Berkeley FFS

- tunefs -p /

  tunefs: ACLs: (-a)   disabled

  tunefs: MAC multilabel: (-l)   disabled

  tunefs: soft updates: (-n)   disabled

  tunefs: maximum blocks per file in a cylinder group: (-e) 2048

  tunefs: average file size: (-f)   16384

  tunefs: average number of files in a directory: (-s)   64

  tunefs: minimum percentage of free space: (-m)   8%

  tunefs: optimization preference: (-o)   time

- Another seminal paper describing a file system that is heavily optimized and used in FreeBSD, Mac OSX (default is HFS)

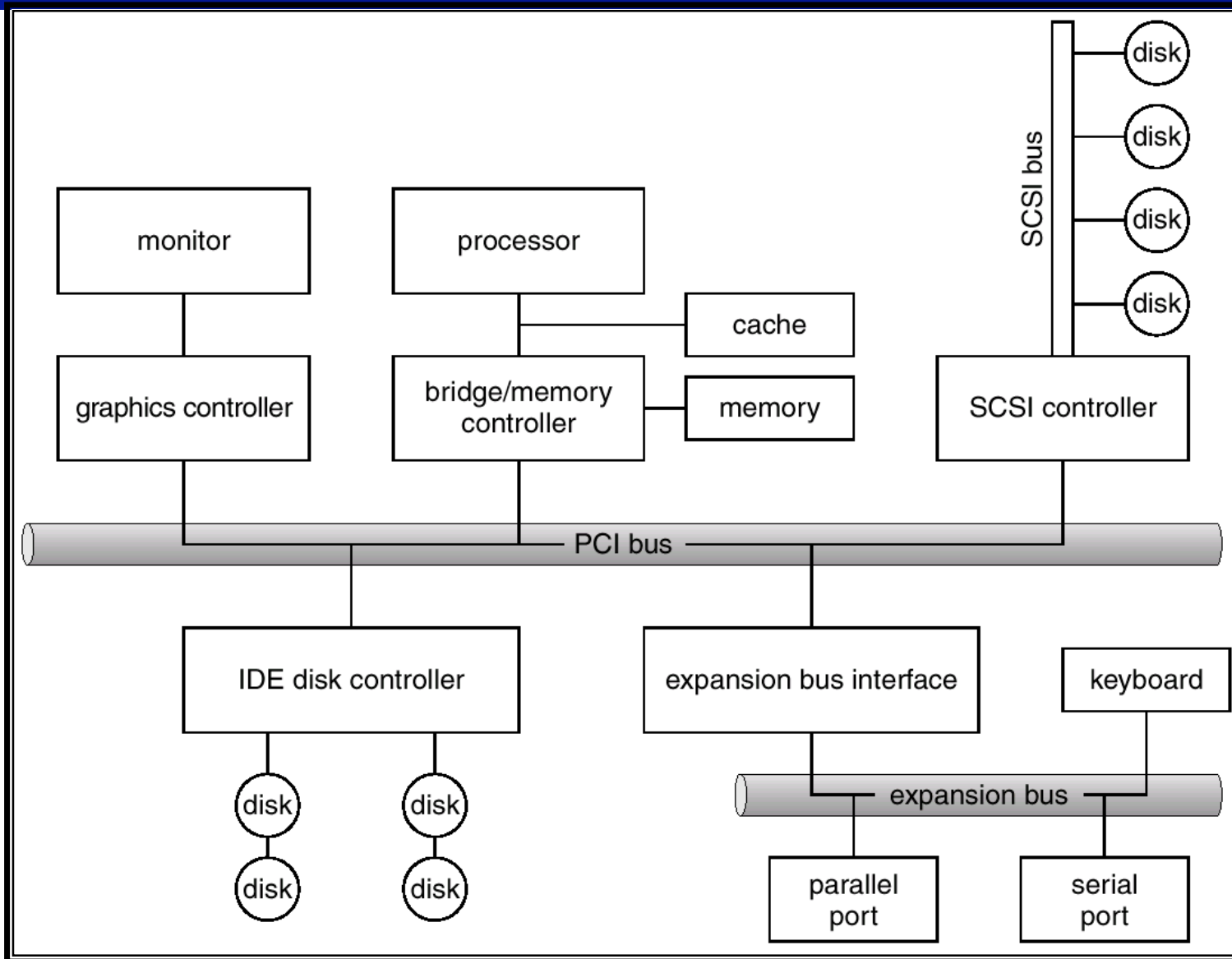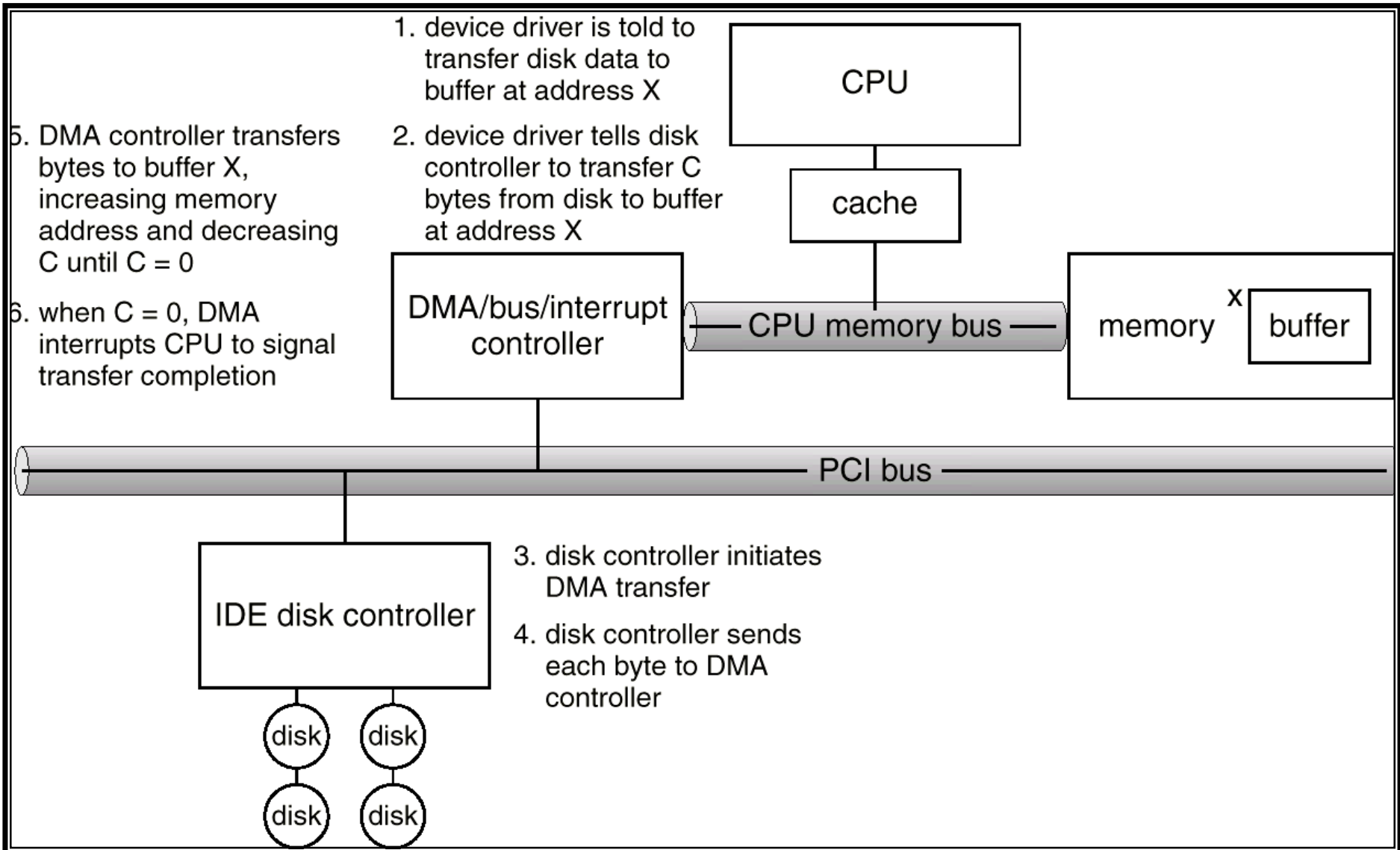- Optimize page placement, and block size to reflect newer usage patterns

LFS

- Files are only written to logs, there is no traditional file system backing up the LFS

- Write performance is much improved, especially for small files

DMA transfer

1. device driver is told to transfer disk data to buffer at address X

2. device driver tells disk controller to transfer C bytes from disk to buffer at address X

5. DMA controller transfers bytes to buffer X, increasing memory address and decreasing C until C = 0

6. when C = 0, DMA interrupts CPU to signal transfer completion

CPU

cache

DMA/bus/interrupt controller

CPU memory bus

memory ^{X} buffer

PCI bus

IDE disk controller

disk   disk

disk   disk

3. disk controller initiates DMA transfer

4. disk controller sends each byte to DMA controller

I/O modes

aspect	variation	example
data-transfer mode	character block	terminal disk
access method	sequential random	modem CD-ROM
transfer schedule	synchronous asynchronous	tape keyboard
sharing	dedicated sharable	tape keyboard
device speed	latency seek time transfer rate delay between operations	
I/O direction	read only write only readÐwrite	CD-ROM graphics controller disk

I/O Systems

- ## Hardware
  - Bus and interconnects & Controller
  - Host adapter
  - Registers and Memory mapped I/O
  - Polling, Interrupt driven with an interrupt controller
  - Maskable and non-maskable interrupts and priorities
  - Direct Memory Access (DMA)

- ## Application Interface
  - Character or Block: terminal or disk
  - Sequential or Random access: modem or CD-ROM
  - Synchronous or asynchronous: tape or keyboard
  - Shared or dedicated: tape or keyboard
  - RW, R and W only: disk, CDROM, graphics controller

Kernel I/O

- I/O Scheduling
  - Buffering; double-buffering
    - Copy semantics for async. I/O
  - Caching
  - Spooling and device reservation
  - Error handling
- Performance
  - Front end processors to off load processing

Improving performance

- Reduce number of context switches

- Reduce data copying

- Reduce interrupts by using large transfers, smart controllers, polling

- Use DMA

- Balance CPU, memory, bus, and I/O performance for highest throughput

Disk scheduling

- Schedule disk accesses to gain performance
  - FCFS - first come first service
  - SSTF - shorted seek time first
    - starvation
  - SCAN
    - Elevator algorithm
  - CSCAN
    - Restarts from the beginning after each cycle
  - LOOK
    - Look till end of direction
  - CLOOK

- Disk scheduling harder with smart disks that can rearrange bad sectors

Disk attachment

- ## Host-attached storage
  - SCSI, Fibre-Channel

- ## Network attached storage (NAS)
  - Device implements a complete file system

- ## Storage-Area Networks
  - High speed interconnect
  - Can dynamically reassign disks to other servers

- ## iSCSI
  - SCSI using IP protocols
  - Security, congestion etc. are issues

- ## Direct Access File System (DAFS)
  - Emerging standard leveraging Remote Direct Memory Access infrastructure
  - http://www.dafscollaborative.org/