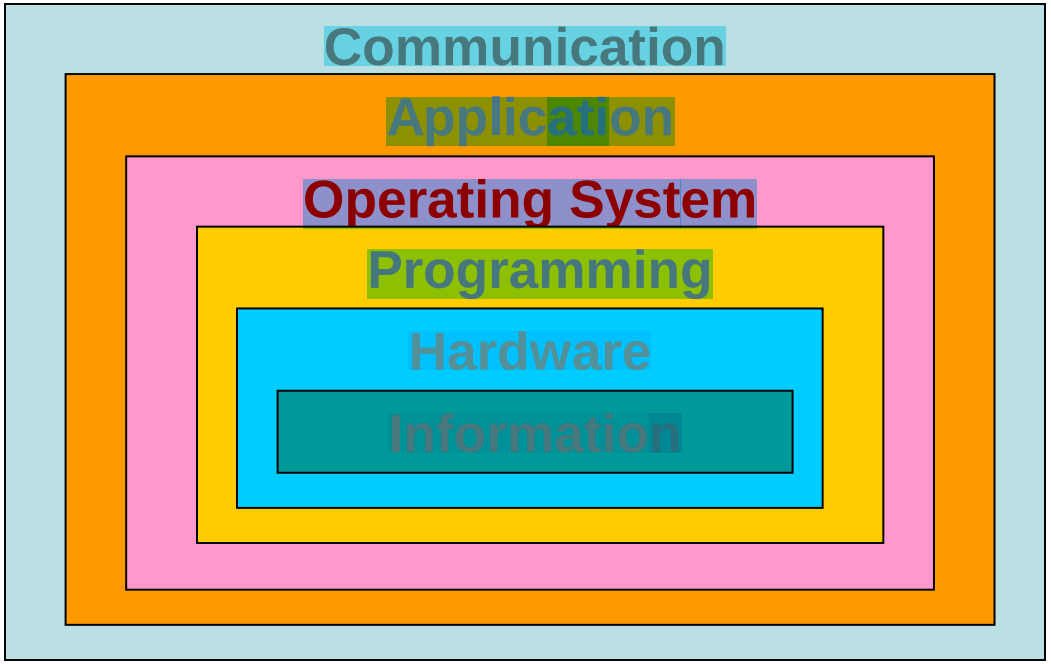
Introduction to Computing

Section 5 – The Operating Systems Layer



Software Categories

SCSE

- **Application software** Software written to address specific needs—to solve problems in the real world (Programs that helps us solve real-world problems)

  Word processing programs, games, inventory control systems, automobile diagnostic programs, and missile guidance programs are all application software

- **System software** Software that manages a computer system at a fundamental level (Programs that manage a computer system and interact with hardware)

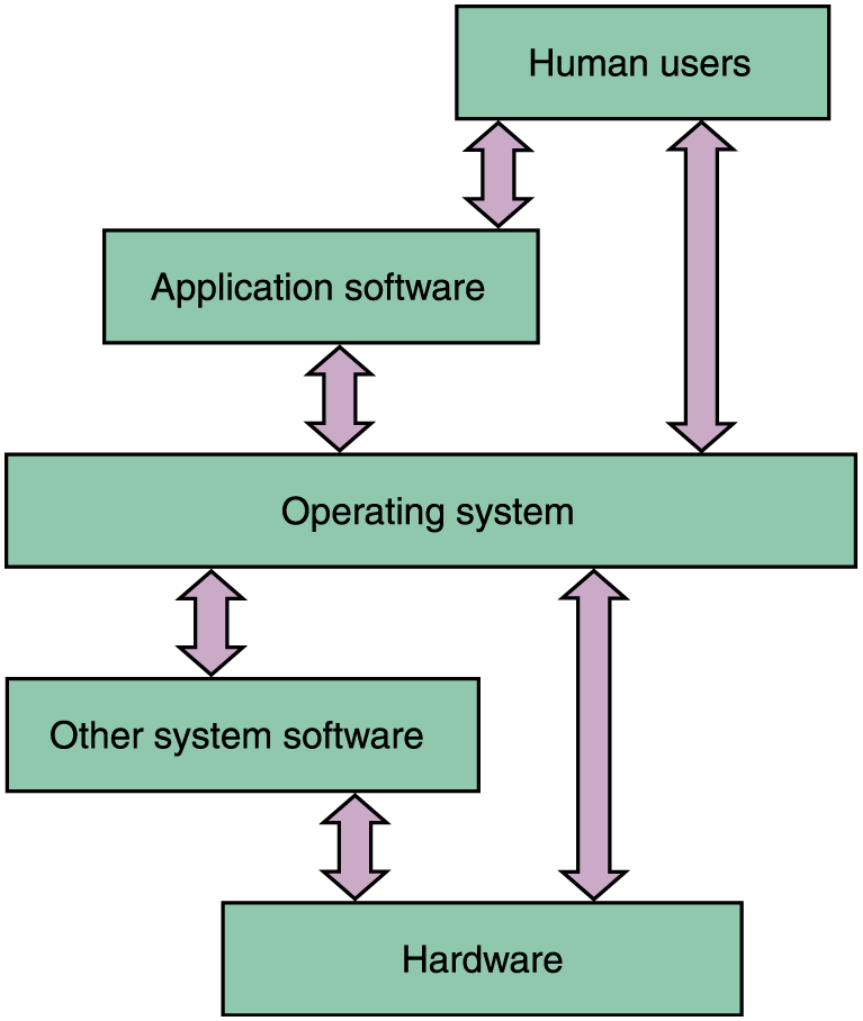
  It provides the tools and an environment in which application software can be created and run

Operating System

SCSE

- An **operating system**
  - manages computer resources, such as memory and input/output devices
  - provides an interface through which a human can interact with the computer
  - allows an application program to interact with these other system resources

Operating System



An operating system interacts with many aspects of a computer system.

Operating System

SCSE

- The various roles of an operating system generally revolve around the idea of "sharing nicely"
- An operating system manages resources, and these resources are often shared in one way or another among programs that want to use them

Resource Management

SCSE

- **Multiprogramming** The technique of keeping multiple programs in main memory at the same time that compete for access to the CPU so that they can execute
- **Memory management** The process of keeping track of what programs are in memory and where in memory they reside

Resource Management

SCSE

- **Process** The dynamic representation of a program during execution.
- The operating system performs **process management** to carefully track the progress of a process and all of its intermediate states
- **CPU scheduling** determines which process in memory is executed by the CPU at any given point

Batch Processing

SCSE

- A typical computer in the 1960s and ‘70s was a large machine
- Its processing was managed by a human *operator*
- The operator would organize various jobs from multiple users into *batches*

Batch Processing

SCSE

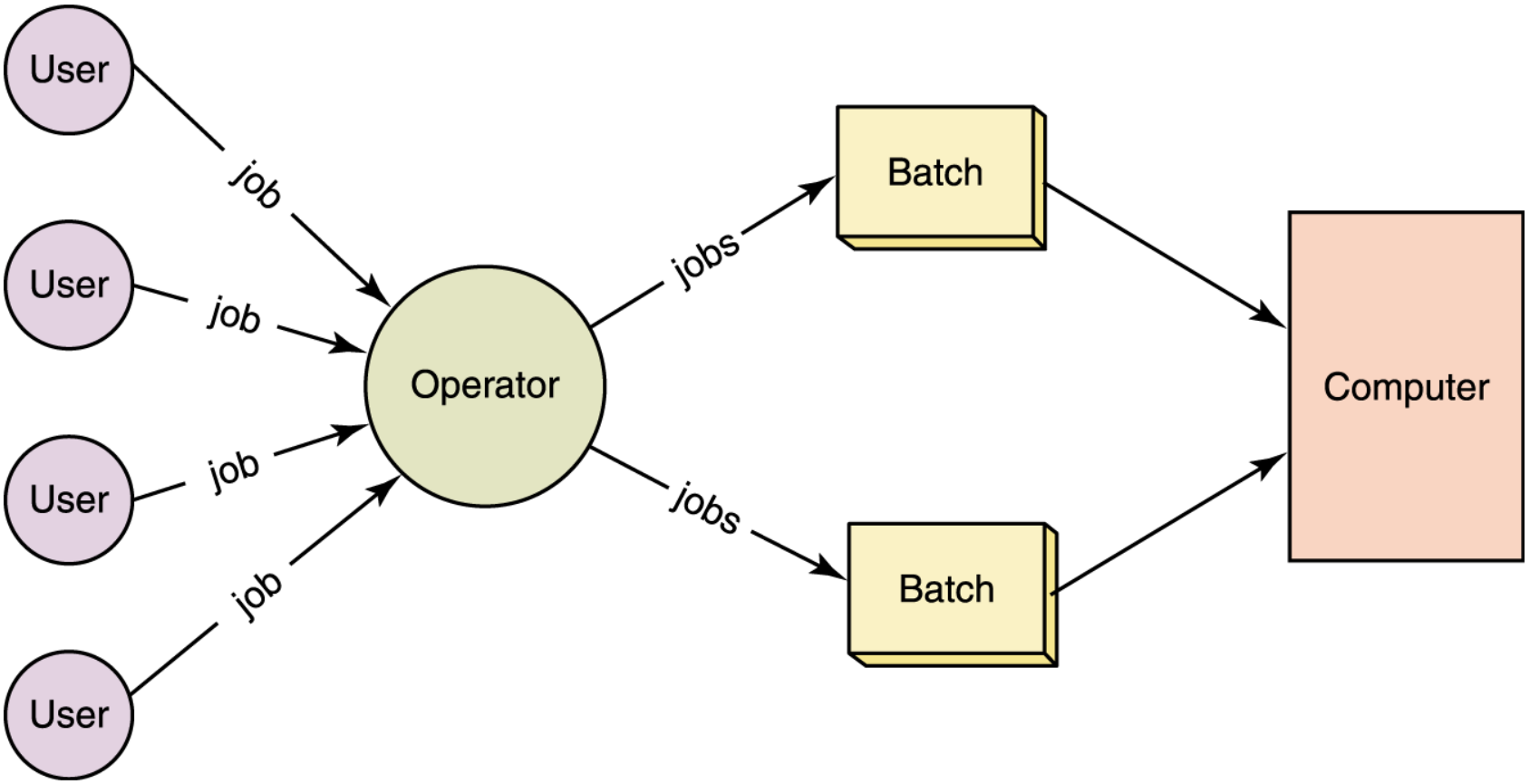


Figure 10.2 **In early systems, human operators would organize jobs into batches**

Timesharing

SCSE

- **Timesharing system** A system that allows multiple users to interact with a computer at the same time
- **Multiprogramming** A technique that allows multiple processes to be active at once, allowing programmers to interact with the computer system directly, while still sharing its resources
- In a timesharing system, each user has his or her own **virtual machine**, in which all system resources are (in effect) available for use

Other Factors

SCSE

- **Real-time System** A system in which response time is crucial given the nature of the application
- **Response time** The time delay between receiving a stimulus and producing a response
- **Device driver** A small program that "knows" the way a particular device expects to receive and deliver information.

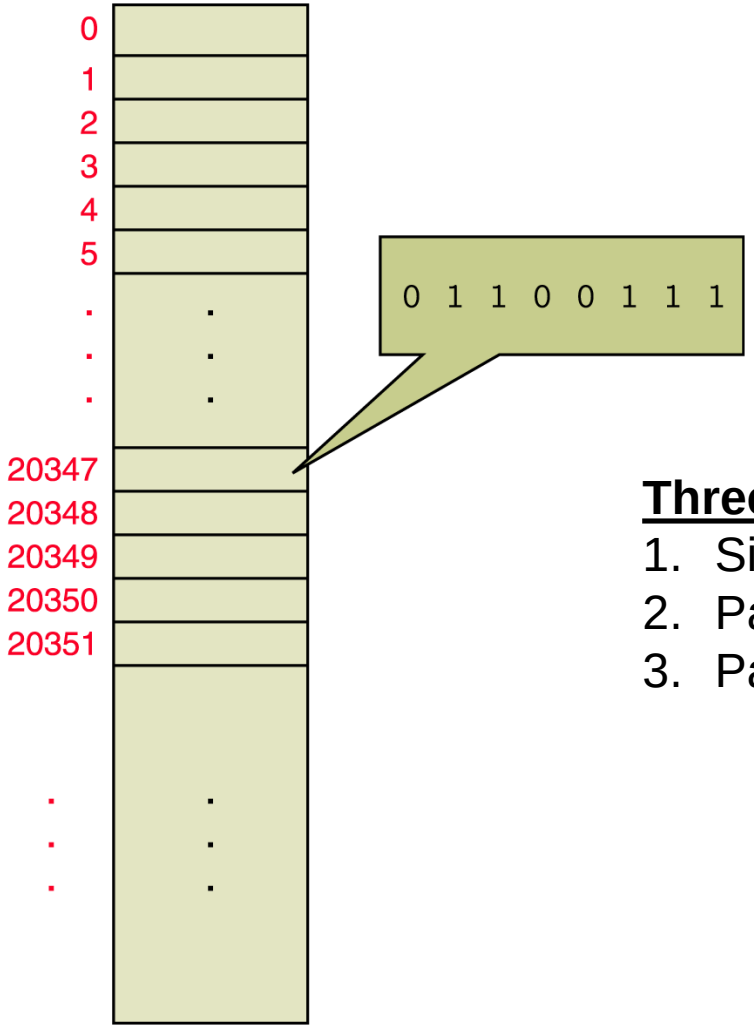
Memory Management

SCSE

- Operating systems must employ techniques to
  - Track where and how a program resides in memory
  - Convert **logical addresses** into actual **addresses**
- **Logical address** (sometimes called a virtual or relative address) A value that specifies a generic location, relative to the program but not to the reality of main memory
- **Physical address** An actual address in the main memory device

The mapping from a logical address to a physical address is called ***address binding***

Memory Management



Three techniques:

1. Single contiguous memory management
2. Partition memory management
3. Paged memory management

Memory is a continuous set of bits referenced by specific addresses

Single Contiguous Memory Management

SCSE

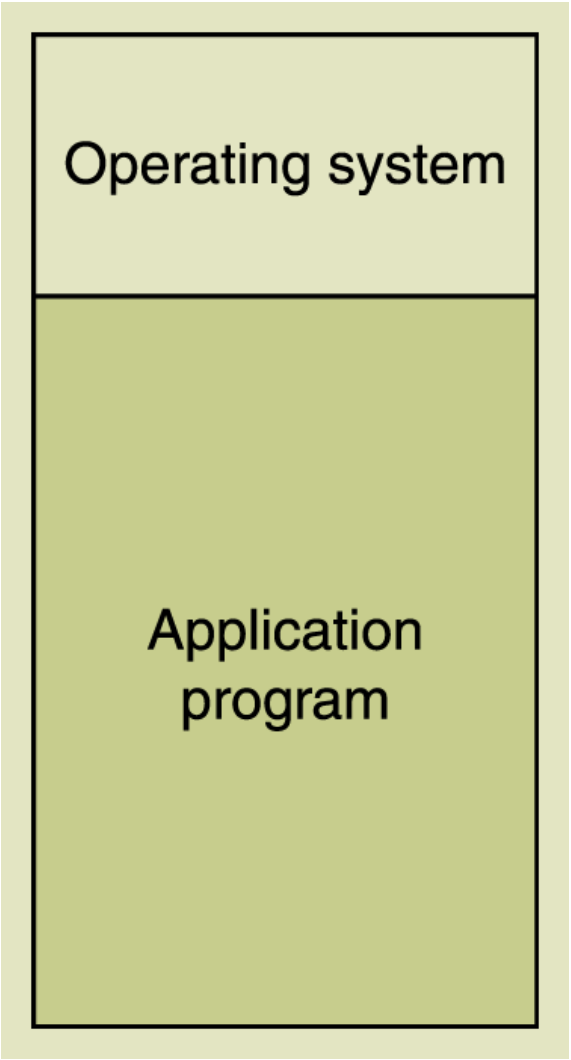


Figure 10.4
Main memory divided into two sections

- There are only two programs in memory
  - The operating system
  - The application program
- This approach is called **single contiguous memory management**

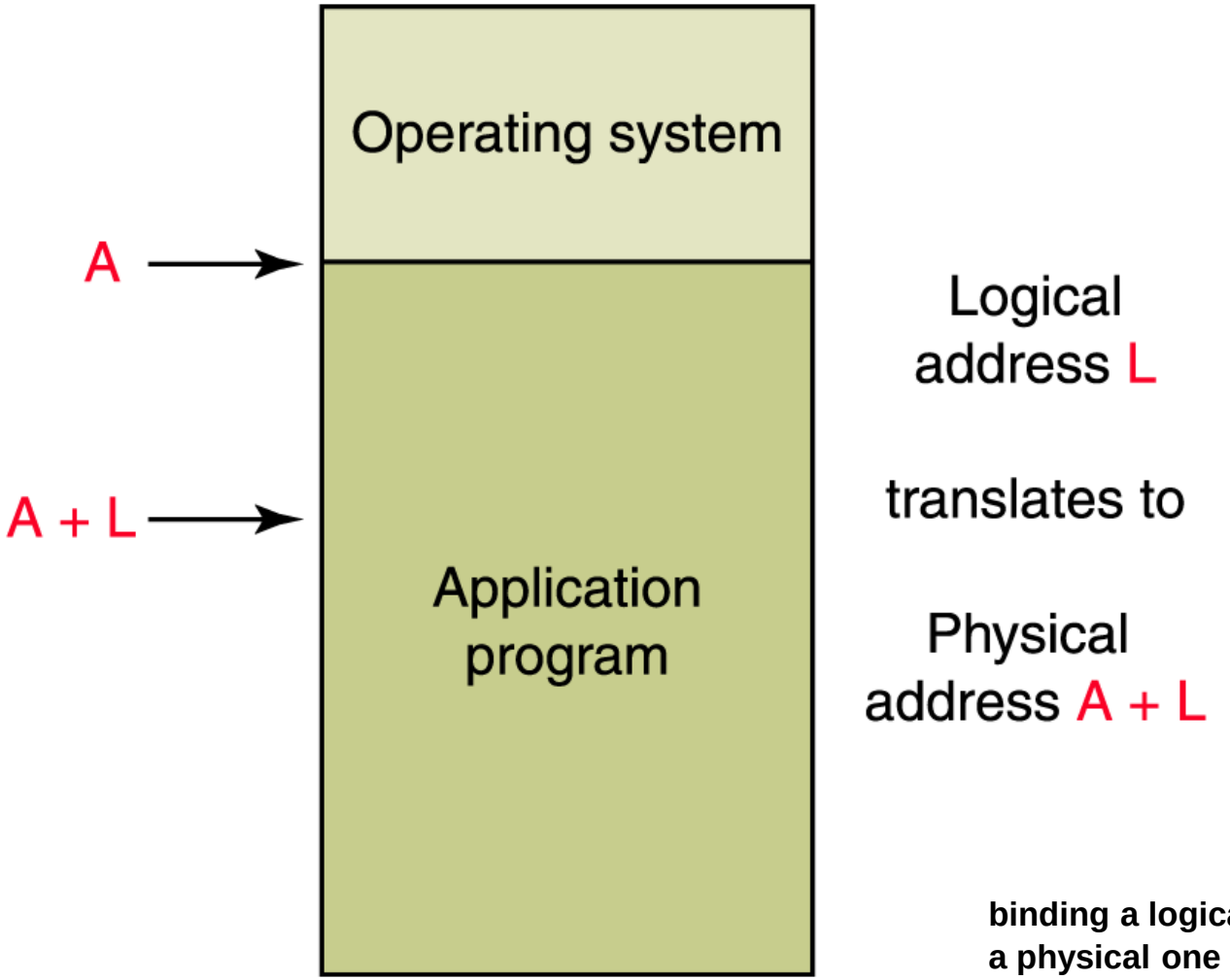
Single Contiguous Memory Management

SCSE

- A logical address is simply an integer value relative to the starting point of the program
- To produce a physical address, we add a logical address to the starting address of the program in physical main memory

Single Contiguous Memory Management

SCSE



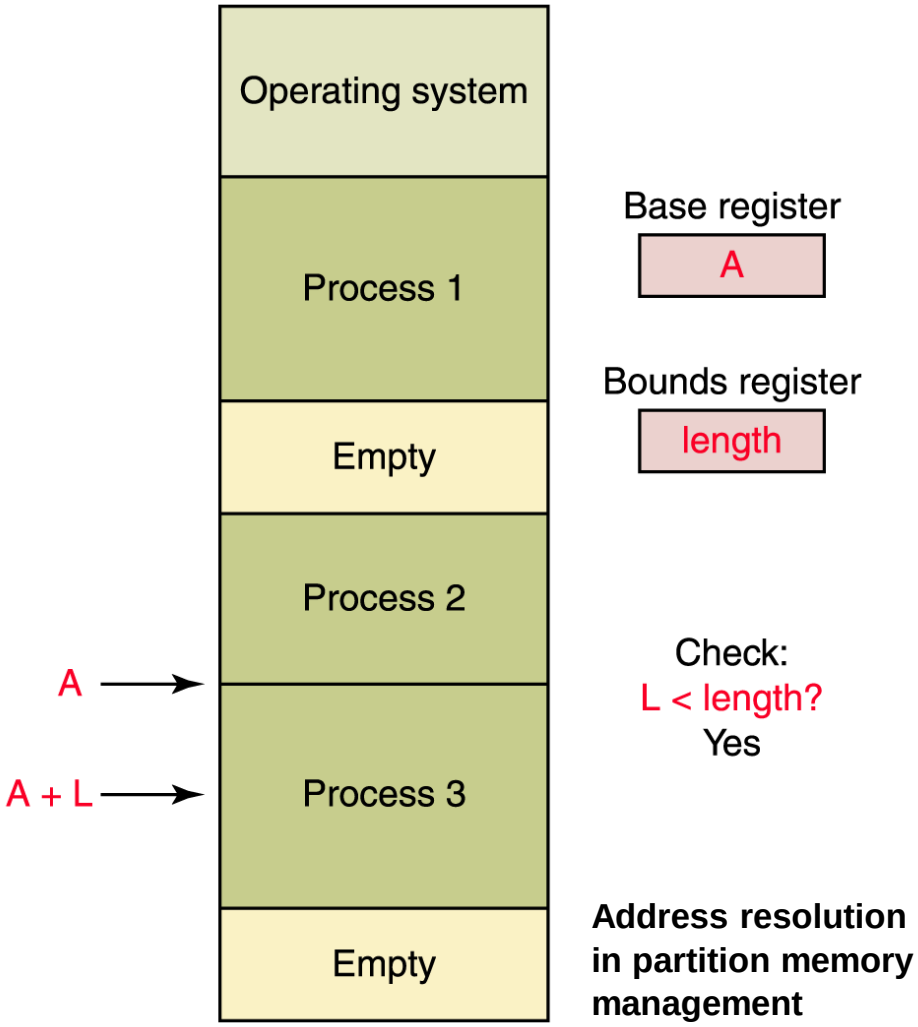
binding a logical address to a physical one

Partition Memory Management

SCSE

- **Fixed partitions** Main memory is divided into a particular number of partitions
- **Dynamic partitions** Partitions are created to fit the needs of the programs

Partition Memory Management



Address resolution in partition memory management

- At any point in time memory is divided into a set of partitions, some empty and some allocated to programs
- **Base register** A register that holds the beginning address of the current partition
- **Bounds register** A register that holds the length of the current partition

Partition Selection Algorithms

SCSE

Which partition should we allocate to a new program?

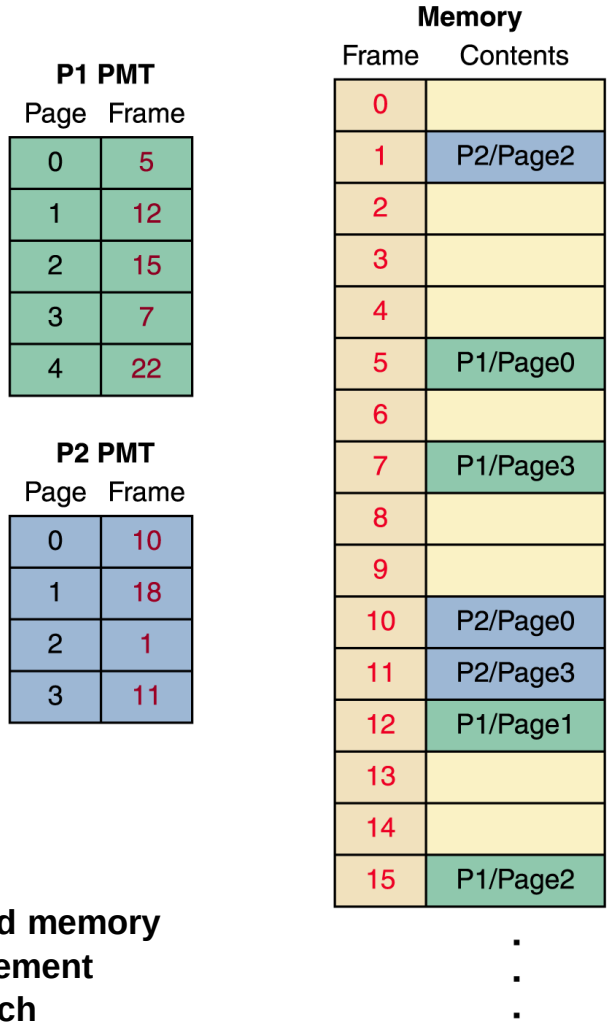
- **First fit** Allocate program to the first partition big enough to hold it
- **Best fit** Allocated program to the smallest partition big enough to hold it
- **Worst fit** Allocate program to the largest partition big enough to hold it

Paged Memory Management

SCSE

- **Paged memory technique** A memory management technique in which processes are divided into fixed-size **pages** and stored in memory **frames** when loaded into memory
  - **Frame** A fixed-size portion of *main memory* that holds a process page
  - **Page** A fixed-size portion of a *process* that is stored into a memory frame
  - **Page-map table** (PMT) A table used by the operating system to keep track of page/frame relationships

Paged Memory Management



A paged memory management approach

- The logical address is often written as <page, offset>.
- To produce a physical address, you first look up the page in the PMT to find the frame number in which it is stored
- Then multiply the frame number by the frame size and add the offset to get the physical address

Ex: if process 1 active, logical address of <1,222>
Page 1 of process 1 is in frame 12
*Physical address: 12*1024+222=12510*

Paged Memory Management

SCSE

- **Demand paging** An important extension of paged memory management
  - Not all parts of a program actually have to be in memory at the same time
  - In demand paging, the pages are brought into memory on demand
- **Page swap** The act of bringing in a page from secondary memory, which often causes another page to be written back to secondary memory

Paged Memory Management

SCSE

- The demand paging approach gives rise to the idea of **virtual memory**, the illusion that there are no restrictions on the size of a program
- Too much page swapping, however, is called **thrashing** and can seriously degrade system performance.

Example

- Given the following PMT

Page	1	1	2	3
Frame	5	2	7	3

If the frame size is 1024, what is the physical address associated with the logical address <2,85>?

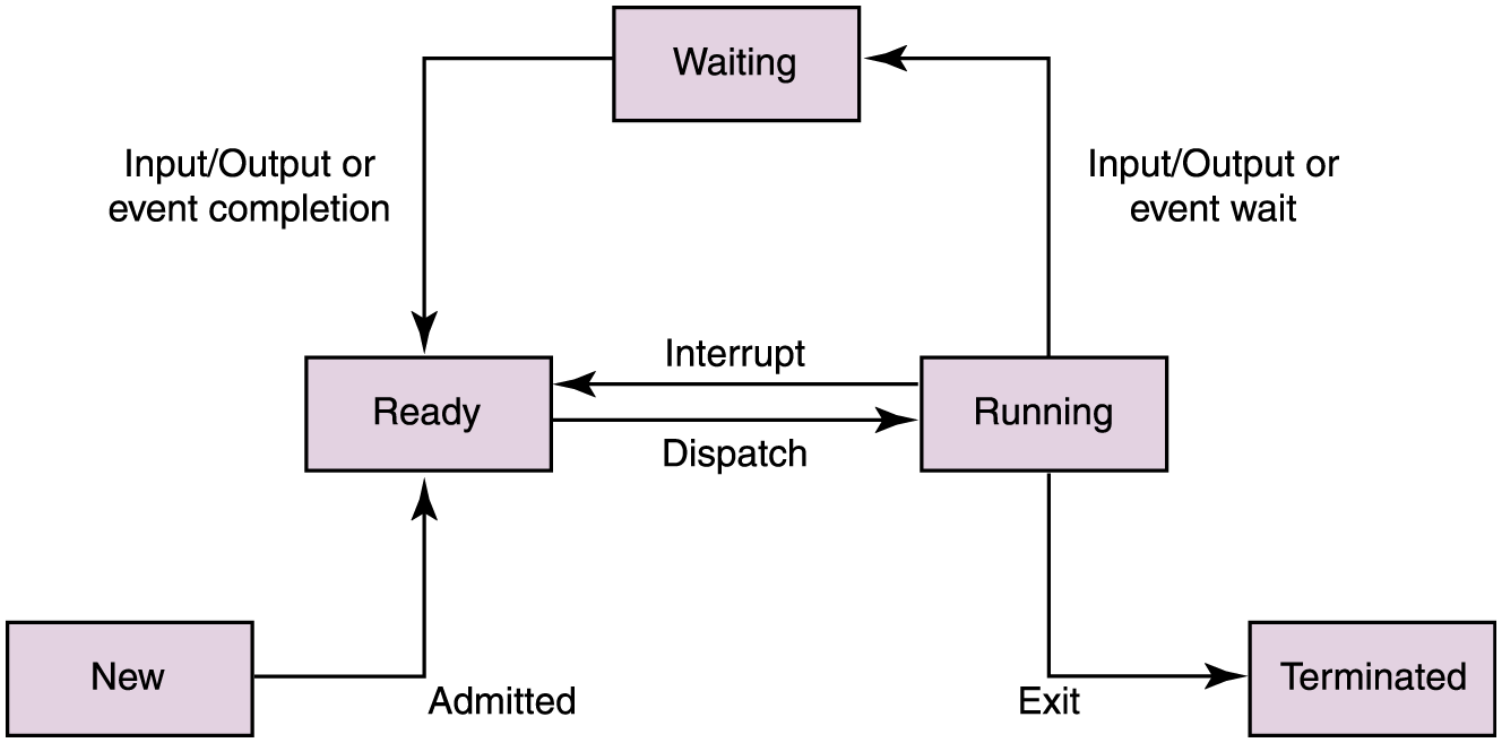
If the frame size is 1024, what is the physical address associated with the logical address <3,555>?

If the frame size is 1024, what is the physical address associated with the logical address <3,1555>?

Process Management

SCSE

- The Process States



The process life cycle

The Process Control Block

SCSE

- The operating system must manage a large amount of data for each active process
- Usually that data is stored in a data structure called a process control block **(PCB)**
- Each state is represented by a list of PCBs, one for each process in that state

The Process Control Block

SCSE

- Keep in mind that there is only one CPU and therefore only one set of CPU registers
  - These registers contain the values for the currently executing process
- Each time a process is moved to the running state:
  - Register values for the currently running process are stored into its PCB
  - Register values of the new running state are loaded into the CPU
  - This exchange of information is called a **context switch**

CPU Scheduling

SCSE

- **CPU Scheduling** The act of determining which process in the *ready* state should be moved to the *running* state
  - Many processes may be in the ready state
  - Only one process can be in the running state, making progress at any one time

CPU Scheduling Algorithms

SCSE

First-Come, First-Served

- Processes are moved to the CPU in the order in which they arrive in the running state

Shortest Job Next

- Process with shortest estimated running time in the ready state is moved into the running state first

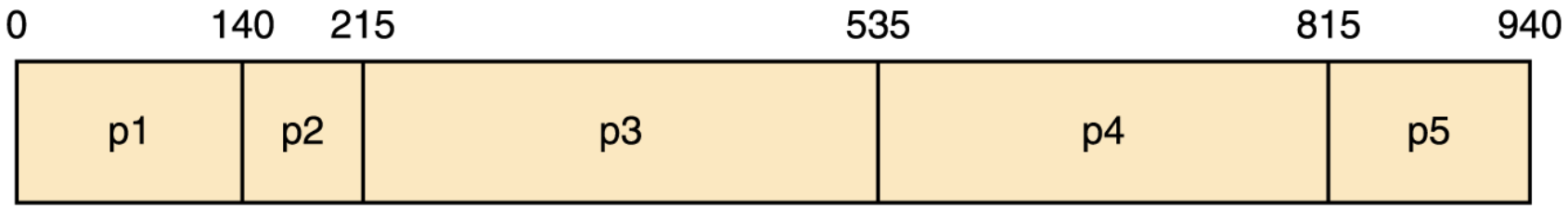
Round Robin

- Each process runs for a specified time slice and moves from the running state to the ready state to await its next turn if not finished

First-Come, First-Served

SCSE

Process	Service time
p1	140
p2	75
p3	320
p4	280
p5	125

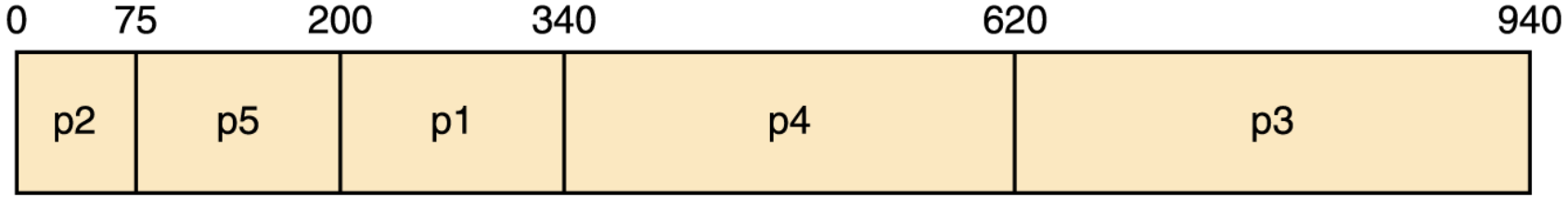


The average turnaround time is: (140+215+535+815+940)/5=529

Shortest Job Next

SCSE

- Looks at all processes in the ready state and dispatches the one with the smallest service time



The average turnaround time is: (75+200+340+620+940)/5=435

Round Robin

SCSE

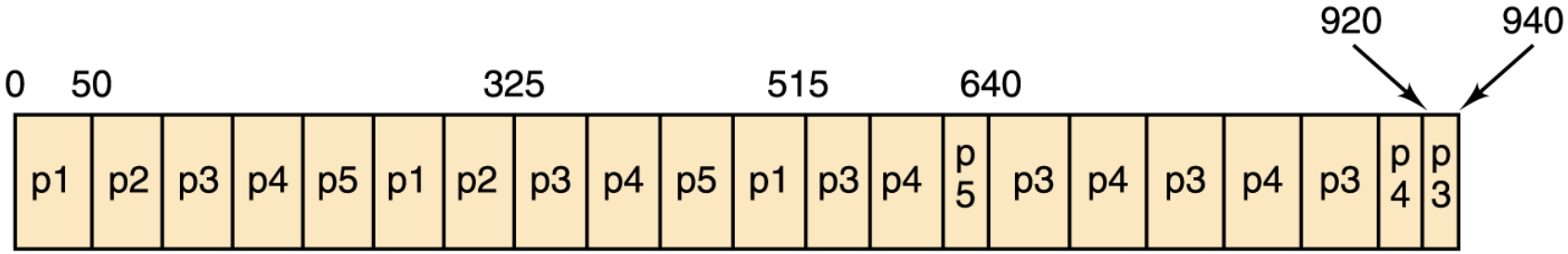
- Distributes the processing time equitably among all ready processes
- The algorithm establishes a particular **time slice** (or time quantum), which is the amount of time each process receives before being preempted and returned to the ready state to allow another process its turn

Round Robin

Process	Service time
p1	140
p2	75
p3	320
p4	280
p5	125

SCSE

- Suppose the time slice was 50



The average turnaround time is: (515+325+940+920+640)/5=668

Example

Given the following table of processes and services time

Process	P1	P2	p3	P4	P5
Service time	120	60	180	50	300

Draw a Gantt chart that show the completion time for each process using:

+ first-come, first-served CPU scheduling
+ the shortest-job-next CPU scheduling
+ round-robin CPU scheduling with a time slice of 60.

Summary

An operating system:

- is the part of the system software,
- manages resources on a computer,
- serves as modulator among human users, applications software, hardware devices in the system.

Summary (cont)

Multiprogramming: is the technique for keeping multiple programs in memory at the same time

A process: is a program in execution.

→**The OS:** performs carefully CPU scheduling, memory management, process management to ensure fair access to the CPU.

Summary (cont)

Batch processing: organizes into batches using the same or similar resources.

Timesharing: allows multiple users to interact with a computer at the same time, creating a virtual machine for each user.

→**The OS:** manages memory to control and monitor where processes are loaded into main memory

Summary (cont)

Memory management technique: defines the manner in which it binds a logical address to a physical one.

The single contiguous approach: allows only one program other than the OS to be in main memory

The partition approach: divides memory into several partitions into which processes are loaded. Fix and dynamic partitions are applied.

The paging approach: devide memory into frames and programs into pages.

Summary (cont)

CPU scheduling algorithms: determine which process gets priority to use the CPU next.

First-come, first-served CPU scheduling: gives priority to the earliest-arriving job.

The shortest-job-next algorithm: gives priority to jobs with short running times.

Round-robin scheduling: rotates the CPU among active processes, giving a little time to each process.