

# Principles of Operating Systems

CS 446/646

## 4. CPU Scheduling

**René Doursat**

*Department of Computer Science & Engineering  
University of Nevada, Reno*

*Spring 2006*

# Principles of Operating Systems

## CS 446/646

- 0. Course Presentation
- 1. Introduction to Operating Systems
- 2. Processes
- 3. Memory Management
- 4. CPU Scheduling**
- 5. Input/Output**
- 6. File System**
- 7. Case Studies**

# Principles of Operating Systems

CS 446/646

## 4. CPU Scheduling

- a. **Concepts of Scheduling**
- b. **Scheduling Algorithms**
- c. Queuing Analysis
- d. Thread Scheduling

# Principles of Operating Systems

CS 446/646

## 4. CPU Scheduling

### a. Concepts of Scheduling

- ✓ Three-level scheduling
- ✓ Purpose of CPU scheduling

### b. Scheduling Algorithms

### c. Queuing Analysis

### d. Thread Scheduling

# 4.a Concepts of Scheduling

## Three-level scheduling

### ➤ Long-term scheduling (mostly in batch)

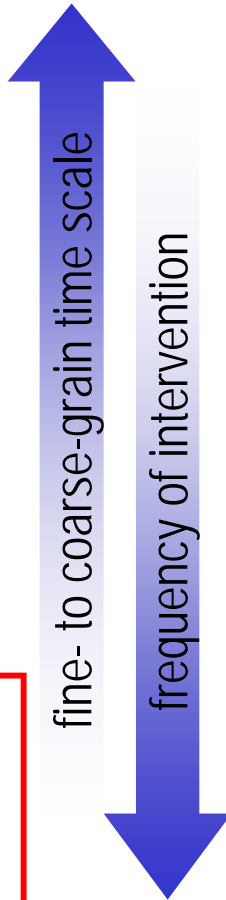
- ✓ the decision to add a program to the pool of processes to be executed: controls the degree of multiprogramming

### ➤ Medium-term scheduling

- ✓ the decision to add to the number of processes that are partially or fully in main memory ("swapping")
- ✓ *not* the same as paging: swapping out means removing all the pages of a process

### ➤ Short-term scheduling = CPU scheduling

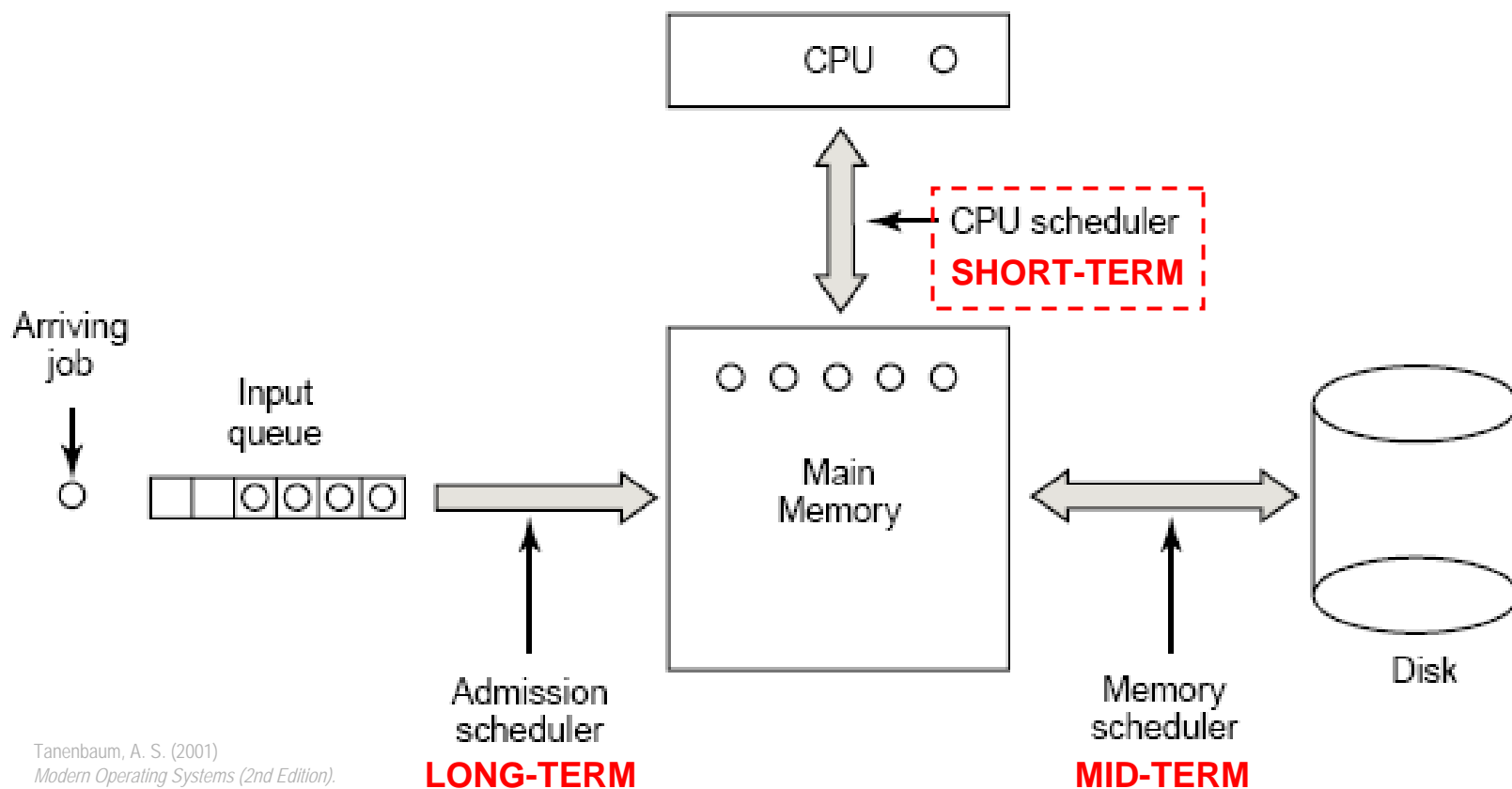
- ✓ the decision as to which available processes in memory are to be executed by the processor ("dispatching")



# 4.a Concepts of Scheduling

## Three-level scheduling

### ➤ Three-level scheduling



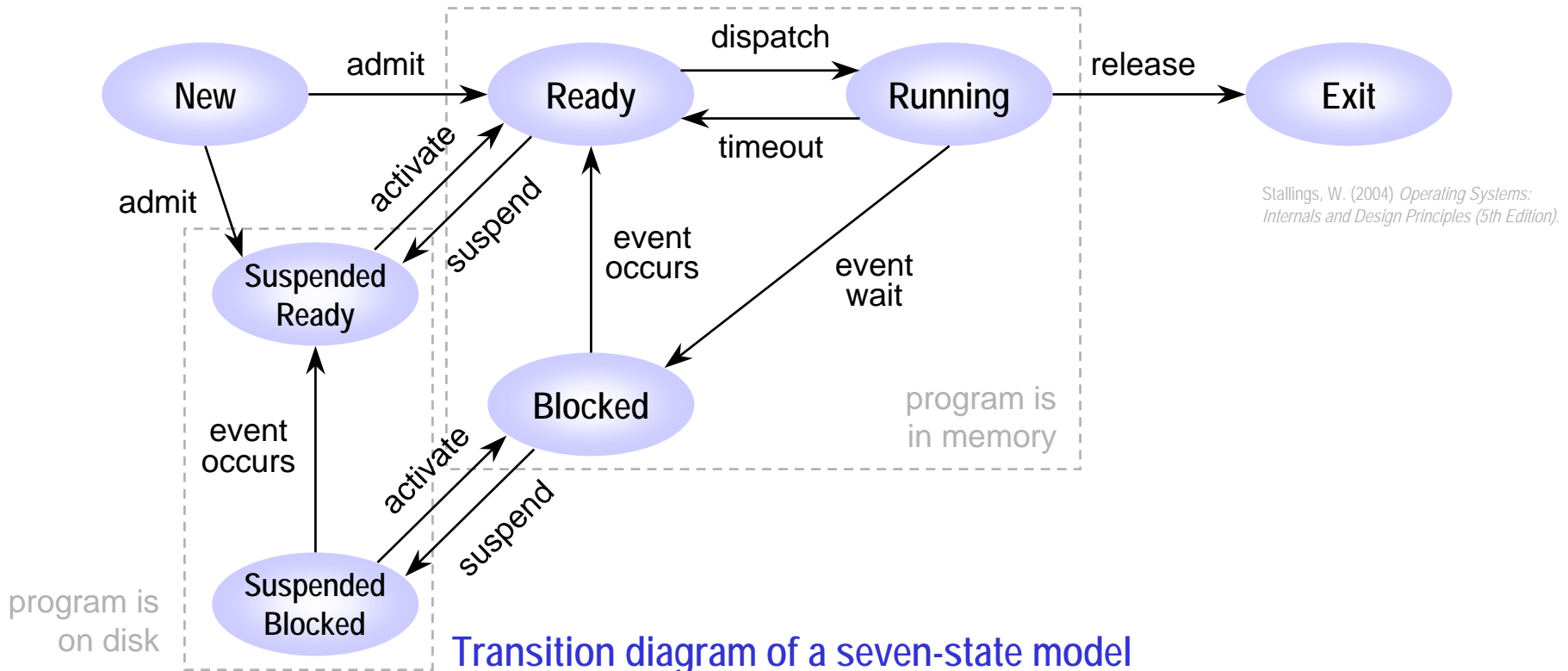
Tanenbaum, A. S. (2001)  
*Modern Operating Systems (2nd Edition)*.

### Three-level scheduling

# 4.a Concepts of Scheduling

## Three-level scheduling

### ➤ Reminder: process states

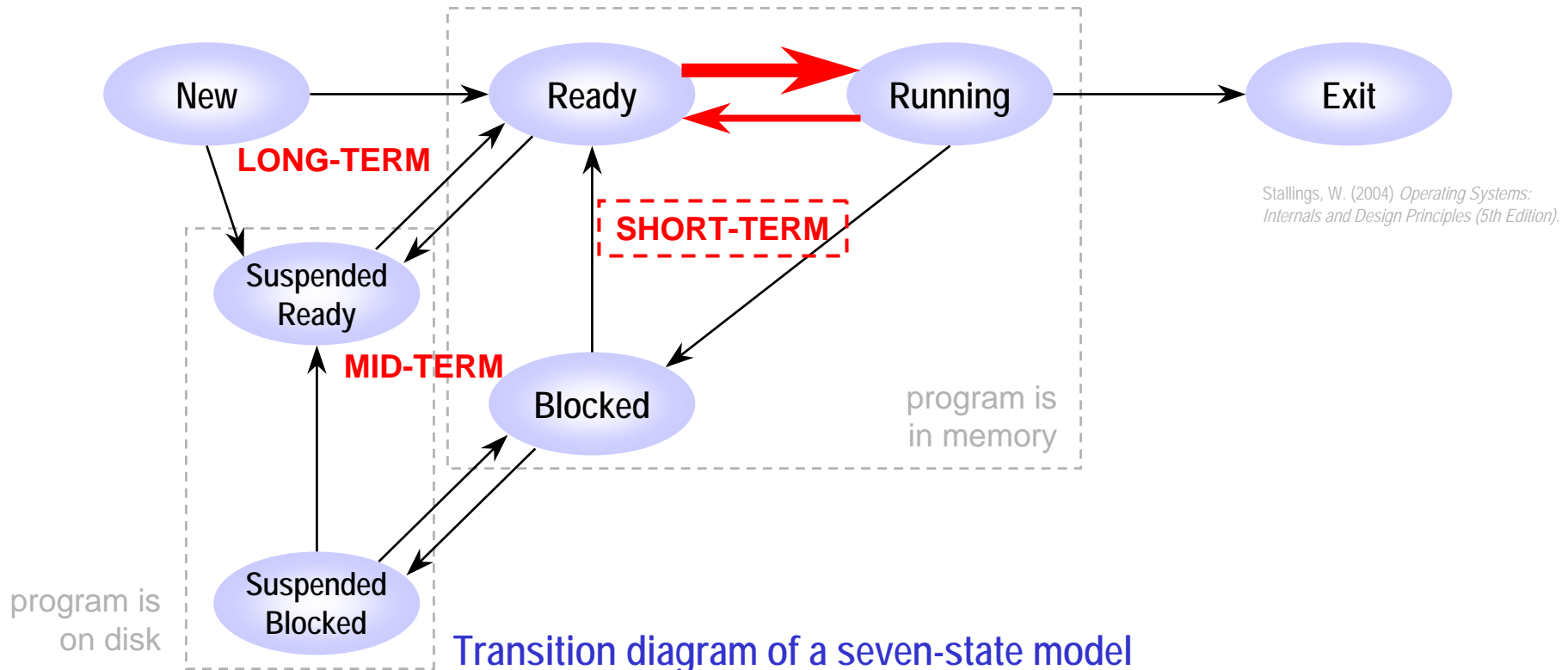


Stallings, W. (2004) *Operating Systems: Internals and Design Principles (5th Edition)*.

# 4.a Concepts of Scheduling

## Three-level scheduling

- In the O/S, the CPU scheduler decides which “Ready” process to run next (and which “Running” to time out)
  - ✓ the discipline it follows is the **scheduling algorithm**



Stallings, W. (2004) *Operating Systems: Internals and Design Principles* (5th Edition).

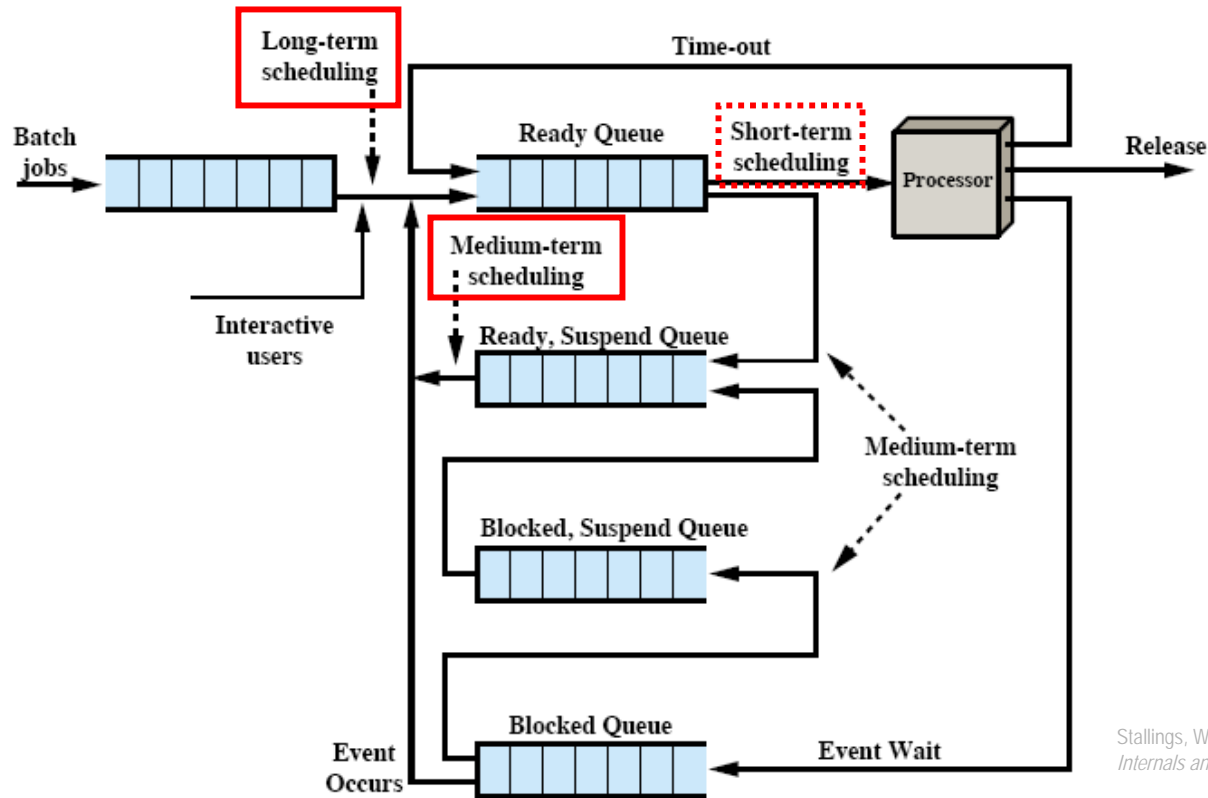


# 4.a Concepts of Scheduling

## Three-level scheduling

### ➤ General queuing system for scheduling

- ✓ in most algorithms, queues are not strictly FIFO: rather “pools”



Stallings, W. (2004) *Operating Systems: Internals and Design Principles (5th Edition)*.

Queuing diagram for scheduling

## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ Why scheduling matters: user service response

- ✓ example: choosing between
    - a process that updates the screen after the user has closed a window
    - a process that sends out queued email
  - ✓ taking 2 seconds to close the window while sending the email would be unacceptable
  - ✓ on the other hand, delaying the email while closing the window would hardly be noticed
- *schedule wisely to match user's expectations*

## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ Why scheduling matters: CPU usage

- ✓ switching processes (contexts) is heavy
  - switch from user mode to kernel mode
  - CPU state must be saved
  - process state must be saved
  - pages and page bits must be saved
  - MMU must be reloaded with new page table
  - etc.

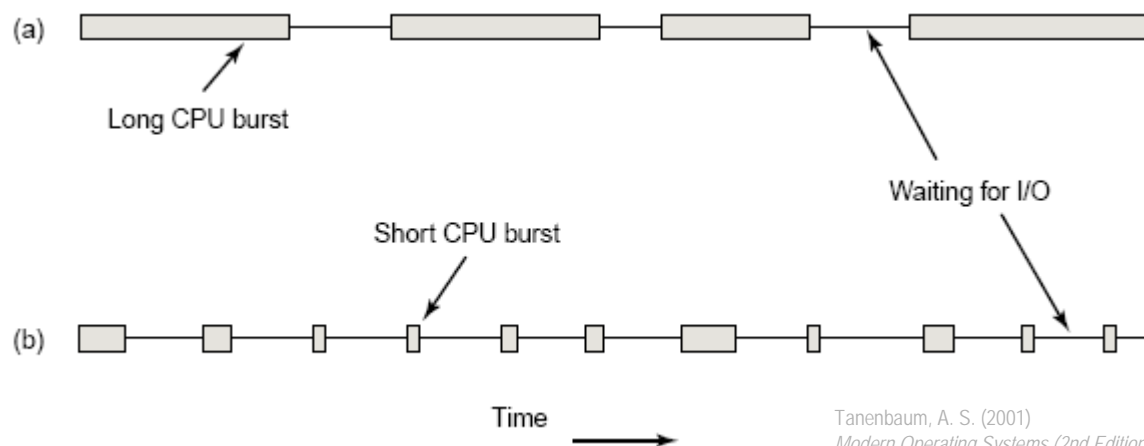
→ *to maximize CPU utilization, interleave but at the same time minimize process switches*

## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ Types of process behavior: CPU-I/O burst cycle

- ✓ processes alternate CPU usage with I/O wait
  - compute-bound processes have long CPU bursts and infrequent I/O
  - I/O-bound processes have short CPU bursts and frequent I/O



Tanenbaum, A. S. (2001)  
*Modern Operating Systems (2nd Edition)*.

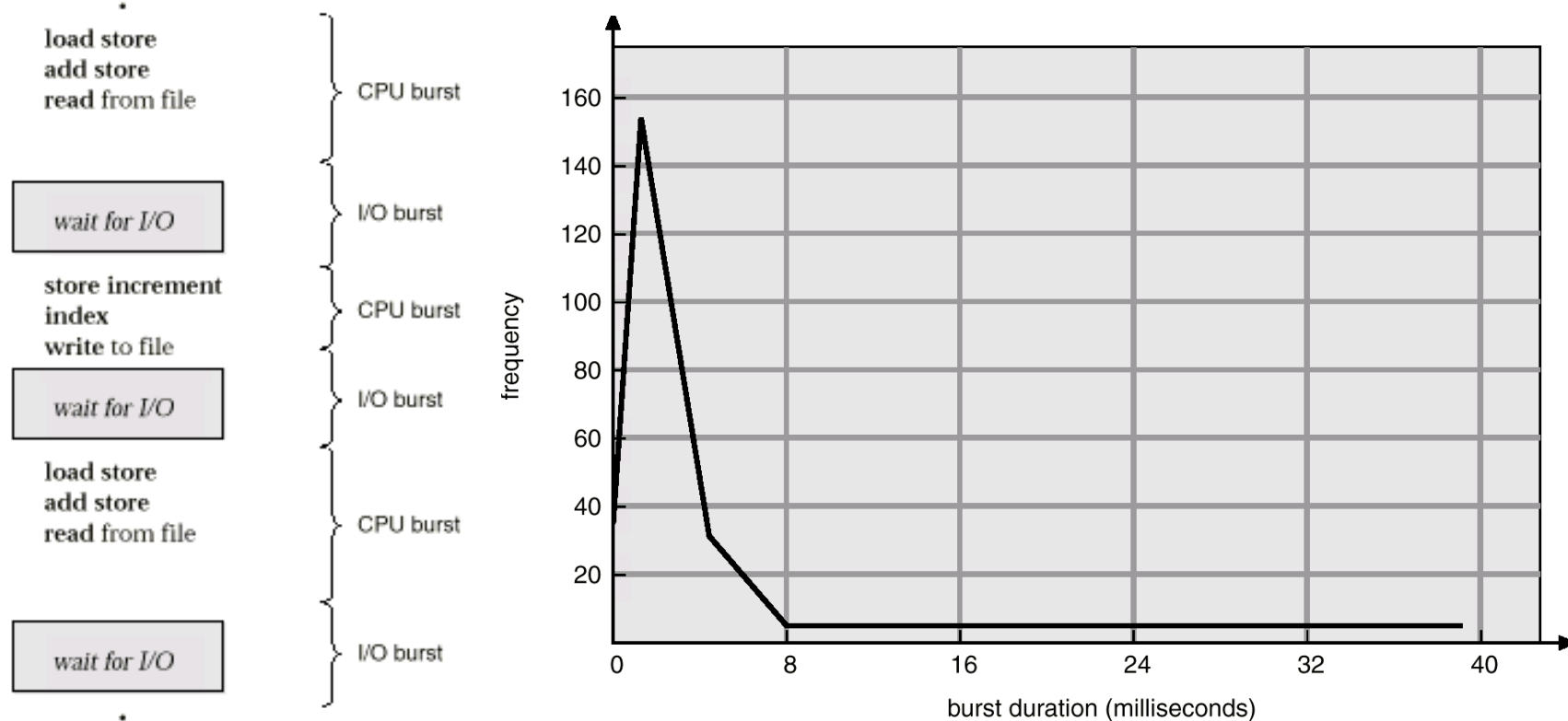
(a) Compute-bound process vs. (b) I/O-bound process

# 4.a Concepts of Scheduling

## Purpose of CPU scheduling

### ➤ Types of process behavior: CPU-I/O burst cycle

✓ power-law: large # of short CPU bursts, small # of large bursts



Silberschatz, A., Galvin, P. B. and Gagne, G. (2003)  
*Operating Systems Concepts with Java (6th Edition).*

Typical histogram of CPU-burst times

## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ I/O-bound processes

- ✓ as CPU speeds increase, processes generally tend to become more and more I/O-bound
- ✓ the scheduling of I/O-bound processes will likely become an important subject in the future
- *basic idea: an I/O-bound process that is "Ready" to run should get the CPU quickly so it can keep the disk busy*

## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ When scheduling decisions must be made

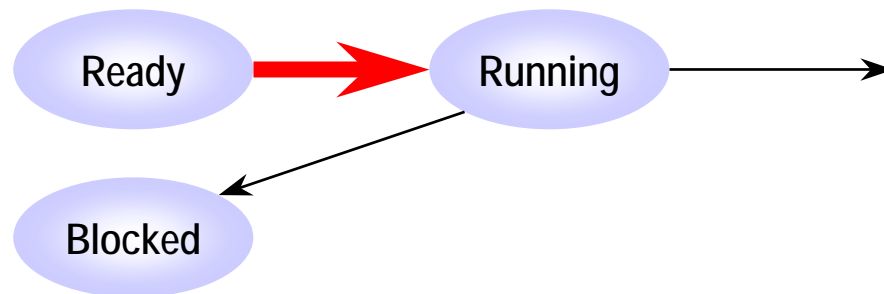
- ✓ when a new process is created — run the child or the parent?
- ✓ when a process exits — who's next?
- ✓ when an I/O interrupt occurs upon finishing an I/O task — should the waiting process be rescheduled? or let the currently running process continue? or pick another process? etc.
- ✓ when a timeout (clock interrupt) occurs — who's next?

## 4.a Concepts of Scheduling

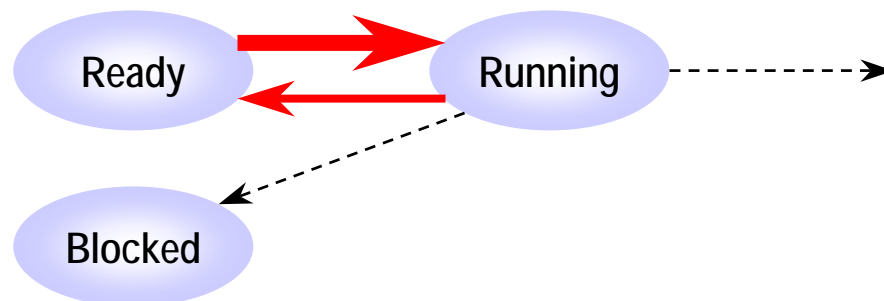
### Purpose of CPU scheduling

#### ➤ Two kinds of CPU-scheduling algorithms

- ✓ **cooperative** scheduling — let a process run until it blocks on I/O, terminates or voluntarily releases the CPU (system call)



- ✓ **preemptive** scheduling — follow clock interrupts (ex: 50Hz) to forcibly switch processes (demote the "Running" to "Ready")





## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ Scheduling algorithm goals — all systems (batch & int.)

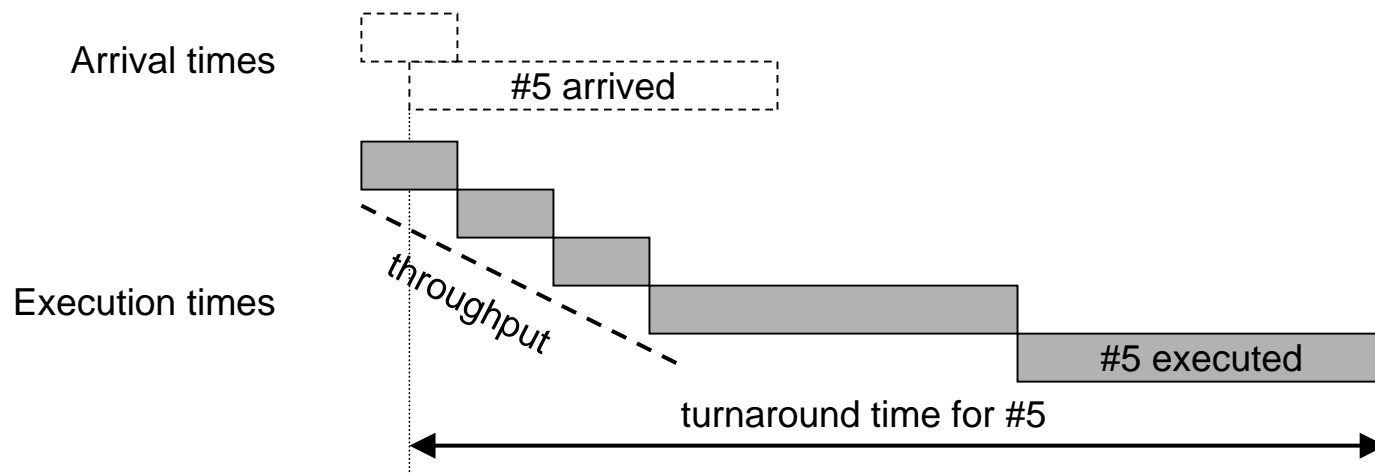
- ✓ **fairness** — comparable processes get comparable service
  - ✓ **compliance to system's policy** — ex: high-priority override low-priority processes (ex: safety control vs. payroll in a nuclear plant)
  - ✓ **keep system busy** — CPU and I/O devices should be utilized fully
    - if all CPU-bound were run first: fight for CPU, I/O idle
    - then all I/O-bound were run: fight for I/O, CPU idle
- *keep a well-balanced mix of CPU-bound and I/O-bound processes, so they can fill in for each other*

## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ Scheduling algorithm goals — batch systems

- ✓ **throughput** — maximize # of completed jobs per time unit
- ✓ **turnaround time (latency)** — minimize time between submission and termination of job
  - high throughput and low turnaround are rarely compatible
  - for ex: supply of short jobs scheduled in front of long jobs: good throughput, bad turnaround time for long jobs



## 4.a Concepts of Scheduling

### Purpose of CPU scheduling

#### ➤ Scheduling algorithm goals — interactive systems

- ✓ **response time** — respond to requests quickly: minimize time between issuing command and getting result
  - ex: a user request to start a new program should take precedence over background work
  - having interactive requests go first will be perceived as good service
- ✓ **proportionality time** — meet users' expectation, even if irrational
  - ex: 45 seconds to establish a modem connection is *perceived* as acceptable, yet 45 seconds to hang up is not
  - whenever possible, take this into account when scheduling

# 4.a Concepts of Scheduling

## Purpose of CPU scheduling

### ➤ Scheduling algorithm goals — summary

#### **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

Tanenbaum, A. S. (2001)  
*Modern Operating Systems (2nd Edition).*

#### **Real-time systems**

Meeting deadlines - avoid losing data

Predictability - avoid quality degradation in multimedia systems

Some goals of CPU scheduling under different circumstances

# Principles of Operating Systems

CS 446/646

## 4. CPU Scheduling

### a. Concepts of Scheduling

- ✓ Three-level scheduling
- ✓ Purpose of CPU scheduling

### b. Scheduling Algorithms

### c. Queuing Analysis

### d. Thread Scheduling