

1. Answer:

- Figures shown in the lecture assume that the time required for the I/O operation is relatively short.
- The more typical case, especially for a slow device such as printer, is that the I/O operation will take much more time than executing a sequence of user instructions. In this case, the user program reaches the second WRITE call before the first I/O operation is completed.
- The result is that the user program is hung up at that point. When the preceding I/O operation completes, the new WRITE call may be processed, and a new I/O operation may be started.
- Though the processor might still need to wait in this situation, the waiting time is noticeably shortened. There is still a gain in efficiency because part of the time during which the I/O operation is underway overlaps with the execution of the user instructions.

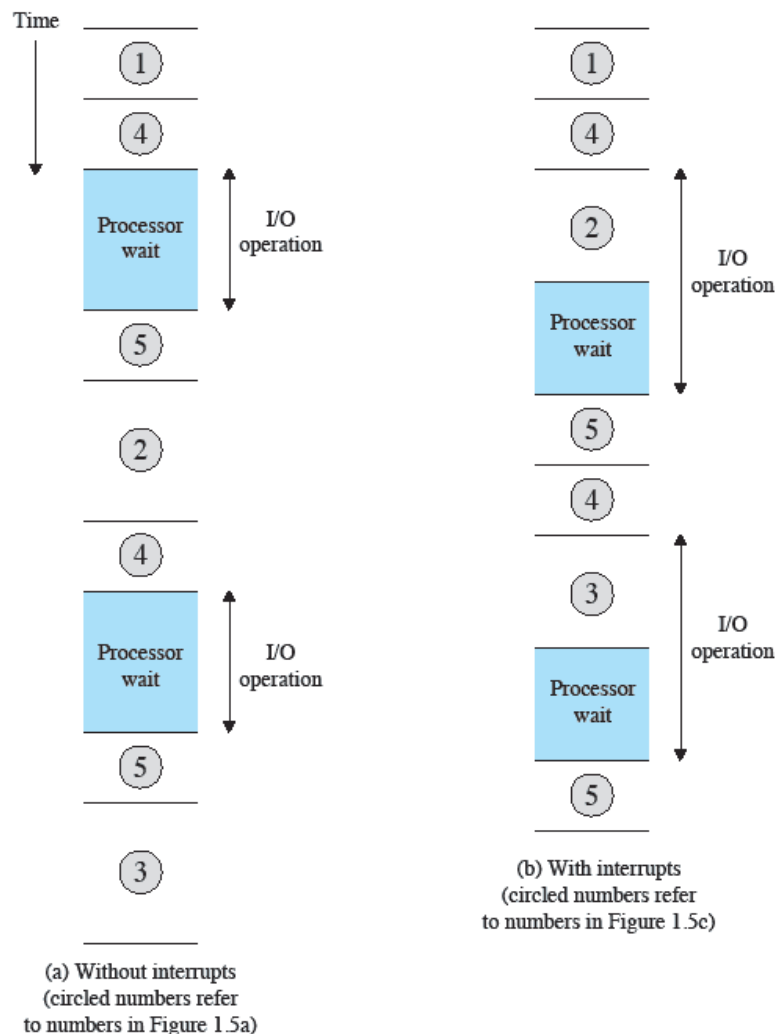


Figure 1.9 Program Timing: Long I/O Wait

2. Answer:

- A multiprogramming system runs more than one program “simultaneously” on one processor. The system attempts to keep several programs resident in main memory and switches the processor rapidly between them. Multiprogramming was developed to improve processor and I/O resource utilization.
- A multiprocessor is a computer system with more than one processor. Multiprocessing was developed in an effort to increase processing speeds by allowing truly parallel computation.

3. Answer:

The hit ratio has to be 0.95. Then the average time to access a word is
 $(0.95) (0.1\mu\text{s}) + (0.05) (0.1\mu\text{s} + 1\mu\text{s}) = 0.095 + 0.055 = 0.15 \mu\text{s}$

4. Answer:

a)

- Spatial locality can be exploited by using larger cache blocks.
- Temporal locality can be exploited by keeping recently used instruction and data values in cache memory.

b)

(i) A reference to the first instruction is immediately followed by a reference to the second.

(ii) The ten accesses to $a[i]$ within the inner for loop occur within a short interval of time.

Self-test

1. D
2. C
3. B
4. D
5. D
6. A