



Operating System

Dr. Satyabrata Das

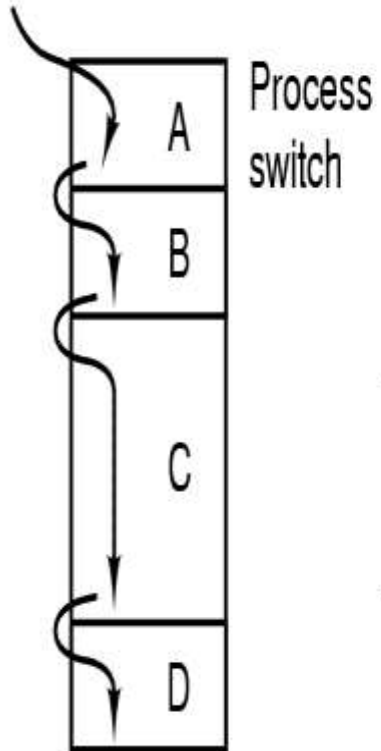
Associate Professor

Dept. of IT, VSSUT, Burla

Process Switch

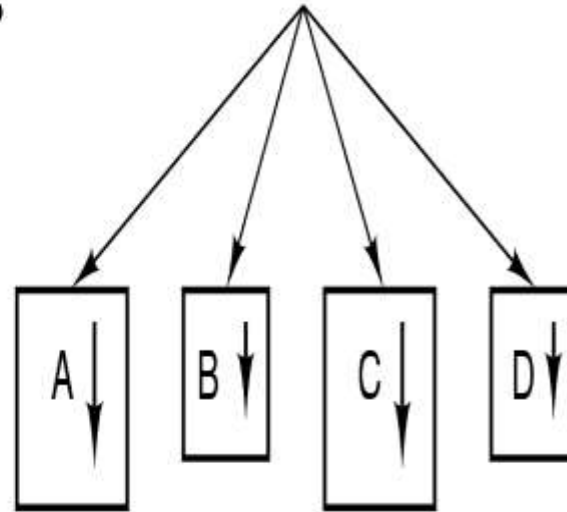


One program counter

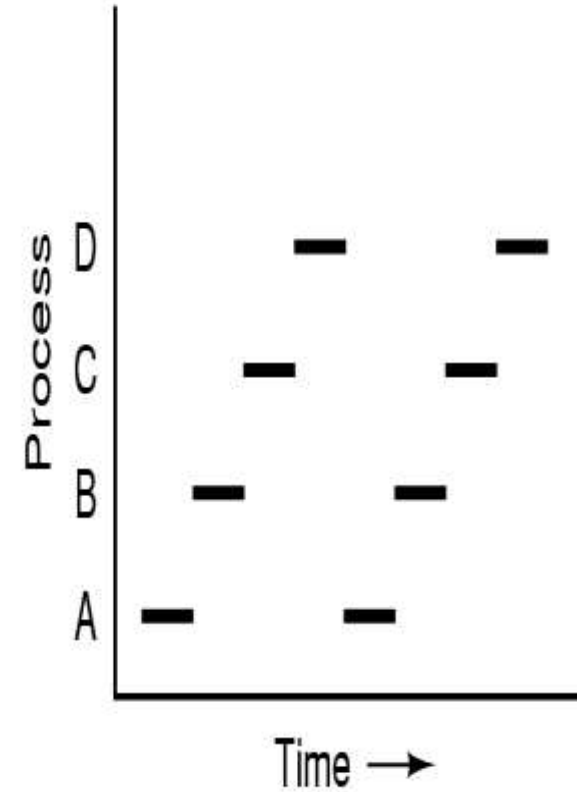


(a)

Four program counters



(b)



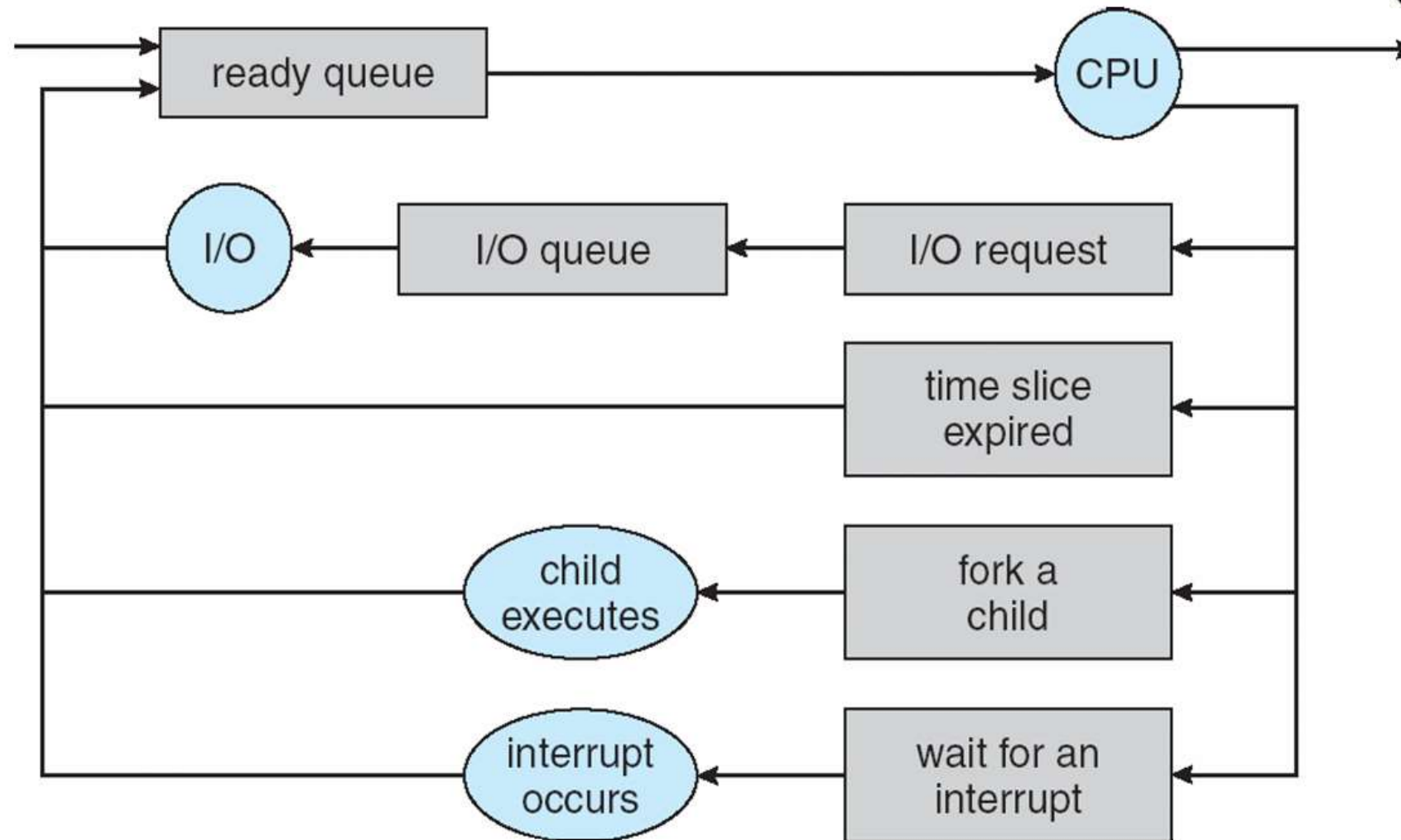
(c)

When to Switch a Process?



- A process switch may occur whenever the OS has gained control of CPU. i.e., when:
 - Supervisor Call
 - explicit request by the program (example: file open) – the process will probably be blocked.
 - Trap
 - an error resulted from the last instruction – it may cause the process to be moved to terminated state.
 - Interrupt
 - the cause is external to the execution of the current instruction – control is transferred to Interrupt Handler.

Reasons for Process Switch

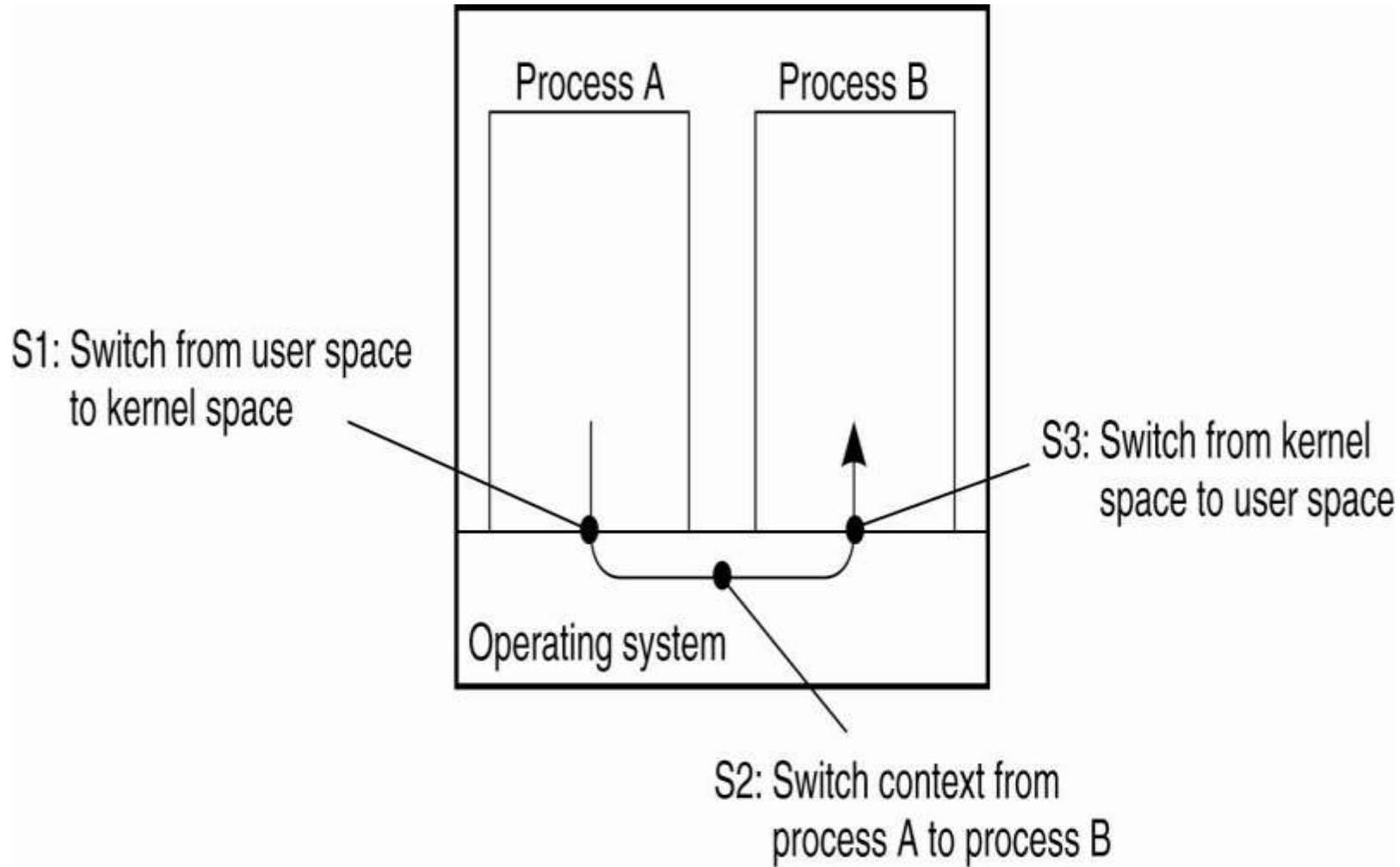




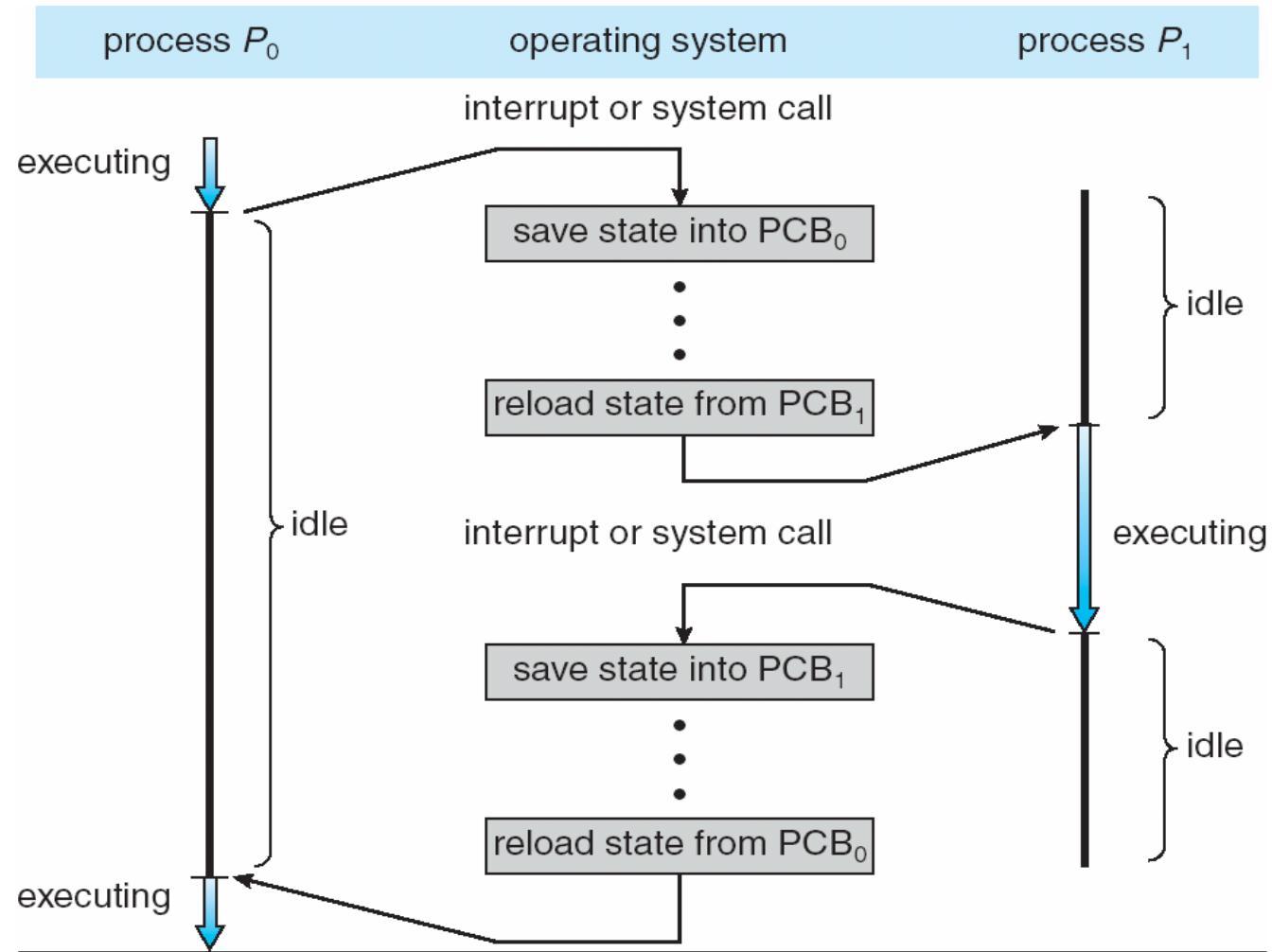
Context Switch

- When CPU switches to another process, the system must save the state of the old process and load the saved state for the new process.
- This is called context switch.
- Context of a process represented in the PCB.
- The time it takes is dependent on hardware support.
- Context-switch time is overhead; the system does no useful work while switching.

Context switch between processes (1)



Context switch between processes (2)



Steps in Context Switch



- Save context of processor including program counter and other registers.
- Update the PCB of the running process with its new state and other associate information.
- Move PCB to appropriate queue – ready, blocked,
- Select another process for execution.
- Update PCB of the selected process.
- Restore CPU context from that of the selected process.

Scheduling Criteria/ Methodology



- CPU Utilization
 - Keep the CPU and other resources as busy as possible
 - Throughput
 - # of processes that complete their execution per time unit.
 - Turnaround time
 - amount of time to execute a particular process from its entry time.
 - The time interval between the submission of a process and the time of completion is the turnaround time.
- TAT=Waiting time in ready queue + executing time + Waiting time in waiting queue for I/O

Scheduling Criteria (cont.)



- Waiting time
 - amount of time a process has been waiting in the ready queue.
- Response Time (in a time-sharing environment)
 - amount of time it takes from when a request was submitted until the first response is produced, NOT output.

Optimization Criteria

- Optimize overall system
 - Max CPU Utilization
 - Max Throughput
- Optimize individual processes' performance
 - Min Turnaround time
 - Min Waiting time
 - Min Response time