SCHEDULING

MAN

AUGUST 2015

SCHEDULING

- When more than one process is ready to run, but only one CPU is available, a choice is to make
- Part of OS that does it is scheduler
- The algorithm it uses is scheduling algorithm

SCHEDULING

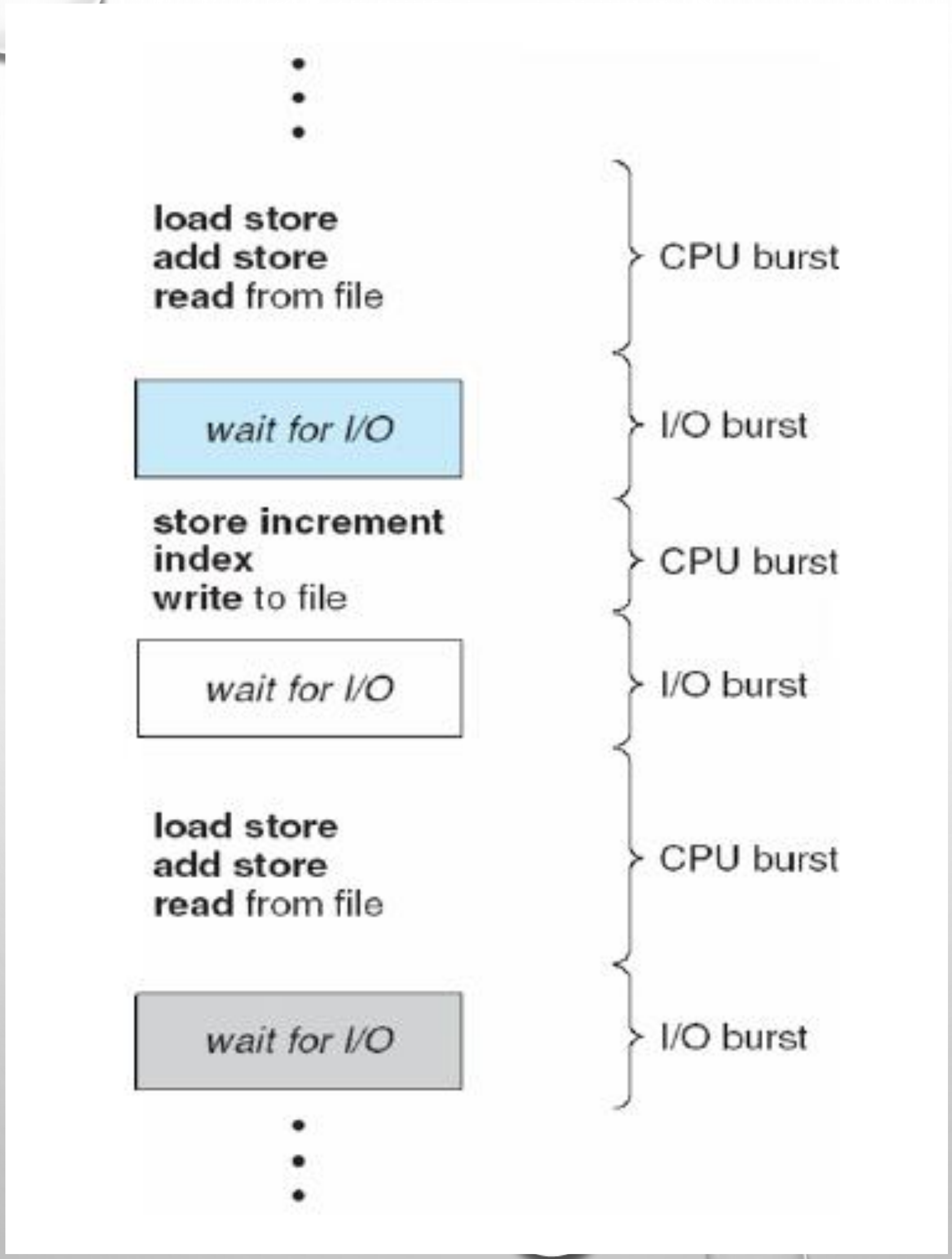
- Efficiency is needed as process switching is costly:
  - Switch from user mode to kernel mode
  - State of current process need to be saved
  - Memory map may be saved
  - A process is selected
  - MMU to be reloaded with memory map of new process
  - New process is started

IMPORTANCE OF SCHEDULING

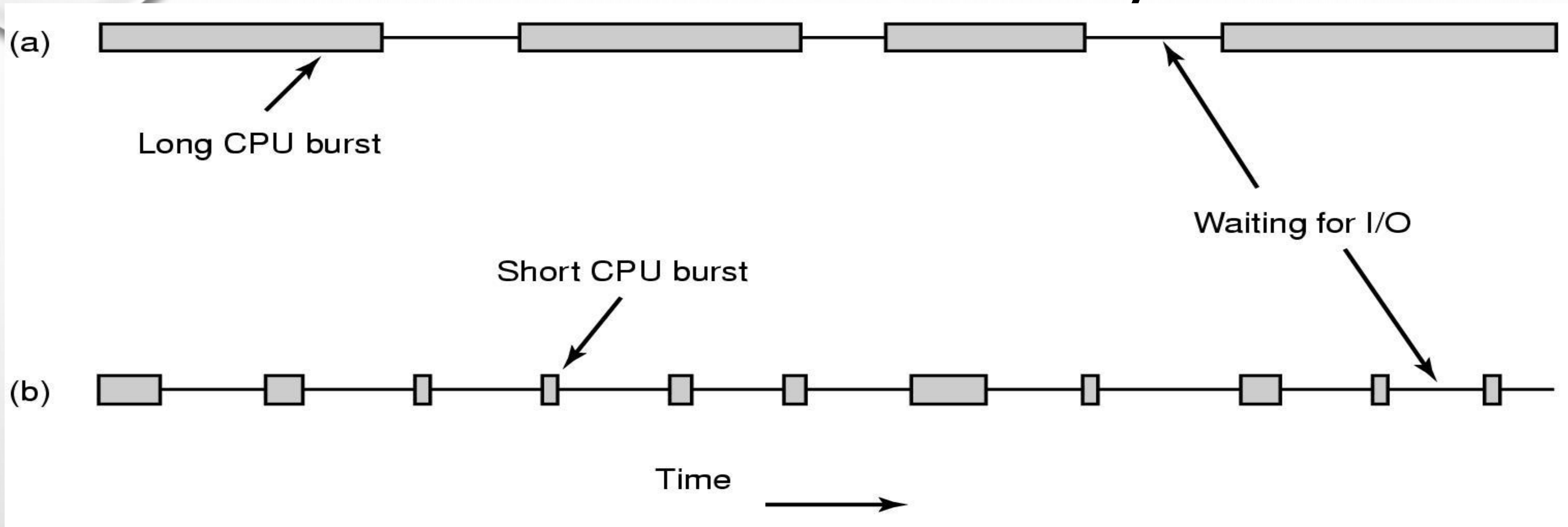
- Good scheduling algorithms can make a big difference
  - Resource utilization
  - Perceived performance & User satisfaction
  - Meeting other system goals (e.g., important tasks being taken care of immediately)

PROCESS BEHAVIOR

- Processes usually alternate bursts of *computing* with *I/O requests.*
- *CPU burst: the amount of time the process uses the processor before it is no longer ready*
- I/O in this sense is when a process enters the blocked state waiting for an external device to complete its work


load store
add store
read from file
CPU burst
wait for I/O
I/O burst
store increment
index
write to file
CPU burst
wait for I/O
I/O burst
load store
add store
read from file
CPU burst
wait for I/O
I/O burst

PROCESS: COMPUTE AND I/O-BOUND



- a CPU-bound process (data encryption/decryption, multimedia encoding)
  - Spend most of the time computing
  - Long CPU bursts => infrequent I/O waits
- an I/O bound process (shell waiting for user commands)
  - Spend most of the time waiting for I/O
  - Short CPU bursts => frequent I/O waits
- Key factor is the **length of CPU burst** not the length of the I/O burst

PROCESS: COMPUTE AND I/O-BOUND

- As the CPUs get faster, processes tend to get more I/O bound: WHY?
- If a I/O bound process is ready, it should get a chance quickly.
  - Increase resource utilization

WHEN TO SCHEDULE

- When a new process is created:
  - Parent or child? Both are Ready
  - which one to run?
- When a process exits:
  - One of the ready processes should be run
- When a process blocks: Another process has to be selected to run
  - Blocking may occur for:
    - I/O
    - Semaphore

WHEN TO SCHEDULE

- When an I/O interrupt occurs:
  - In case of an interrupt of an I/O device having completed its work, some blocked process may now be ready
- If a h/w clock provides periodic interrupt: A scheduling decision can be made at each (or kth) clock interrupt

PREEMPTIVE & NON-PREEMPTIVE

Classification of Scheduling Algorithm depending on dealing with clock interrupt

- Non-preemptive: Picks a process to run and lets it run until it **blocks** or voluntarily releases the CPU. In effect at each clock interrupt, no scheduling is done.
- Preemptive: Picks a process and lets it run for a maximum of some fixed time. If still running, it is suspended and another is picked.
- Preemptive scheduling requires having a clock interrupt occur at the end of the time interval to give control of the CPU back to the scheduler

DIFFERENT SYSTEMS, DIFFERENT FOCUSES

All systems

Fairness - giving each process a fair share of the CPU
Policy enforcement - seeing that stated policy is carried out
Balance - keeping all parts of the system busy

Batch systems

Throughput - maximize jobs per hour
Turnaround time - minimize time between submission and termination
CPU utilization - keep the CPU busy all the time

Interactive systems

Response time - respond to requests quickly
Proportionality - meet users' expectations

Real-time systems

Meeting deadlines - avoid losing data
Predictability - avoid quality degradation in multimedia systems

BATCH SYSTEMS

- Users submit their job to the batch system
- Batch system starts user job when appropriate
- User gets notification that job is done
  - No interaction in between
- No users impatiently waiting at terminals for a quick response to a short request
- Used in business world such as Profit calculation at banks, claims processing at insurance companies…

BATCH SYSTEMS

- Common performance metrics
  - Throughput: number of jobs completed per hour
  - Turnaround time: average time between the submission and completion of a job
- Maximizing Throughput may not necessarily minimize Turnaround time

BATCH SYSTEMS

Algorithms used:

- Non-preemptive
- Preemptive algorithms with long time periods are often acceptable
  - Reduces process switches and improves performance

Representative algorithms:

1. First Come First Serve (FCFS)
2. Shortest Job First
3. Shortest Remaining Time First

FIRST COME FIRST SERVE (FCFS)

- Process that requests the CPU FIRST is allocated the CPU FIRST.
- Also called FIFO
- non-preemptive
- Used in Batch Systems
- Real life analogy?
  - Transaction at Sonali Bank
- Implementation
  - FIFO queues
  - A new process enters the tail of the queue
  - The schedule selects from the head of the queue.

FCFS EXAMPLE

Process	Duration	Order	Arrival Time
P1	24	1	0
P2	3	2	0
P3	4	3	0

The final schedule:

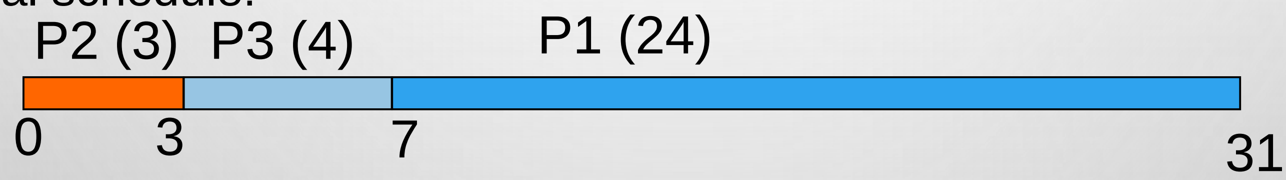
P1 turnaround: 24
P2 turnaround: 27
P3 turnaround: 31

The average turnaround:
(24+27+31)/3 = 27.33

FCFS EXAMPLE 2

Process	Duration	Order	Arrival Time
P1	24	3	0
P2	3	1	0
P3	4	2	0

The final schedule:



P1 turnaround: 31
P2 turnaround: 3
P3 turnaround: 7

The average turnaround:
$(31+3+7)/3 = 13.67$

ADVANTAGE

- Easy to understand and implement
- Fair for equivalent processes

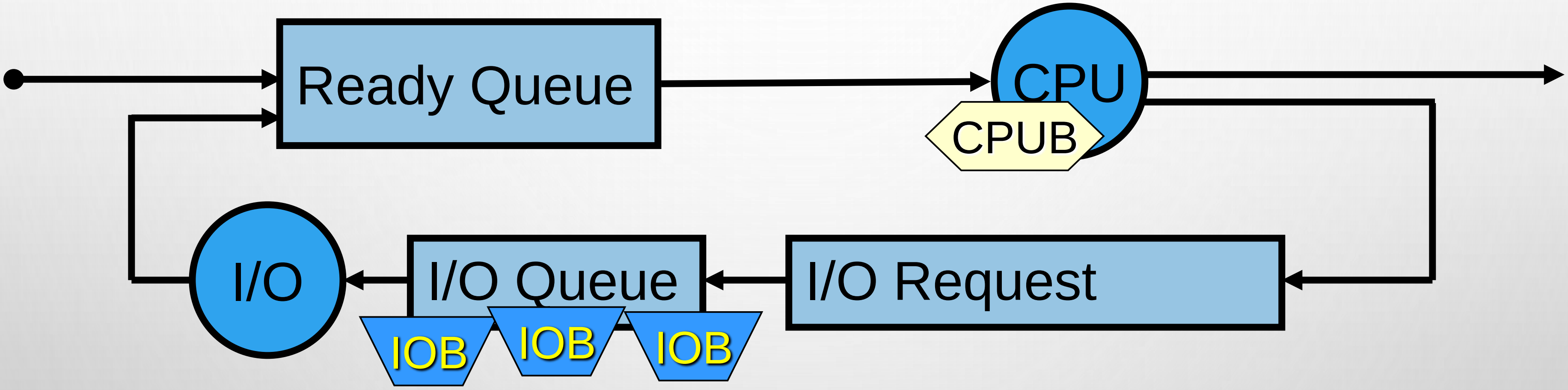
PROBLEMS WITH **FCFS**

- Non-preemptive
- Non optimal turnaround
- Cannot **utilize** resources in parallel:
  - Assume 1 process CPU bounded and many I/O bounded processes
  - result: **Convoy effect**,
    - low CPU **and** I/O Device utilization
  - Why?

CONVOY EFFECT

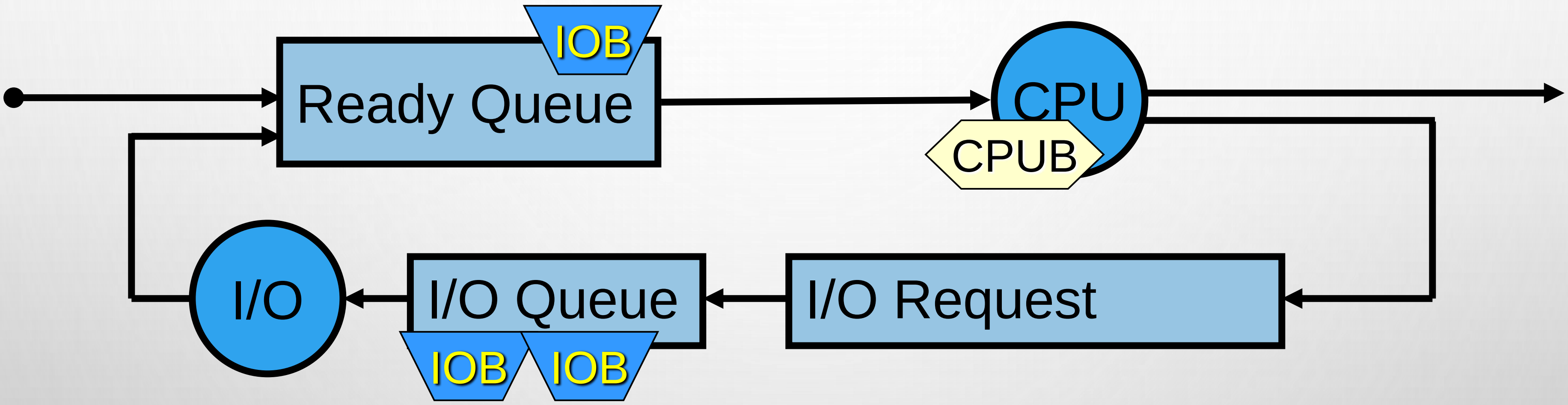
- When the CBP uses the CPU
  - IBPs finish their I/O and move into the ready queue, waiting for the CPU
  - the I/O devices are idle
- When the CBP finally relinquishes the CPU,
  - CBP moves to an I/O device
  - the IBPs pass through the CPU quickly and move back to the I/O queues
  - the CPU is idle
- The cycle repeats itself when the CBP gets back to the ready queue

CONVOY EFFECT



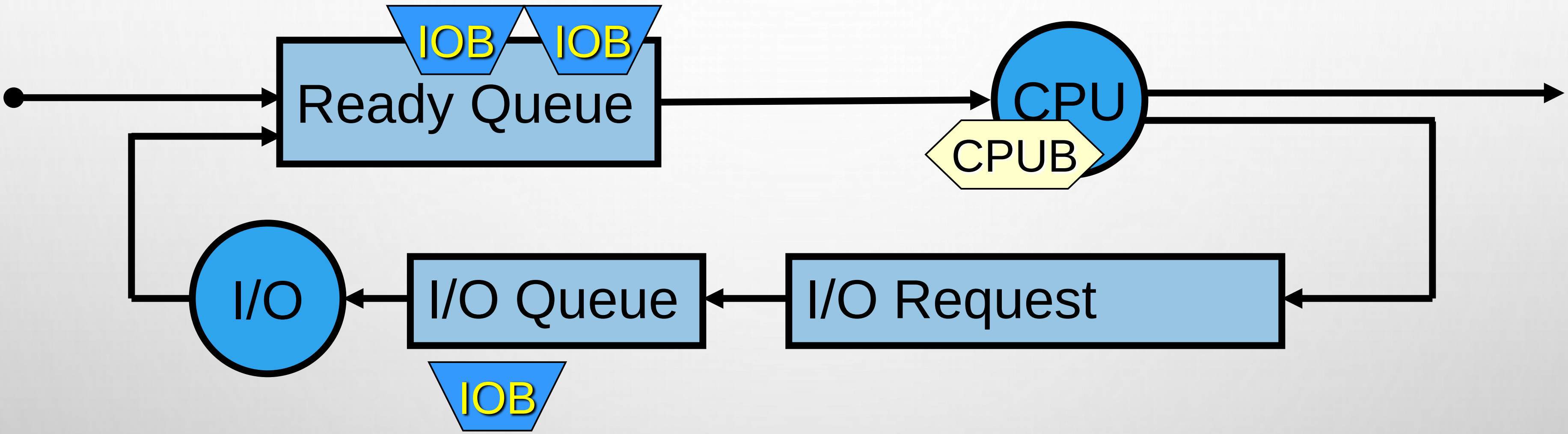
CPU is running CPUB

CONVOY EFFECT



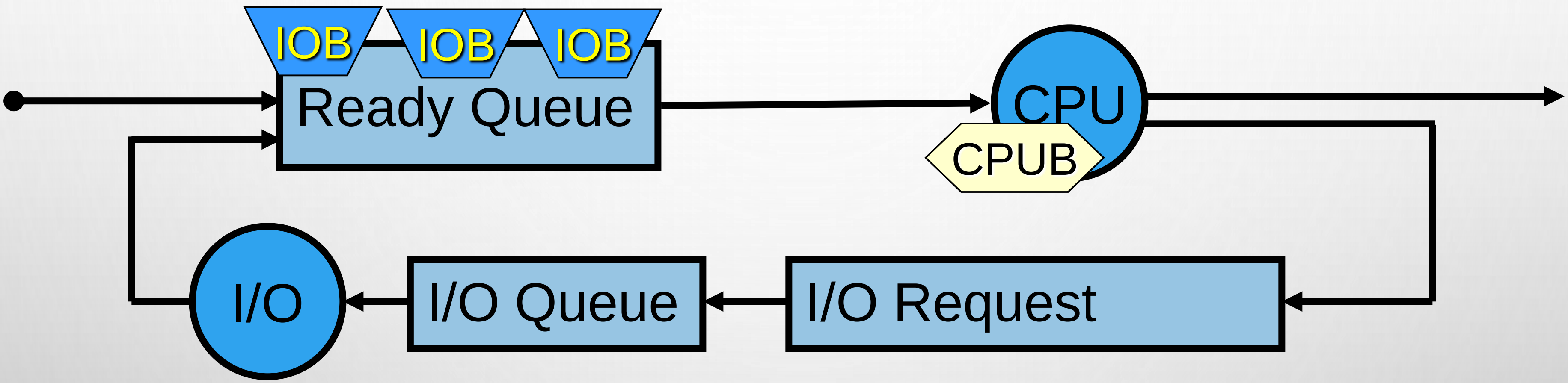
CPU is running CPUB

CONVOY EFFECT



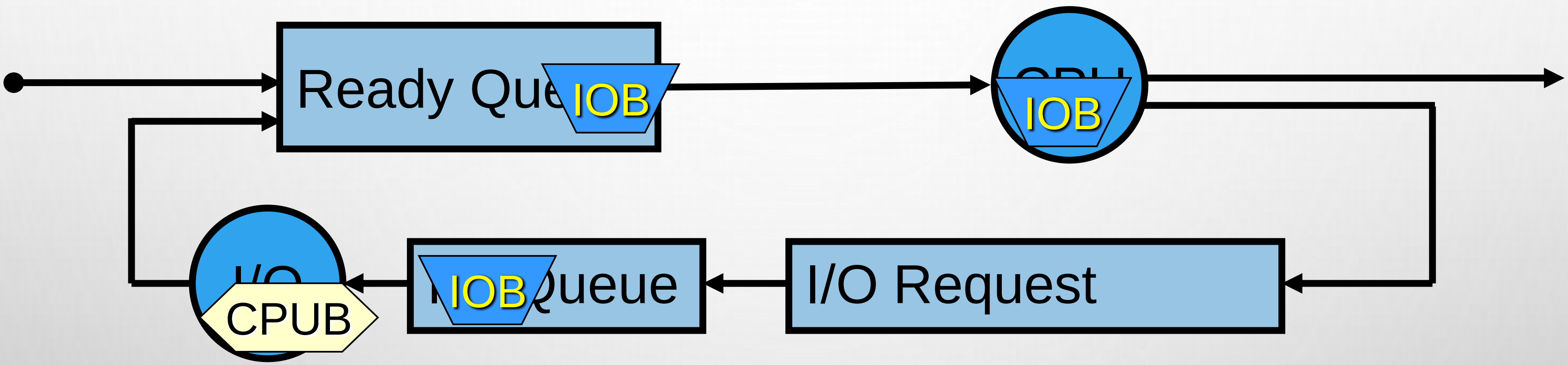
CPU is running CPUB

CONVOY EFFECT



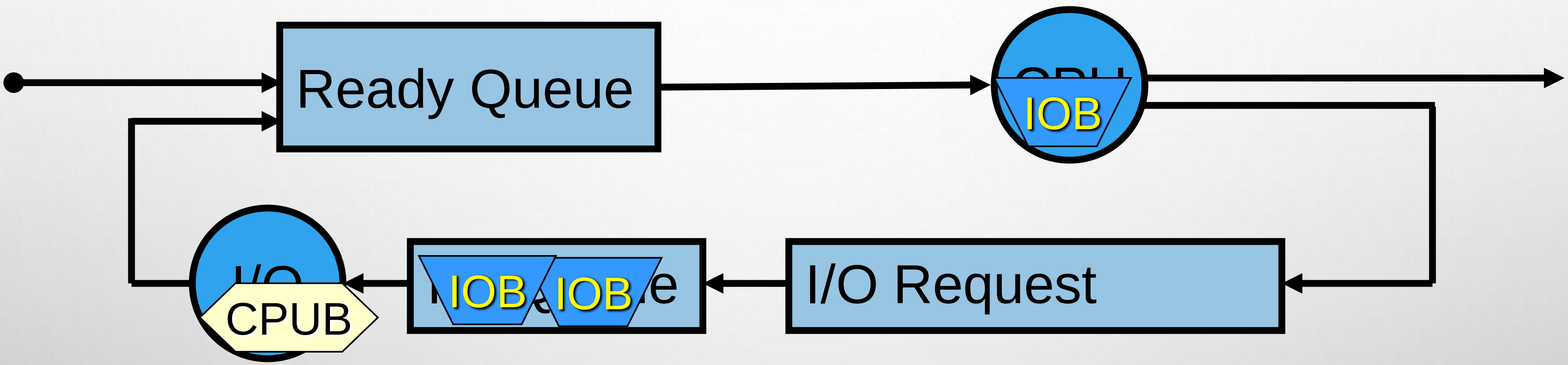
CPU is running CPUB
I/O devices idle

CONVOY EFFECT



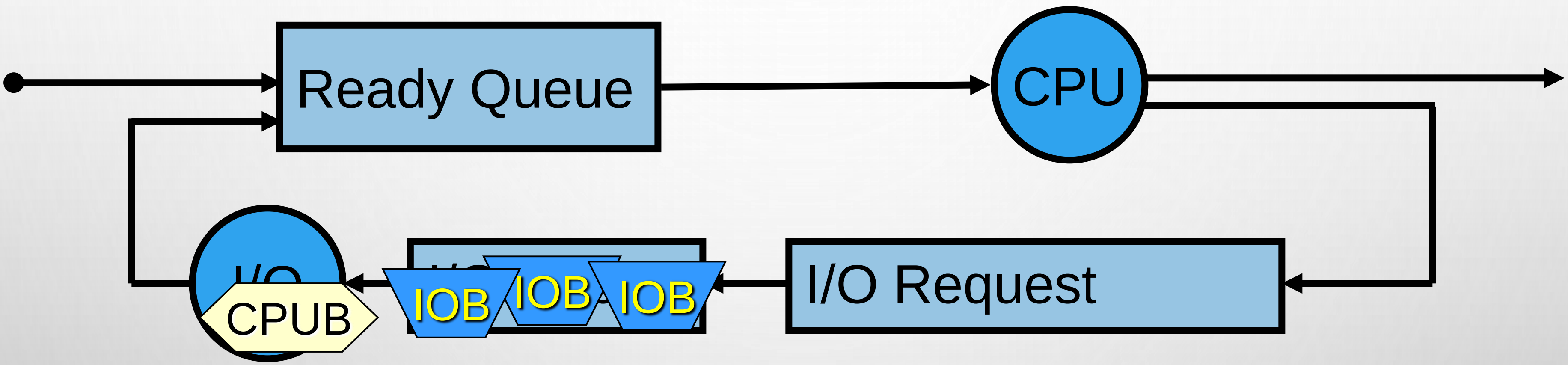
CPUB moves to I/O device

CONVOY EFFECT



I/O Bound jobs take very small amount of CPU time and go for I/O

CONVOY EFFECT



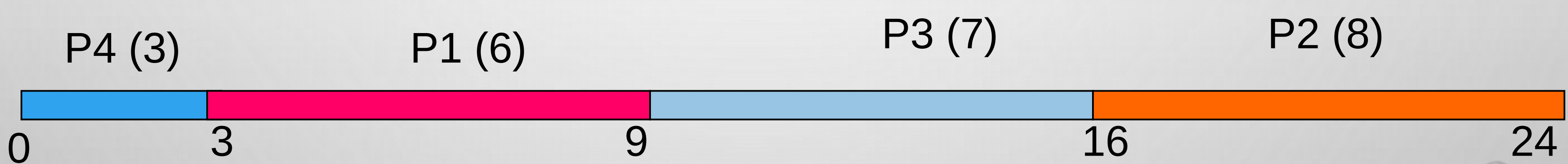
CPU idle

SHORTEST JOB FIRST (SJF)

- Scheduling algorithm in batch systems
- Schedule the job with the shortest run time first
- Requirement: the run time needs to be known in **advance**
- SJF is optimal in terms of turnaround, if **all** jobs arrive at **same time**

SJF: EXAMPLE

Process	Duration	Order	Arrival Time
P1	6	1	0
P2	8	2	0
P3	7	3	0
P4	3	4	0



Do it yourself

P4 turnaround: 3
P1 turnaround: 9
P3 turnaround: 16
P2 turnaround: 24

Total execution time: 24
The average turnaround:
(3+9+16+24)/4 = 13

COMPARING TO FCFS

Process	Duration	Order	Arrival Time
P1	6	1	0
P2	8	2	0
P3	7	3	0
P4	3	4	0

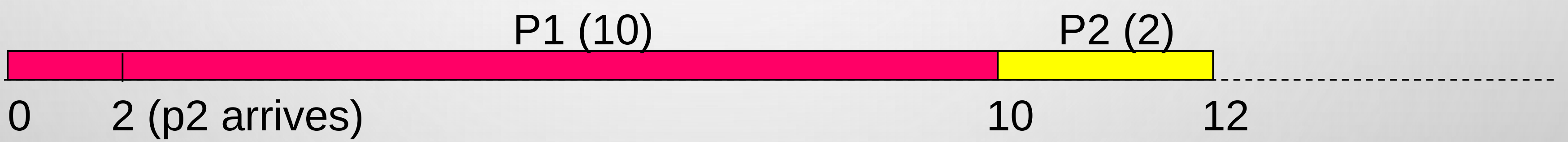
P1 turnaround: 6
P2 turnaround: 14
P3 turnaround: 21
P4 turnaround: 24

The total time is the same.
The average turnaround:
(6+14+21+24)/4 = 16.25
(comparing to 13)

SJF IS NOT ALWAYS OPTIMAL

- SJF OPTIMAL ONLY IF ALL JOBS HAVE ARRIVED AT SCHEDULING TIME

Process	Duration	Order	Arrival Time
P1	10	1	0
P2	2	2	2



P1 turnaround: 10
P2 turnaround: 10

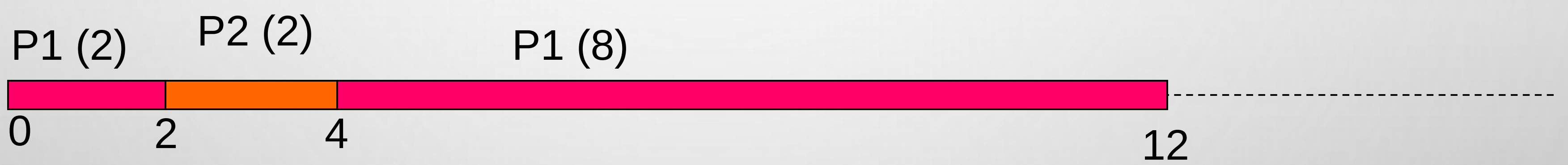
The average turnaround (AWT):
(10+10)/2 = 10

PREEMPTIVE SJF

- Also called Shortest Remaining Time Next
  - Schedule the job with the shortest remaining time required to complete
  - When new job arrives, compare its total time with the remaining time of the running job
  - If the new job needs less time the current job is suspended and the new job started
- Requirement: the run time needs to be known in advance

PREEMPTIVE SJF: SAME EXAMPLE

Process	Duration	Order	Arrival Time
P1	10	1	0
P2	2	2	2



P1 turnaround: 12
P2 turnaround: 2

The average turnaround:
(2+12)/2 = 7

PROBLEM WITH PREEMPTIVE SJF?

- Starvation
  - In some condition, a job is waiting for ever
  - Example: Preemptive SJF
    - Process A with run time of 1 hour arrives at time 0
    - But every 1 minute, a short process with run time of 1 minute arrives
    - Result of Preemptive SJF: A never gets to run

INTERACTIVE SYSTEM

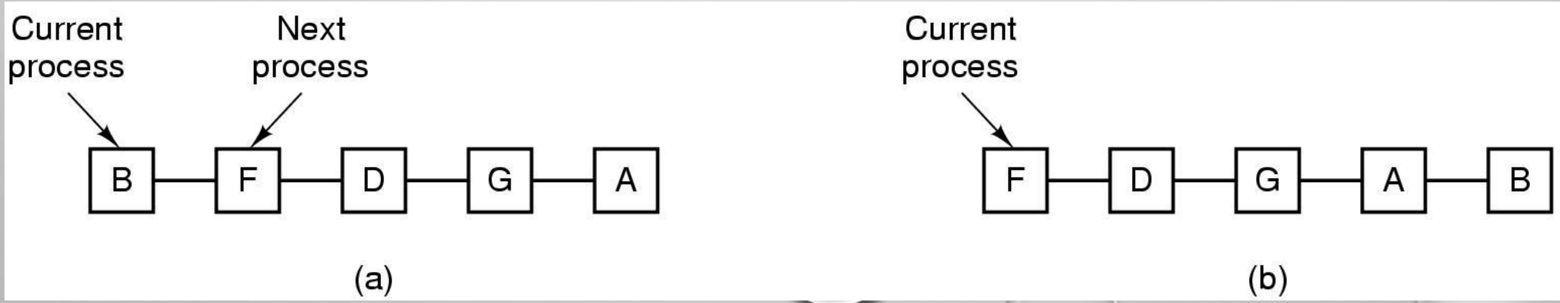
- Example: Servers
  - Serve multiple remote users all of whom are in a big hurry
- Performance Criteria
  - Min response time:
    - amount of time it takes from when a request was submitted until the first response is produced, not output
    - respond to requests quickly

INTERACTIVE SYSTEM

- Algorithms used here usually preemptive
  - Time is **sliced** into quantum (time intervals)
  - Scheduling decision is also made at the beginning of each quantum
- Representative algorithms:
  - Round-robin
  - Priority-based
  - Shortest process time
  - Guaranteed Scheduling
  - Lottery Scheduling
  - Fair Sharing Scheduling

ROUND ROBIN

- **Round Robin (RR)**
  - Often used for timesharing
  - Each process is given a time slice called a ***quantum***
  - It is run for the quantum or until it blocks
  - RR allocates the CPU uniformly (fairly) across participants from ready queue.
- Problem:
  - Do not consider priority
  - Context switch overhead



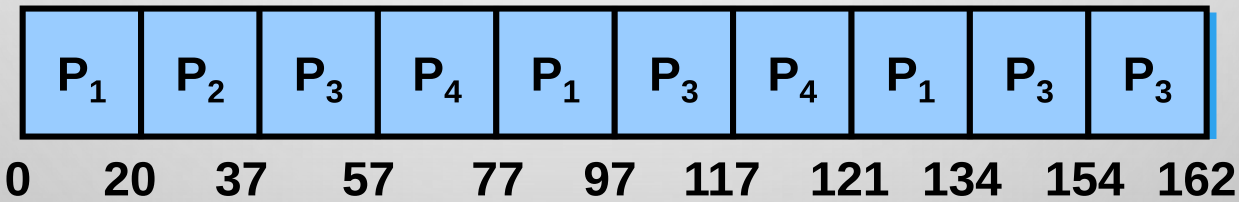
IMPLEMENTING ROUND ROBIN

- Keep the ready queue as a FIFO queue of processes.
- New processes are added to the tail of the ready queue.
- The scheduler
  - picks the first process from the ready queue
  - sets a timer to interrupt after 1 time quantum, and
  - Starts the process.
- When the quantum is over
  - The running process will be put at the **tail** of the ready queue.

RR WITH TIME QUANTUM = 20

Process	Run Time
P_1	53
P_2	17
P_3	68
P_4	24

- All processes arrive at time 0
- The Gantt chart is



- Higher average turnaround than SJF
- But better response time

RR: CHOICE OF TIME QUANTUM

- Performance depends on length of the timeslice
  - Context switching isn’t a free operation.
  - If timeslice time is set too high
    - attempting to amortize context switch cost, you get FCFS.
      - i.e. processes will finish or block before their slice is up anyway
      - Poor response time
  - If it’s set too low
    - you’re spending all of your time context switching between threads.

PRIORITY SCHEDULING

- Each job is assigned a priority
- Select highest priority job to run next
- Rational: higher priority jobs are more important
  - Example: simulation vs. auto save a document
- Problems:
  - Low priority process may starve
- Solution:
  - Priority need to be **adjusted** depending on the situation

ASSIGN PRIORITY

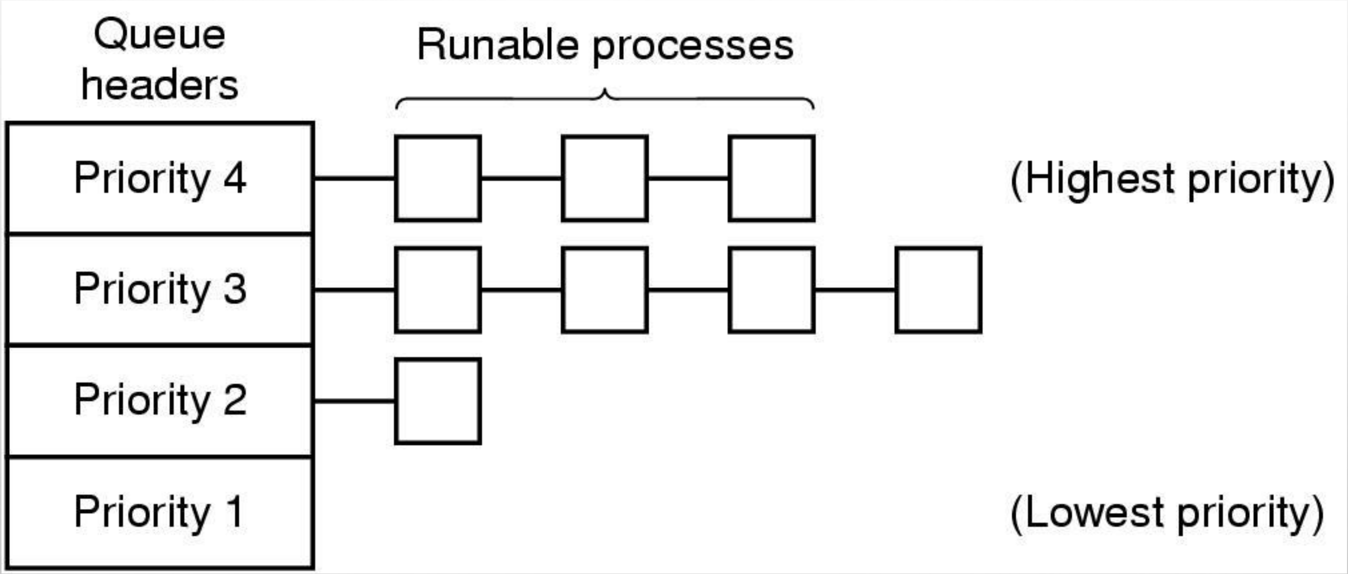
- Two approaches
  - Static (for system with well known and regular application behaviors)
  - Dynamic (otherwise)
- Priority may be based on:
  - Cost to user.
  - Importance of user
  - Percentage of CPU time used in last X hours

EXAMPLE: DYNAMIC PRIORITY ASSIGNMENT

- Whenever highly I/O bound processes wants the CPU it should be given the CPU immediately.
- Why?
- A simple algorithm for giving priority to I/O bound processes is to set the priority to $1/f$
  - f is the fraction of the last quantum used by a process
  - A process that used only 1 msec of its 50 msec quantum would get priority 50
  - A process that used 25 msec of its 50 msec quantum would get priority 2

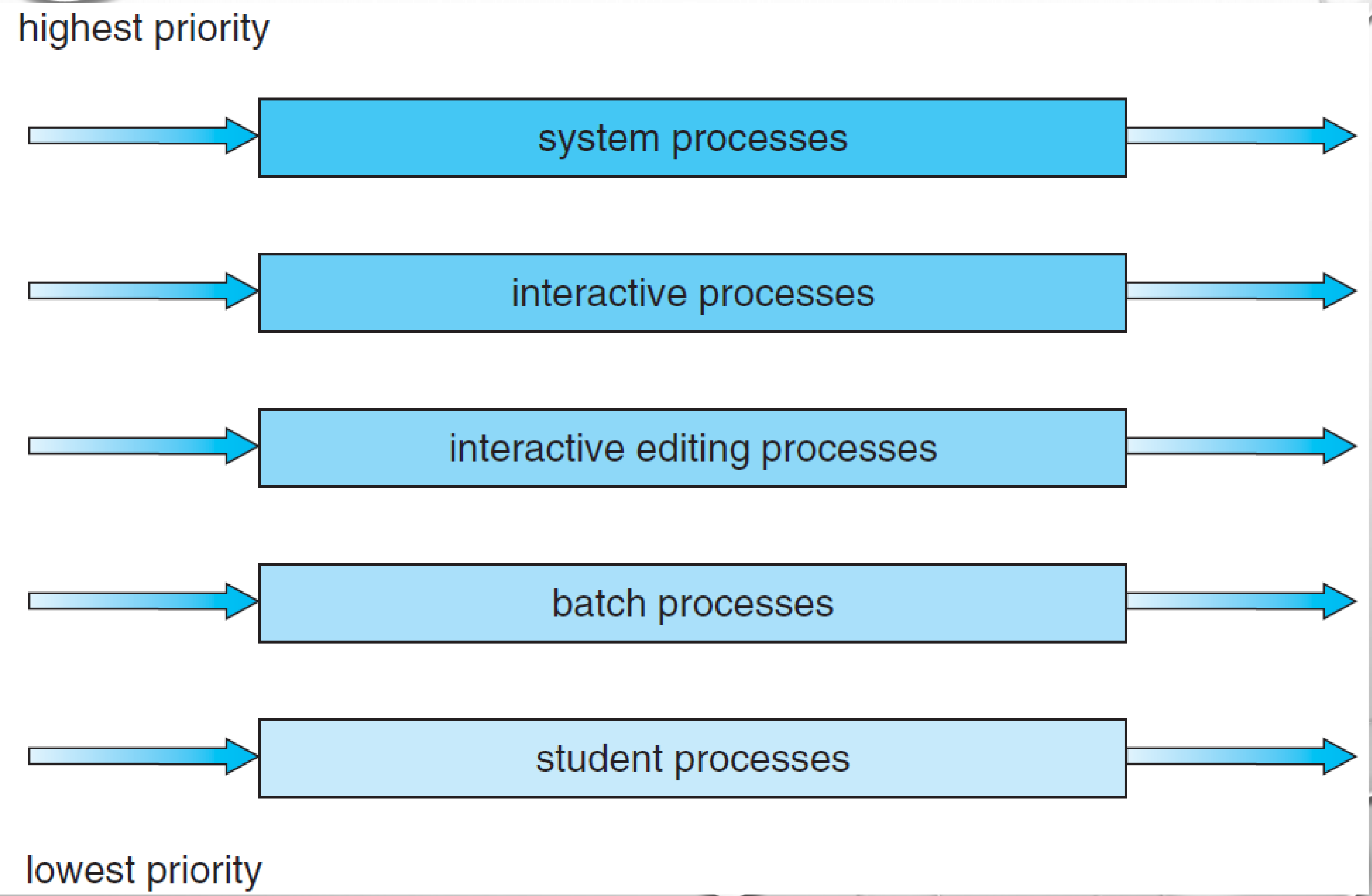
PRIORITY CLASS

- It is often convenient to group processes into priority classes and use priority scheduling among the classes but RR within each class



- If priorities are not adjusted occasionally, lower priority classes may all starve to death

PRIORITY CLASS



SHORTEST PROCESS NEXT

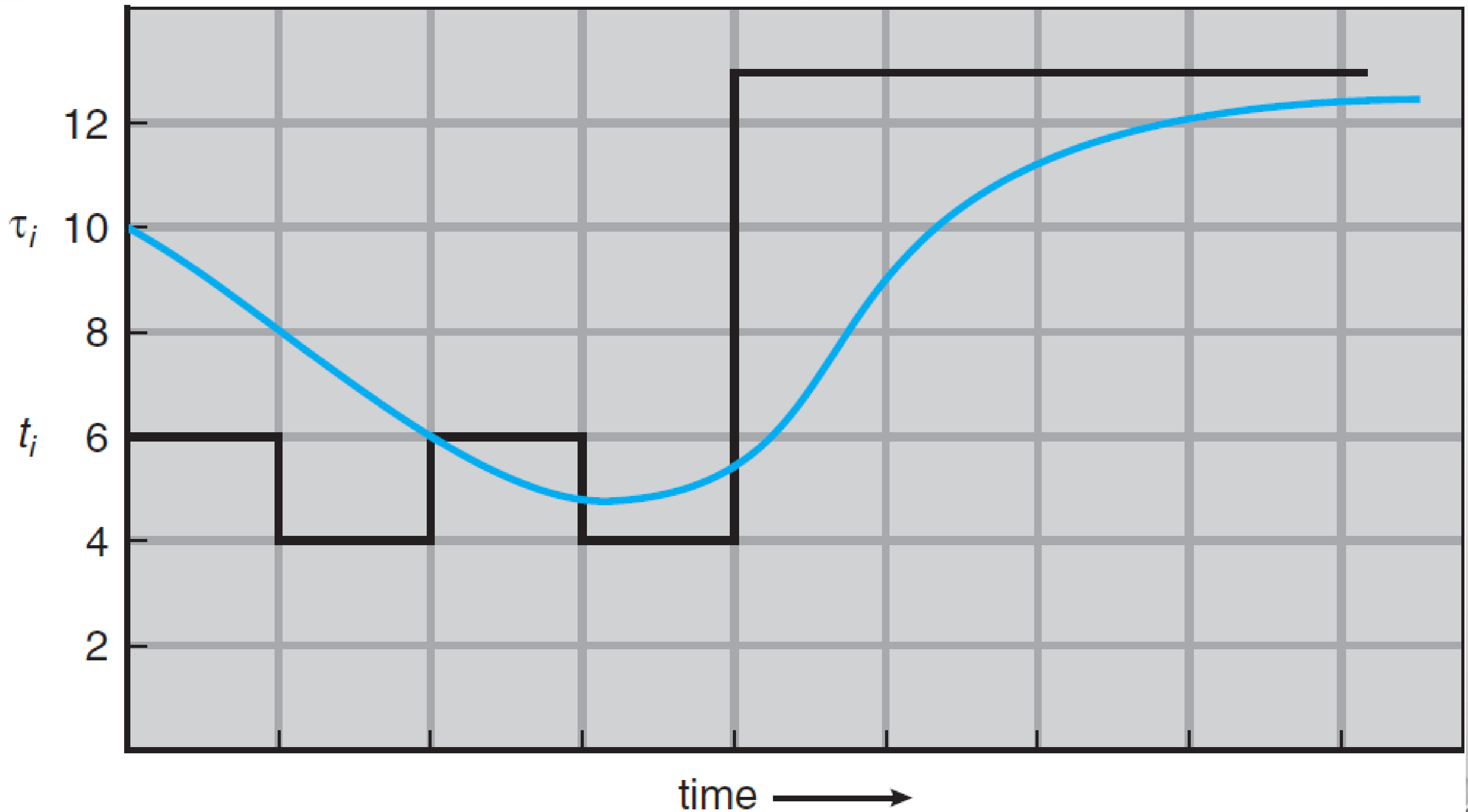
- Let’s apply SJF for interactive processes
- General pattern of a interactive process: CPU burst, I/O burst, …
- Let’s regard the execution of each CPU burst as a separate “job”
- Now we can minimize overall response time by running the process with shortest “job” first

SHORTEST PROCESS NEXT

- How to know the length of the next CPU burst?
- A possible answer: Exponential averaging
- Make estimate based on past behavior and run the process with the shortest estimated CPU burst
  - Let the **current** estimated CPU burst is τ_n
  - length of the nth CPU burst t_n
  - predicted value for the next CPU burst $\tau_{n+1} = \alpha t_n + (1-\alpha)\tau_n$

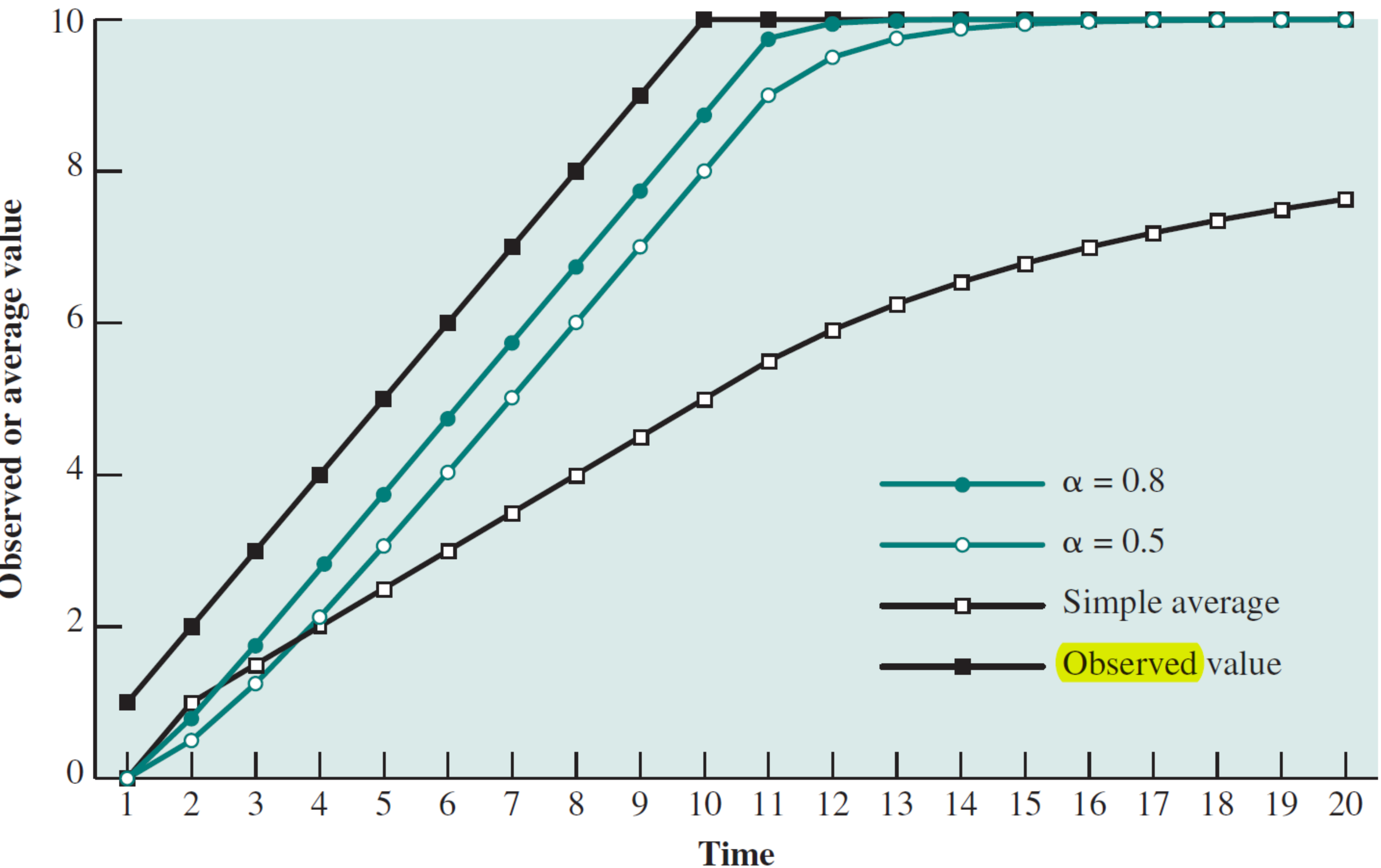
$$\tau_{n+1} = \alpha t_n + (1 - \alpha)\alpha t_{n-1} + \cdots + (1-\alpha)^j \alpha t_{n-j} + \cdots + (1-\alpha)^{n+1}\tau_0.$$

EXPONENTIAL AVERAGING



CPU burst (t_i)		6	4	6	4	13	13	13	…
"guess" (τ_i)	10	8	6	6	5	9	11	12	…

EXPONENTIAL AVERAGING



GUARANTEED SCHEDULING

- Make promises to users about performance & then meet those promises
- With n processes running, each one should get 1/n of the CPU cycles
- Calculate ratio for each process
- $\frac{Amount\ of\ CPU\ time\ process\ has\ had\ since\ its\ creation}{Amount\ of\ CPU\ time\ process\ should\ have\ since\ creation}$
- Run the process with the **lowest** ratio until its ratio has moved above its closest competitor
- Problem:
  - Implementation is difficult

THANKS FOR YOUR PATIENCE