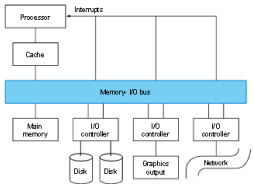
Chapter 8

I/O

Interfacing Processors and Peripherals

- I/O Design affected by many factors (expandability, resilience)
- Performance:
  - — access latency
  - — throughput
  - — connection between devices and the system
  - — the memory hierarchy
  - — the operating system
- A variety of different users (e.g., banks, supercomputers, engineers)



I/O Devices

- Very diverse devices
  - — behavior (i.e., input vs. output)
  - — partner (who is at the other end?)
  - — data rate

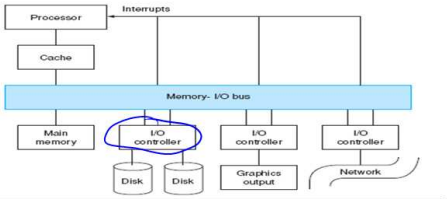
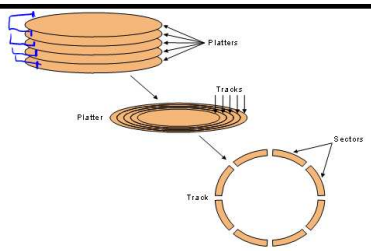


FIGURE 8.2 A typical collection of I/O devices. The connections between the I/O devices, processor, and memory are usually called *buses*. Communication among the devices and the processor use both interrupts and protocols on the bus, as we will see in this chapter. Figure 8.11 on page 585 shows the organization for a desktop PC.

I/O Example: Disk Drives



- To access data:
  - — seek: position head over the proper track (3 to 14 ms. avg.)
  - — rotational latency: wait for desired sector (.5 / RPM)
  - — transfer: grab the data (one or more sectors) 30 to 80 MB/sec

Exemplo

- Qual o tempo médio para ler ou escrever num disco com:
  - 10.000 RPM
  - Seek time médio de 6ms
  - Taxa de transferência de 50 MB/s
  - Overhead do controlador de 0,2ms

9.2 ms

- Setor 512 bytes

1) Latência Rotacional: 0.5 / 10.000 RPM = 3 ms
2) Seek médio: 6 ms
3) Transferência: 0.5 KB / 50 MB/s = 0,01 ms
4) Overhead: 0.2 ms

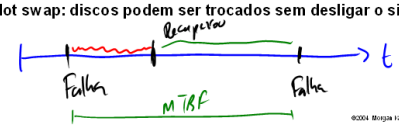
Falhas

- Permanentes
  - Confiabilidade
- Intermitentes
  - Disponibilidade

1,200,000 (MTTF)

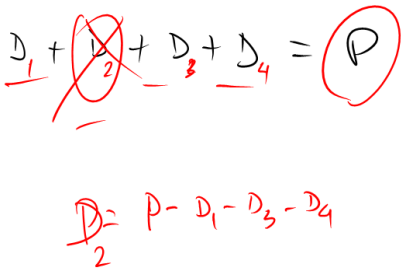
- MTTF: Mean time to failure
- MTTR: Mean time to repair
- MTBF: Mean time between failures (= MTTF + MTTR)

- Hot swap: discos podem ser trocados sem desligar o sistema



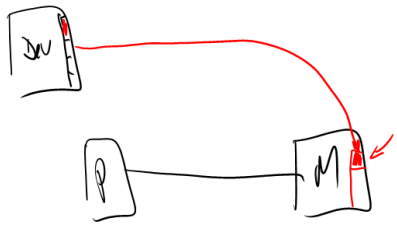
RAID - Redundant Arrays of Inexpensive Disks

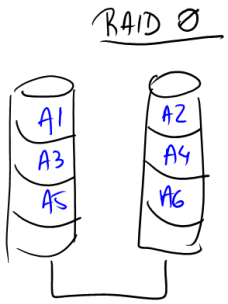
- RAID 0
  - Sem redundância, apenas divisão dos dados entre discos
- RAID 1
  - Redundância total, espelhamento dos discos
- RAID 2
  - Inclui detecção e correção de erro (caiu em desuso)
- RAID 3
  - Um disco extra para armazenar a paridade dos dados
- RAID 4
  - Similar ao 3, com organização da paridade em blocos nos discos
- RAID 5
  - Similar ao 4, com paridade distribuída entre os discos
- RAID 6
  - Inclui um disco extra para recuperação de um segundo erro

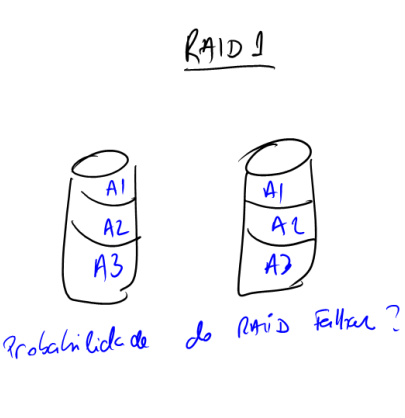


Comunicação com os dispositivos

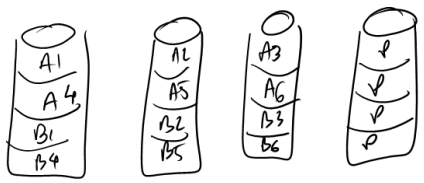
- Enviar comandos
  - I/O mapeada em memória
  - Instruções específicas de I/O
- Comunicação
  - Polling
  - Interrupção
    - Prioridades
  - DMA
    - Memória Virtual
    - Cache
    - Overhead
    - Independência





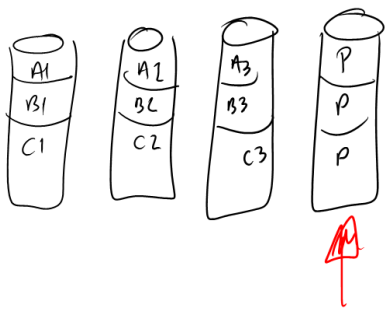


RAID 3

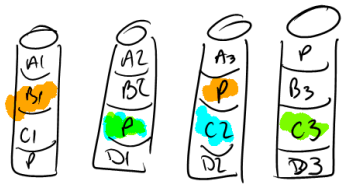


byte-level stripping

RAID 4



RAID 5



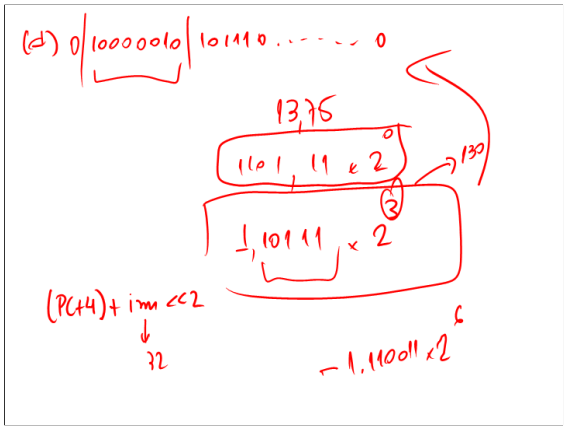
Cálculo de Paridade

- Leia o bloco antigo
- Leia a paridade antiga
- Compare o bloco antigo c/ o sendo escrito
  - ↳ Atualize paridade
- Escreva novo dado
- Escreva nova paridade

RAID 6

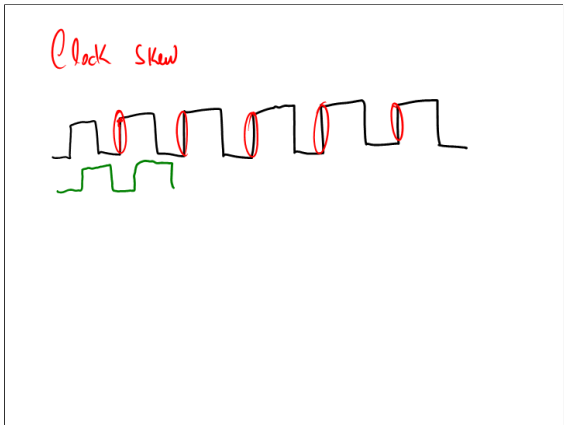
- Paridade adicional
- Penalidade em escritas
- Continuar funcionando em caso de falha dupla

RAID 6

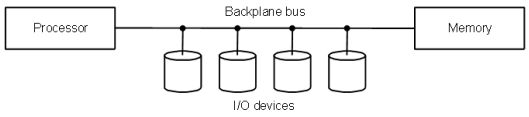


I/O Example: Buses

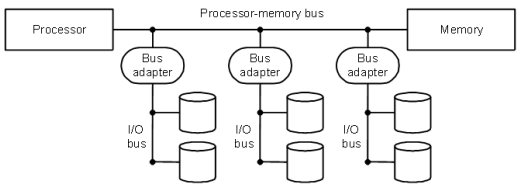
- Shared communication link (one or more wires)
- Difficult design:
  - may be bottleneck
  - length of the bus
  - number of devices
  - tradeoffs (buffers for higher bandwidth increases latency)
  - support for many different devices
  - cost
- Types of buses:
  - processor-memory (short high speed, custom design)
  - backplane (high speed, often standardized, e.g., PCI)
  - I/O (lengthy, different devices, e.g., USB, Firewire)
- Synchronous vs. Asynchronous
  - use a clock and a synchronous protocol, fast and small but every device must operate at same rate and clock skew requires the bus to be short
  - don't use a clock and instead use handshaking



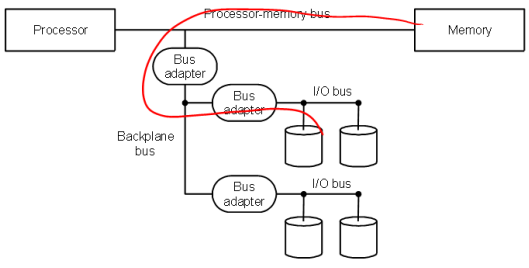
Organização



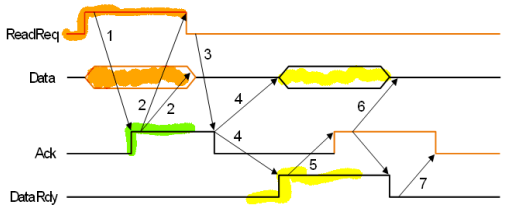
Organização



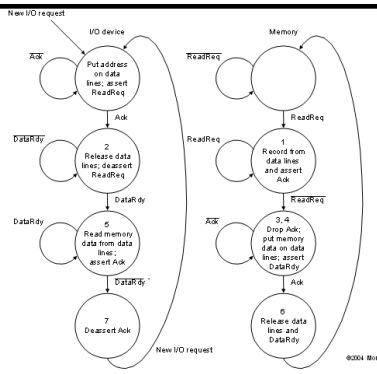
Organização



Barramentos Assíncronos



Máquina de Estados



I/O Bus Standards

- **Today we have two dominant bus standards:**

Characteristic	Firewire (1394)	USB 2.0
Bus type	I/O	I/O
Basic data bus width (signals)	4	2
Clocking	asynchronous	asynchronous
Theoretical peak bandwidth	50 MB/sec (Firewire 400) or 100 MB/sec (Firewire 800)	0.2 MB/sec (low speed), 1.5 MB/sec (full speed), or 60 MB/sec (high speed)
Hot pluggable	yes	yes
Maximum number of devices	63	127
Maximum bus length (copper wire)	4.5 meters	5 meters
Standard name	IEEE 1394, 1394b	USE Implementors Forum

FIGURE 8.9 Key characteristics of two dominant I/O bus standards.

Other important issues

- **Bus Arbitration:**
  - **— daisy chain arbitration (not very fair)**
  - **— centralized arbitration (requires an arbiter), e.g., PCI**
  - **— collision detection, e.g., Ethernet**
- **Performance Analysis techniques:**
  - **— queuing theory**
  - **— simulation**
  - **— analysis, i.e., find the weakest link (see "I/O System Design")**
- **Many new developments**

Considere o sistema:
- CPU: 3 bilhões instr/seg e 100.000 instr de OS por I/O
- Barramento memória: 1000 MB/seg
- Controlador SCSI Ultra320: 320MB/seg para até 7 discos
- Discos: 75 MB/sec leitura e escrita
  6ms seek+latência rotacional

Workload: 64KB de leitura (bloco sequencial numa trilha); o programa executa 200.000 instr por operação de I/O.

Qual a taxa máxima sustentável de I/O, o número de discos e de controladores necessários? Assuma que as leituras podem sempre ser feitas em um disco idle, se existir um.

- Quais os componentes fixos?
barram. Memória, CPU
- Qual o gargalo?

$I_B = \frac{1000 \times 10^6}{64 \times 10^3} = 15{,}625$ I/O/s

$I_{CPU} = \frac{3 \times 10^9}{(200+100) \times 10^3} = 10\,000$ I/O/s

10000 I/O/s

1 Disco:

Tempo 1 I/O: $6ms + \frac{64KB}{75\,MB/sec} = 6.9ms$

Em 1s: $\frac{1000ms}{6.9ms} \approx 145$

discos: $\frac{10000}{145} = 69$ discos $\Rightarrow$ 10 controladores

Exemplo: Impacto de E/S

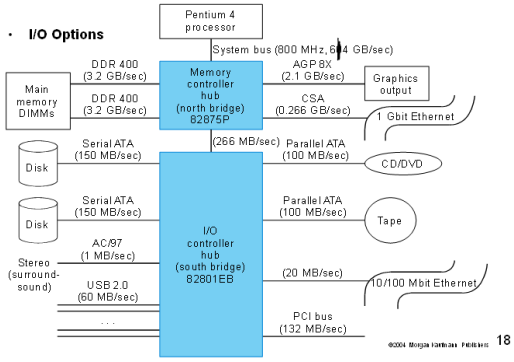
- Benchmark original gasta 100s, sendo 90s de processamento e 10s de E/S. Se o processador fica 50% mais rápido por ano e a E/S não melhora, qual será a melhora de desempenho dentro de 5 anos?

Anos	Processador	E/S	Total	% de E/S	Speedup
0	90s	10s	100s	10%	0%
1	60s	10s	70s	14%	43%
2	40s	10s	50s	20%	100%
3	27s	10s	37s	27%	170%
4	18s	10s	28s	36%	257%
5	12s	10s	22s	45%	455%

Obs.: 90s/12s = 7,5

Pentium 4

- **I/O Options**



Pentium 4

	875P chip set	845GL chip set
Target segment	Performance PC	Value PC
System bus (64 bit)	800/533 MHz	400 MHz
Memory controller hub ("north bridge")		
Package size, pins	42.5 × 42.5 mm, 1005	37.5 × 37.5 mm, 760
Memory speed	DDR 400/333/266 SDRAM	DDR 266/200, PC133 SDRAM
Memory buses, widths	2 × 72	1 × 64
Number of DIMMs, DRAM Mbit support	4, 128/256/512 Mbits	2, 128/256/512 Mbits
Maximum memory capacity	4 GB	2 GB
Memory error correction available?	yes	no
AGP graphics bus, speed	yes, 8X or 4X	no
Graphics controller	external	internal (Extreme Graphics)
CSA Gigabit Ethernet interface	yes	no
South bridge interface speed (8 bit)	266 MHz	266 MHz
I/O controller hub ("south bridge")		
Package size, pins	31 × 31 mm, 460	31 × 31 mm, 421
PCI bus: width, speed, masters	32-bit, 33 MHz, 6 masters	32-bit, 33 MHz, 6 masters
Ethernet MAC controller, interface	100/10 Mbit	100/10 Mbit
USB 2.0 ports, controllers	8, 4	6, 3
ATA 100 ports	2	2
Serial ATA 150 controller, ports	yes, 2	no
RAID 0 controller	yes	no
AC-97 audio controller, interface	yes	yes
I/O management	SMBus 2.0, GPIO	SMBus 2.0, GPIO

FIGURE 8.12 Two Pentium 4 I/O chip sets from Intel. The 845GL north bridge uses many fewer pins than the 875 by having just one memory bus and by omitting the AGP bus and the Gigabit Ethernet interface. Note that the serial nature of USB and Serial ATA means that two more USB ports and two more Serial ATA ports need just 39 more pins in the south bridge of the 875 versus the 845GL chip sets.

Fallacies and Pitfalls

- **Fallacy: the rated mean time to failure of disks is 1,200,000 hours, so disks practically never fail.**
- **Fallacy: magnetic disk storage is on its last legs, will be replaced.**
- **Fallacy: A 100 MB/sec bus can transfer 100 MB/sec.**
- **Pitfall: Moving functions from the CPU to the I/O processor, expecting to improve performance without analysis.**

