I/O Systems

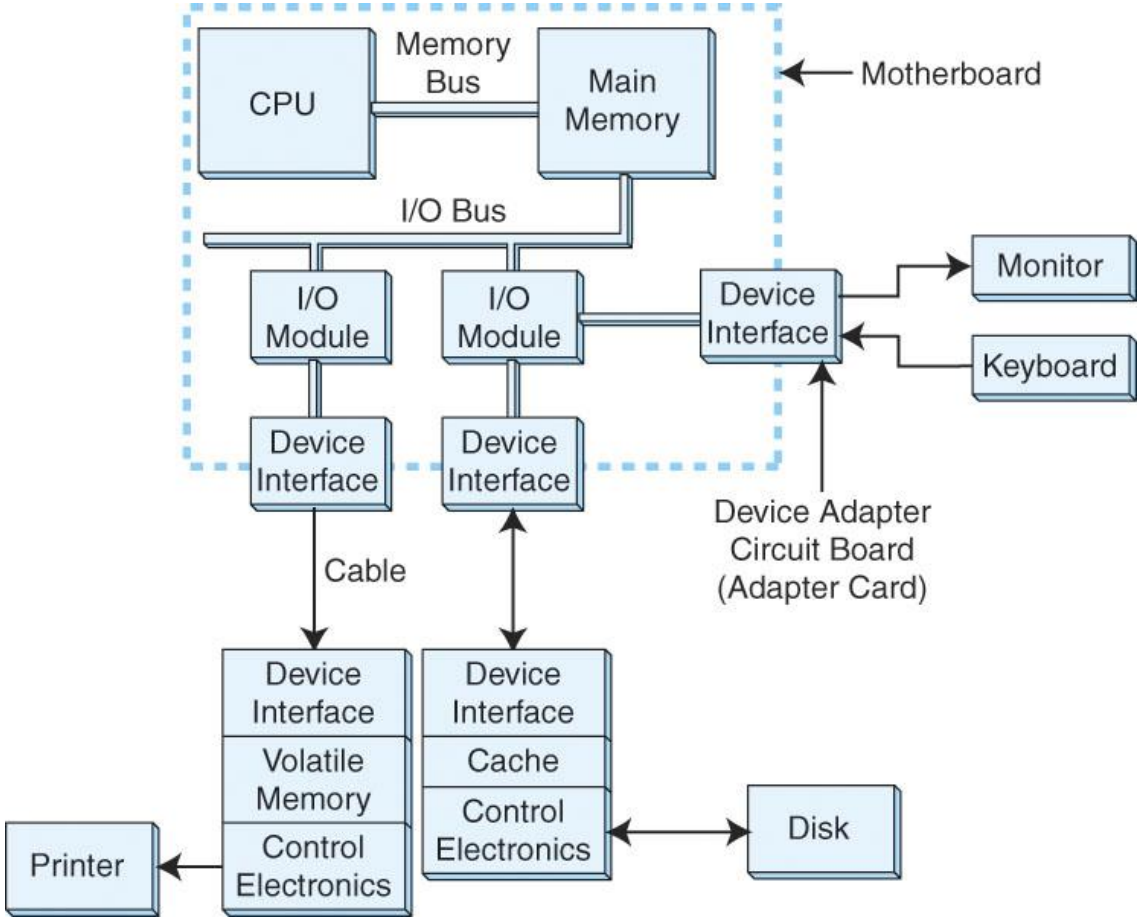
- Modern computers use various types of I/O devices. The devices are used in different environments and designing interfaces to these devices is very difficult. These devices are managed by special software called device drivers.
- The functions of the I/O systems are to manage variety of device drivers, to schedule I/O requests to devices, to allocate devices to requesting processes, to match the speed of the I/O devices with CPU/Memory using buffering /caching data etc.

- The I/O subsystem is treated as an independent unit in the computer
  - The CPU initiates I/O commands generically
    - Read, write, scan, etc
    - This simplifies the CPU
  - I/O modules are components that connect an I/O device to the I/O bus
    - The I/O module is an intermediary between CPU and the I/O device, and possibly between memory and the I/O device
  - This allows us to tailor I/O devices to specific uses without having to worry about how the CPU might be able to handle that new type of device
  - In addition, while the CPU may initiate an I/O operation, once begun, the I/O module takes over so that the CPU can get back to doing whatever it was doing

- I/O subsystem will typically include
  - Blocks of memory dedicated to I/O buffering
  - I/O bus(es)
  - I/O devices
  - Specialized interfaces (for instance for keyboard and monitor) and interface cards
  - Possibly other connections (network, cable, etc)

- Main tasks of I/O system**:**
  - Present **logical** (abstract) view of devices
    - Hide details of hardware interface
    - Hide error handling
  - Facilitate **efficient** use
    - Overlap CPU and I/O
  - Support **sharing** of devices
    - Protection when device is shared (disk)
    - Scheduling when exclusive access needed (printer)

I/O Architectures



I/O Hardware

- Any I/O device is hosted by one and only one I/O bus which is the data path connecting the CPU and an I/O device.
- I/O bus is connected to an I/O device by a hierarchy of hardware components, including I/O ports, interfaces, and device controllers.

- An I/O controller is a peripheral device that enables the main processor to transfer data between the host system and the I/O devices.
- It is a special purpose processor and carries out I/O operations in parallel with CPU execution of programs.
- The CPU interacts with an I/O controller through a set of controller interface registers called I/O ports.
- The ports constitute the I/O address space of a controller and I/O address spaces of all I/O controllers together constitute the I/O address space of the computer system.

- I/O ports are classified into four groups: Command, Status, Input and Output.
- Command and status registers are used to start and stop all devices connected to the controller, to initialize them and to diagnose any problems encountered with them.

The steps to start an I/O operation:

1. CPU gives appropriate commands to ports and input data to Input ports.
2. Controller executes the command and writes back the command execution status to the status ports and output data to output ports.
3. CPU reads the status and output ports respectively to find the command completion status and the output produced.

- The devices have different physical organizations with various physical characteristics such as the recording medium, storage capacity, interface operations and the access speed.
- Examples of I/O devices are disks, tapes, CDs, printers, network interface cards, keyboards, monitors etc.
- Some devices are purely input devices like keyboard, some are purely output devices like monitors and some are both like disks.

Classification of devices

- Human readable
- Machine readable
- Communication

Categories of I/O Devices

- Human readable
  - Used to communicate with the user
  - Printers
  - Video display terminals
    - Display
    - Keyboard
    - Mouse

Categories of I/O Devices

- Machine readable
  - Used to communicate with electronic equipment
  - Disk and tape drives
  - Sensors
  - Controllers
  - Actuators

Categories of I/O Devices

- Communication
  - Used to communicate with remote devices
  - Digital line drivers
  - Modems

Another way of classifying I/O devices based on the nature of their function is

- Storage devices - Data are stored persistently
- Communication devices - Data are transferred from one hardware component to another.

- The devices may also be categorized based on the physical attributes:
- Character or Block: A device may transfer one character/byte or one block (multiple bytes) of data at a time. Byte is smallest unit of data transfer for a character device and fixed block size is the smallest unit for a block device.
- Read-only, write-only or read-write: Some devices such as keyboards allow the system to read characters whereas printers are write-only which allow system to perform write operation; Devices such as disks allow the system to load as well as store data.

- Speed: Some devices such as keyboards have very low transferring speed (in terms of bytes per second) and some devices are faster like Ethernet cards that transfer millions of bytes per second.
- Transient or durable: Transient devices store data for shorter duration as in Ethernet cards where as disks store persistent data.
- Sharable or exclusive: Some devices such as disks are sharable and can be accessed by processes concurrently. Some devices are exclusive in nature like Graphical plotters.

- Access pattern:
  - Sequential access devices: Data blocks are only accessible in a sequential order. e.g. Tape devices
  - Direct access devices: Any data block is directly accessible in any arbitrary order by providing its position. e.g. Disks
  - Random access devices: Data block are directly accessible but the access time is independent of the position of data. e.g. Flash devices.

Differences in I/O Devices

- Data rate
  - May be differences of several orders of magnitude between the data transfer rates
- Application
  - Disk used to store files requires file management software
  - Disk used to store virtual memory pages needs special hardware and software to support it
  - Terminal used by system administrator may have a higher priority
- Complexity of control
- Unit of transfer
  - Data may be transferred as a stream of bytes for a terminal or in larger blocks for a disk
- Data representation
  - Encoding schemes
- Error conditions
  - Devices respond to errors differently

I/O Buffering

- Reasons for buffering
  - Processes must wait for I/O to complete before proceeding
  - Certain pages must remain in main memory during I/O

I/O Buffering

- Block-oriented
  - Information is stored in fixed sized blocks
  - Transfers are made a block at a time
  - Used for disks and tapes
- Stream-oriented
  - Transfer information as a stream of bytes
  - Used for terminals, printers, communication ports, mouse and other pointing devices, and most other devices that are not secondary storage

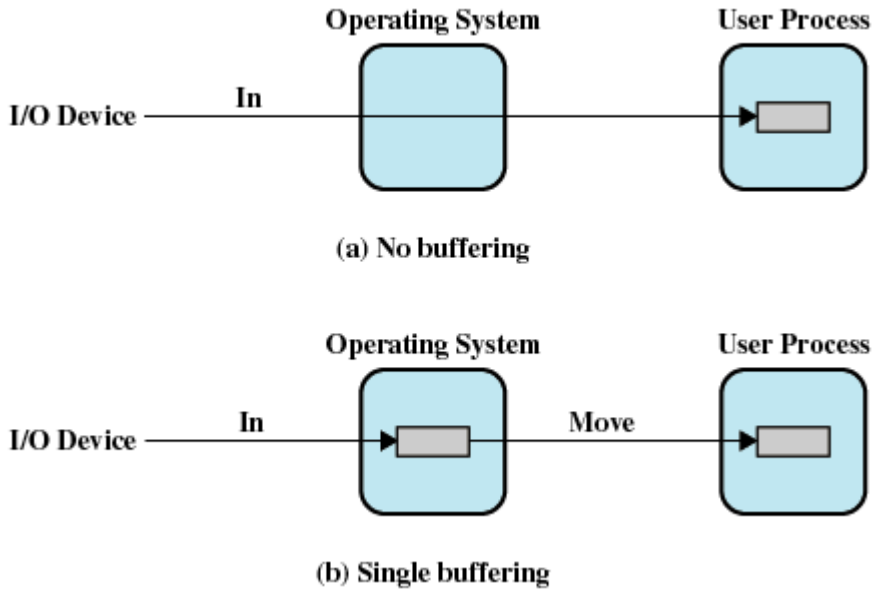
Single Buffer

- Operating system assigns a buffer in main memory for an I/O request
- Block-oriented
  - Input transfers made to buffer
  - Block moved to user space when needed
  - Another block is moved into the buffer
    - Read ahead
  - User process can process one block of data while next block is read in
  - Swapping can occur since input is taking place in system memory, not user memory
  - Operating system keeps track of assignment of system buffers to user processes

Single Buffer

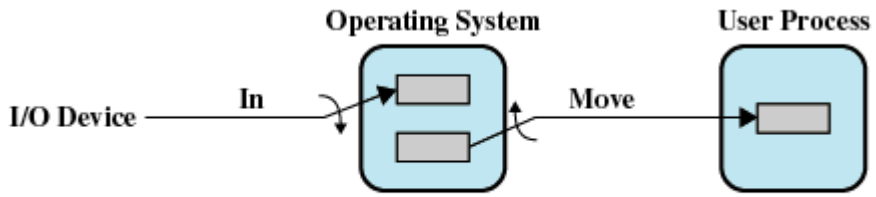
- Stream-oriented
  - Used a line at time
  - User input from a terminal is one line at a time with carriage return signaling the end of the line
  - Output to the terminal is one line at a time

I/O Buffering



Double Buffer

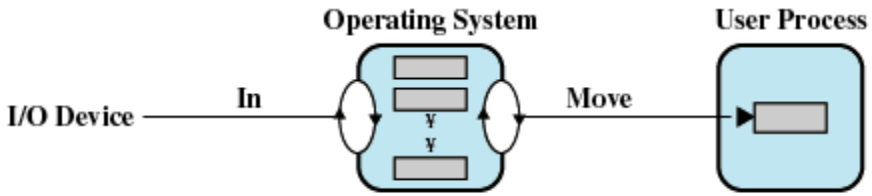
- Use two system buffers instead of one
- A process can transfer data to or from one buffer while the operating system empties or fills the other buffer



(c) Double buffering

Circular Buffer

- More than two buffers are used
- Each individual buffer is one unit in a circular buffer
- Used when I/O operation must keep up with process



(d) Circular buffering