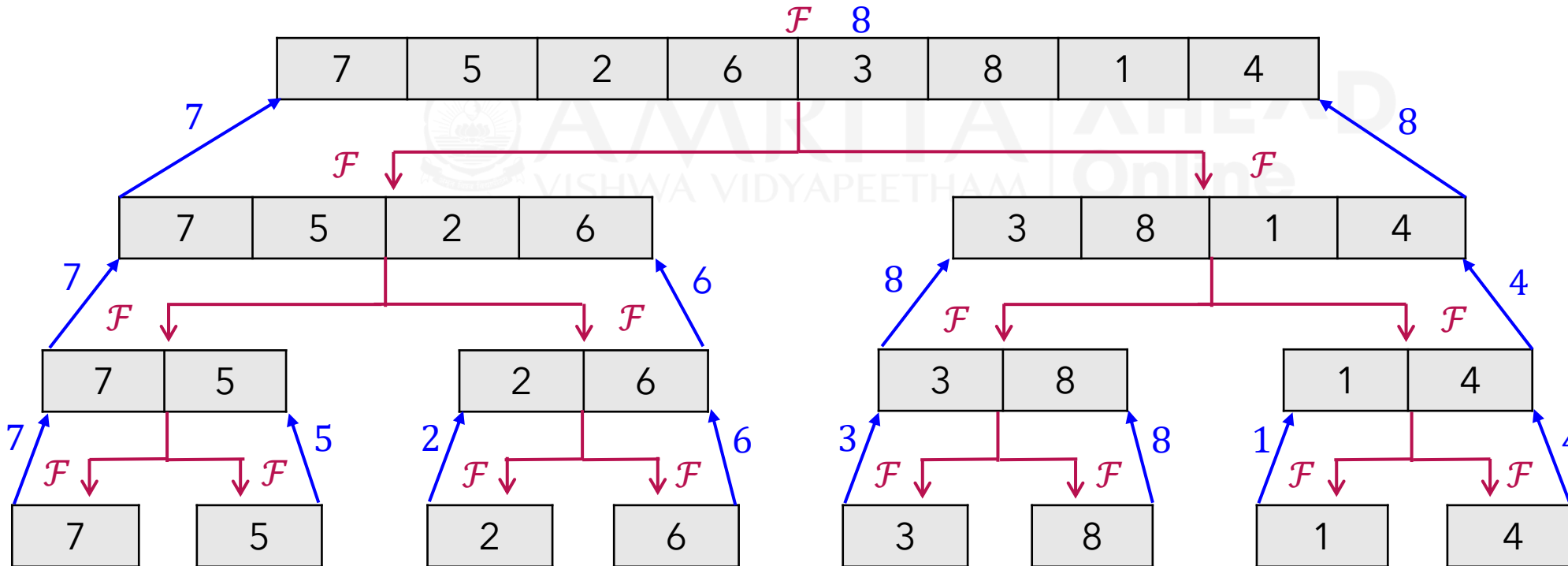
```
}
```

1. Finding max of an array

Place your
Webcam Video here
Size 38%

$F = \text{FindMax}(s, e)$

- Here, $n = 8$. Initially, $s = 1$ and $e = n$ or $(s, e) = (0, n-1)$.
- Divide: Split the array into two halves.
- Combine: Compare left & right results, return bigger.



Array $a[1..n]$

$F(a, s, e) \{$

if $(s == e)$

Base case

return $a[s]$

Divide step

$m = (s + e) / 2$

$l = F(a, s, m)$

$r = F(a, m+1, e)$

Combine step

return $l > r ? l : r$

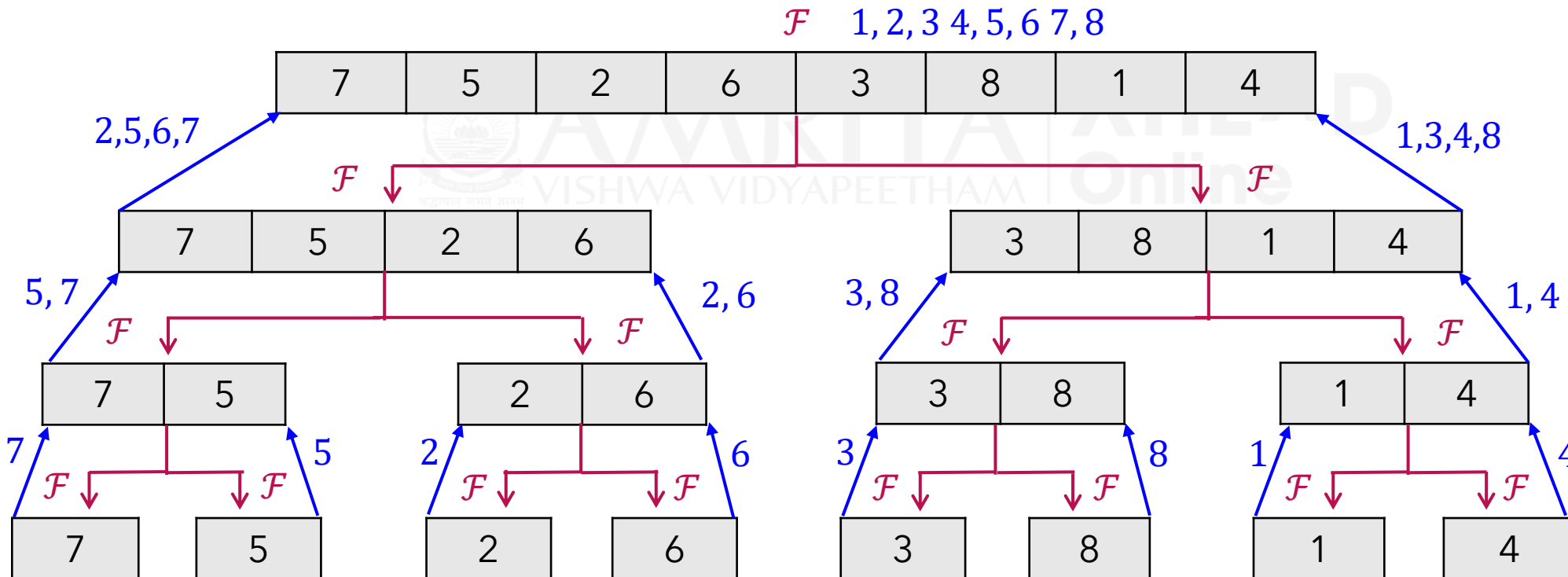
}

2. Merge sort

Place your
Webcam Video here
Size 38%

$F = \text{MergeSort}(s, e)$

- Here, $n = 8$. Initially, $s = 1$ and $e = n$ or $(s, e) = (0, n-1)$.
- Divide: Split the array into two halves.
- Combine: Merge 2 sorted subarrays.



Array $a[1..n]$

```
F(a, s, e) {  
  if (s == e)  
    Base case  
    return a[s]
```

Divide step
 $m = (s + e)/2$

```
l = F(a, s, m)  
r = F(a, m+1, e)
```

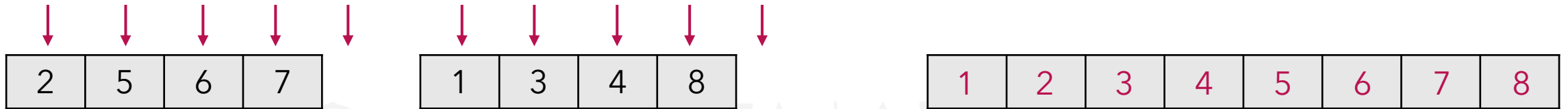
Combine step
return merge(l, r)
}

The merge step in Merge sort

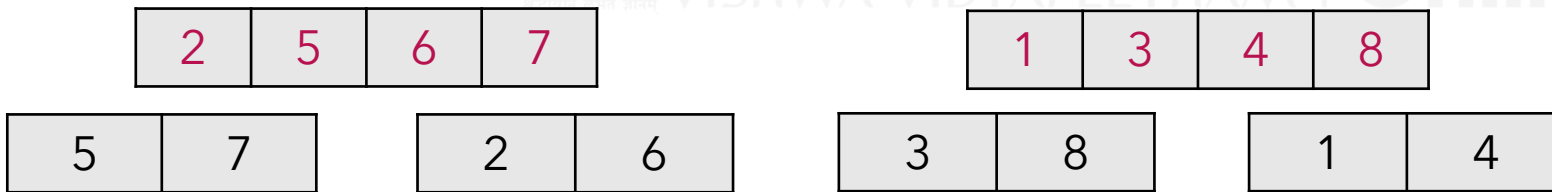
Place your
Webcam Video here
Size 38%

Given two sorted lists, how many comparisons does it take to merge them into a single sorted list?

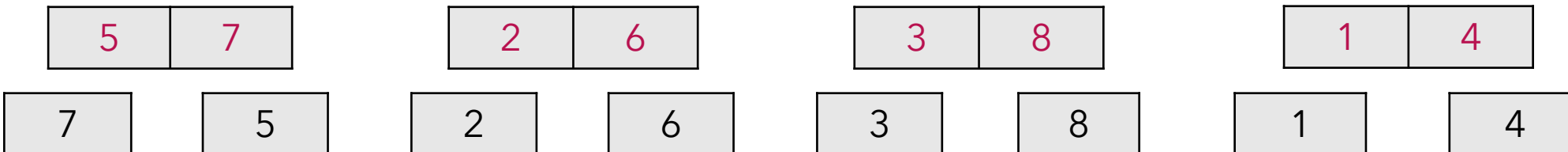
Ignoring constants, at most n comparisons at each level



7 comparisons. In general, $n - 1$ comparisons



6 comparisons. In general,
 $(\frac{n}{2} - 1) * 2$ comparisons



4 comparisons
 $(\frac{n}{4} - 1) * 4$
comparisons

Solving Recurrence for Merge sort

Place your
Webcam Video here
Size 38%

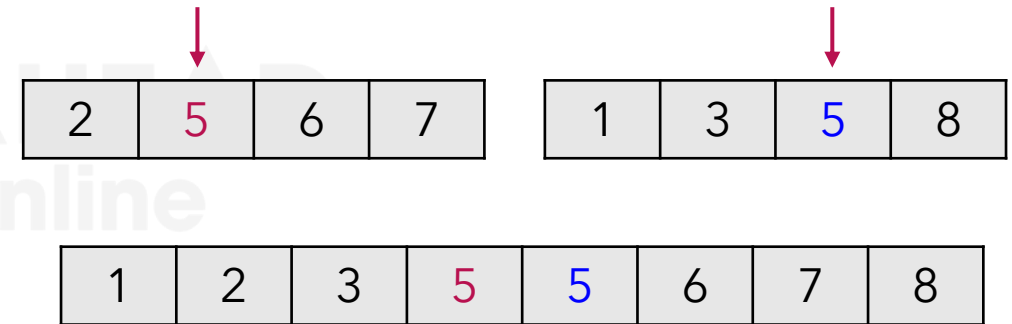
Let's use substitution method to solve the recurrence.

$$T(n) = \begin{cases} 1 & \text{if } n = 1 \\ 2T\left(\frac{n}{2}\right) + n & \text{otherwise} \end{cases}$$

$$\begin{aligned} T(n) &= 2T\left(\frac{n}{2}\right) + n &= 2T\left(\frac{n}{2}\right) + 1.n \\ &= 2 \left[2T\left(\frac{n}{4}\right) + n/2 \right] + n &= 2^2 T\left(\frac{n}{2^2}\right) + 2.n \\ &= 2 \left[2^2 T(n/8) + n/4 \right] + 2n &= 2^3 T\left(\frac{n}{2^3}\right) + 3.n \\ &= \dots\dots\dots &= 2^{\log n} T(1) + \log n.n \\ &= 2 \left[2^{\log n - 1} T(n/2^{\log n}) + \log n.n \right] &= n + n \log n \\ & &= n \log n \end{aligned}$$

Properties of merge sort

Stable, but not in-place



Complexity

$$\begin{aligned} T(n) &= \Theta(n \log n) \\ S(n) &= \Theta(n) \end{aligned}$$

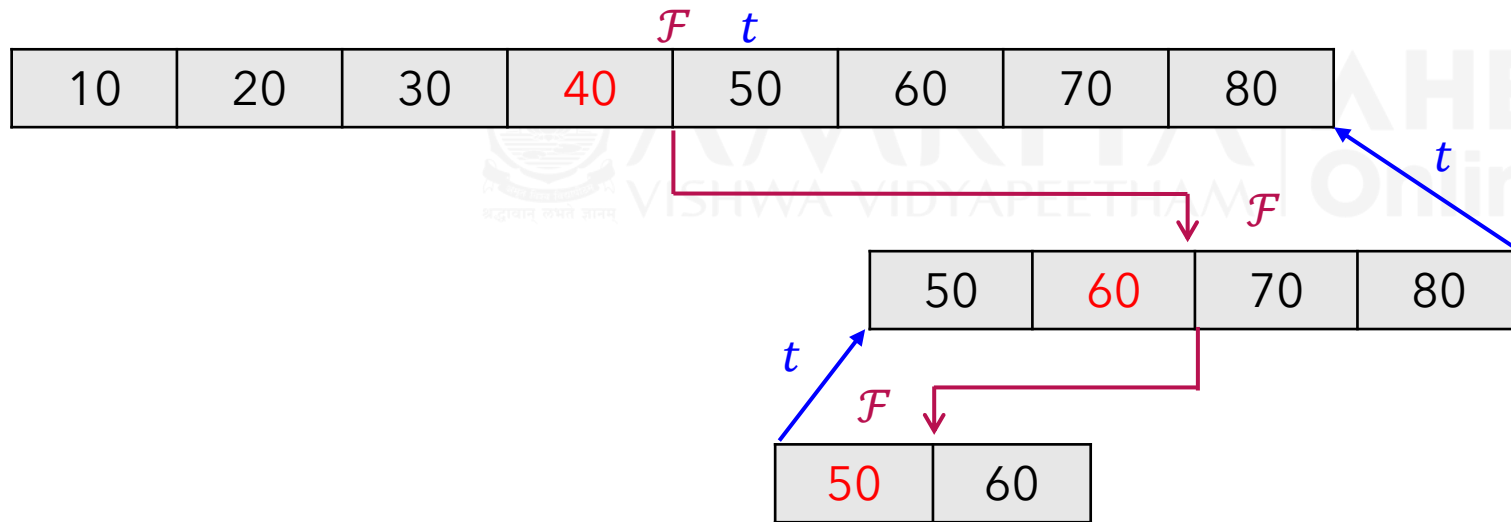
3. Binary Search on Sorted Array

Place your
Webcam Video here
Size 38%

Divide-and-Conquer way to search a sorted array.

$F = \text{BinarySearch}(a, s, e, k)$

- Consider the below example with $k = 50$.



Array $a[1..n]$

$F(a, s, e, k) \{$

$m = (s + e)/2$

if ($k == a[m]$) // Extra
return true // check

if ($s == e$)
return false // Base case

Divide and recurse

if ($k < a[m]$)
return $F(a, s, m, k)$
else
return $F(a, m+1, e, k)$

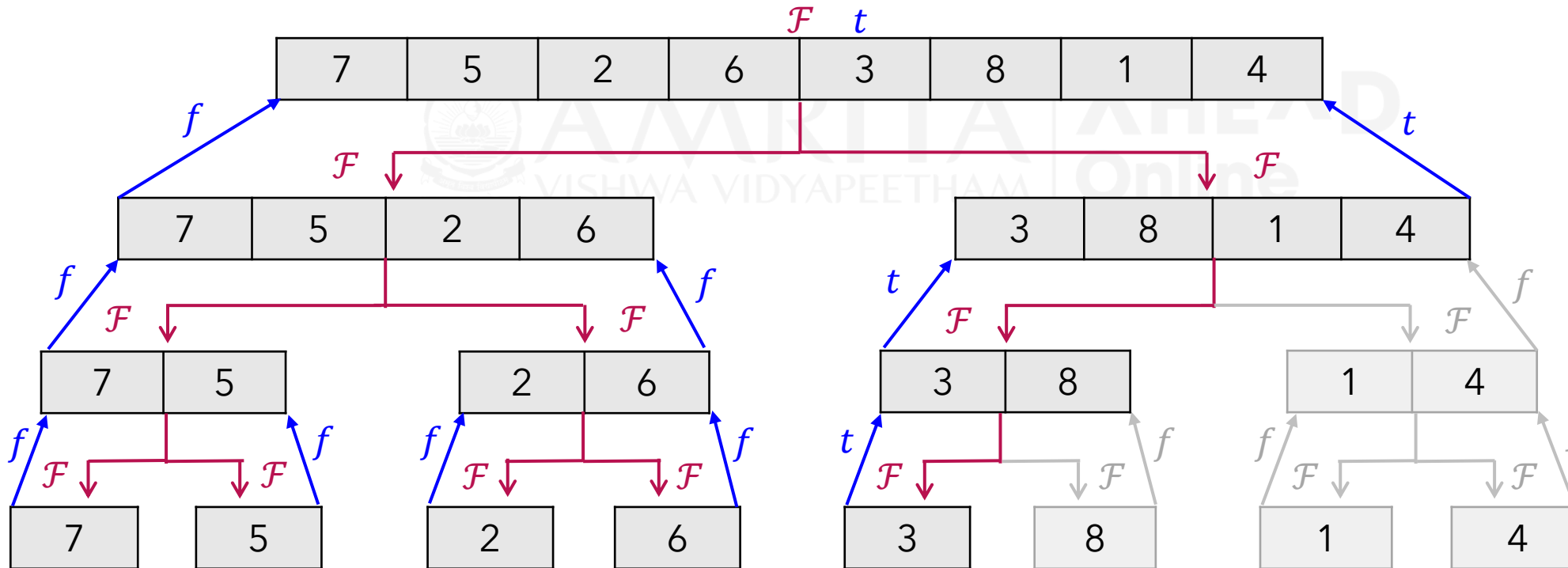
}

4. Binary Search on Unsorted Array

Place your
Webcam Video here
Size 38%

$F = \text{DnCSearch}(s, e, k)$. In the example, let $\text{key} = 3$.

- If found, rest of the search can be abandoned.
- If not, we must compare with every element (worst case).
- So, no advantage due to divide & conquer strategy.



Array $a[1..n]$

$$F(a, s, e, k) \{$$

```
if (s == e ∧ a[s]
== k)
```

Base case
return false

Divide step
 $m = (s + e)/2$

$$l = F(a, s, m, k)$$
$$r = F(a, m+1, e, k)$$

Combine step
return $l \vee r$

}

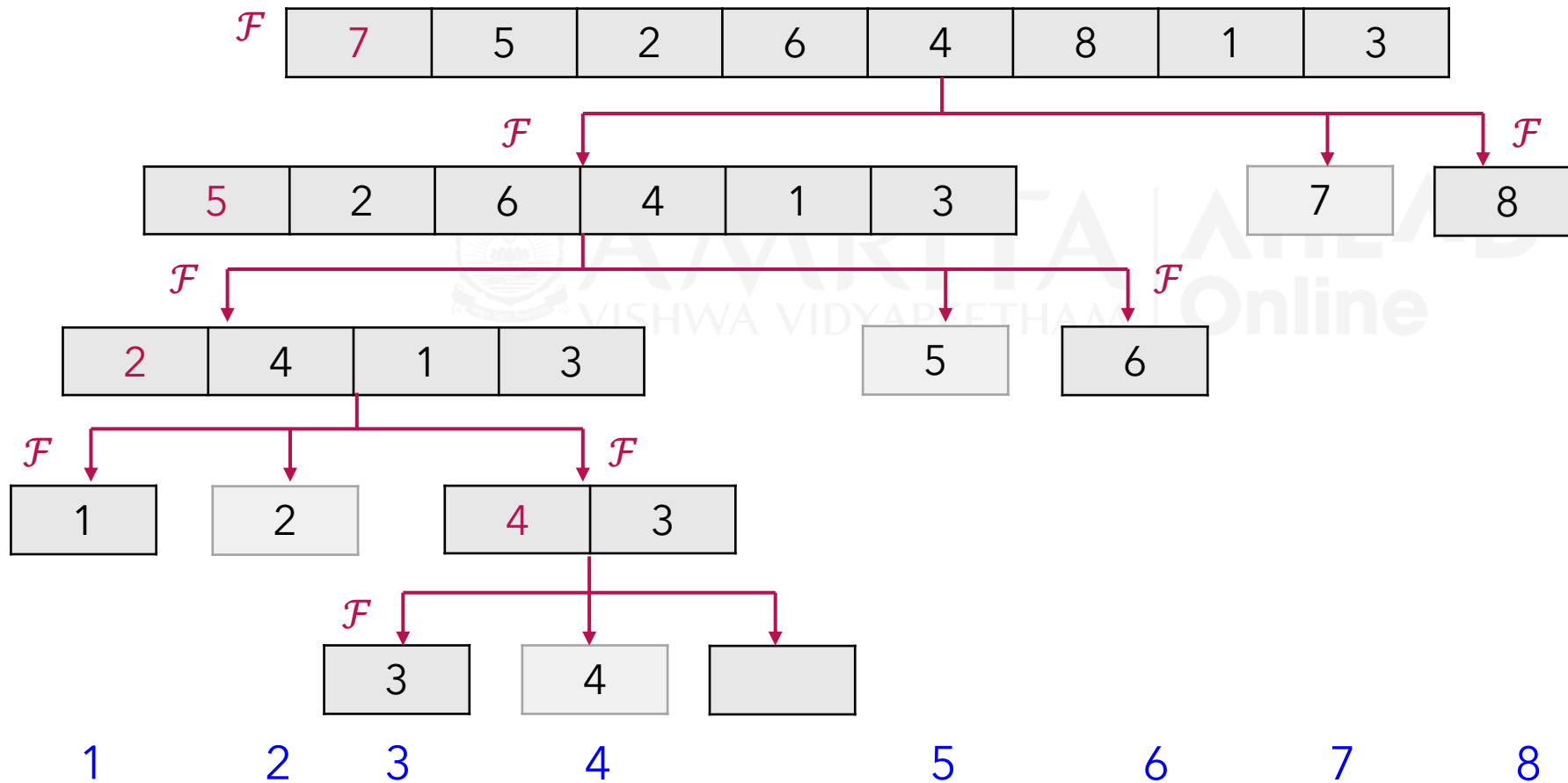


5. Quick Sort

Place your
Webcam Video here
Size 38%

$F = \text{QuickSort}(a, s, e).$

- Pick a pivot p . Divide a into $\{ < p \}, p, \{ \geq p \}$



```
Array a[1..n]
F(a, s, e) {
  if (size(a) == 1)
    Base case
    return a[s]
```

```
Divide step
for i  $\leftarrow$  s to e do
  p = a[s]
  l = {a[i]  $\ni$  a[i] < p}
  r = {a[i]  $\ni$  a[i]  $\geq$  p}
```

```
Combine step
return F(l) o p o F(r)
}
```

Properties of Quick Sort

Place your
Webcam Video here
Size 38%

1. The **divide** step can leave the left and right arrays to be imbalanced.

- **Worst case:** $\text{size}(l) = n-1$, $\text{size}(r) = 0$ or vice-versa.

- $T(n) = T(n-1) + n \Rightarrow \Theta(n^2)$

Pivot is max or min at each step.

- **Best case:** $\text{size}(l) = \frac{n-1}{2}$, $\text{size}(r) = \frac{n-1}{2}$

- $T(n) \leq 2 T(\frac{n}{2}) + n \Rightarrow \Theta(n \log n)$

Pivot is the median at each step.

2. Seems like time and space inefficient algorithm.

But not really so!

3. Fastest sorting algorithm in practice.

- **Average case:** $\Theta(n \log n)$. Worst case rarely happens.

Rigorous math behind.

- **In-place** sorting by optimal implementation of inner loop.

We will see next.

4. Space complexity $S(n) = \Theta(n)$.

5. Not a **stable** sort.

Summary

- We looked at Divide-and-conquer approach and how it applies to sorting and searching.



Place your
Webcam Video here
Size 100%



AHEAD Online

Design and Analysis of Algorithms

Dr. Swaminathan J

Department of Computer Science

Amrita Vishwa Vidyapeetham

Objectives

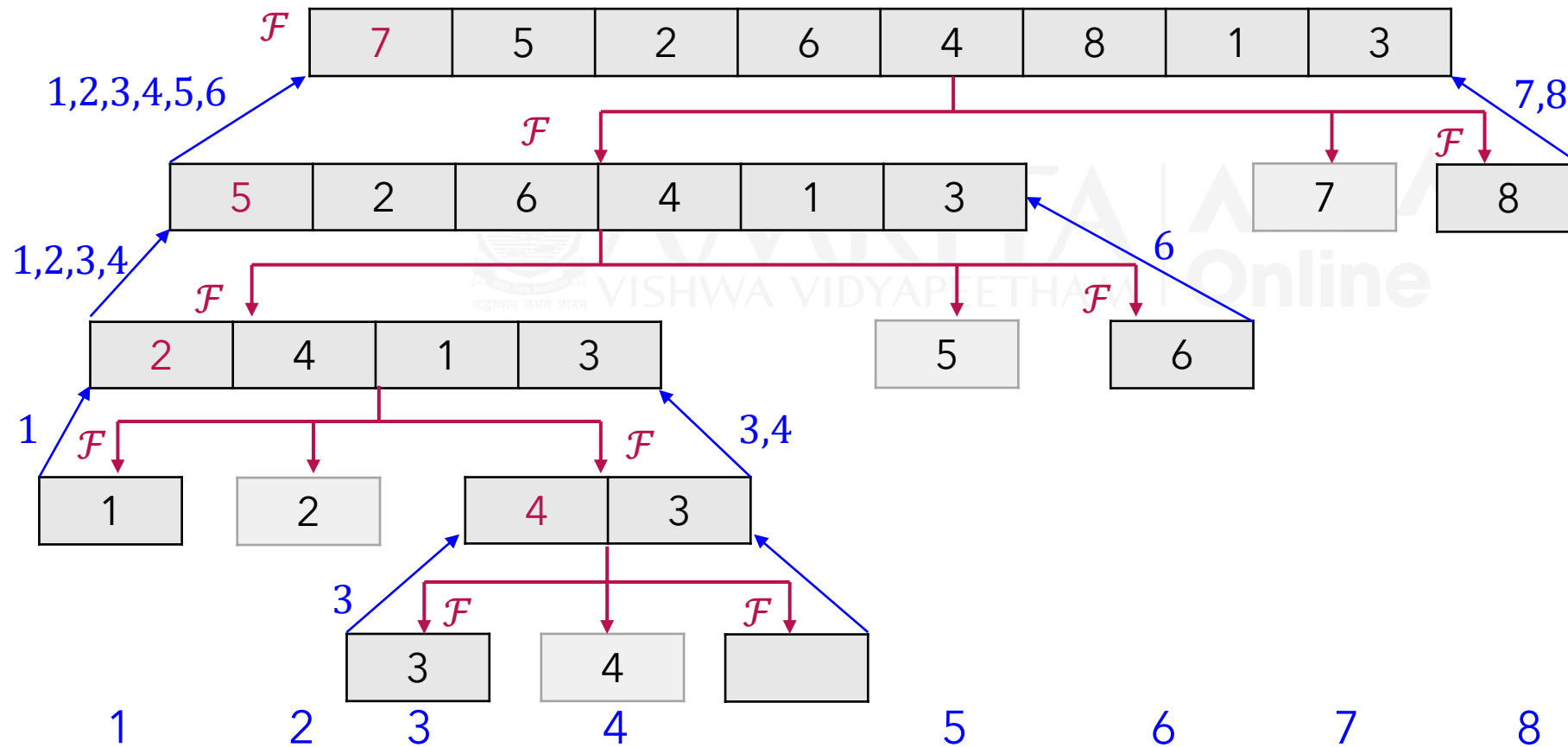
- To introduce the quick sort algorithm and discuss its properties.
- To take a deeper look at in-place implementation of quick sort.
- To discuss divide-and-conquer approach for quick select and maximum sum subarray problems.
- To show Master's method to determine complexity.

Place your
Webcam Video here
Size 100%

5. Quick Sort

$F = \text{QuickSort}(a, s, e).$

- Pick a pivot p . Divide a into $\{ < p \}, p, \{ \geq p \}$



Place your
Webcam Video here
Size 38%

Array $a[1..n]$
 $F(a, s, e) \{$
 if (size(a) == 1)
 Base case
 return $a[s]$

Divide step
 for $i \leftarrow s$ to e do
 $p = a[s]$
 $l = \{a[i] \mid a[i] < p\}$
 $r = \{a[i] \mid a[i] \geq p\}$

Recurse & combine
 return $F(l) \circ p \circ F(r)$
 $\}$

Properties of Quick Sort

Place your
Webcam Video here
Size 38%

1. The **divide** step can leave the left and right arrays to be imbalanced.

- **Worst case**: $\text{size}(l) = n-1$, $\text{size}(r) = 0$ or vice-versa.

- $T(n) = T(n-1) + n \Rightarrow \Theta(n^2)$

Pivot is max or min at each step.

- **Best case**: $\text{size}(l) = \frac{n-1}{2}$, $\text{size}(r) = \frac{n-1}{2}$

- $T(n) \leq 2 T(\frac{n}{2}) + n \Rightarrow \Theta(n \log n)$

Pivot is the median at each step.

2. Seems like time and space inefficient algorithm.

But not really so!

3. Fastest sorting algorithm in practice.

- **Average case**: $\Theta(n \log n)$. Worst case rarely happens.

Rigorous math behind.

- **In-place** sorting by optimal implementation of inner loop.

The divide step.

4. Space complexity $S(n) = \Theta(n)$.

5. Not a **stable** sort.

Which repeat element is the pivot?

Time complexity

Place your
Webcam Video here
Size 38%

- Let's examine the worst and best cases.

{1, 2, 3, 4, 5, 6, 7}

Worst case rarely happens. So, average case is considered.
 $O(n \log n)$.

{1, 2, 3, 4, 5, 6, 7}

{ } 1 {2, 3, 4, 5, 6, 7}

{1, 2, 3}

4

{5, 6, 7}

{ } 2 {3, 4, 5, 6, 7}

{1}

2

{3}

{5}

6

{7}

{ } 3 {4, 5, 6, 7}

{ } 4 {5, 6, 7}

Worst case

$$\begin{aligned} T(n) &= T(n-1) + n \\ &= \Theta(n^2) \end{aligned}$$

{ } 5 {6, 7}

{ } 6 {7}

Best case

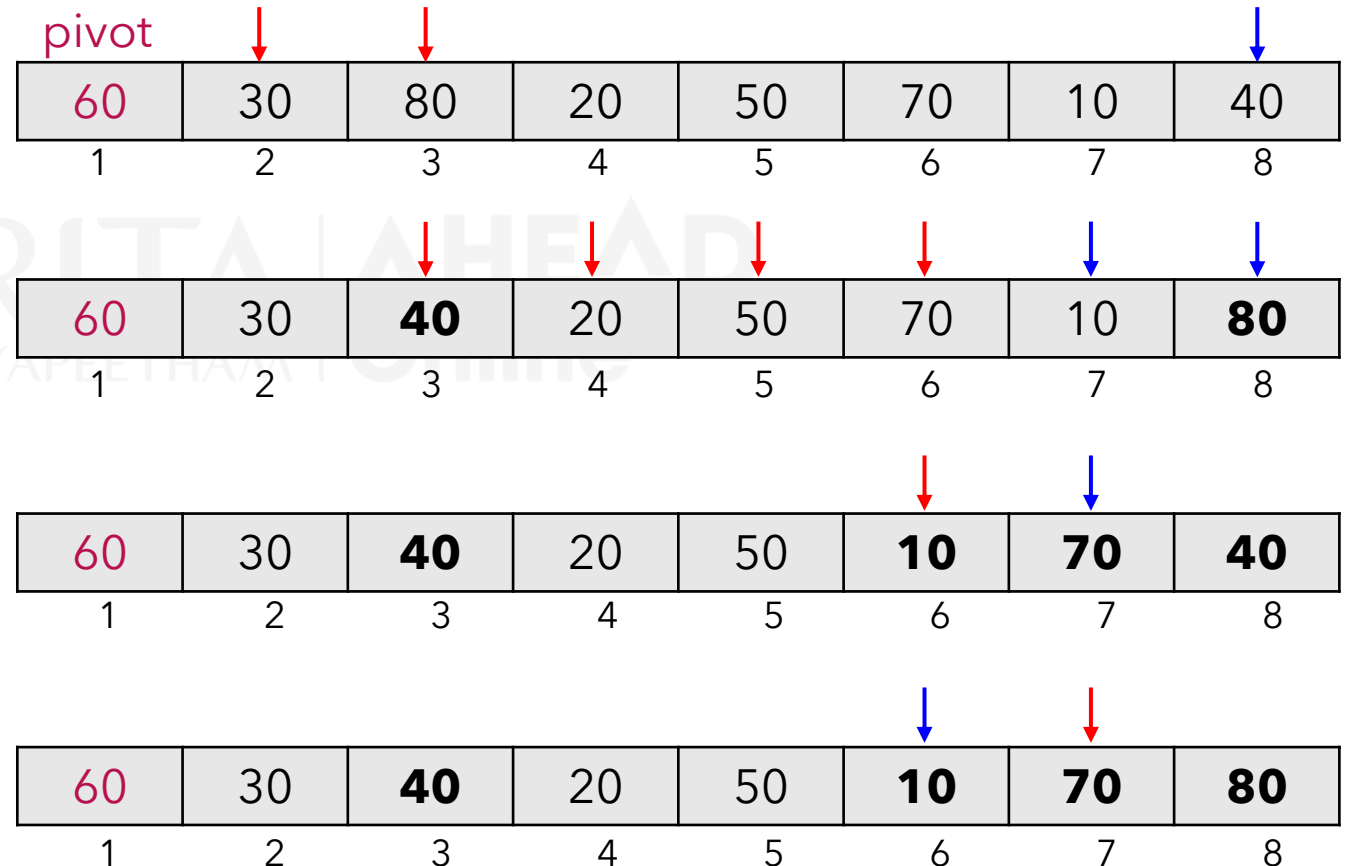
$$\begin{aligned} T(n) &= 2T\left(\frac{n-1}{2}\right) + n \\ &\leq 2T\left(\frac{n}{2}\right) + n \\ &= O(n \log n) \end{aligned}$$

In-place implementation

Place your
Webcam Video here
Size 38%

- As the pointers **left** and **right** move towards each other, smaller element at right is swapped with larger one at left.

- left** = 2
- left**++ → **left** = 3
- right** = 8
- right** remains
- swap_pos(3, 8)
- left**++, **right**--
- **left** = 4, **right** = 7
- left**++ → **left** = 5
- left**++ → **left** = 6
- right** remains
- swap_pos(6, 7)
- left**++, **right**--
- **left** = 7, **right** = 6



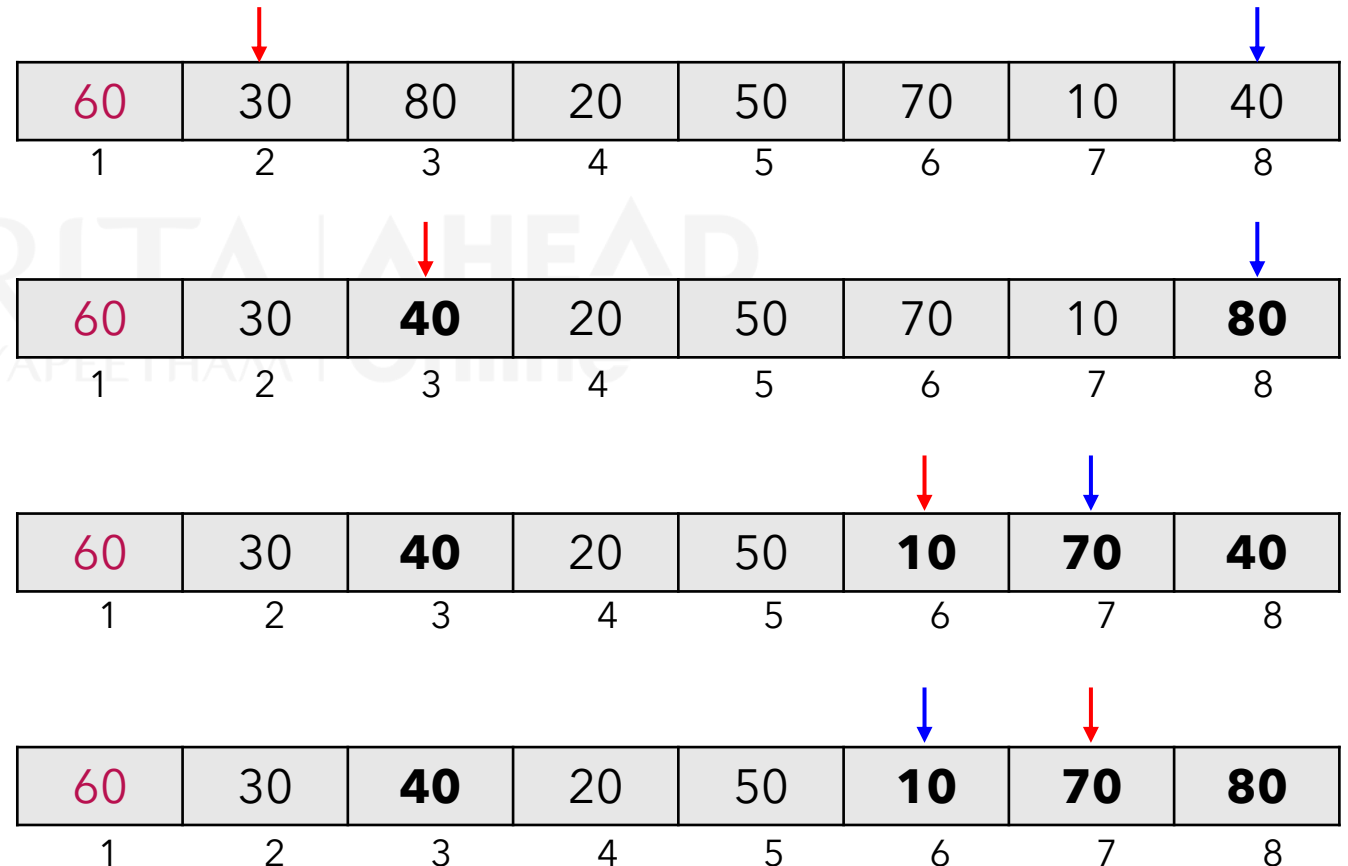
In-place implementation

Place your
Webcam Video here
Size 38%

- As the pointers **left** and **right** move towards each other, smaller element at right is swapped with larger one at left.

```
pivot ← a[1]
left ← 2
right ← n
while left < right do
  while a[left] < pivot do
    left++
  while a[right] ≥ pivot do
    right--
  swap_pos(left, right)
  left++; right--
```

$$T(n) = S(n) = \Theta(n)$$



Pivot selection and Stability

Place your
Webcam Video here
Size 38%

Better pivot => faster algorithm. **Randomized pivot** selection is statistically proved to work well.

```
pivot ← a[1]
left ← 2
right ← n
while left < right do
    while a[left] < pivot do
        left++
    while a[right] ≥ pivot do
        right--
    swap_pos(left, right)
    left++; right--
```

Not stable

$$T(n) = S(n) = \Theta(n)$$

60	30	80	20	50	70	10	40
1	2	3	4	5	6	7	8

Median of left, middle and right is a popular pivot choice strategy too. Here, $\text{median}(60, 20, 40) = 40$.

Question: How will you implement in-place strategy if pivot is not in first position?

Is quicksort stable? No. Because pivot can either be the 1st or 2nd occurrence of a duplicate element.

60	30	80	20	50	70	30	40
----	-----------	----	----	----	----	-----------	----

Selection problem – Quick select

Place your
Webcam Video here
Size 38%

Finding the k^{th} ranked element of an array.

- Pivot based strategy works well here.
- Let's say we are interested in finding 6th ranked element.

1. Best case (pivot)

{6, 3, 2, 4, 8, 1, 7, 5}

~~{3, 2, 4, 1, 5}~~

6

~~{8, 7}~~

size(left) = 5
($k-1$)

Rank
 k

2. In the left

{6, 3, 2, 4, 8, 1, 7, 5}

{3, 2, 4, 1, 5, 6}

~~7~~

~~{8}~~

size(left) ≥ 6
($\geq k$)

select(left, 6)

3. In the right

{6, 3, 2, 4, 8, 1, 7, 5}

~~{2, 1}~~

~~3~~

{6, 4, 8, 7, 5}

size(left) < 6
($< k$)

select(right, 3)
 $6 - 2 - 1 = 3$
rank-size(left)-1=3

A special case of selection problem is finding the median. i.e. $k = n/2$.

- Randomized quick select takes $O(n)$ time to find the median statistically.

6. Maximum sum subarray

Place your
Webcam Video here
Size 38%

Given an array $a[1..n]$, you need to determine the subarray whose sum is the maximum.

$\{-3, \boxed{5, -1, 9}, -6, 2, -8, \boxed{7}\}$
13 7

$$\text{MSS}(a[1..n]) = \text{MSS}(a[1..\frac{n}{2}])$$

Result comes
from left

$\{-3, \boxed{5, -1, 9}, -6, -2, \boxed{8, 7}\}$
13 15

$$\text{MSS}(a[1..n]) = \text{MSS}(a[\frac{n}{2} + 1..n])$$

Result comes
from right

$\{-3, \boxed{5, -1, 9}, \boxed{6, 2}, -8, 7\}$
13 8
21

$$\text{MSS}(a[1..n]) = \text{MSS}(a[1..\frac{n}{2}]) + \text{MSS}(a[\frac{n}{2} + 1..n])$$

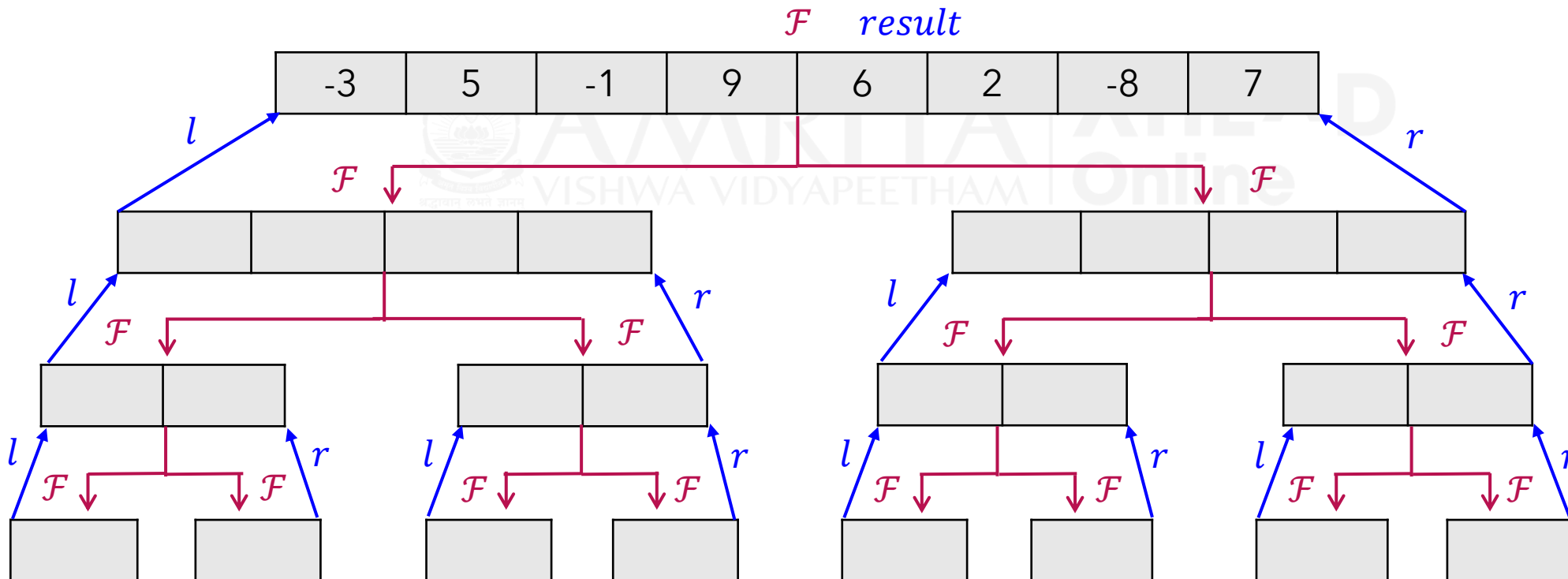
Result is the
sum of both
left and right

6. Maximum sum subarray

Place your
Webcam Video here
Size 38%

$F = \text{MaxSumSubarray}(s, e)$

- Here, $n = 8$. Initially, $s = 1$ and $e = n$ or $(s, e) = (0, n-1)$.
- Divide: Split the array into two halves.
- Combine: Merge 2 sorted subarrays.



Array $a[1..n]$

```
F(a, s, e) {  
  if (s == e)  
    Base case  
    return a[s]
```

Divide step
 $m = (s + e)/2$

```
  l = F(a, s, m)  
  r = F(a, m+1, e)  
  m = .....
```

Combine step
return $\max(l, r, m)$
}

Master's method

Place your
Webcam Video here
Size 38%

A quick method to determine the time complexity for divide-and-conquer solution.

Consider a generalized recurrence equation: $T(n) = a T(\frac{n}{b}) + n^k$

- a denotes the **number** of subproblems
- b denotes the **size** of the subproblem
- n^k denotes the effort for **divide + combine** steps.

$$1. n^{\log_b a} > n^k$$

$$T(n) = \Theta(n^{\log_b a})$$

$$2. n^{\log_b a} < n^k$$

$$T(n) = \Theta(n^k)$$

$$3. n^{\log_b a} = n^k$$

$$T(n) = \Theta(n^k \log_b a)$$



Master's method examples

Place your
Webcam Video here
Size 38%

Three examples are demonstrated below.

$$n^{\log_b a} > n^k$$

$$T(n) = \Theta(n^{\log_b a})$$

$$1. T(n) = 9 T\left(\frac{n}{3}\right) + n$$

- $a = 9, b = 3, k = 1$
- $\log_b a = \log_3 9 = 2$
- Here, $\log_b a > k$
- Hence, $\Theta(n^2)$

$$n^{\log_b a} < n^k$$

$$T(n) = \Theta(n^k)$$

$$2. T(n) = 3 T\left(\frac{n}{4}\right) + n\sqrt{n}$$

- $a = 3, b = 4, k = 1.5$
- $\log_b a = \log_4 3 < 1$
- Here, $\log_b a < k$
- Hence, $\Theta(n^{1.5})$

$$n^{\log_b a} = n^k$$

$$T(n) = \Theta(n^k \log_b a)$$

$$3. T(n) = T\left(\frac{2n}{3}\right) + 1$$

- $a = 1, b = 3/2, k = 0$
- $\log_{3/2} 1 = 0$
- Here, $\log_b a = k$
- Hence, $\Theta(\log_{3/2} n)$

Apply this for the algorithms discussed so far.

Summary

- With this we complete the discussion on divide-and-conquer strategy.



Place your
Webcam Video here
Size 100%



AHEAD Online

Design and Analysis of Algorithms

Dr. Swaminathan J, Ms. Deepa Sreedhar
Department of Computer Science
Amrita Vishwa Vidyapeetham

Objectives

- To look at few optimization problems and recursive formulation of their solutions.



Place your
Webcam Video here
Size 100%



Optimization Problems

Place your
Webcam Video here
Size 38%

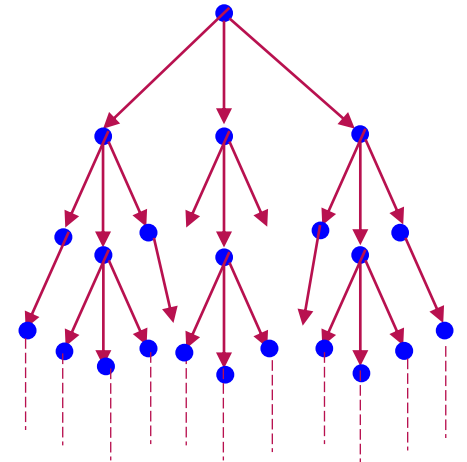
In an optimization problem, the goal is to maximize or minimize an objective subject to given constraints.

1. Make an amount with fewest currencies possible.
 - **Denominations:** ₹1, ₹2, ₹5, ₹10, ₹20, ₹50, ₹100. **Make** ₹492.
$$4 \times ₹100 + 1 \times ₹50 + 2 \times ₹20 + 1 \times ₹2 [8]$$
2. Schedule tasks to minimize average finish time.
 - **Task duration** $T[1..5]$: {8, 3, 5, 11, 6}
$$T[2], T[3], T[5], T[1], T[4]$$
3. Finding the longest increasing subsequence.
 - **Sequence** $S[1..12]$: 2, 5, -1, 8, 0, 4, 6, 11, 7, 9, 12, 10, 14
$$-1, 0, 4, 6, 7, 9, 12, 14$$
4. Finding a subarray that makes the maximum sum.
 - **Array** $A[1..8]$: {4, -7, 5, 1, -3, -2, 8, -6}
$$A[3..7] \text{ yields the sum } 9$$
5. Determine the longest common subsequence of 2 strings.
 - **Strings:** AGGCCTCCAGAATTGA, CCAAACACATATCGAG
$$AAAATTGA$$

Recursive formulation

Place your
Webcam Video here
Size 38%

- An indispensable method to **recursively breakdown** an **optimization problem** into subproblem.
 - Let F be the function that solves a **problem of size n** . The problem is broken down into **subproblem of size $n-k$** .
 - $F(n) = F(n-k) + \text{extra ops}$ (Usually $k = 1$)
 - Recursion continues: $F(n-k) \rightarrow F(n-2k) \rightarrow \dots \rightarrow F(1)$
- There may be **multiple ways** to breakdown.
 - One of them would lead to the optimal solution.
 - You may or may not know which one leads to solution.
- Recursion leads to a **tree of choices**/possibilities.
 - The tree reveals **certain properties** which is what we are interested in. They help choose the best strategy to solve.

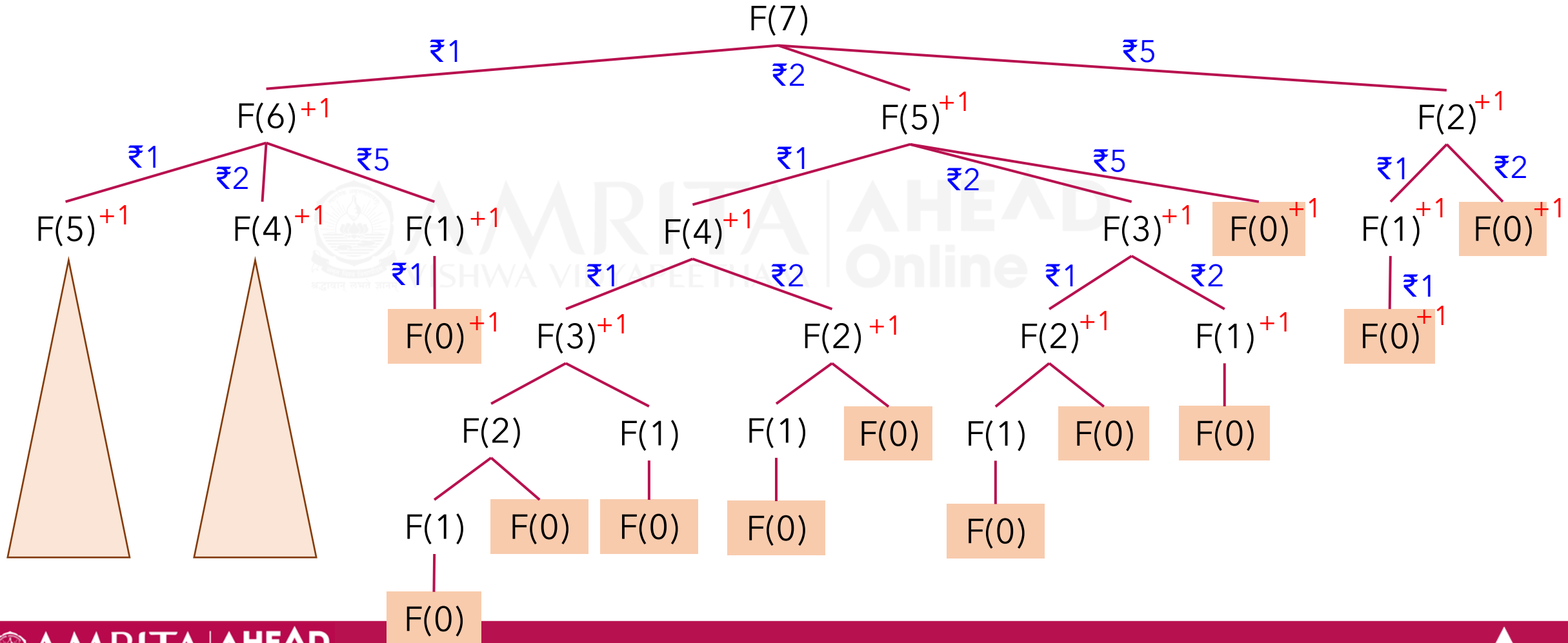


1. Optimal substructure
2. Greedy choice
3. Overlapping subproblems

1. Currency exchange - Example

Place your
Webcam Video here
Size 38%

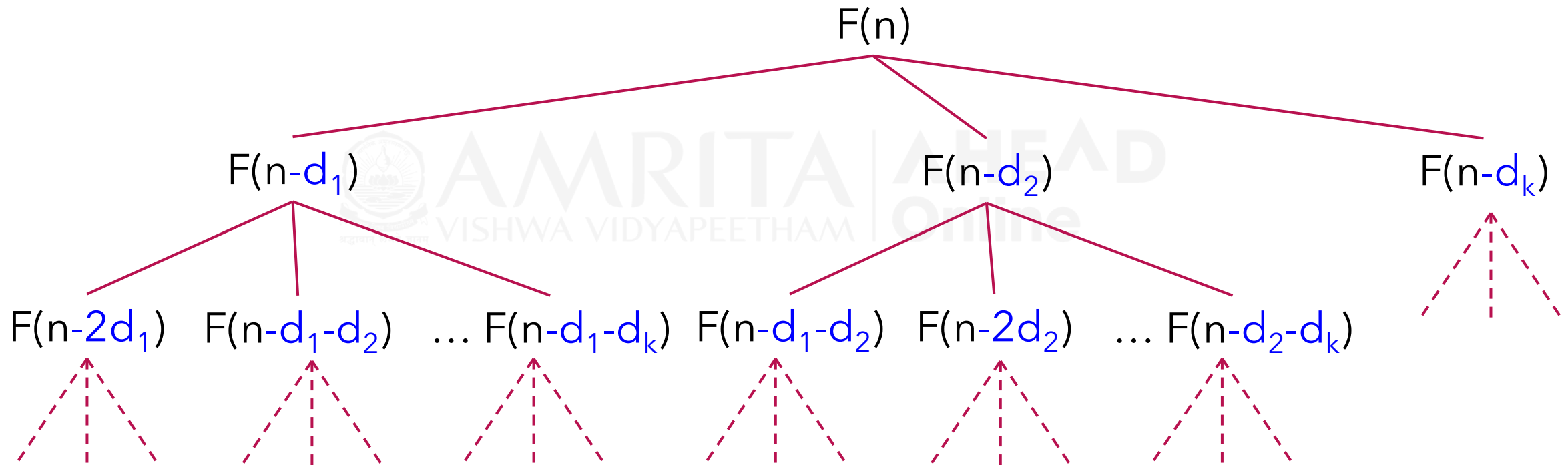
- Denominations: ₹1, ₹2, ₹5. $F = \text{MakeChange}(7)$.



1. Currency exchange problem

Place your
Webcam Video here
Size 38%

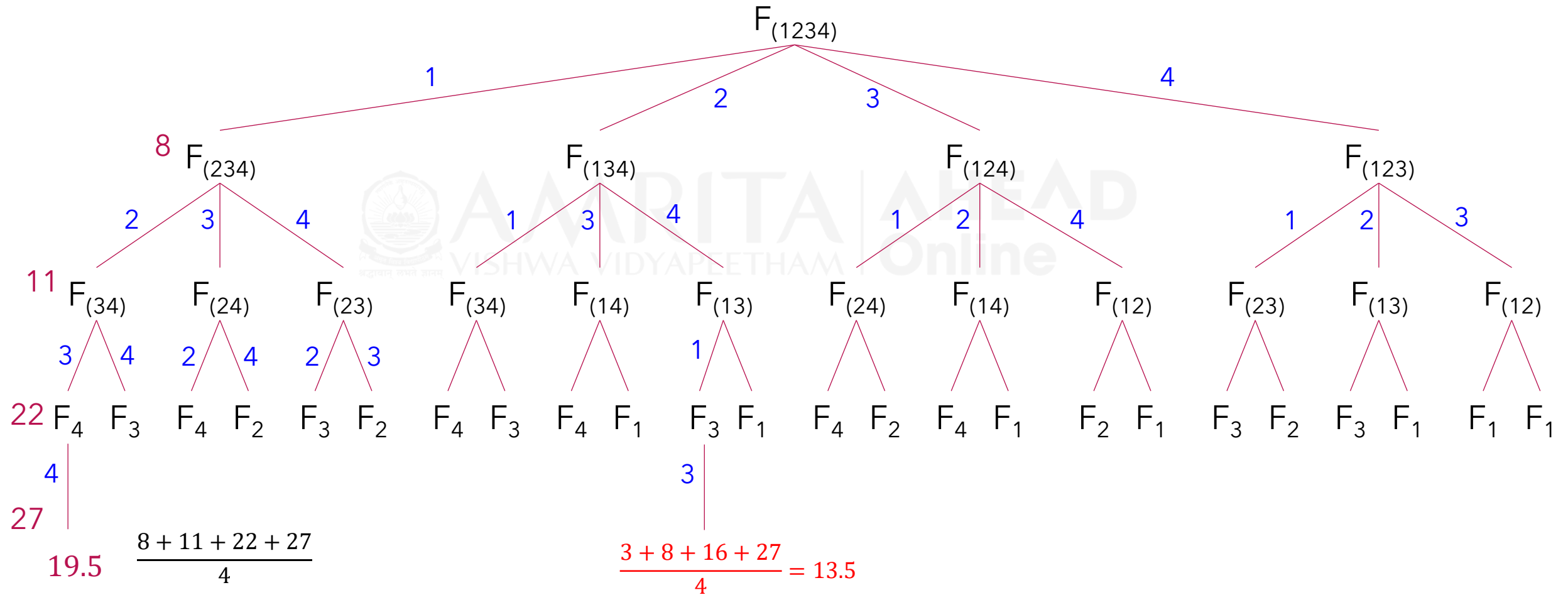
- **Denominations:** d_1, d_2, \dots, d_k . $F = \text{MakeChange}(n)$



2. Task scheduling - Example

Place your
Webcam Video here
Size 38%

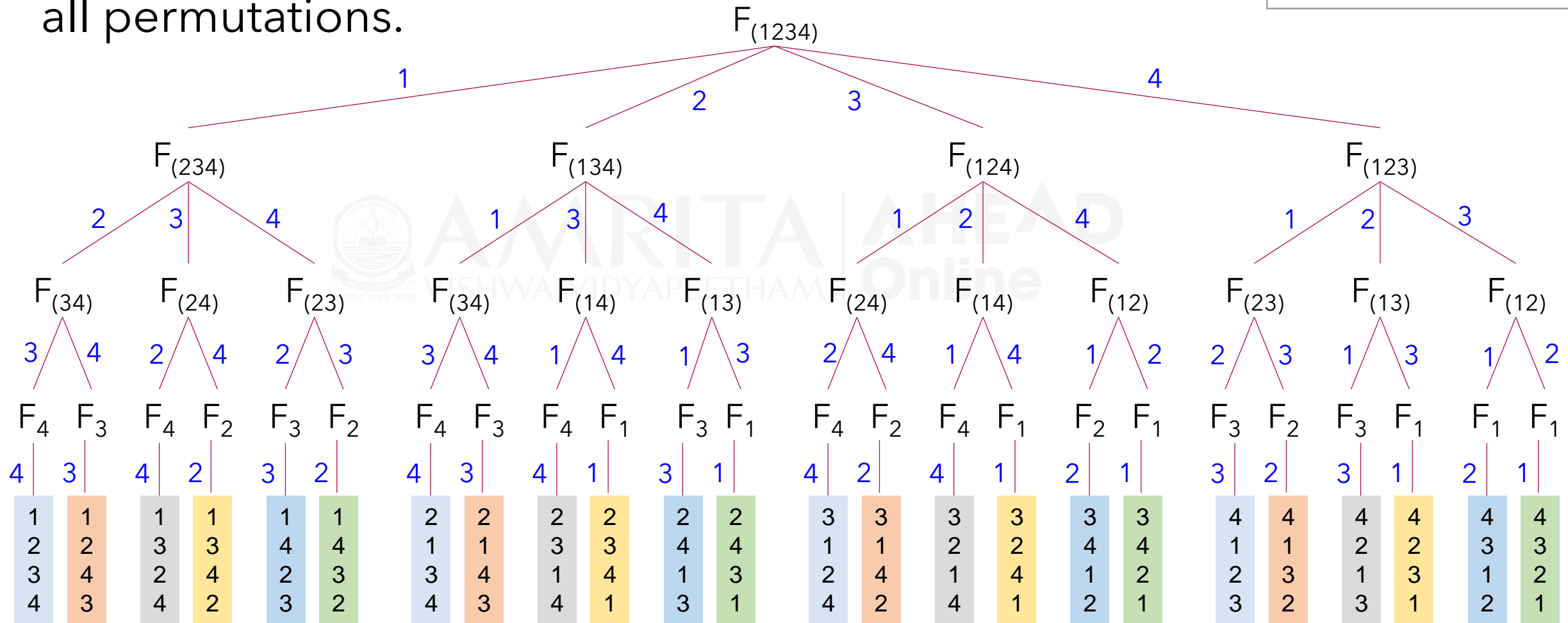
Task duration: {8, 3, 11, 5}. $F = \text{TaskSchedule}(1..4)$.



2. Generating all permutations

Place your
Webcam Video here
Size 38%

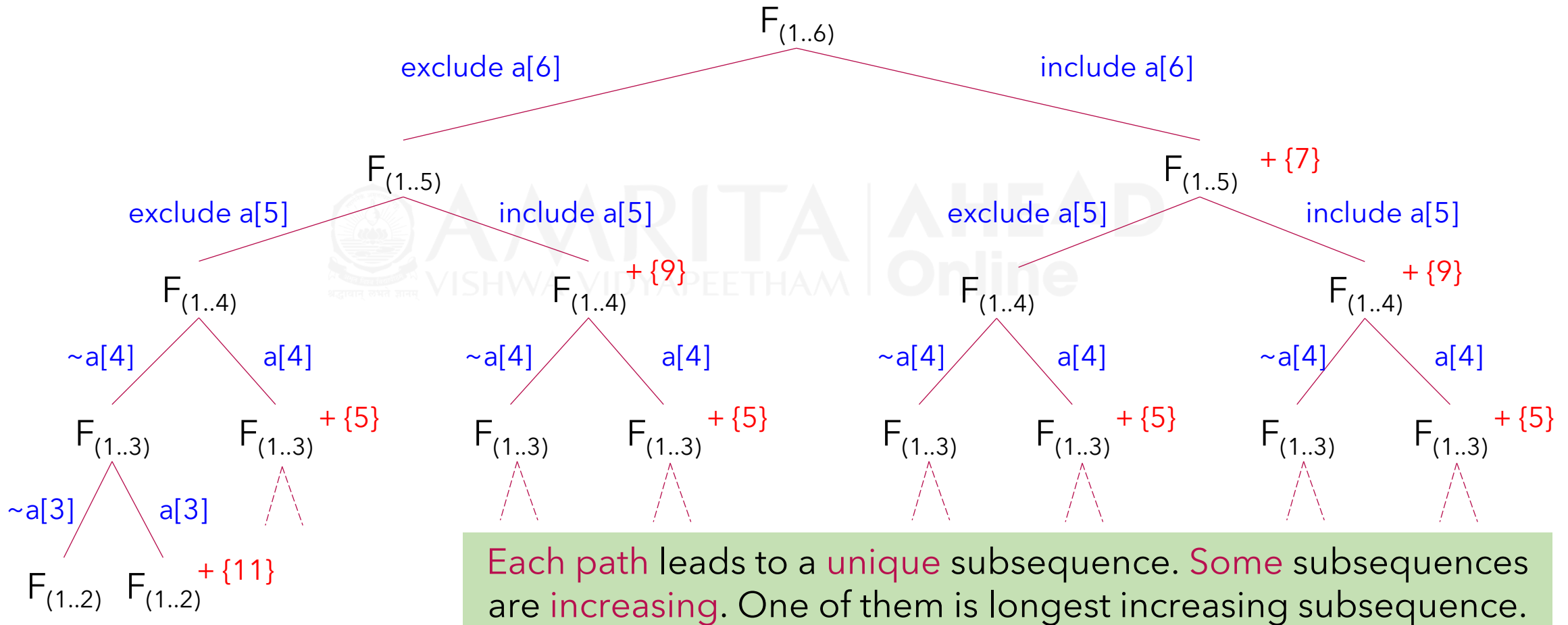
Recursive formulation provides a method to generate all permutations.



3. Longest Increasing Subsequence

Place your
Webcam Video here
Size 38%

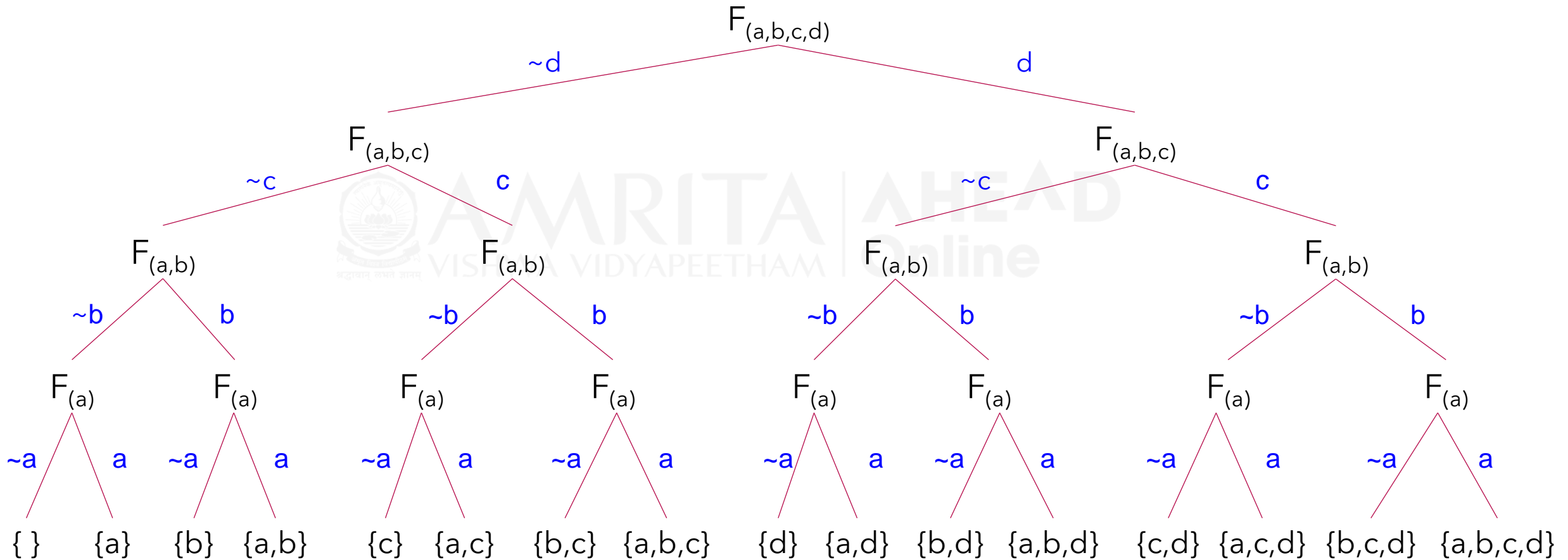
Array $a[1..6] = \{8, 3, 11, 5, 9, 7\}$. $F = \text{LIS}(1..6)$.



Generating all sub sequences

Place your
Webcam Video here
Size 38%

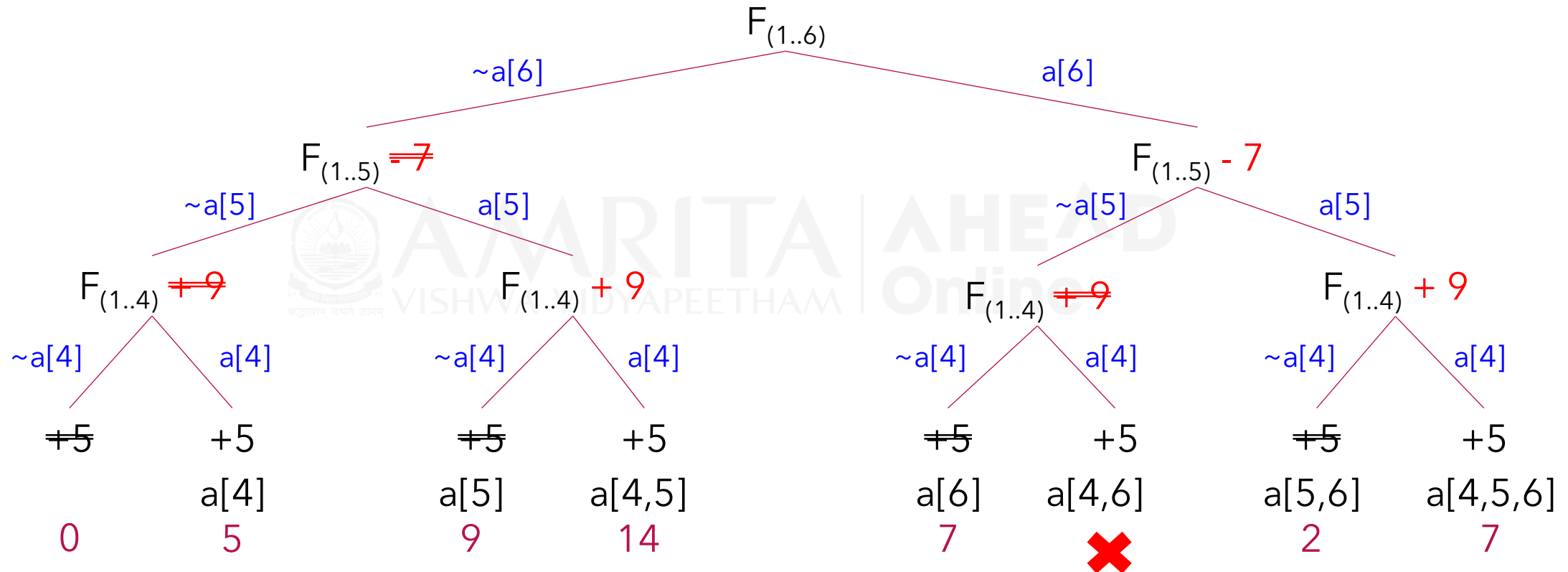
Array $a[1..n] = \{a, b, c, d\}$. $F = \text{LIS}(1..4)$.



4. Maximum Sum Subarray

Place your
Webcam Video here
Size 38%

Array $a[1..6] = \{-8, 3, -11, 5, 9, -7\}$. $F = \text{MSS}(1..6)$.



5. Longest Common Subsequence

Place your
Webcam Video here
Size 38%

Given two strings $S[1..n]$ and $T[1..m]$, find $\text{LCS}(S, T)$. Let's work out few examples.

1. $\text{LCS}(\text{"algae"}, \text{"league"}) = \text{"age"} \text{ or } \text{"lge"} \text{ (length 3)}$
 - $S[n] == T[m] \rightarrow \text{LCS}(\text{"alga"}, \text{"leagu"}) + 'e'$
2. $\text{LCS}(\text{"algaex"}, \text{"league"}) = \text{"age"} \text{ or } \text{"lge"} \text{ (length 3)}$
 - $S[n] \neq T[m]$, but $S[n-1] == T[m] \rightarrow \text{LCS}(\text{"algae"}, \text{"league"})$
3. $\text{LCS}(\text{"algae"}, \text{"leaguey"}) = \text{"age"} \text{ or } \text{"lge"} \text{ (length 3)}$
 - $S[n] \neq T[m]$, but $S[n] == T[m-1] \rightarrow \text{LCS}(\text{"algae"}, \text{"league"})$

Both $S[n]$ and $T[m]$ are useful

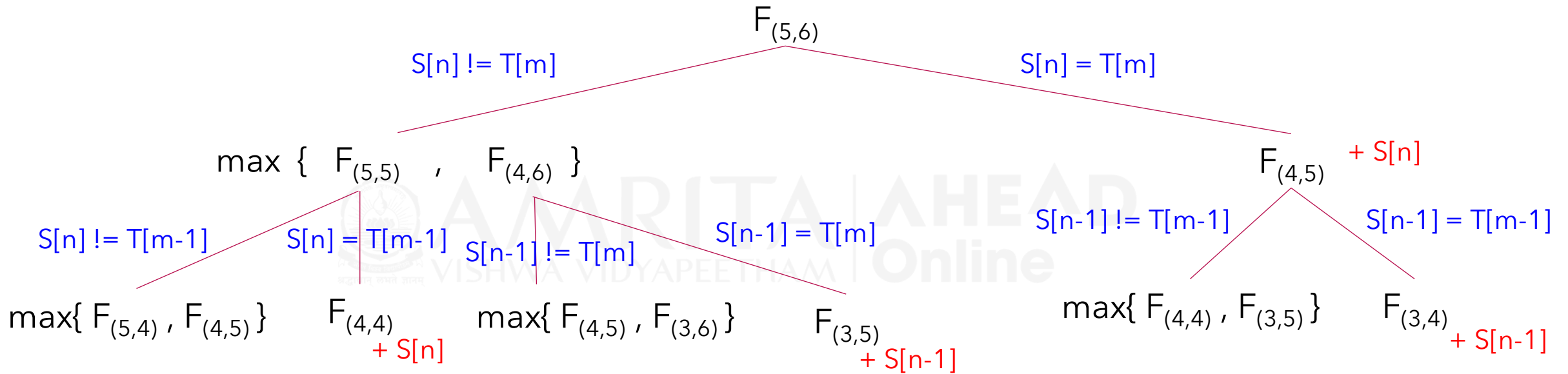
$T[m]$ can be useful

$S[n]$ can be useful

LCS generalized

Place your
Webcam Video here
Size 38%

$S[1..n] = \text{"algae"}, T[1..m] = \text{"league"}$. $F = \text{LCS}(n, m)$.



Summary

- We introduced few optimization problems and discussed the recursive formulation of solutions to the problems.

Place your
Webcam Video here
Size 100%



AHEAD Online

Design and Analysis of Algorithms

Dr. Swaminathan J
Department of Computer Science
Amrita Vishwa Vidyapeetham

Objectives

- To critically analyze the recursive formulation.
- To ascertain the properties necessary for applying greedy strategy.
- To discuss problems for which greedy strategy can be applied.

Place your
Webcam Video here
Size 100%

Optimization Problems Recap

Place your
Webcam Video here
Size 38%

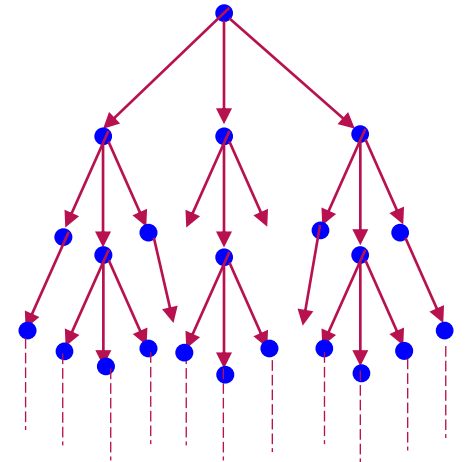
In an optimization problem, the goal is to maximize or minimize an objective subject to given constraints.

1. Make an amount with fewest currencies possible.
 - **Denominations:** ₹1, ₹2, ₹5, ₹10, ₹20, ₹50, ₹100. **Make** ₹492.
$$4 \times ₹100 + 1 \times ₹50 + 2 \times ₹20 + 1 \times ₹2 [8]$$
2. Schedule tasks to minimize average finish time.
 - **Task duration** $T[1..5]$: {8, 3, 5, 11, 6}
$$T[2], T[3], T[5], T[1], T[4]$$
3. Finding the longest increasing subsequence.
 - **Sequence** $S[1..12]$: 2, 5, -1, 8, 0, 4, 6, 11, 7, 9, 12, 10, 14
$$-1, 0, 4, 6, 7, 9, 12, 14$$
4. Finding a subarray that makes the maximum sum.
 - **Array** $A[1..8]$: {4, -7, 5, 1, -3, -2, 8, -6}
$$A[3..7] \text{ yields the sum } 9$$
5. Determine the longest common subsequence of 2 strings.
 - **Strings:** AGGCCTCCAGAATTGA, CCAAACACATATCGAG
$$AAAATTGA$$

Recursive formulation Recap

Place your
Webcam Video here
Size 38%

- An indispensable method to recursively breakdown an optimization problem into subproblem.
 - Let F be the function that solves a problem of size n . The problem is broken down into subproblem of size $n-k$.
 - $F(n) = F(n-k) + \text{extra ops}$ (Usually $k = 1$)
 - Recursion continues: $F(n-k) \rightarrow F(n-2k) \rightarrow \dots \rightarrow F(1)$
- There may be multiple ways to breakdown.
 - One of them would lead to the optimal solution.
 - You may or may not know which one leads to solution.
- Recursion leads to a tree of choices/possibilities.
 - The tree reveals certain properties which is what we are interested in. They help choose the best strategy to solve.



1. Optimal substructure
2. Greedy choice
3. Overlapping subproblems

The 3 properties

Place your
Webcam Video here
Size 38%

- You have a recursive formulation of a problem which contains layers of subproblems.
- You look for the following properties.
 1. **Optimal Substructure**: The optimal solution to a problem contains within itself the optimal solution to its subproblems.
 2. **Greedy choice**: Picking locally optimal choice all the way leads to globally optimal solution.
 3. **Overlapping subproblems**: Breaking down a problem leads to repeated subproblems.

1 & 2 holds →
Apply greedy technique

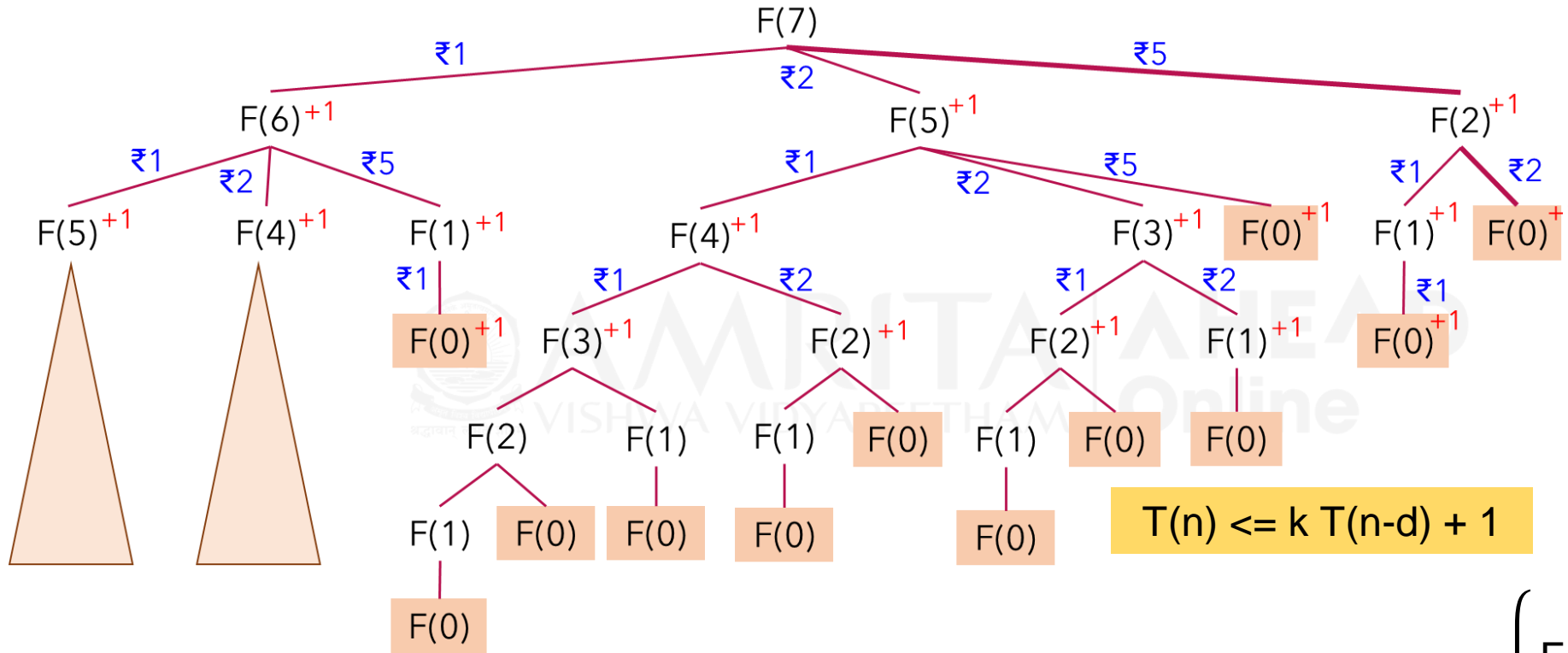
1 & 3 holds →
Apply dynamic programming

Let's now look for these on the problems we introduced.

1. Currency exchange

Let's revisit the example.

Place your
Webcam Video here
Size 38%



- ✓ Optimal substructure
- ✓ Greedy choice
- ✓ Overlapping subproblems

$$T(n) \leq k T(n-d) + 1$$

Denominations: d_1, \dots, d_k
 $F(n) = \text{MakeChange}(n)$

Minimize $\sum_{i=1}^k c_i$ such
 that $\sum_{i=1}^k c_i * d_i = n$

$$F(n) = \min \begin{cases} 0 & n = 0 \\ F(n - d_1) + 1 & n > d_1 \\ F(n - d_2) + 1 & n > d_2 \\ \dots & \\ F(n - d_k) + 1 & n > d_k \end{cases}$$

Greedy Algorithm

Implementation of greedy algorithm.

Denominations: d_1, \dots, d_k
 $F(n) = \text{MakeChange}(n)$

*Minimize $\sum_{i=1}^k c_i$ such that $\sum_{i=1}^k c_i * d_i = n$*

Place your
Webcam Video here
Size 38%

Algorithm: MakeChange

Input: $n, d[1..k]$

Output: $c[1..k]$

$\text{sort}(d[1..k])$

$c[1] \leftarrow c[2] \leftarrow \dots \leftarrow c[k] \leftarrow 0$

$\text{denom} \leftarrow k$

while $\text{denom} > 0$ **do**

$c[\text{denom}] \leftarrow n / d[\text{denom}]$

$n \leftarrow n \bmod d[\text{denom}]$

$\text{denom} \leftarrow \text{denom} - 1$

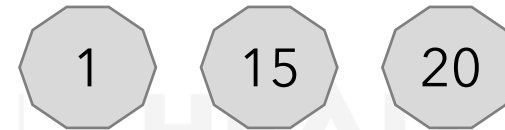
return $c[1..k]$

$T(n) = k \log k + k = O(k \log k)$

The main effort is to **sort** the denominations.

Maxheap can also be used.

Does greedy choice property hold?



MakeChange(30)

No!

Greedy can result in a suboptimal solution.

Variants of the problem

1. Given $d[1..k]$ with limited number of $d[i]$'s, determine the minimum number of currency notes required to make change for amount n .
2. Given $d[1..k]$, can you make a change for n ?

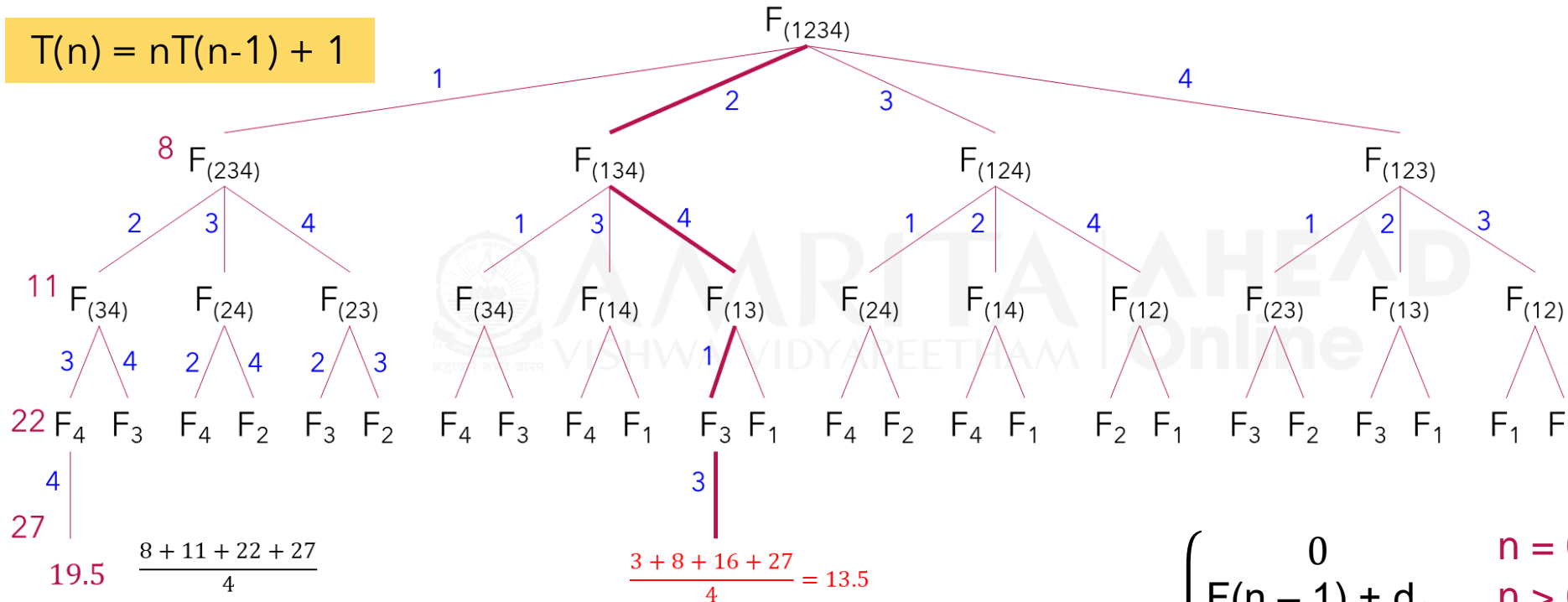
Currency exchange is a particular case of **Knapsack** problem. Similar problems include **subset sum**, **set partition** problems.

2. Task scheduling

Task duration: {8, 3, 11, 5}.

Task durations: d_1, \dots, d_n
 $F(n) = \text{TaskSchedule}(n)$

$$T(n) = nT(n-1) + 1$$



Place your
Webcam Video here
Size 38%

- ✓ Optimal substructure
- ✓ Greedy choice
- ✓ Overlapping subproblems

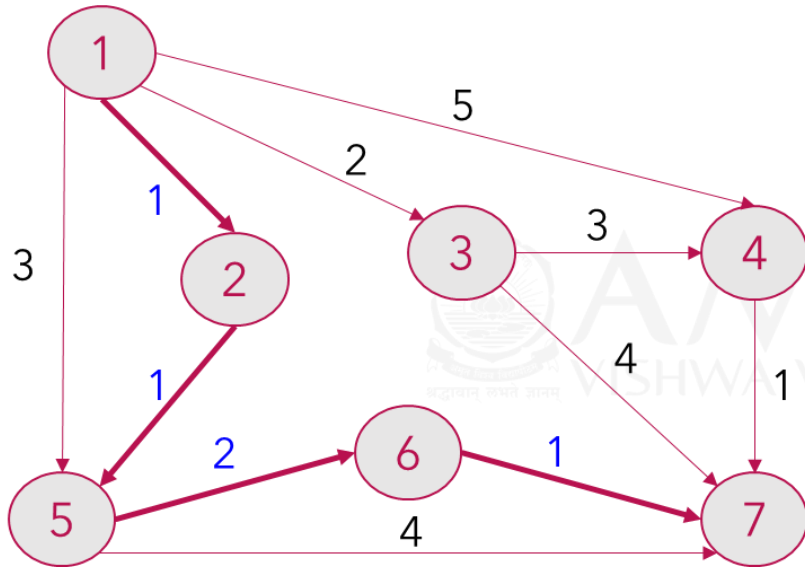
Greedy: Sort by duration, pull out tasks
 $T(n) = O(n \log n)$

$$F(n) = \min \begin{cases} 0 & n = 0 \\ F(n-1) + d_1 & n > 0 \wedge 1^{\text{st}} \text{ not scheduled} \\ F(n-1) + d_2 & n > 0 \wedge 2^{\text{nd}} \text{ not scheduled} \\ \vdots & \\ F(n-1) + d_n & n > 0 \wedge n^{\text{th}} \text{ not scheduled} \end{cases}$$

3. Single source shortest path

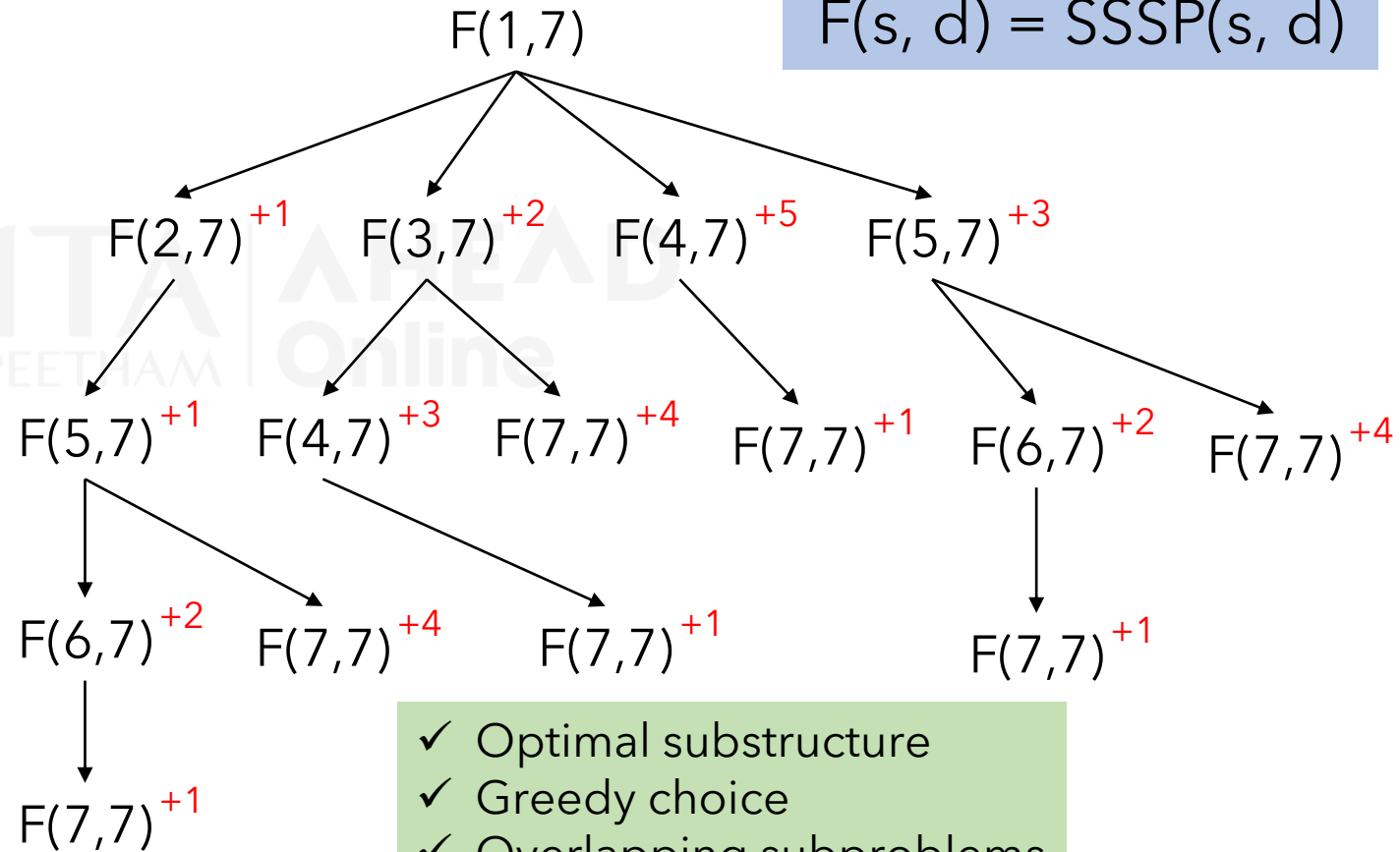
Place your
Webcam Video here
Size 38%

This is a recap from DSA course. We will only look at the recursive formulation and the properties.



$$F(s, d) = \min \begin{cases} 0 & \text{if } s = d \\ \text{adj}[s, v_1] + F(v_1, d) \\ \text{adj}[s, v_2] + F(v_2, d) \\ \dots \end{cases}$$

$$F(s, d) = \text{SSSP}(s, d)$$



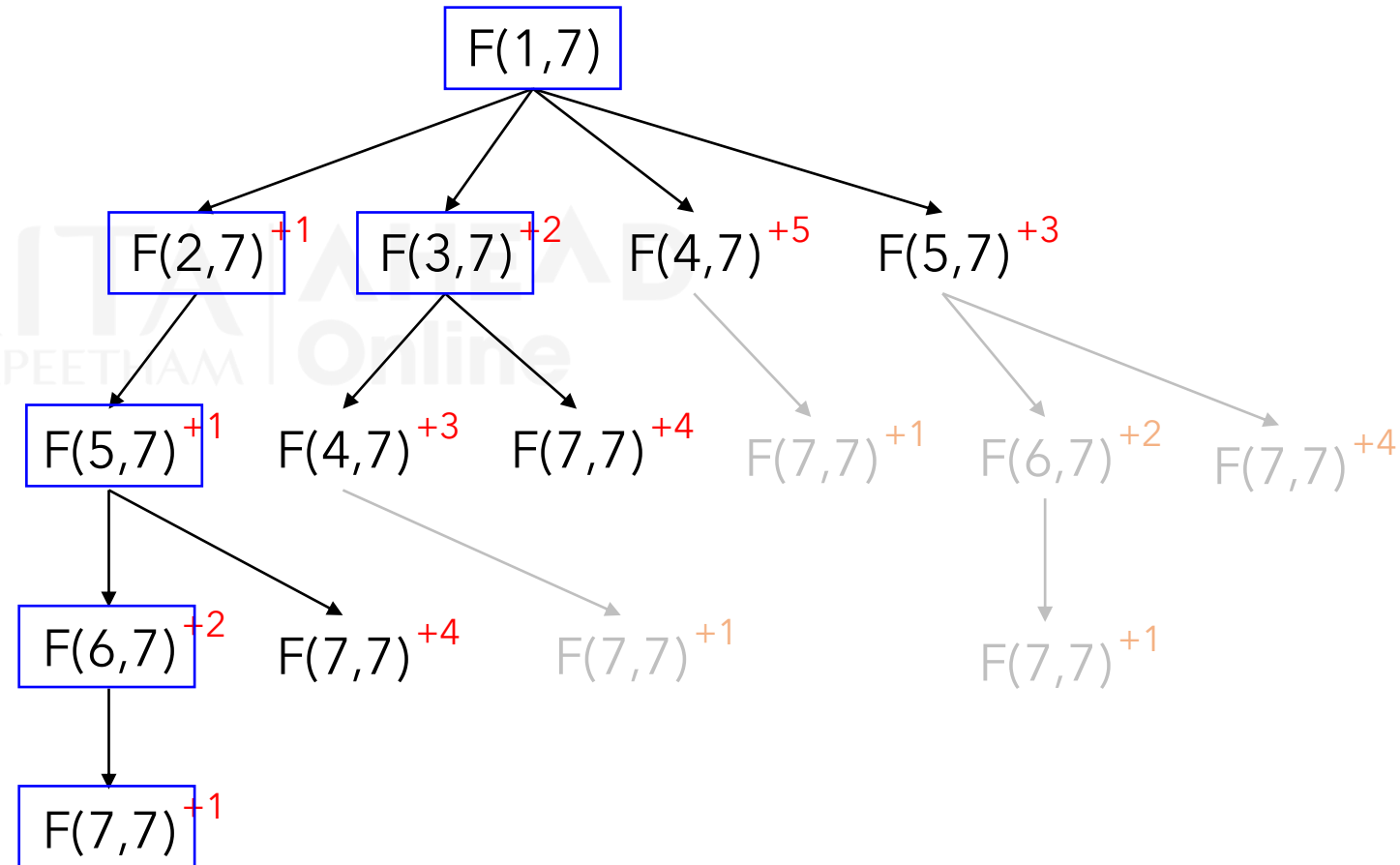
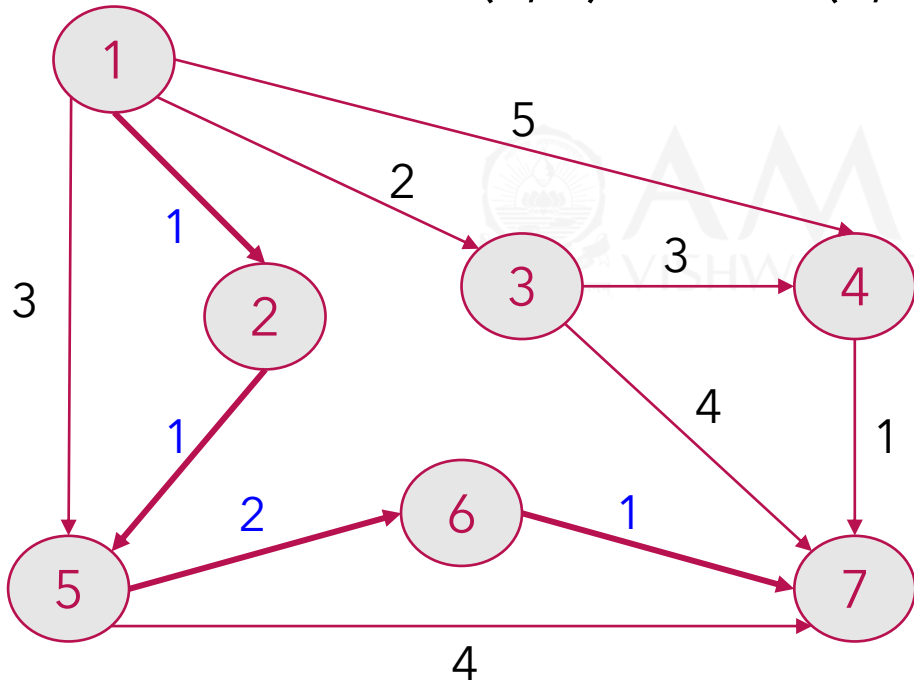
- ✓ Optimal substructure
- ✓ Greedy choice
- ✓ Overlapping subproblems

Dijkstra's Greedy Algorithm

Place your
Webcam Video here
Size 38%

Here is the demonstration of Dijkstra's algorithm runs.

$$F(s,d) = \text{SSSP}(s,d)$$

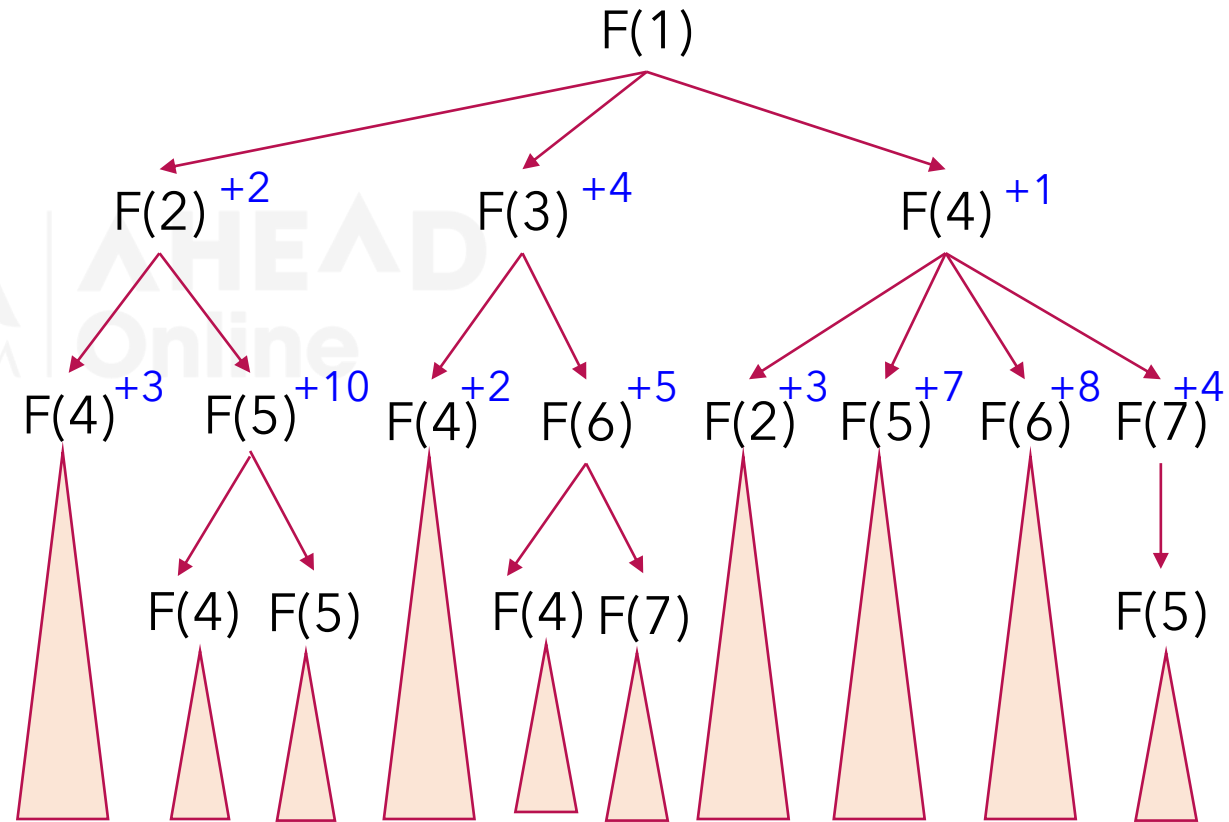
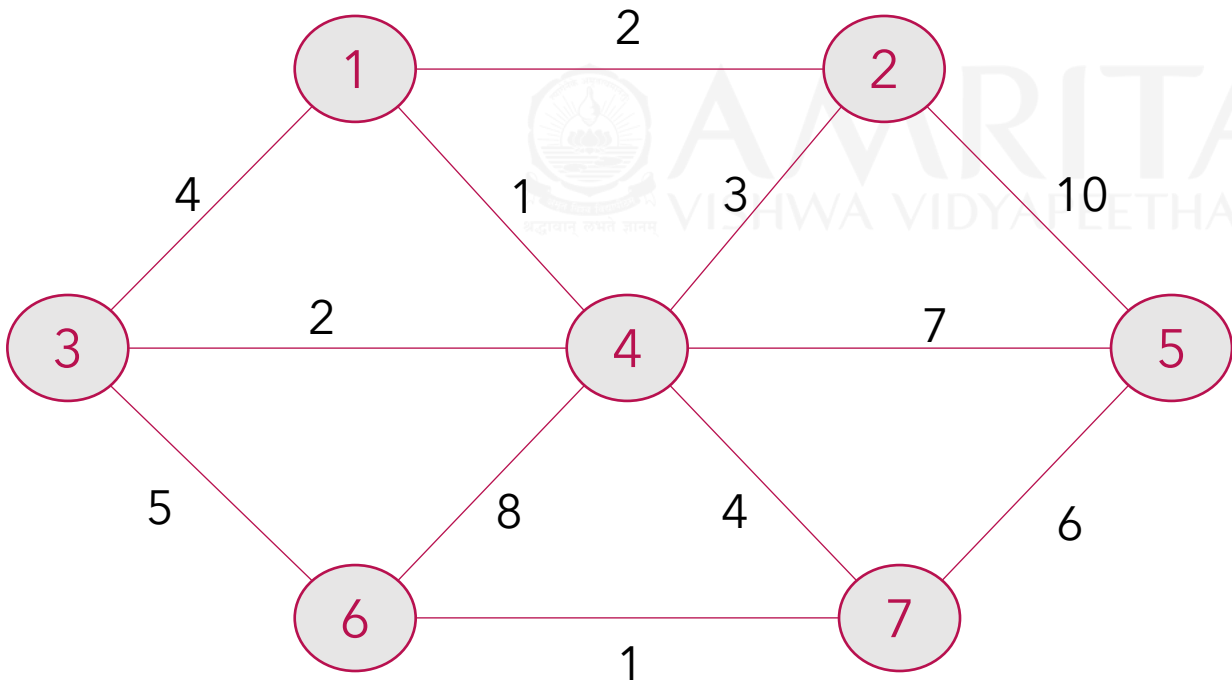


4. Minimum Spanning Tree

Place your
Webcam Video here
Size 38%

This is a recap from DSA course. Formulating the recursive solution and ascertaining properties left as exercise.

$$F(G, G^{\text{mst}}) = \text{MST}(G, G^{\text{mst}})$$

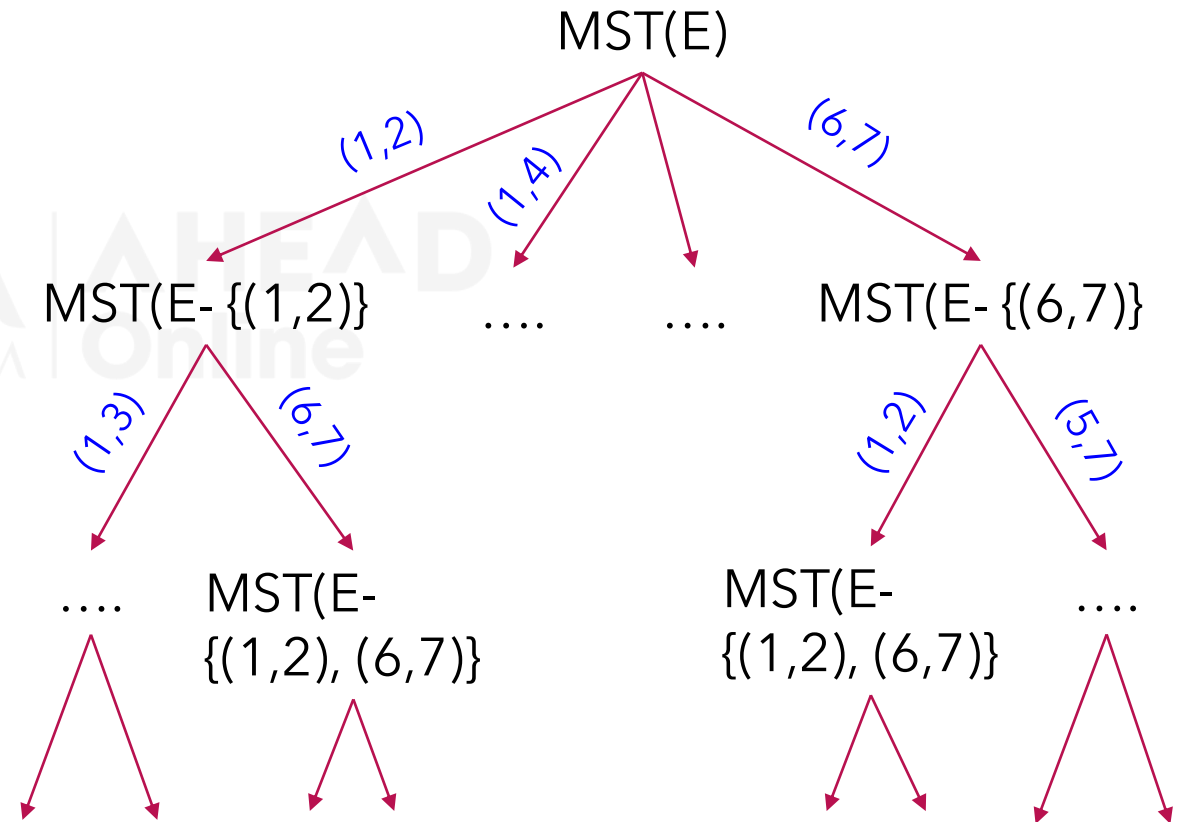
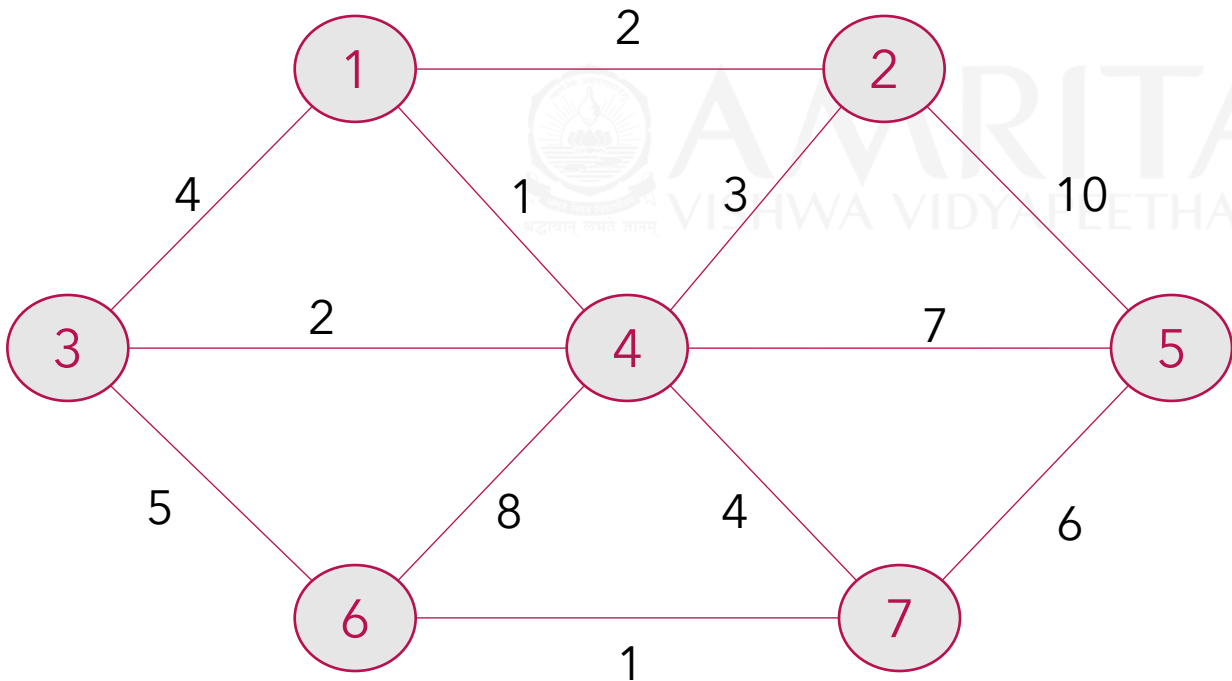


4. Minimum Spanning Tree

Place your
Webcam Video here
Size 38%

This is a recap from DSA course. Formulating the recursive solution and ascertaining properties left as exercise.

$$F(E, E^{\text{mst}}) = \text{MST}(E, E^{\text{mst}})$$



Huffman's Encoding - Introduction

Place your
Webcam Video here
Size 38%

A greedy algorithm to compress text made up of chars.

Let's say alphabet contains 4 chars. {a, b, c, d}.

- 2 bits are sufficient to represent them.
- $a \rightarrow 00$ | $b \rightarrow 01$ | $c \rightarrow 10$ | $d \rightarrow 11$

2) Consider another text file with 100 chars with 60 a's, 25 b's, 10 c's and 5 d's. The total size of text is still $2 \times 100 = 200$ bits.

Can we do better? YES. But how?

Use fewer bits for more frequent chars.

Consider the following representation.

- $a \rightarrow 1$ | $b \rightarrow 01$ | $c \rightarrow 001$ | $d \rightarrow 000$

Text size = $60 \times 1 + 25 \times 2 + 10 \times 3 + 5 \times 3$
 $= 60 + 50 + 30 + 15 = 155$ bits.

1) Consider a text file with 100 chars with 25 a's, 25 b's, 25 c's and 25 d's.

The total size of text = $2 \times 100 = 200$ bits.

Is this representation valid? Why or why not?

- $a \rightarrow 0$ | $b \rightarrow 1$ | $c \rightarrow 01$ | $d \rightarrow 10$

Try decoding the sequence: 0110000111101.

A very important question is: Can we correctly decode the text back?

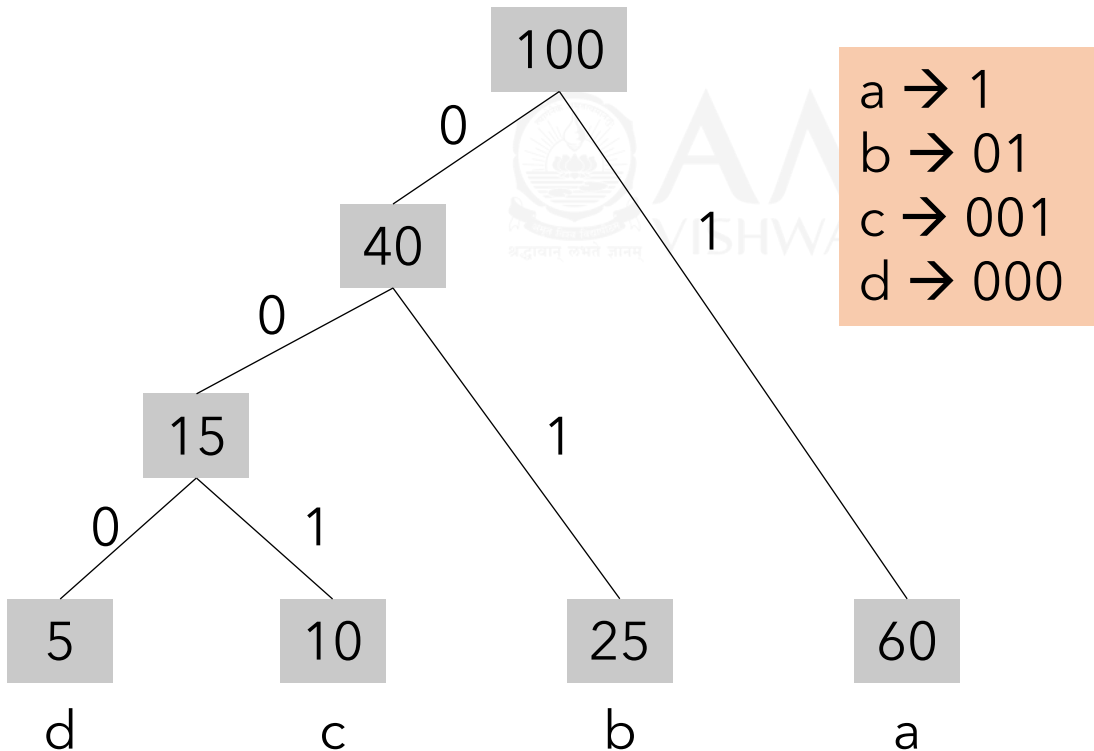
- It depends on how encoding was achieved. Consider the bit sequence: 0110000111101.
- This can be separated as: 0110000111101.
- And equivalent decoding is: badbaab

Huffman's Encoding – Algorithm

Place your
Webcam Video here
Size 38%

Let's work with the examples to compute the encoding.

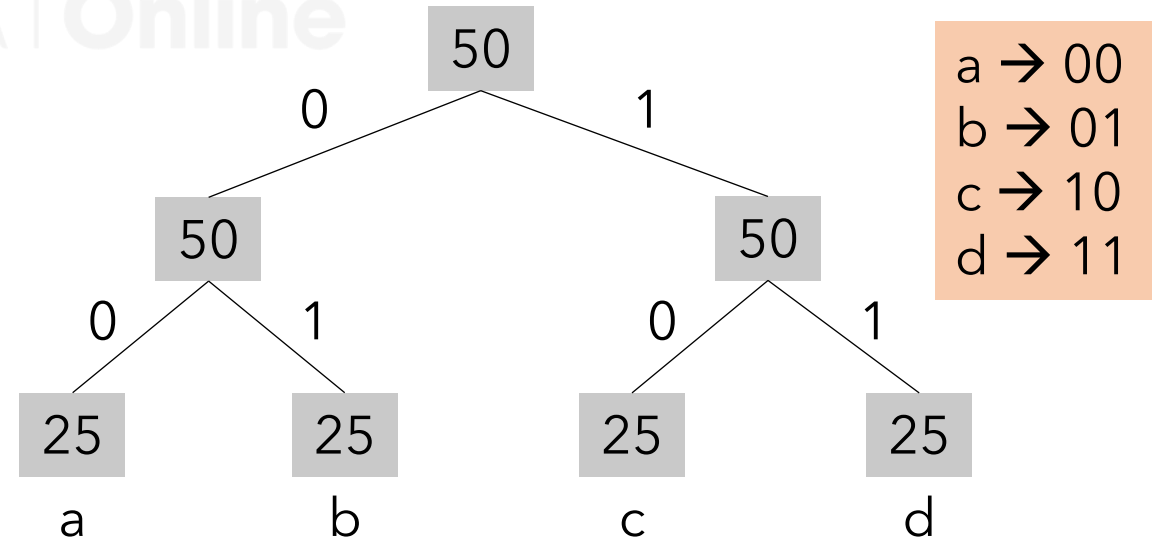
100 chars with 60 a's, 25 b's, 10 c's, 5 d's
• $a \rightarrow 1$ | $b \rightarrow 01$ | $c \rightarrow 001$ | $d \rightarrow 000$



Algorithm: Huffman(occurrence count/%age of symbols)

- Sort the symbols based on occurrence count.
- Make a forest of single node trees - one per symbol.
- Merge two smallest weighted trees until there one tree.

$$T(n) = n \log n + n + n - 1 = O(n \log n)$$



Summary

- We discussed greedy strategy for few optimization problems by analyzing the recursive formulation worked out.

Place your
Webcam Video here
Size 100%