CS241 System Programming
CPU Scheduling

Klara Nahrstedt

Lecture 10

2/10/2006

Content

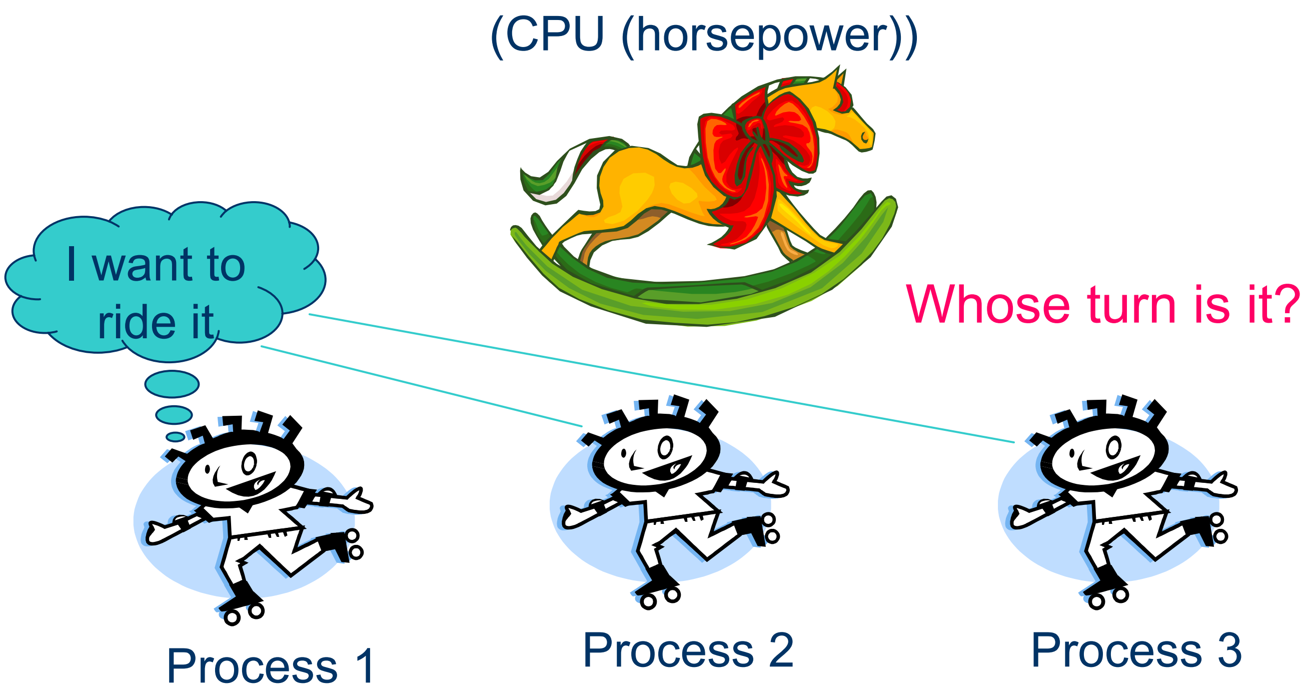
- Why Scheduling?
- Scheduling Levels
- Basic Scheduling Algorithm (FCFS)
- Summary

Administrative

- MP1 due February 13
- Read: T: 2.4 (Scheduling)

Scheduling

- Deciding which process/thread should occupy the resource (CPU, disk, etc)



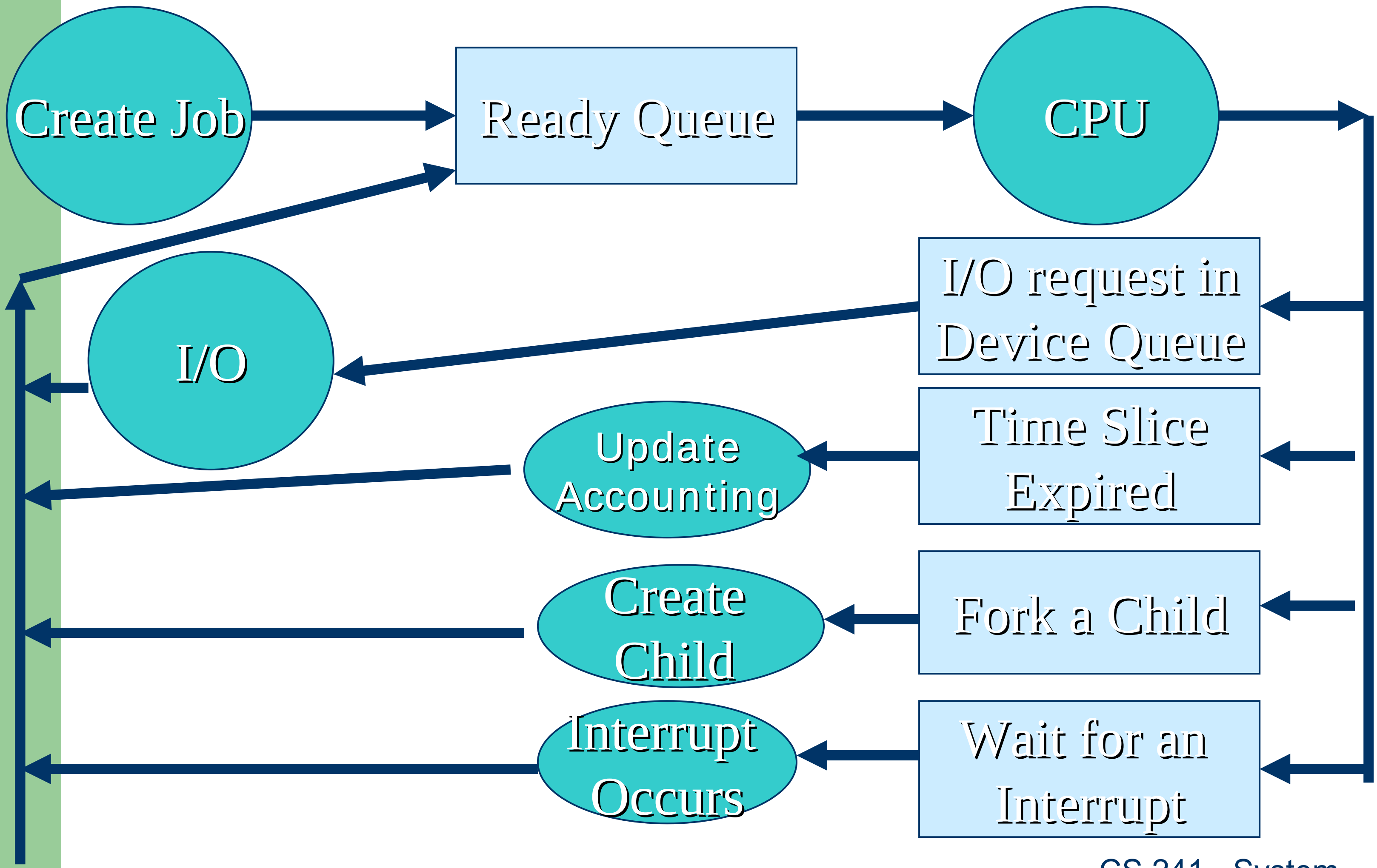
When to schedule?

- A new process starts
- The running process exits
- The running process is blocked
- I/O interrupt (some processes will be ready)
- Clock interrupt (every 10 milliseconds)

Scheduling Queues

- Processes in ready state enter READY QUEUE
  - This queue is generally stored as a linked list
  - Ready-queue header contains pointers to the first and last PCBs in the list
  - Each PCB has a pointer field that points to the next process in he ready queue
- Other Queues
  - I/O Queue

Queuing Diagram for Processes

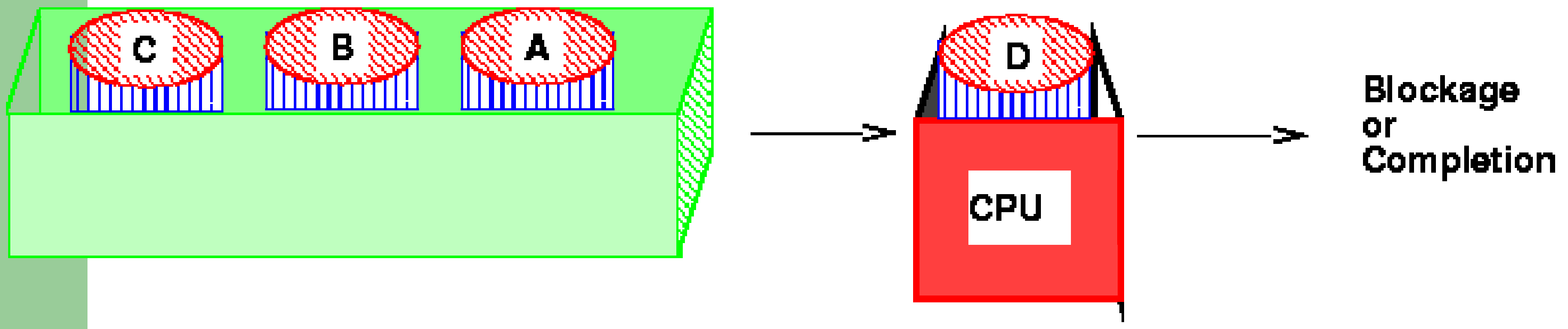


CPU Scheduler

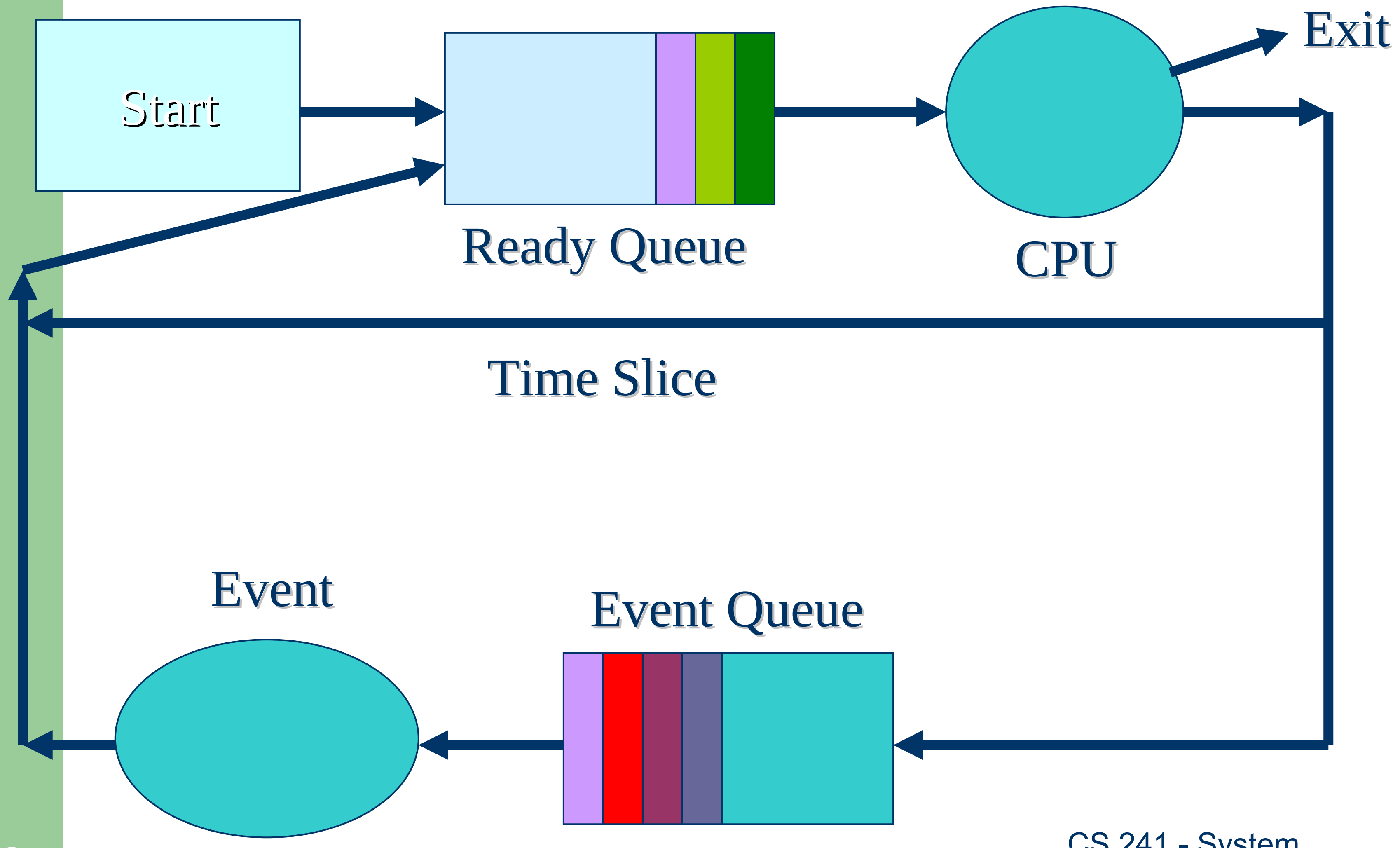
- CPU scheduler selects one of the processes from the ready queue and allocates CPU to one of them;
- Ready queue:= FIFO queue, tree queue, or unordered linked list, or priority queue;
- Elements in the ready queue are PCBs (Process Control Block) of processes

CPU Scheduler Example

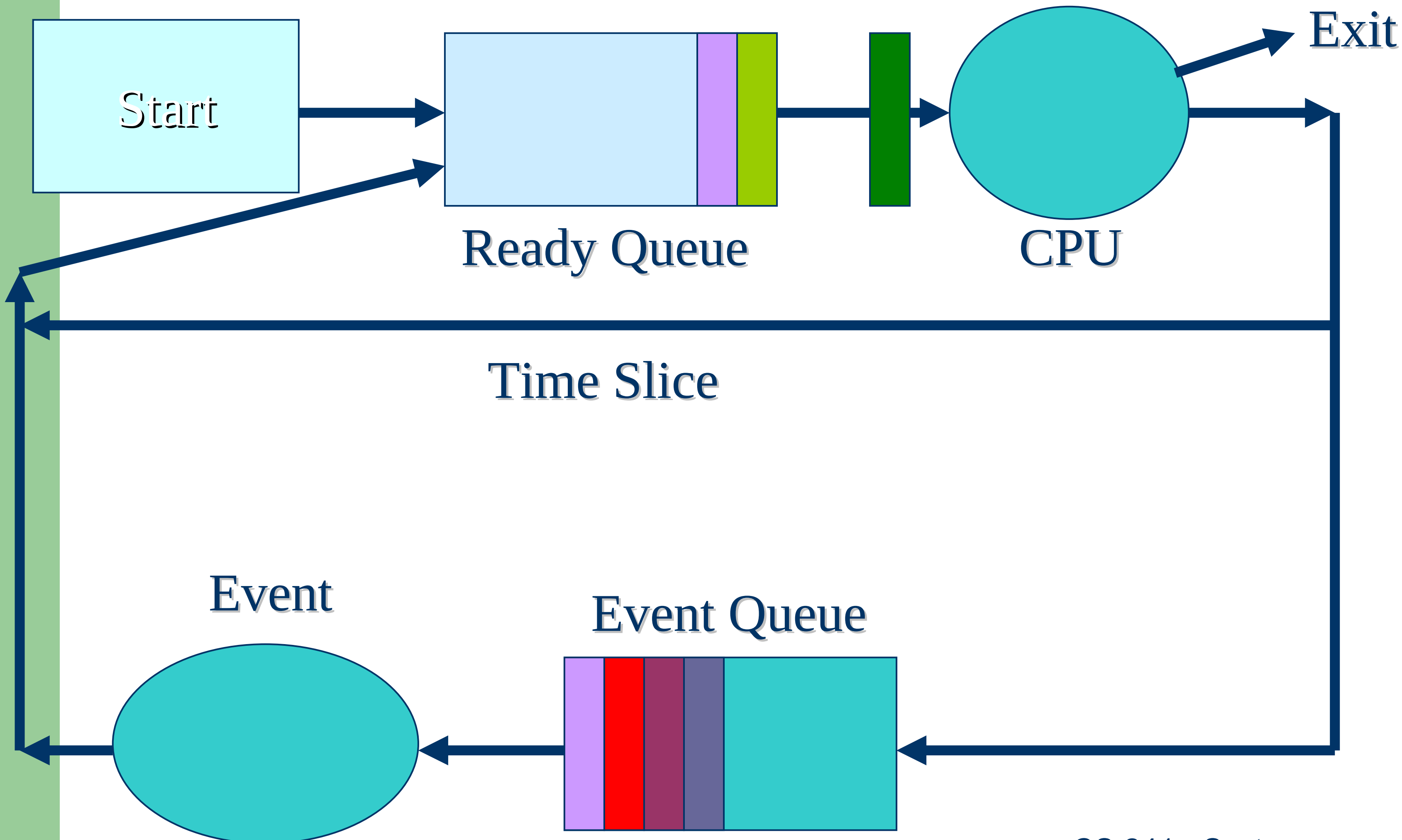
- Process 1 burst computes for 14 time units
- Process 2 burst computes for 8 time units
- Process 3 burst computes for 8 time units



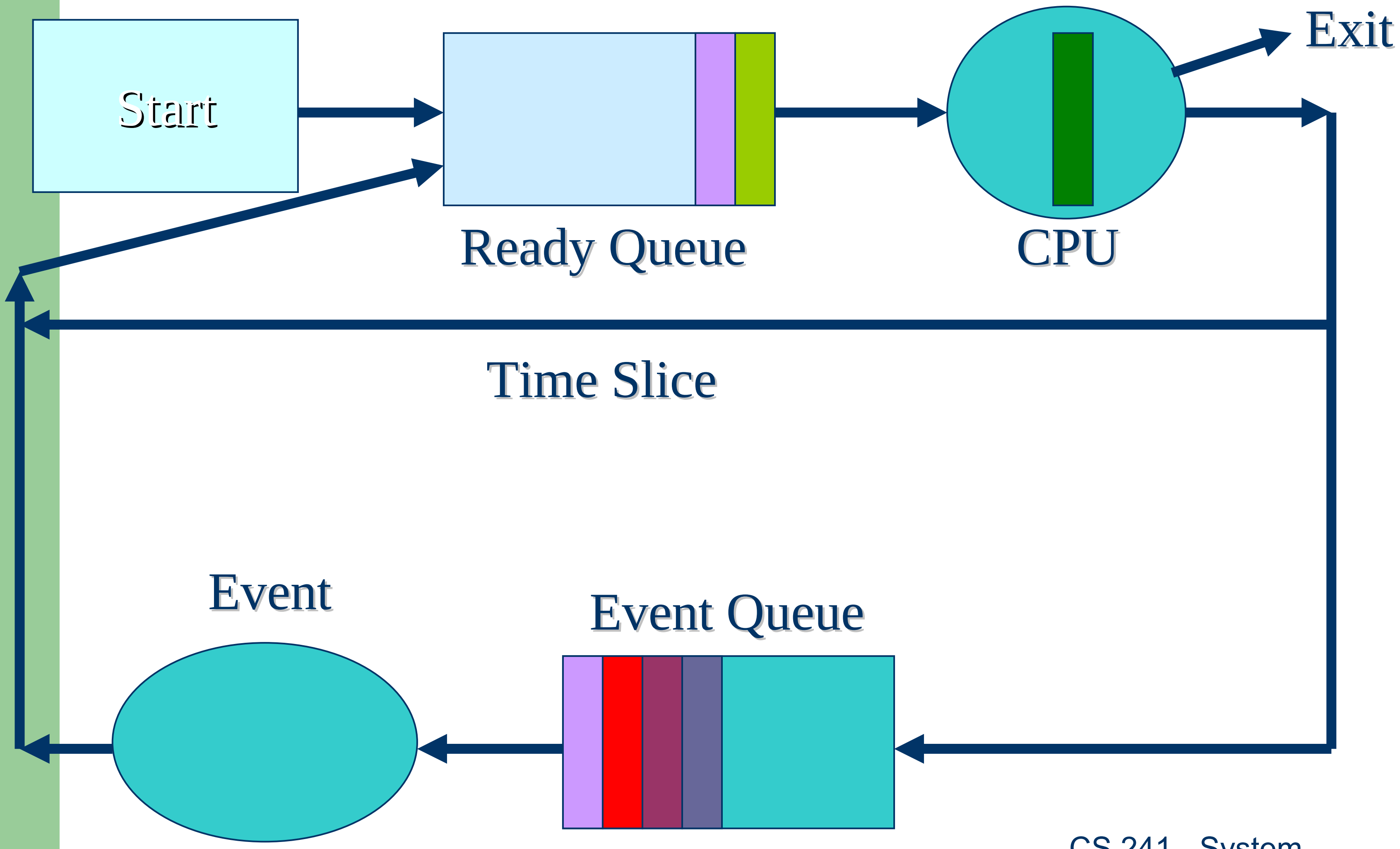
Queuing Diagram for Processes



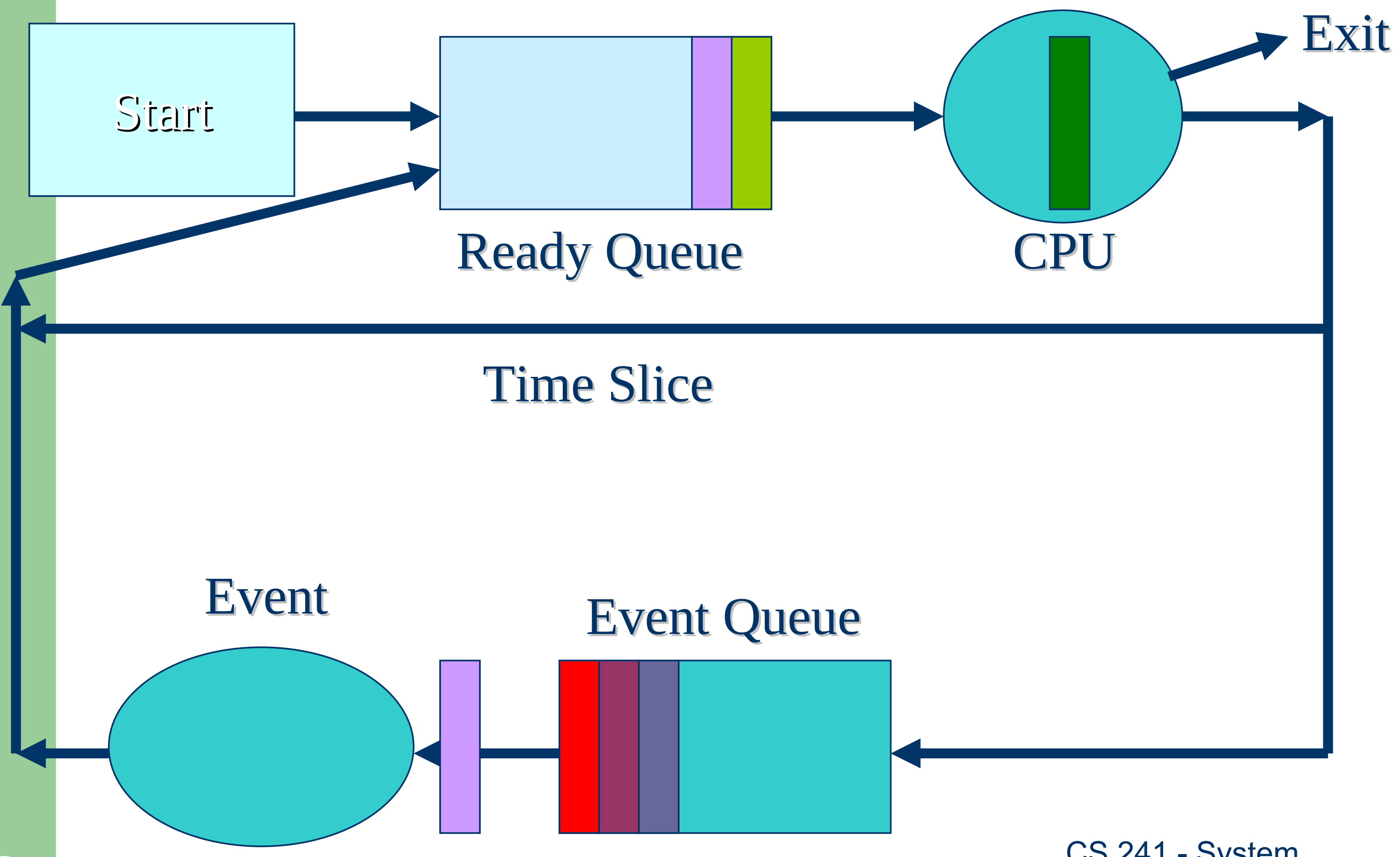
Queuing Diagram for Processes



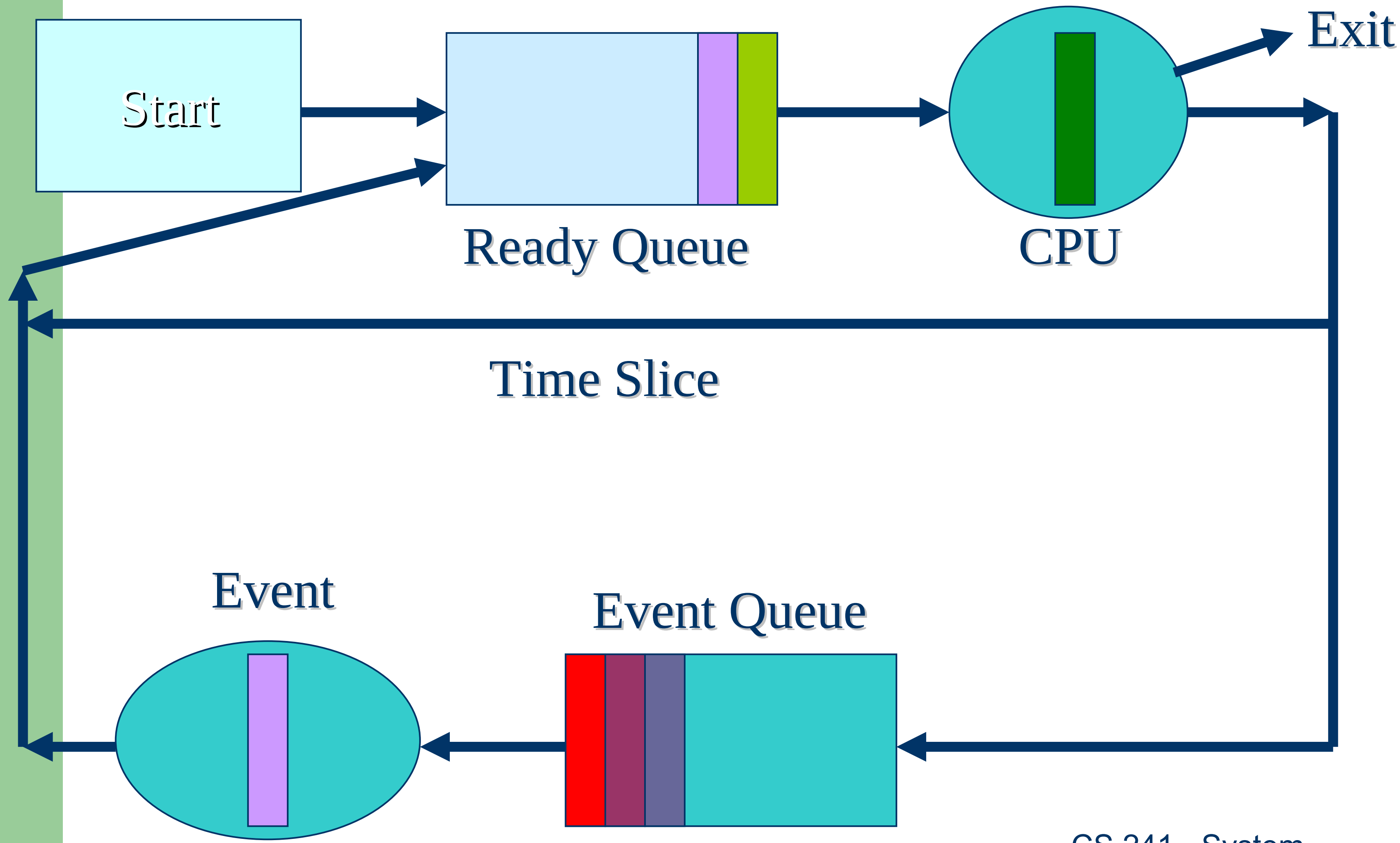
Queuing Diagram for Processes



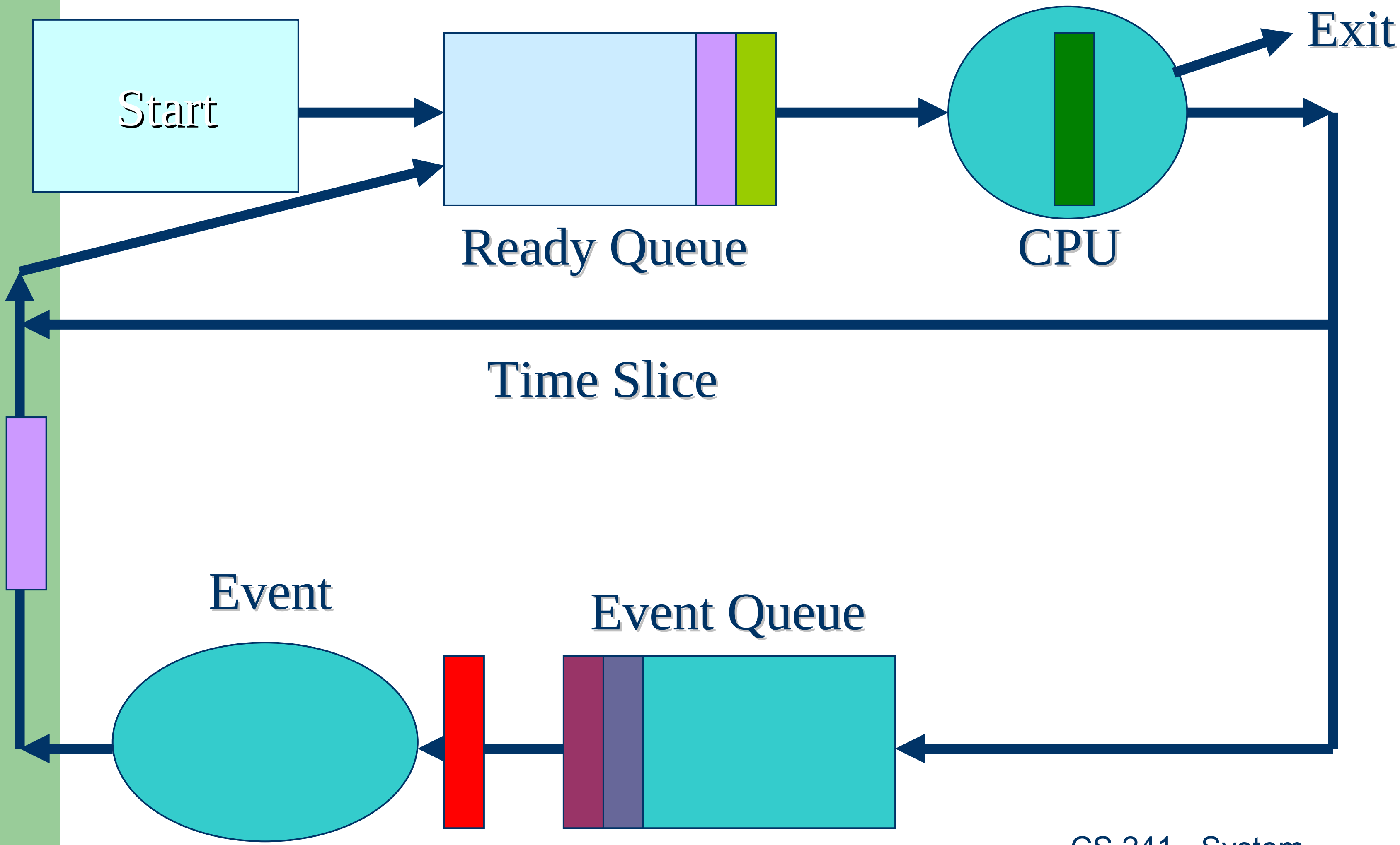
Queuing Diagram for Processes



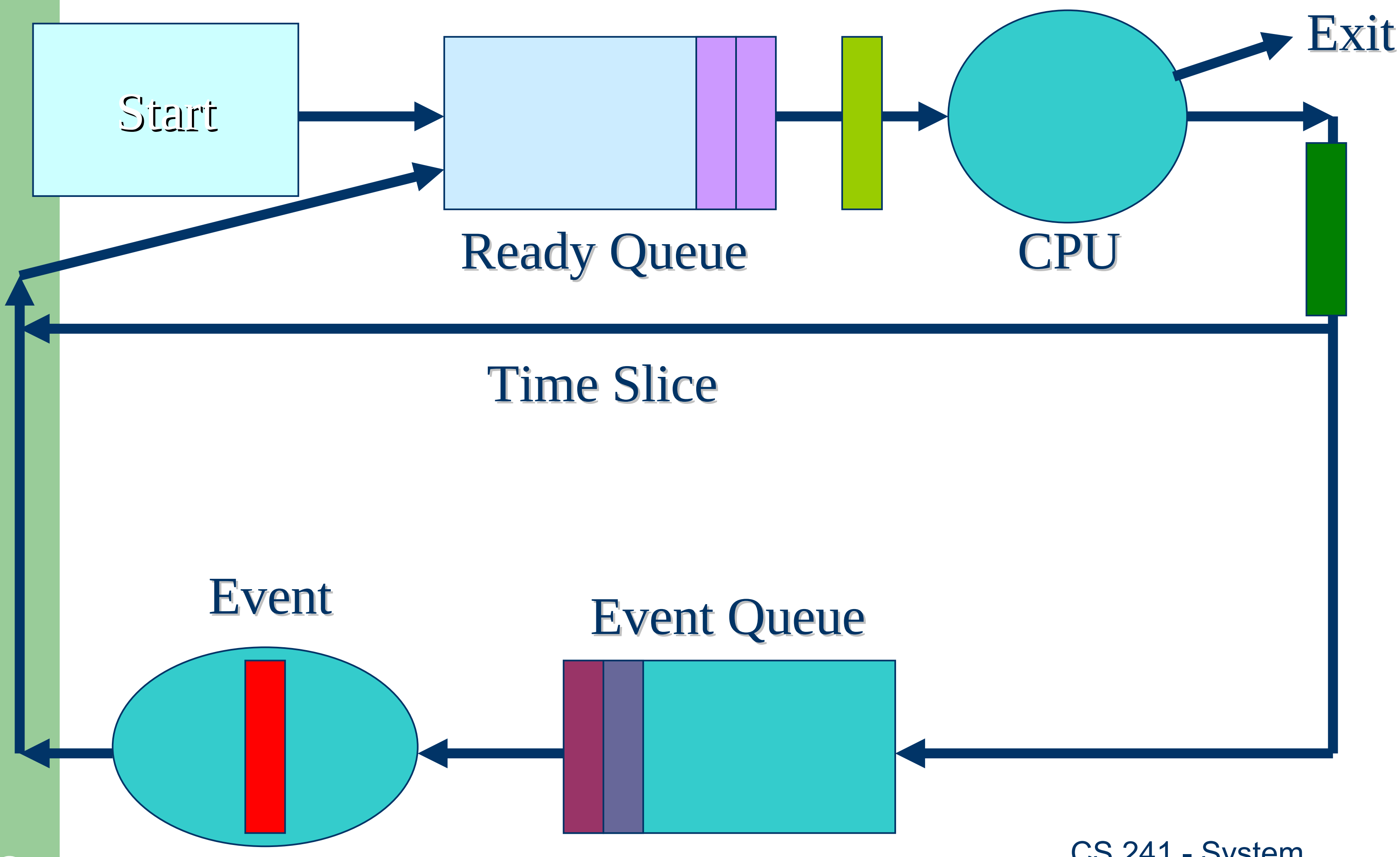
Queuing Diagram for Processes



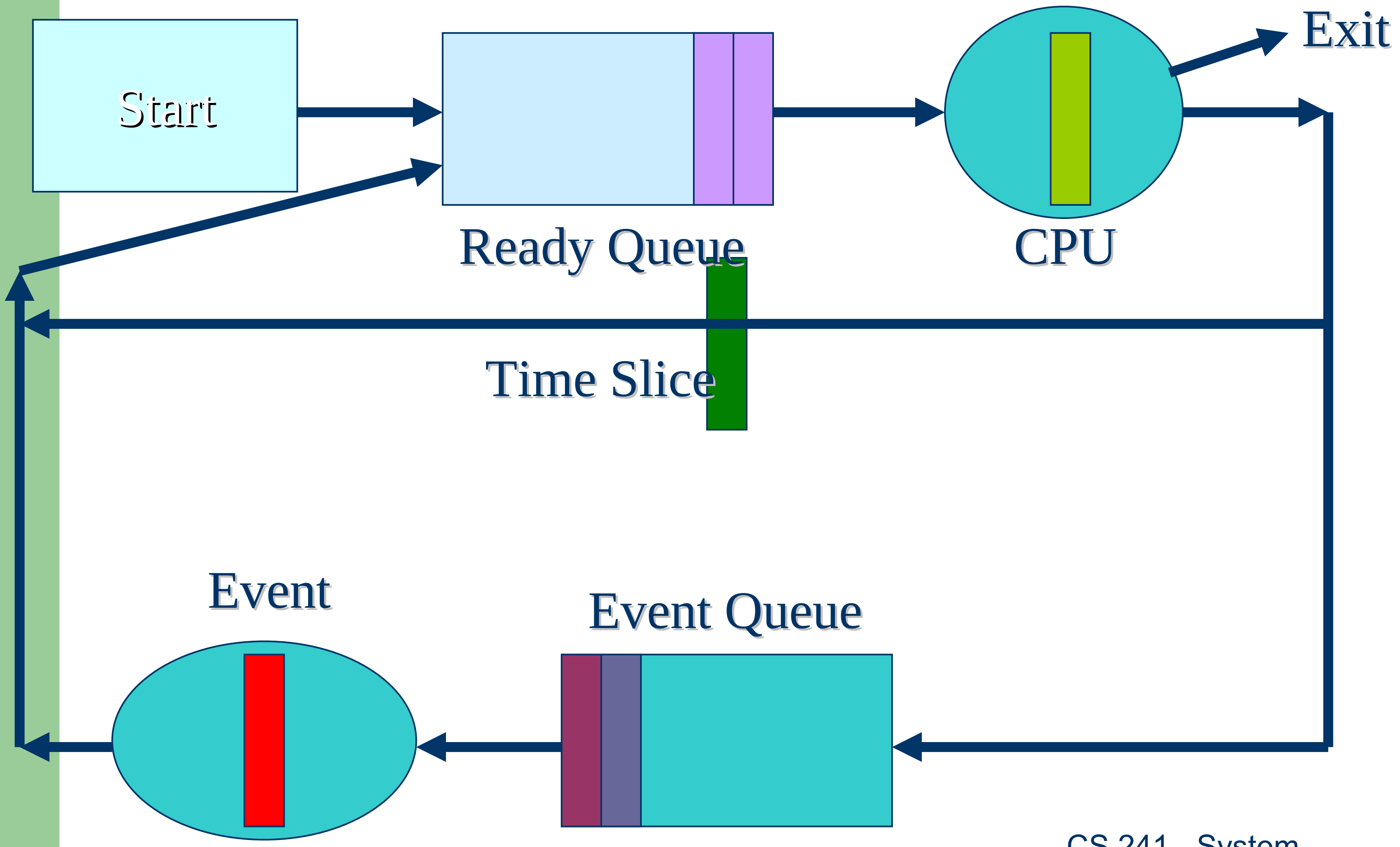
Queuing Diagram for Processes



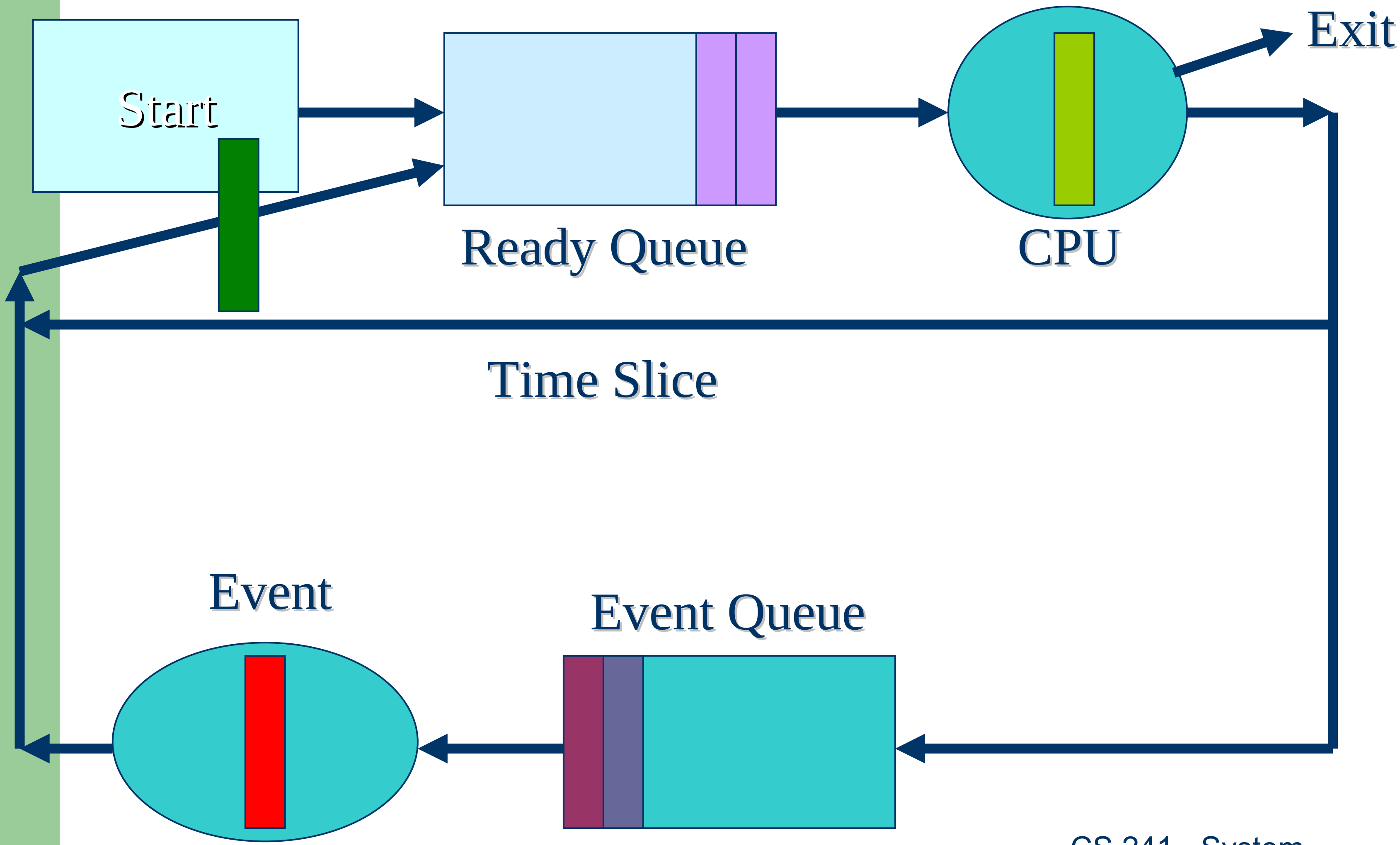
Queuing Diagram for Processes



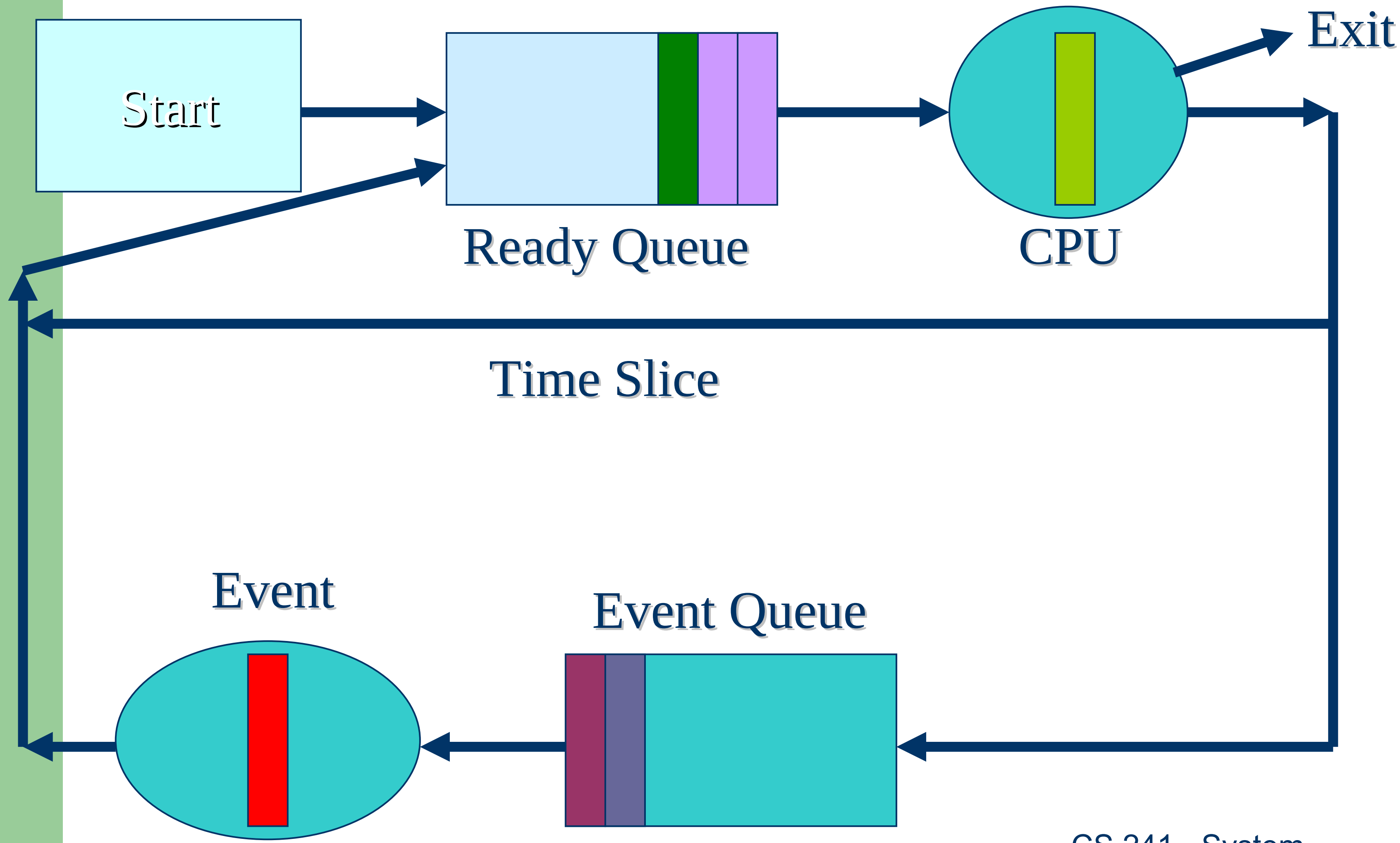
Queuing Diagram for Processes



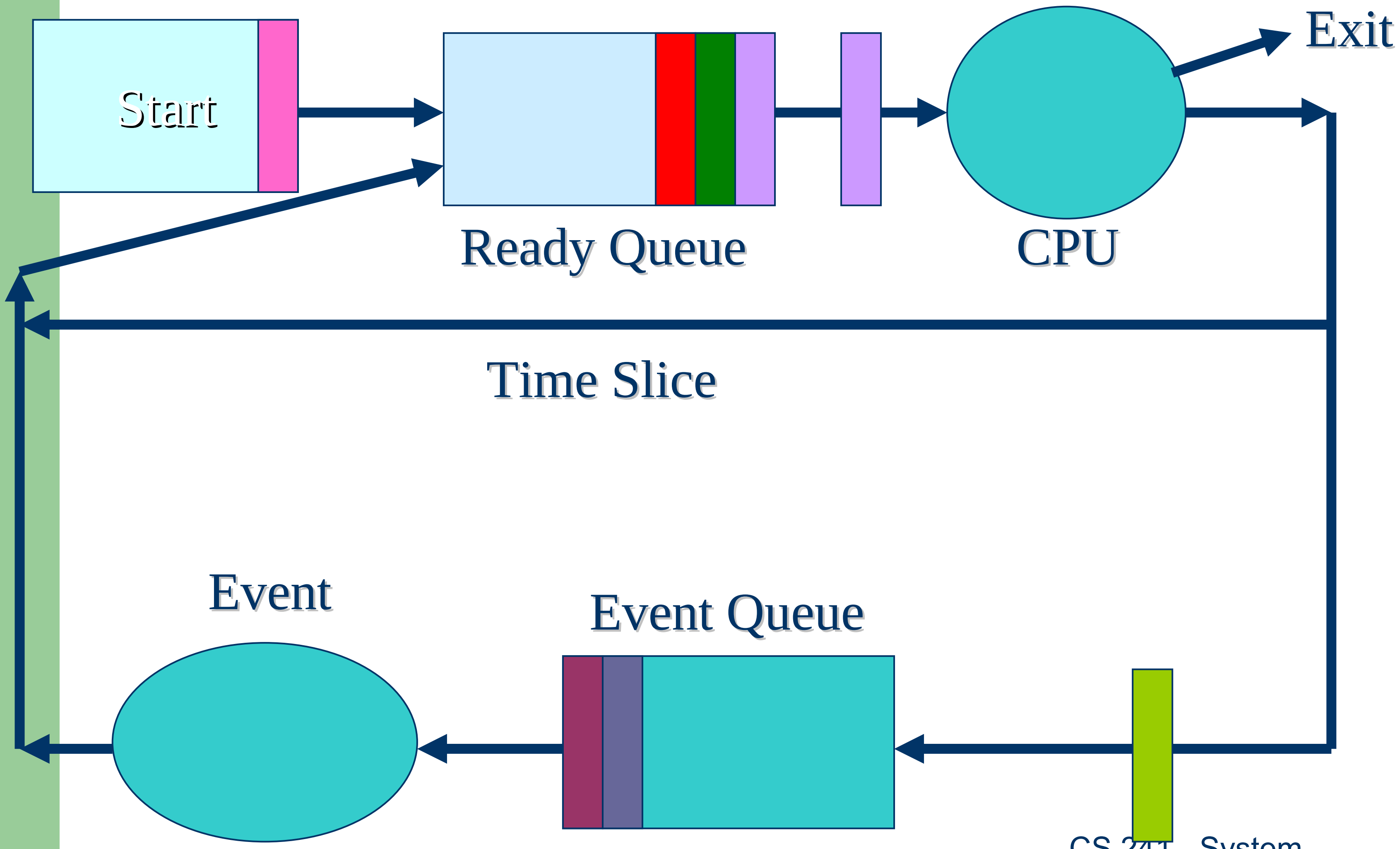
Queuing Diagram for Processes



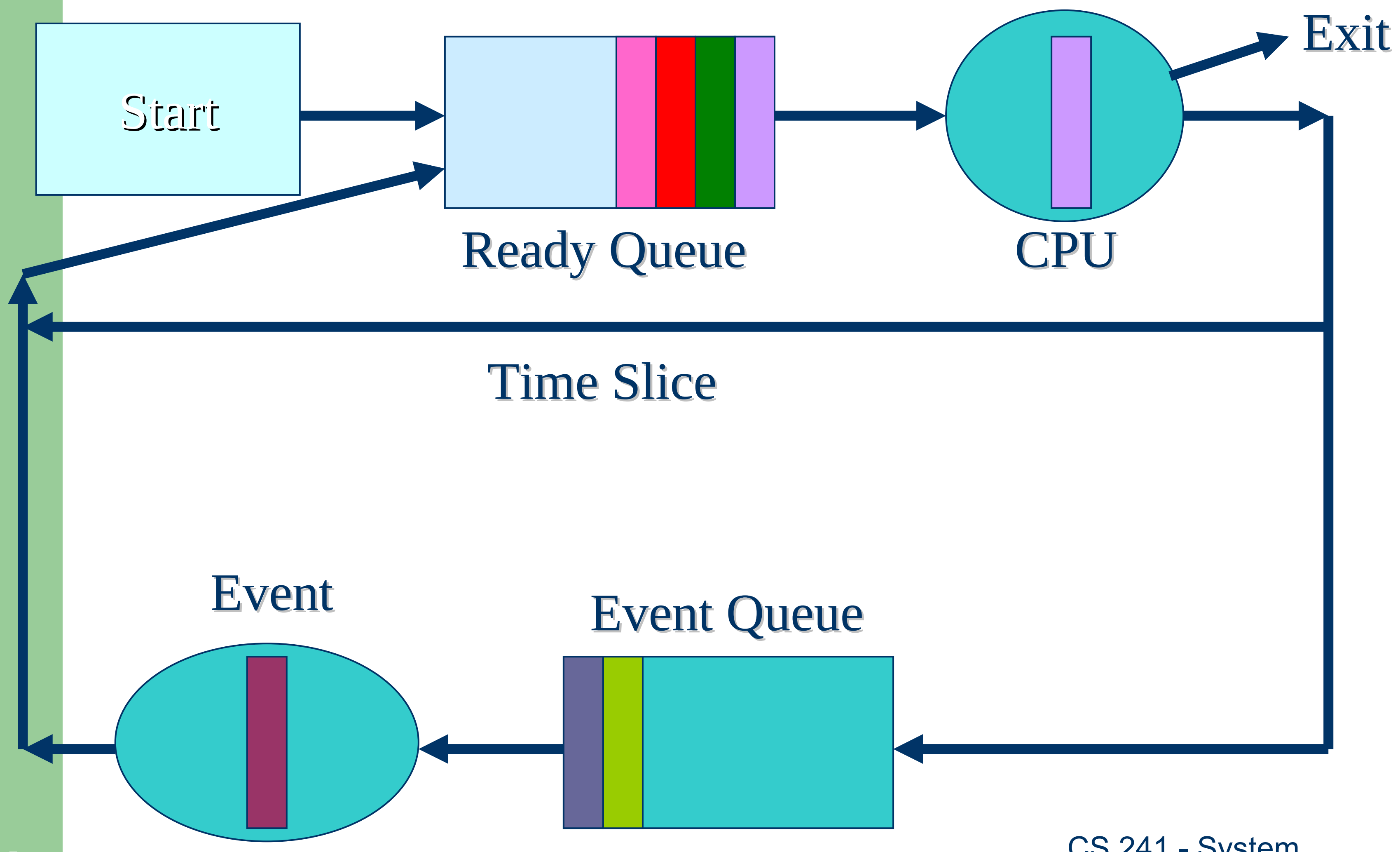
Queuing Diagram for Processes



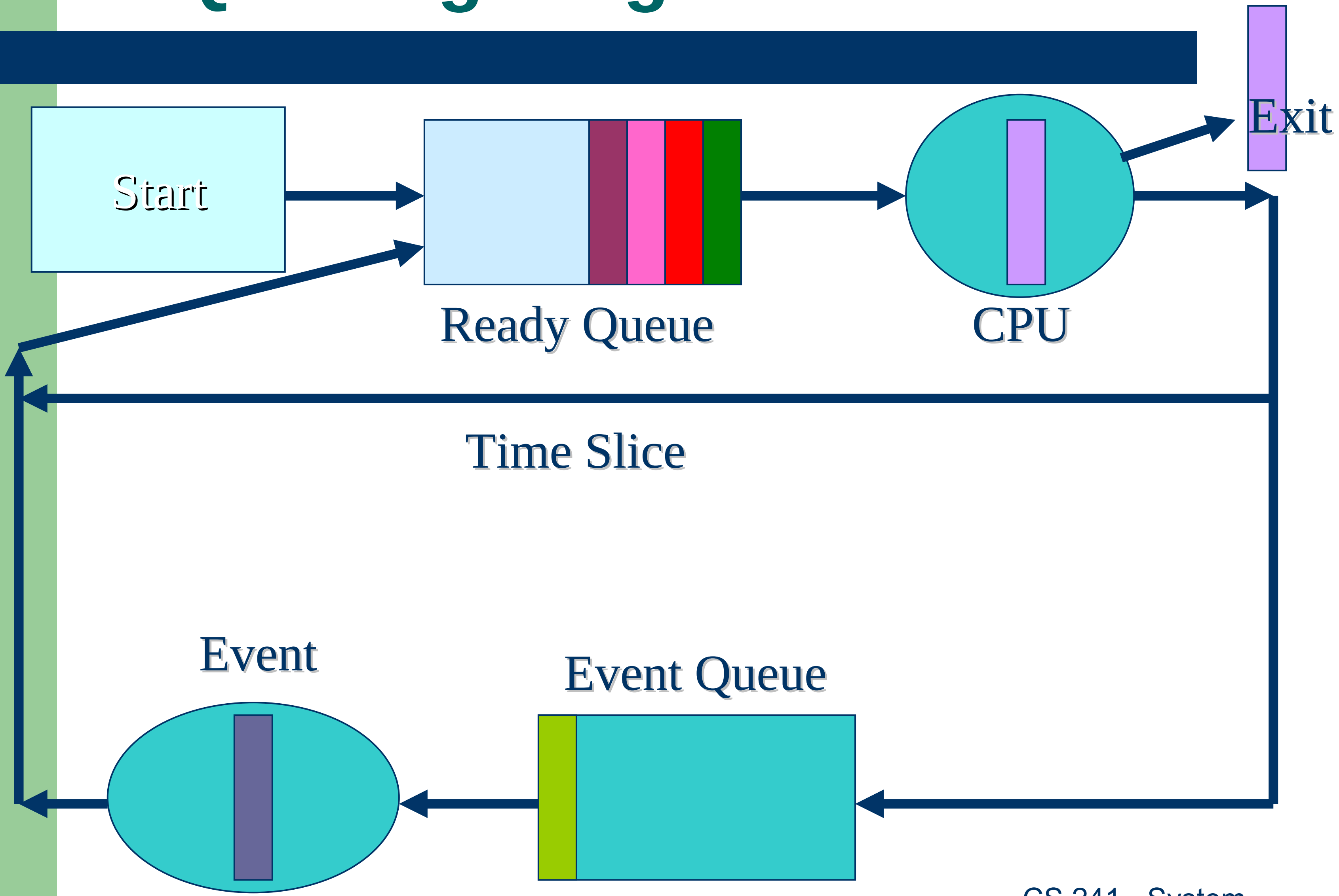
Queuing Diagram for Processes



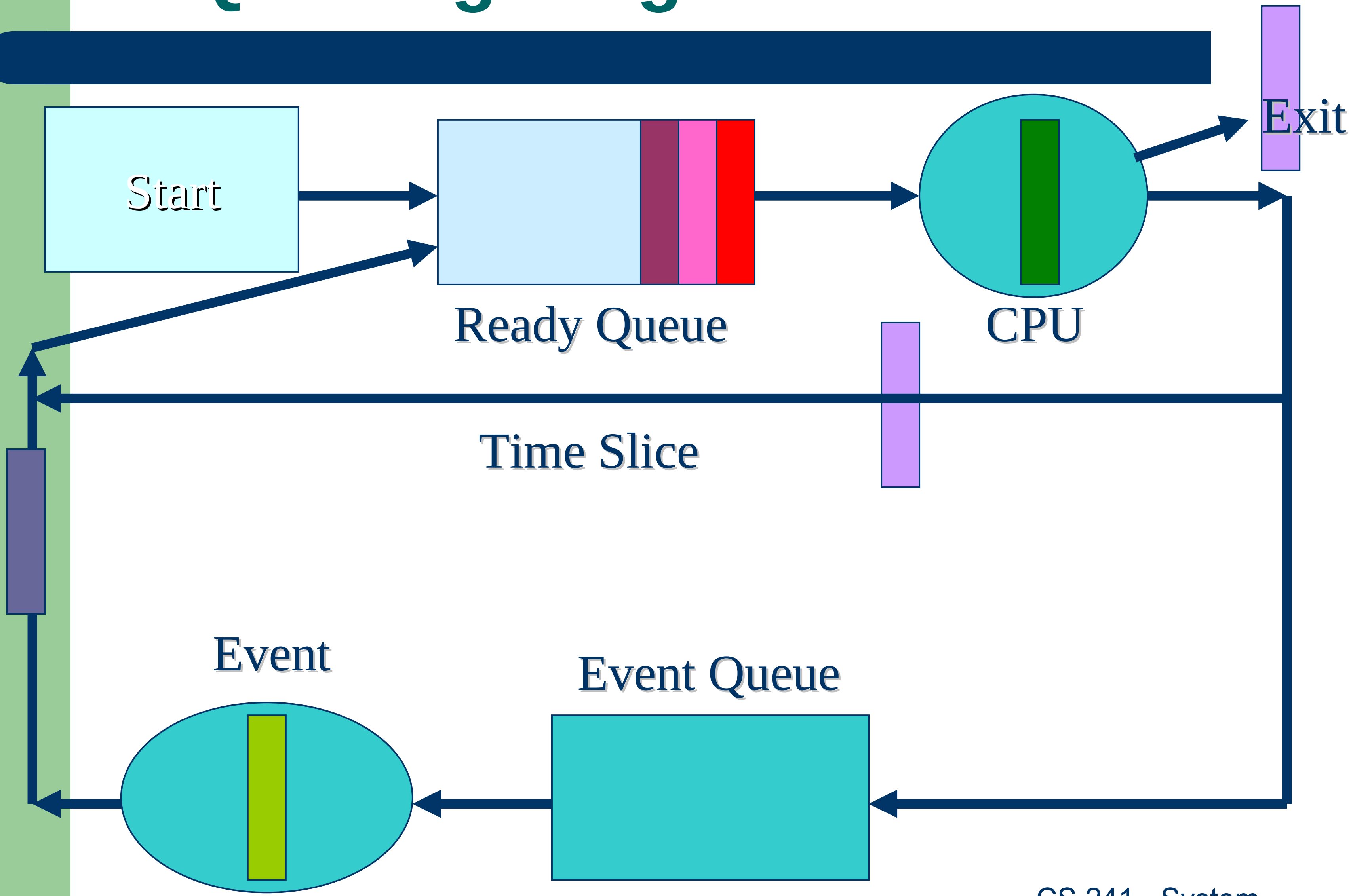
Queuing Diagram for Processes



Queuing Diagram for Processes



Queuing Diagram for Processes



Queueing Theory

- Model
- Poisson Distribution and Random Arrival rates

Problem

- 7 Jobs arrive on average every second
- 8 Jobs are processed by system on average every second

1) How long does it take for a job to be serviced?

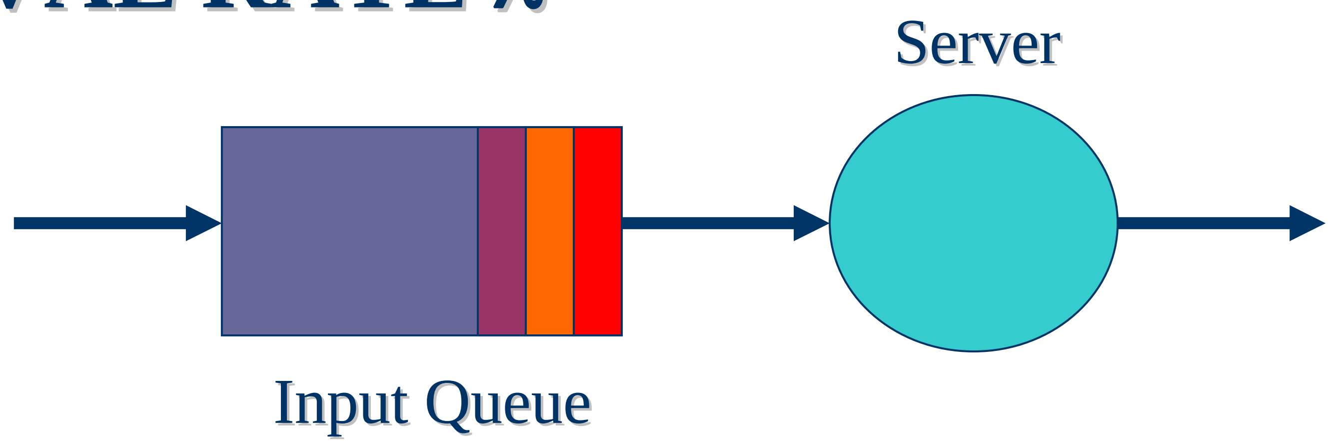
2) How many jobs are waiting in queue to be serviced?

Random Events

- Poisson Distribution
- Each event independent of other events
- Mean event rate, SD is same as mean
- Exponential distribution

Queuing Theory

ARRIVAL RATE λ



SERVICE RATE μ

Queueing Theory

- Steady state
- Poisson arrival with λ constant arrival rate (customers per unit time) each arrival is independent.
- $P(\tau \leq t) = 1 - e^{-\lambda t}$

Analysis of Queueing Behavior

- Probability n customers arrive in time interval t is:

  $e^{-\lambda t} (\lambda t)^n / n!$

- Assume random service times (also Poisson): μ constant service rate (customers per unit time)

Useful Facts From Queuing Theory

- W_q= mean time a customer spends in the queue λ = arrival rate
- L_q = number of customers in queue
- W = mean time a customer spends in the system
- L = number of customers in the system

Useful Facts From Queuing Theory

- $L_q = \lambda w_q$
- $L = \lambda W$ (Little's theorem)

Analysis of Single Server Queue

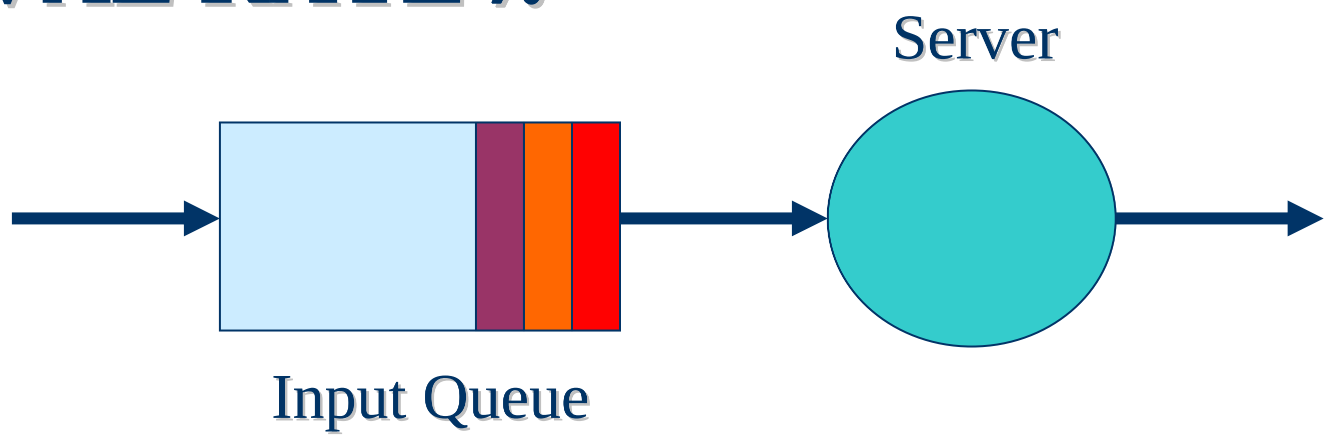
- Server Utilization: $$\rho = \frac{\lambda}{\mu}$$
- Time in System: $$W = \frac{1}{\mu - \lambda}$$
- Time in Queue: $$W_q = \frac{\rho}{\mu - \lambda}$$
- Number in Queue (Little): $$L_q = \frac{\rho^2}{1 - \rho}$$

Example

- Example
  - Arrival 2 jobs/sec
  - Service 3 jobs/sec
  - Utilization 66.66%
  - Time in system 1 sec
  - Time in queue .6666 sec
  - Length of queue 1.3333

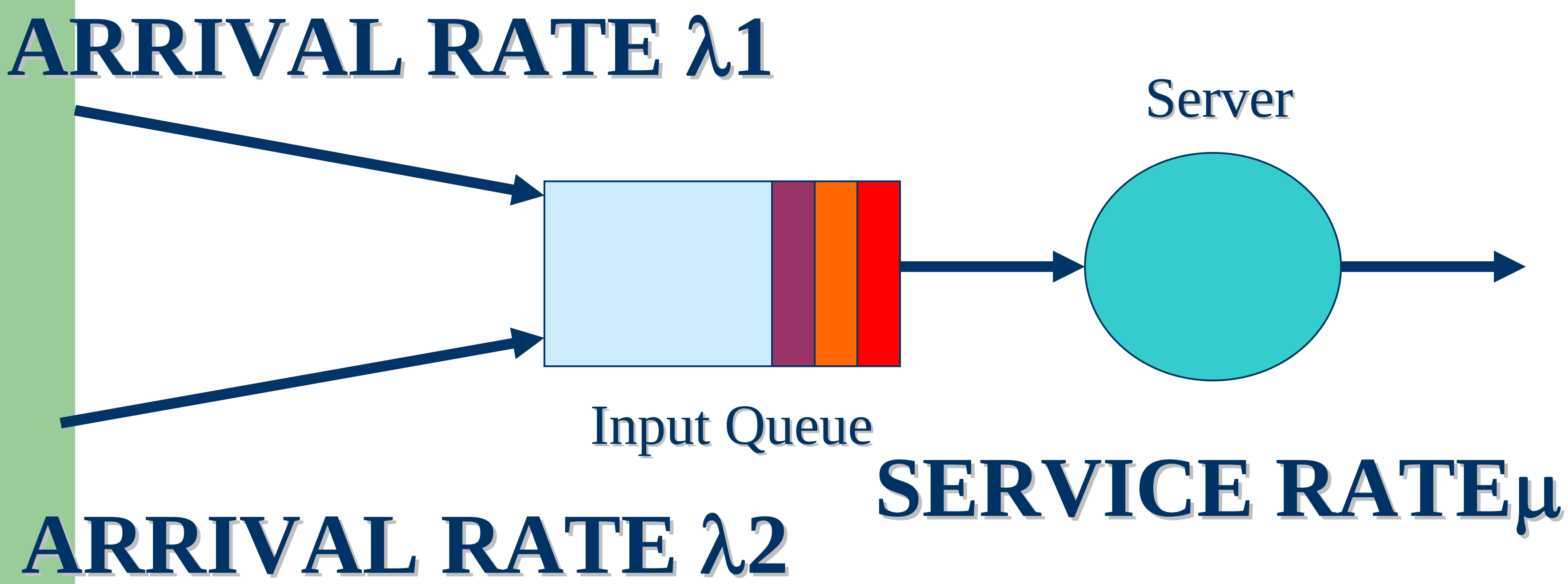
Queuing Theory

ARRIVAL RATE λ



SERVICE RATE μ

Queuing Theory

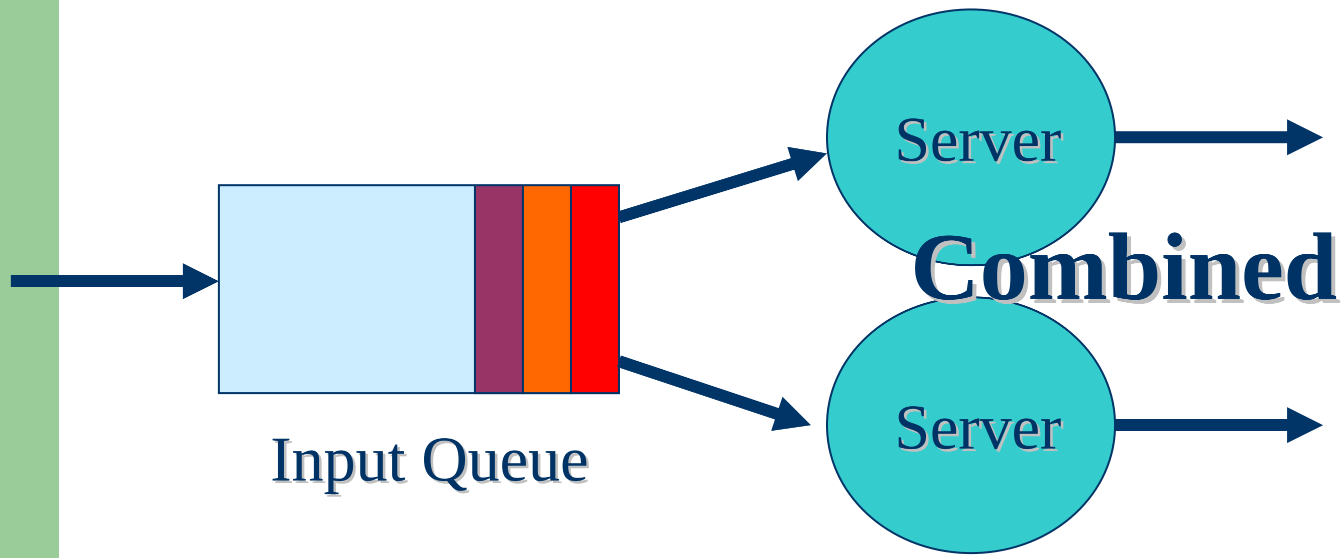


$$\lambda = \lambda 1 + \lambda\, 2$$

Queuing Theory

ARRIVAL RATE λ

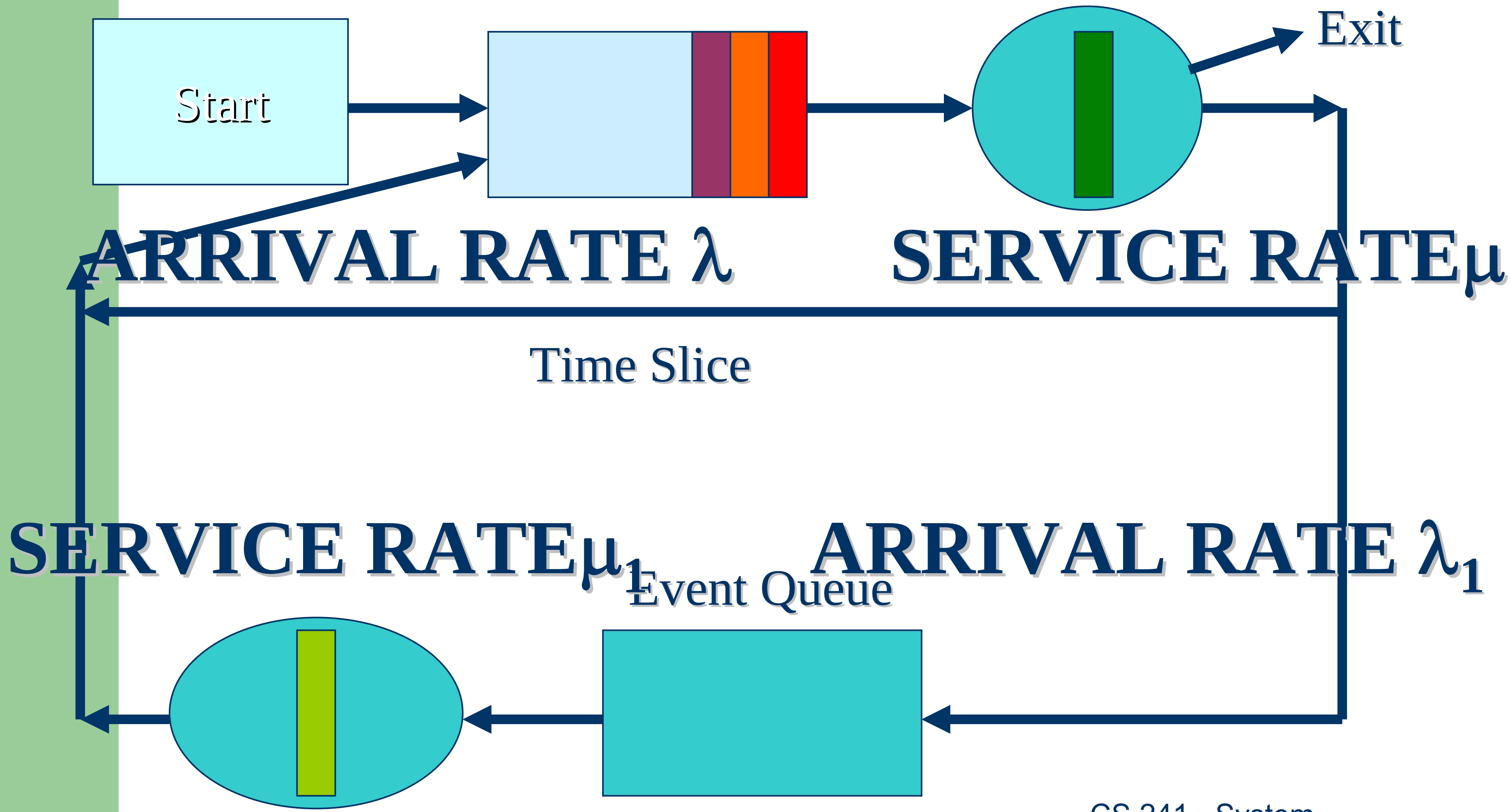
SERVICE RATE μ_1



Combined $\mu = \mu_1 + \mu_2$

SERVICE RATE μ_2

Queuing Diagram for Processes



Example

Arrival 1 job/sec from Start

Arrival 2 jobs/sec from Event queue

Service 4 jobs/sec

- Utilization=$\rho=\lambda/\mu=(1+2)/4=0.75$
- Time in system=$1/(\mu-\lambda)=1$
- Time in queue=$\rho/(\mu-\lambda)=.75$
- Length of queue=$\rho*\rho/(1-\rho)=2.25$

Summary

- What is scheduling
- Scheduling objectives
- CPU Scheduling
- Next lecture:
  - FCFS
  - Queuing Theory
  - Short Job First
  - Round Robin
  - Priority
  - Multi-Queue