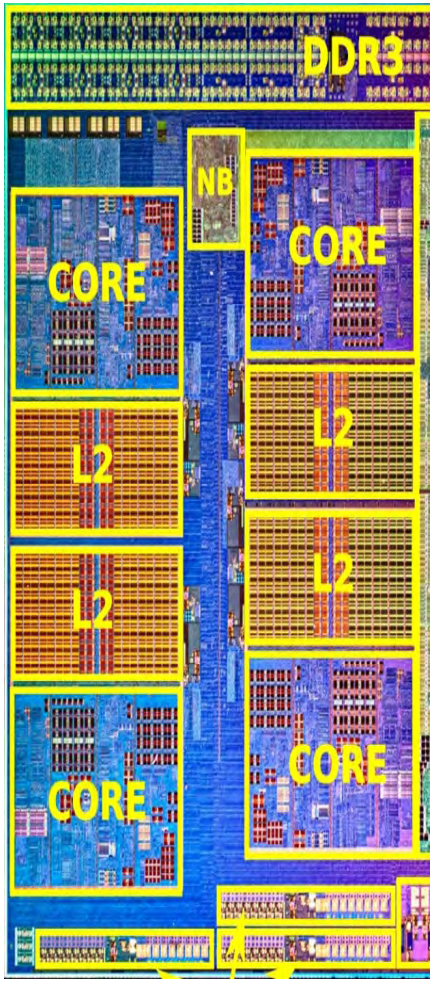
0273359 – Arquitetura e Organização de Computadores 1



Fonte: http://www.techspot.com/article/904-history-of-the-personal-computer-part-5/

Subsitemas de I/O

Luciano de Oliveira Neris
luciano@dc.ufscar.br

Adaptado de slides do prof. Marcio Merino Fernandes
Figuras: David Patterson, John Hennessy
Arquitetura e Organização de Computadores – 4Ed, Elsevier, 2014

Departamento de Computação
Universidade Federal de São Carlos

Introdução

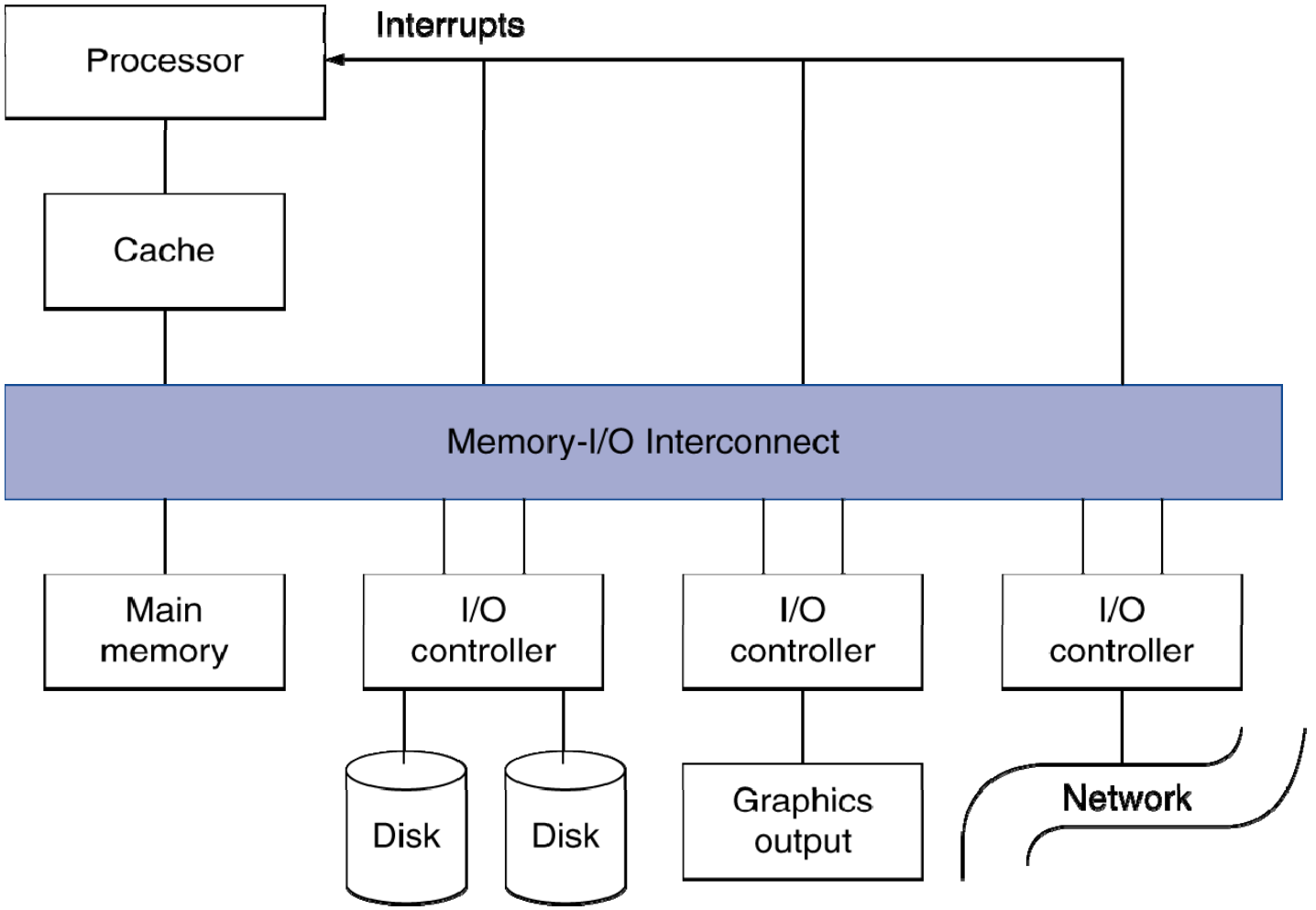
- I/O: permite aos seres humanos interagir c/ o computador
- I/O: permite comunicação entre computadores
- I/O: permite comunicação entre computadores e outros dispositivos
  - Sensores e Atuadores

Introdução

- Dispositivos de I/O podem ser caracterizados de acordo com:
  - Comportamento: entrada, saída, armazenamento
  - Parceiro: homem, máquina
  - Taxa de transferência de dados: bytes/sec, transfers/sec
  - Barramentos de conexão de I/O

Introdução

- Dispositivos de I/O



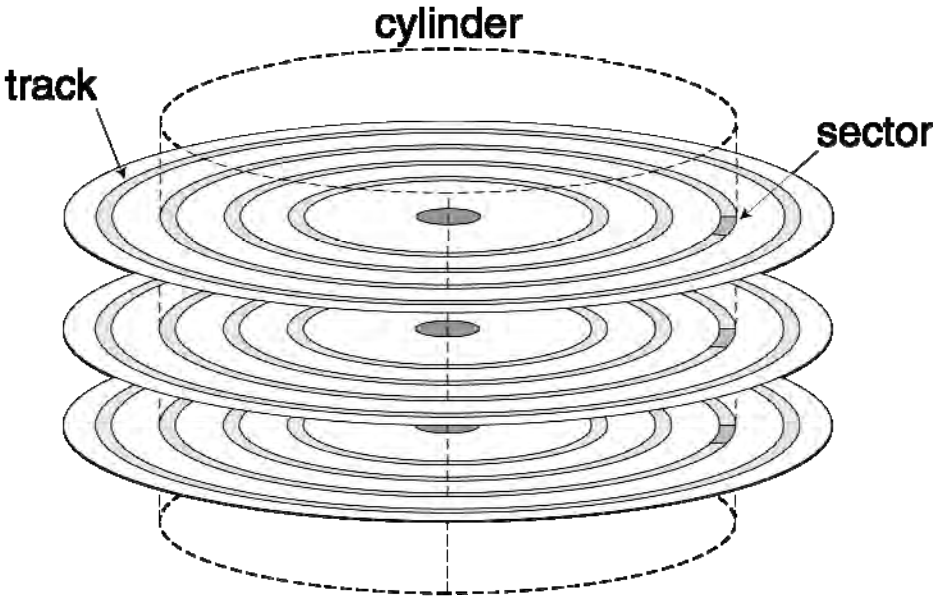
Alguns Dispositivos de I/O

Device	Behavior	Partner	Data rate (Mbit/sec)
Keyboard	Input	Human	0.0001
Mouse	Input	Human	0.0038
Voice input	Input	Human	0.2640
Sound input	Input	Machine	3.0000
Scanner	Input	Human	3.2000
Voice output	Output	Human	0.2640
Sound output	Output	Human	8.0000
Laser printer	Output	Human	3.2000
Graphics display	Output	Human	800.0000–8000.0000
Cable modem	Input or output	Machine	0.1280–6.0000
Network/LAN	Input or output	Machine	100.0000–10000.0000
Network/wireless LAN	Input or output	Machine	11.0000–54.0000
Optical disk	Storage	Machine	80.0000–220.0000
Magnetic tape	Storage	Machine	5.0000–120.0000
Flash memory	Storage	Machine	32.0000–200.0000
Magnetic disk	Storage	Machine	800.0000–3000.0000

FIGURE 6.2 The diversity of I/O devices. I/O devices can be distinguished by whether they serve as input, output, or storage devices; their communication partner (people or other computers); and their peak communication rates. The data rates span eight orders of magnitude. Note that a network can be an input or an output device, but cannot be used for storage. Transfer rates for devices are always quoted in base 10, so that 10 Mbit/sec = 10,000,000 bits/sec.

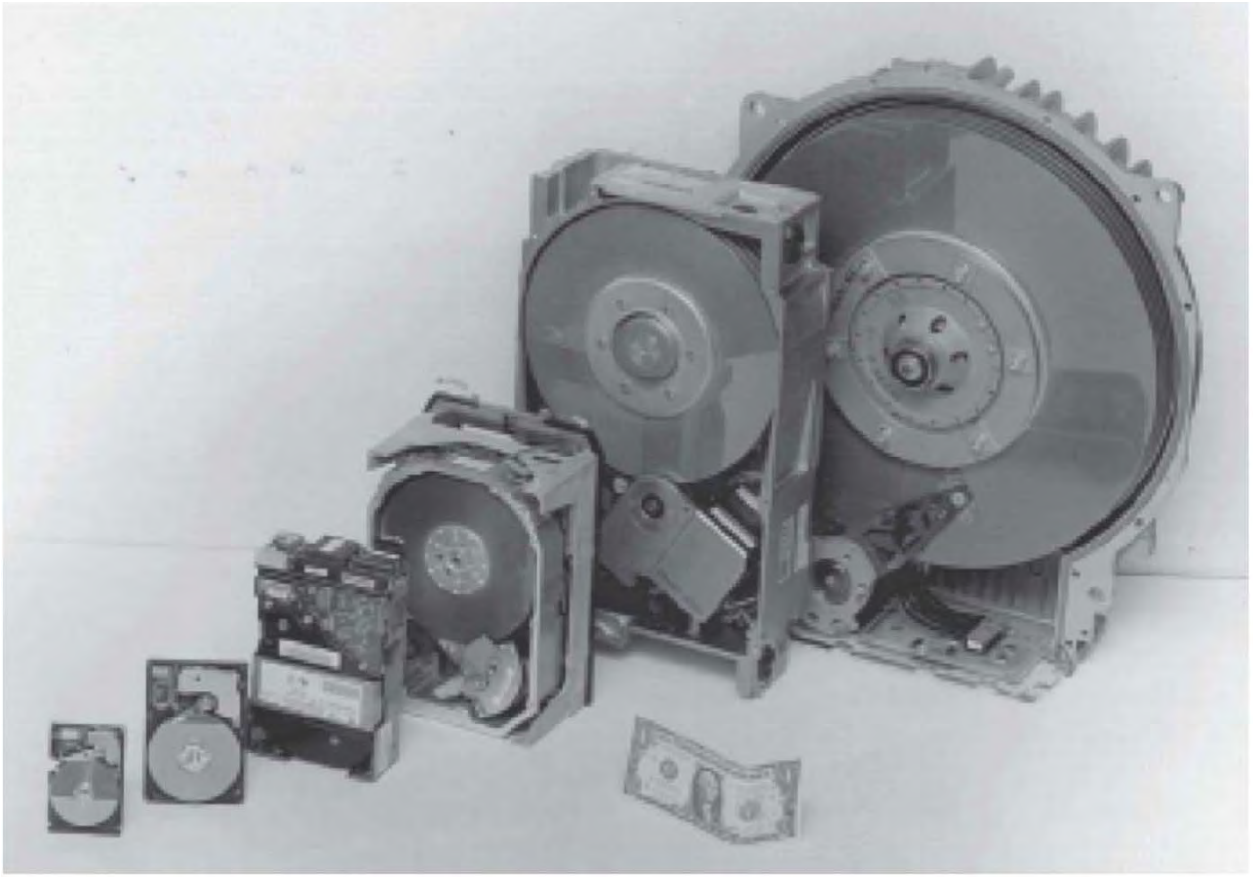
Dispositivos de Armazenamento: Discos

- Armazenamento magnético, não-volátil, rotacional.



Dispositivos de Armazenamento: Discos

- Tecnologia altamente bem sucedida, e c/ grandes avanços ao longo do tempo.



Dispositivos de Armazenamento: Discos

- Cada setor armazena
  - Sector ID
  - Data (512 bytes, 4096 bytes propostos)
  - Código de correção de erros (ECC)
    - Usado p/ lidar c/ defeitos e erros de gravação.
  - Campos p/ sincronização e "gaps" entre setores.

- Acesso a um setor envolve:
  - Fila de espera devido a outros acessos pendentes
  - Seek (Busca): posicionamento da cabeça de leitura
  - Latencia Rotacional
  - Transferência de Dados
  - Overhead da Controladora de disco

Acesso a Disco: Exemplo

- Dados:
  - 512B sector, 15,000rpm, 4ms average seek time, 100MB/s transfer rate, 0.2ms controller overhead, idle disk.

Acesso a Disco: Exemplo

- Dados:
  - 512B sector, 15,000rpm, 4ms average seek time, 100MB/s transfer rate, 0.2ms controller overhead, idle disk.

- Average read time:
  - Assumindo 4ms seek time
    + $\frac{1}{2}$ / (15,000/60) = 2ms rotational latency
    + 512 / 100MB/s = 0.005ms transfer time
    + 0.2ms controller delay
    = 6.2ms

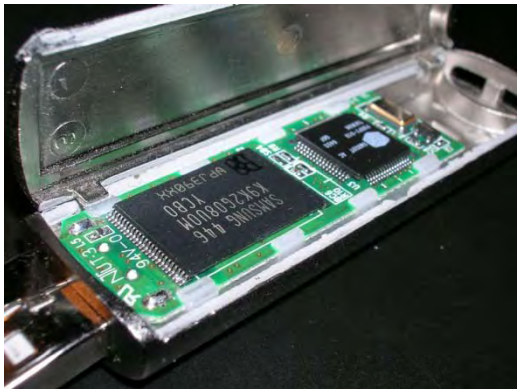
- OBS: Se "average seek time" fosse igual a 1ms:
  - Average read time = 3.2ms

Desempenho de Discos: Exemplo

Characteristics	Seagate ST3300655SS	Seagate ST31000340NS	Seagate ST973451SS	Seagate ST9160821AS
Disk diameter (inches)	3.50	3.50	2.50	2.50
Formatted data capacity (GB)	147	1000	73	160
Number of disk surfaces (heads)	2	4	2	2
Rotation speed (RPM)	15,000	7200	15,000	5400
Internal disk cache size (MB)	16	32	16	8
External interface, bandwidth (MB/sec)	SAS, 375	SATA, 375	SAS, 375	SATA, 150
Sustained transfer rate (MB/sec)	73–125	105	79–112	44
Minimum seek (read/write) (ms)	0.2/0.4	0.8/1.0	0.2/0.4	1.5/2.0
Average seek read/write (ms)	3.5/4.0	8.5/9.5	2.9/3.3	12.5/13.0
Mean time to failure (MTTF) (hours)	1,400,000 @ 25°C	1,200,000 @ 25°C	1,600,000 @ 25°C	—
Annual failure rate (AFR) (percent)	0.62%	0.73%	0.55%	—
Contact start-stop cycles	—	50,000	—	>600,000
Warranty (years)	5	5	5	5
Nonrecoverable read errors per bits read	<1 sector per 10^{16}	<1 sector per 10^{15}	<1 sector per 10^{16}	<1 sector per 10^{14}
Temperature, shock (operating)	5°–55°C, 60 G	5°–55°C, 63 G	5°–55°C, 60 G	0°–60°C, 350 G
Size: dimensions (in.), weight (pounds)	1.0" × 4.0" × 5.8", 1.5 lbs	1.0" × 4.0" × 5.8", 1.4 lbs	0.6" × 2.8" × 3.9", 0.5 lbs	0.4" × 2.8" × 3.9", 0.2 lbs
Power: operating/idle/ standby (watts)	15/11/—	11/8/1	8/5.8/—	1.9/0.6/0.2
GB/cu. in., GB/watt	6 GB/cu.in., 10 GB/W	43 GB/cu.in., 91 GB/W	11 GB/cu.in., 9 GB/W	37 GB/cu.in., 84 GB/W
Price in 2008, $/GB	~ $250, ~ $1.70/GB	~ $275, ~ $0.30/GB	~ $350, ~ $5.00/GB	~ $100, ~ $0.60/GB

Dispositivos de Armazenamento: FLASH

- Armazenamento não volátil em dispositivos semicondutores.
  - 100× – 1000× mais rápido que discos
  - Menor tamanho físcio, menor consumo de energia, mais robusto ☺
  - Maior custo $ / GB, menor durabilidade ☹
  - Possibilidade real de substituir discos no futuro
    - Várias outras tecnologias fracassaram no passado.



Dispositivos de Armazenamento: FLASH

- NOR flash: bit cell semelhante a NOR gate
  - Random read/write access
  - Usado p/ memória de instruções em sistemas embarcados, BIOS, etc.
- NAND flash: bit cell semelhante a NAND gate
  - Maior densisdade (bits/area), porém acesso a blocos.
  - Menor custo $ / GB
  - Usada em PenDrivs, cartões de armazenamento (SD), etc.
- Flash bits se desgastam após 100.000 gravações.
  - Não são apropriados p/ substituir a memória RAM
  - Não são apropriados p/ substituir totalmente os discos
    - Porém, tecnologias estão sendo desenvolvidas p/remapear os acessos, melhorando a confiabilidade.

Desempenho de Memórias Flash: Exemplo

Characteristics	Kingston SecureDigital (SD) SD4/8 GB	Transend Type I CompactFlash TS16GCF133	RiDATA Solid State Disk 2.5 inch SATA
Formatted data capacity (GB)	8	16	32
Bytes per sector	512	512	512
Data transfer rate (read/write MB/sec)	4	20/18	68/50
Power operating/standby (W)	0.66/0.15	0.66/0.15	2.1/—
Size: height × width × depth (inches)	0.94 × 1.26 × 0.08	1.43 × 1.68 × 0.13	0.35 × 2.75 × 4.00
Weight in grams (454 grams/pound)	2.5	11.4	52
Mean time between failures (hours)	> 1,000,000	> 1,000,000	> 4,000,000
GB/cu. in., GB/watt	84 GB/cu.in., 12 GB/W	51 GB/cu.in., 24 GB/W	8 GB/cu.in., 16 GB/W
Best price (2008)	~ $30	~ $70	~ $300

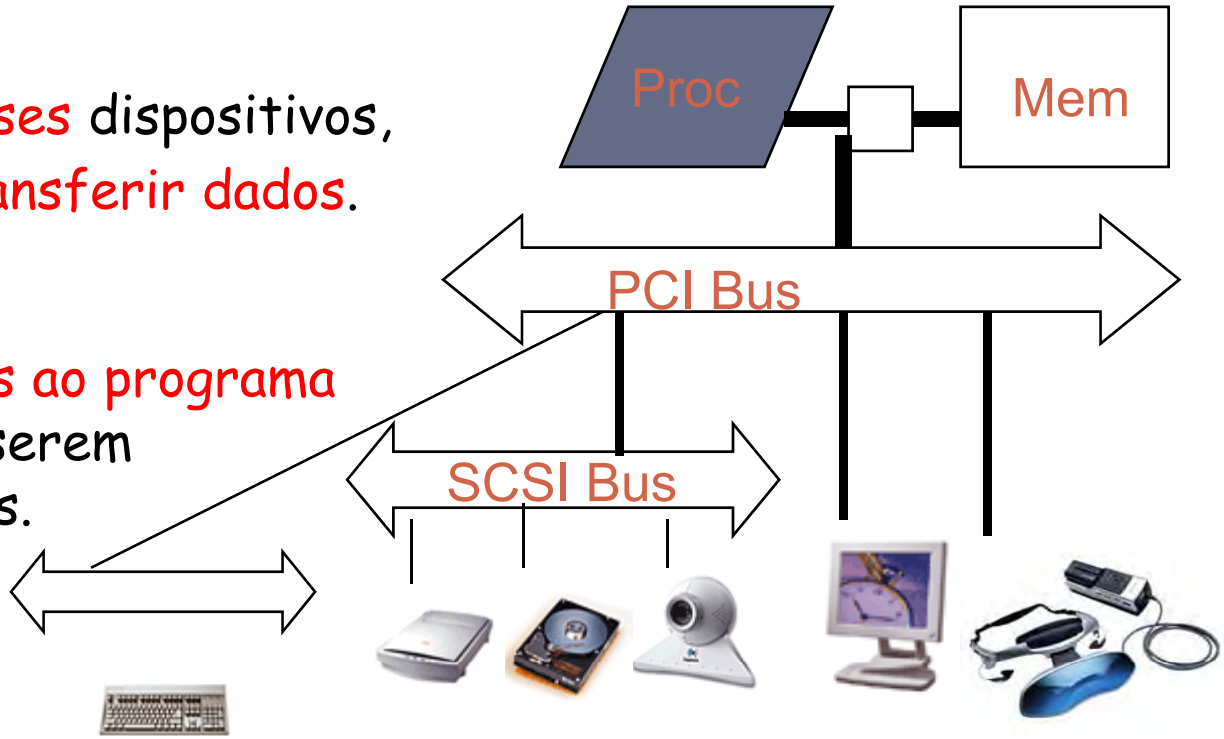
FIGURE 6.7 Characteristics of NOR versus NAND flash memory in 2008. These devices can read bytes and 16-bit words despite their large access sizes.

GAP de Velocidade: Processador / I/O

- CPU de 1GHz pode executar 1 bilhão de instruções de load ou store por segundo → taxa de dados de 4,000,000 KB/s
  - Taxa de dados de dispositivos de I/O → 0.01 KB/s a 125,000 KB/s
- Entrada: o dispositivo pode não estar pronto para enviar dados ao processador assim que requisitado pela CPU
- Saída: dispositivo pode não estar pronto para receber dados na velocidade que a CPU os envia.
- O que fazer ?

Interconectando Componentes

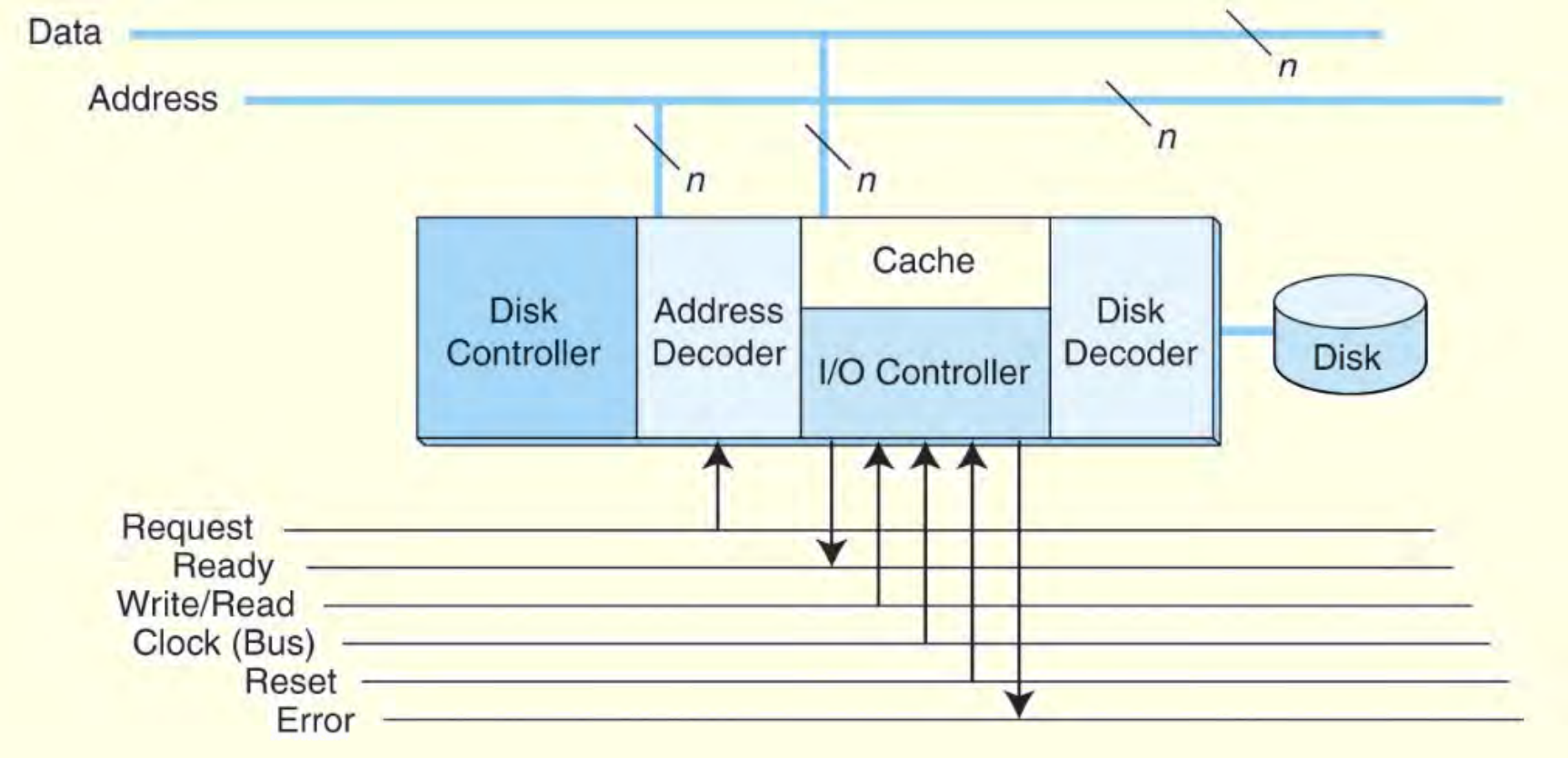
- É preciso conectar vários tipos de dispositivos de I/O ao processador e memória
- É preciso controlar esses dispositivos, responder a eles, e transferir dados.
- É preciso apresentá-los ao programa do usuário, de modo a serem efetivamente utilizados.



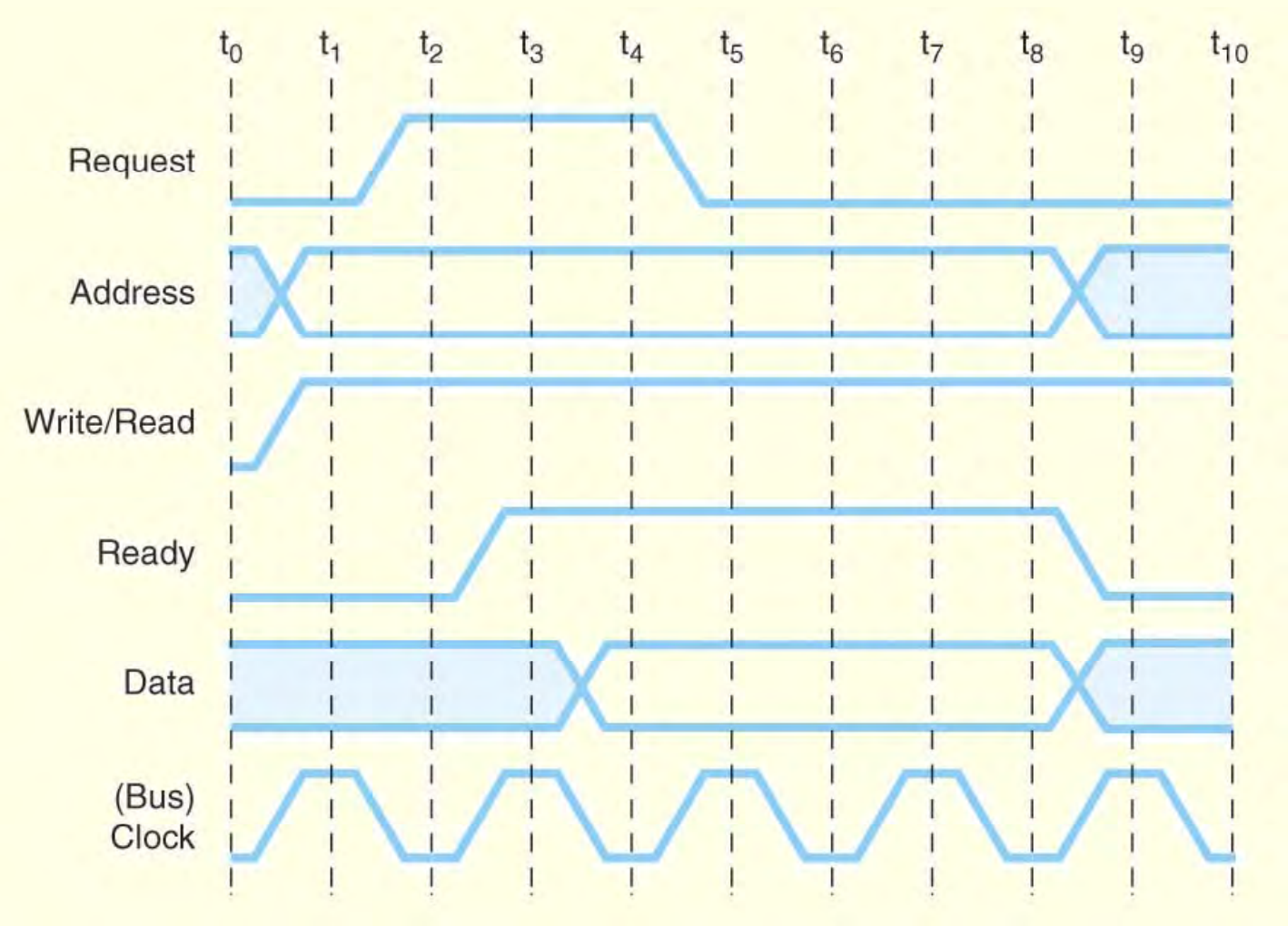
Interconectando Componentes

- Conexões são necessárias entre:
  - CPU, memória, controladores de I/O
- (Barramento (Bus): canal de comunicação compartilhado (shared)
  - Conjunto de fios paralelos usado para transferência de dados e sincronização.
  - Pode ser um gargalo…
- Desempenho limitado por fatores físicos:
  - Comprimento dos fios, número de conexões
- Alternativa mais moderna: conexões seriais de alta velocidade + switches (semelhante a redes)

Ex: Conexão entre barramento e controlador de disco



Ex: Conexão entre barramento e controlador de disco



Tipos de Barramentos

- Barramento Processador-Memória
  - Curto, alta velocidade
  - Projetado de acordo com a organização da memória do computador.

- Barramentos de I/O
  - Mais longos, permitindo conexões múltiplas
  - Projetado de acordo com padrões de interoperabilidade
  - Conectado ao barramento processador-memória por meio de um "barramento-ponte" (bridge)

Barramentos: Sinais e Sincronização

- Linhas de dados (data lines)
  - Transportam endereços e dados (multiplexado ou separados)
- Linhas de controle (control lines)
  - Indicam o tipo de dados trafegando no barramento
  - Sincronizam transações no barramento
- Transações Síncronas
  - Controlada pelo clock do barramento (independente do clock da cpu)
- Transações Assíncronas
  - Utiliza linhas de request/acknowledge p/ "handshaking"

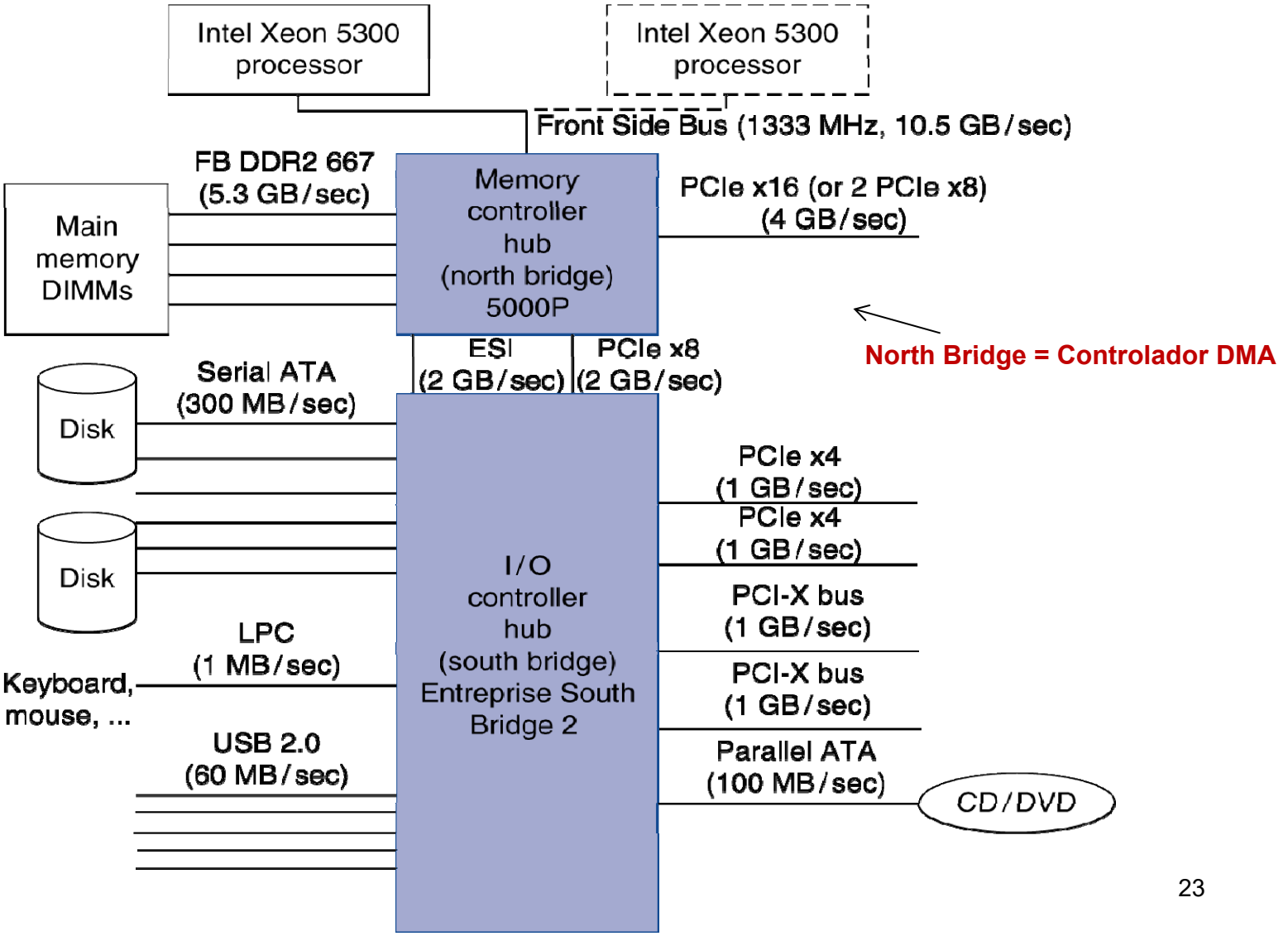
Ex: Barramentos de I/O

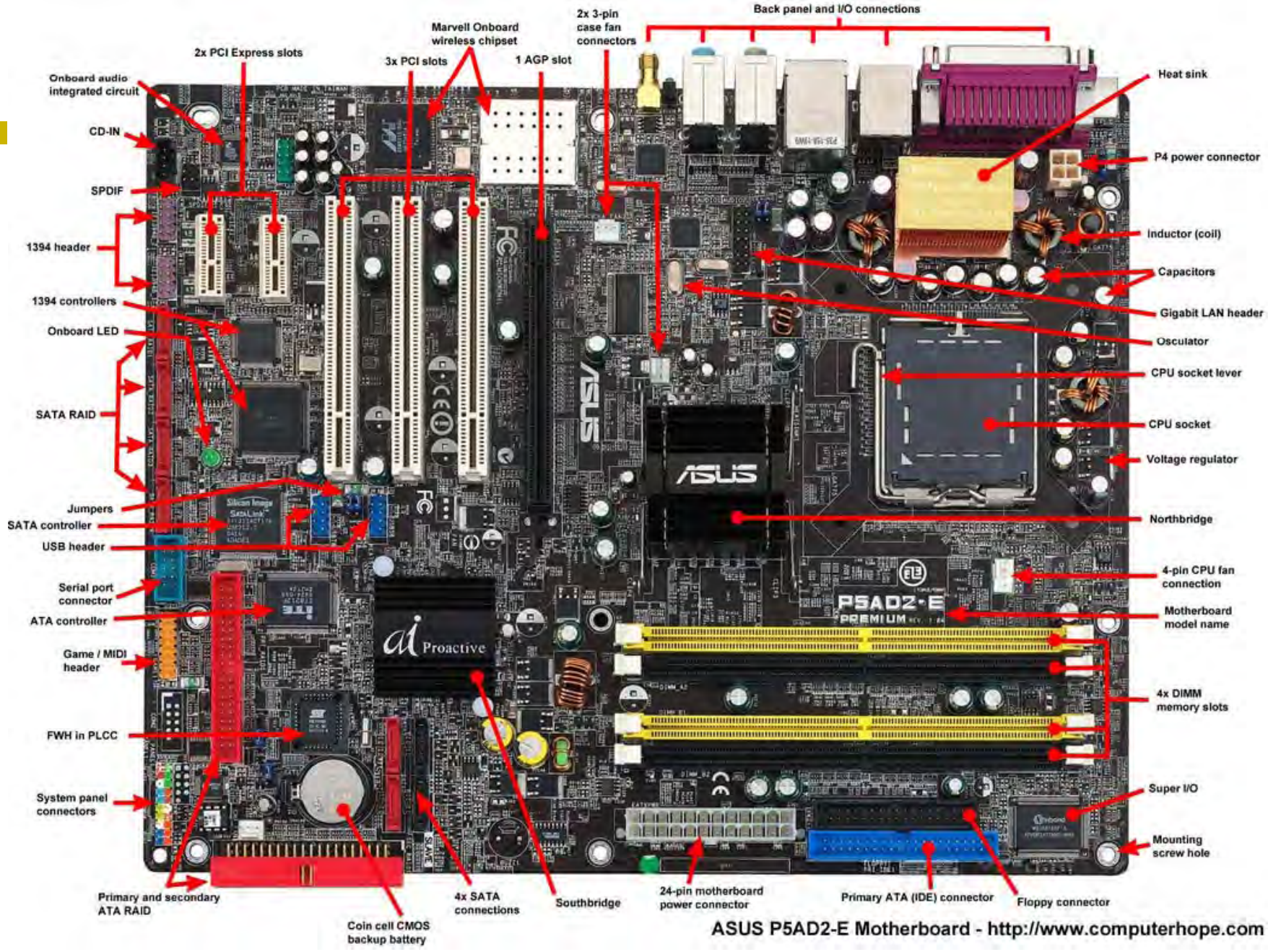
Characteristic	Firewire (1394)	USB 2.0	PCI Express	Serial ATA	Serial Attached SCSI
Intended use	External	External	Internal	Internal	External
Devices per channel	63	127	1	1	4
Basic data width (signals)	4	2	2 per lane	4	4
Theoretical peak bandwidth	50 MB/sec (Firewire 400) or 100 MB/sec (Firewire 800)	0.2 MB/sec (low speed), 1.5 MB/sec (full speed), or 60 MB/sec (high speed)	250 MB/sec per lane (1x); PCIe cards come as 1x, 2x, 4x, 8x, 16x, or 32x	300 MB/ sec	300 MB/sec
Hot pluggable	Yes	Yes	Depends on form factor	Yes	Yes
Maximum bus length (copper wire)	4.5 meters	5 meters	0.5 meters	1 meter	8 meters
Standard name	IEEE 1394, 1394b	USB Implementors Forum	PCI-SIG	SATA-IO	T10 committee

Todos Assíncronos

FIGURE 6.8 Key characteristics of fi ve dominant I/O standards. The intended use column indicates whether it is designed to be used with cables external to the computer or just inside the computer with short cables or wire on printed circuit boards. PCIe can sup port simultaneous reads and writes, so some publications double the bandwidth per lane assuming a 50/50 split of read versus write band width.

Ex: Chipset / Barramentos




Back panel and I/O connections
2x 3-pin case fan connectors
Marvell Onboard wireless chipset
2x PCI Express slots
3x PCI slots
1 AGP slot
Onboard audio integrated circuit
CD-IN
SPDIF
1394 header
1394 controllers
Onboard LED
SATA RAID
Jumpers
SATA controller
USB header
Serial port connector
ATA controller
Game / MIDI header
FWH in PLCC
System panel connectors
Primary and secondary ATA RAID
Coin cell CMOS backup battery
4x SATA connections
Southbridge
24-pin motherboard power connector
Primary ATA (IDE) connector
Floppy connector
Mounting screw hole
Super I/O
4x DIMM memory slots
Motherboard model name
4-pin CPU fan connection
Northbridge
Voltage regulator
CPU socket
CPU socket lever
Osculator
Gigabit LAN header
Capacitors
Inductor (coil)
P4 power connector
Heat sink
ASUS P5AD2-E Motherboard - http://www.computerhope.com

Ex: Chipset / Barramentos

	Intel 5000P chip set	Intel 975X chip set	AMD 580X CrossFiret
Target segment	Server	Performance PC	Server/Performance PC
Front Side Bus (64 bit)	1066/1333 MHz	800/1066 MHz	—
Memory controller hub ("north bridge")			
Product name	Blackbird 5000P MCH	975X MCH	
Pins	1432	1202	
Memory type, speed	DDR2 FBDIMM 667/533	DDR2 800/667/533	
Memory buses, widths	4 × 72	1 × 72	
Number of DIMMs, DRAM/DIMM	16, 1 GB/2 GB/4 GB	4, 1 GB/2 GB	
Maximum memory capacity	64 GB	8 GB	
Memory error correction available?	Yes	No	
PCIe/External Graphics Interface	1 PCIe x16 or 2 PCIe x	1 PCIe x16 or 2 PCIe x8	
South bridge interface	PCIe x8, ESI	PCIe x8	
I/O controller hub ("south bridge")			
Product name	6321 ESB	ICH7	580X CrossFire
Package size, pins	1284	652	549
PCI-bus: width, speed	Two 64-bit, 133 MHz	32-bit, 33 MHz, 6 masters	—
PCI Express ports	Three PCIe x4		Two PCIe x16, Four PCI x1
Ethernet MAC controller, interface	—	1000/100/10 Mbit	—
USB 2.0 ports, controllers	6	8	10
ATA ports, speed	One 100	Two 100	One 133
Serial ATA ports	6	2	4
AC-97 audio controller, interface	—	Yes	Yes
I/O management	SMbus 2.0, GPIO	SMbus 2.0, GPIO	ASF 2.0, GPIO

FIGURE 6.10 Two I/O chip sets from Intel and one from AMD. Note that the north bridge functions are included on the AMD microprocessor, as they are on the more recent Intel Nehalem.

Gerenciamento de I/O

- Operações de I/O são gerenciadas pelo Sistema Operacional
  - Múltiplos programas compartilham recursos de I/O
  - Isso exige Proteção e Escalonamento
- I/O causa interrupções assíncronas
- Programação de I/O é detalhista
  - S.O. fornece abstrações aos programas

Comandos de I/O

- Dispositivos de I/O são gerenciados pela sua respectiva contraladora de I/O (hardware)
  - Transfere dados de/para o dispositivo
  - Syncroniza operações com o software
- Registradores de Comandos
  - Faz com que o dispositivo de I/O efetue alguma ação.
- Registradores de Status
  - Indicam o que o dispositivo está fazendo, ou a ocorrência de erros.
- Registradores de Dados
  - Write: transfere dados para o dispositivo
  - Read: transfere dados do dispositivo

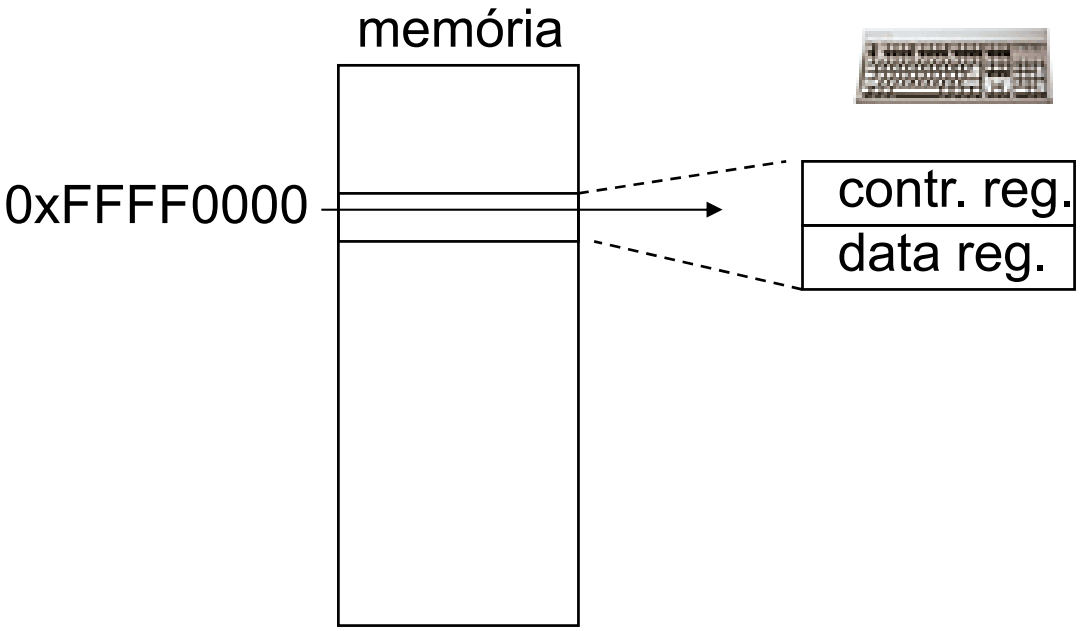
Interação CPU ←→ I/O

- O que o processador deve fazer p/ suportar I/O?
  - Leitura: lê uma sequencia de bytes
  - Saída: escreve uma sequência de bytes
- A) Instruções de I/O
  - Alguns processadores possuem instruções especiais p/ I/O
  - Instruções específicas para acessar registradores de I/O
    - Ex: x86
  - Podem ser acessadas apenas em "kernel mode"
  - Esquema mais antigo....

Interação CPU ←→ I/O

- O que o processador deve fazer p/ suportar I/O?
  - Leitura: lê uma sequencia de bytes
  - Saída: escreve uma sequência de bytes
- B) I/O Mapeado na Memória ("Memory Mapped I/O"):
  - Alternativa às instruções de I/O
  - Usa load p/ input, store p/ output
  - Uma porção do espaço de endereços é usada exclusivamente como canais de comunicação entre a cpu e os dispositivos de I/O.
  - O decodificador de endereços (memória virtual) faz a distinção entre endereços convencionais e endereços de I/O, não permitindo que programas do usuário acessem esses endereços.
  - Os endereços de I/O são na verdade registradores em controladores e/ou dispositivos.
  - Ex: MIPS

I/O Mapeado na Memória



Pesquisa (Polling)

- Como a CPU sabe que um comando de I/O deve ser atendido ?
- A) Pesquisa (Polling): O S.O. periodicamente verifica o registrador de status de um dispositivo de I/O
  - Se estiver pronto, executa a operação
  - Se ocorreu algum erro, execute ações de correção
- Esquema mais usado em sistemas embarcados simples, exigindo pouco desempenho, ou sistemas de tempo real, exigindo garantia no tratamento de I/O.
  - Requisitos de tempo são previsíveis
  - Baixo custo de hardware
- ……mas, no geral, polling é ineficiente, pois desperdiça o tempo da CPU.

Interrupções de I/O

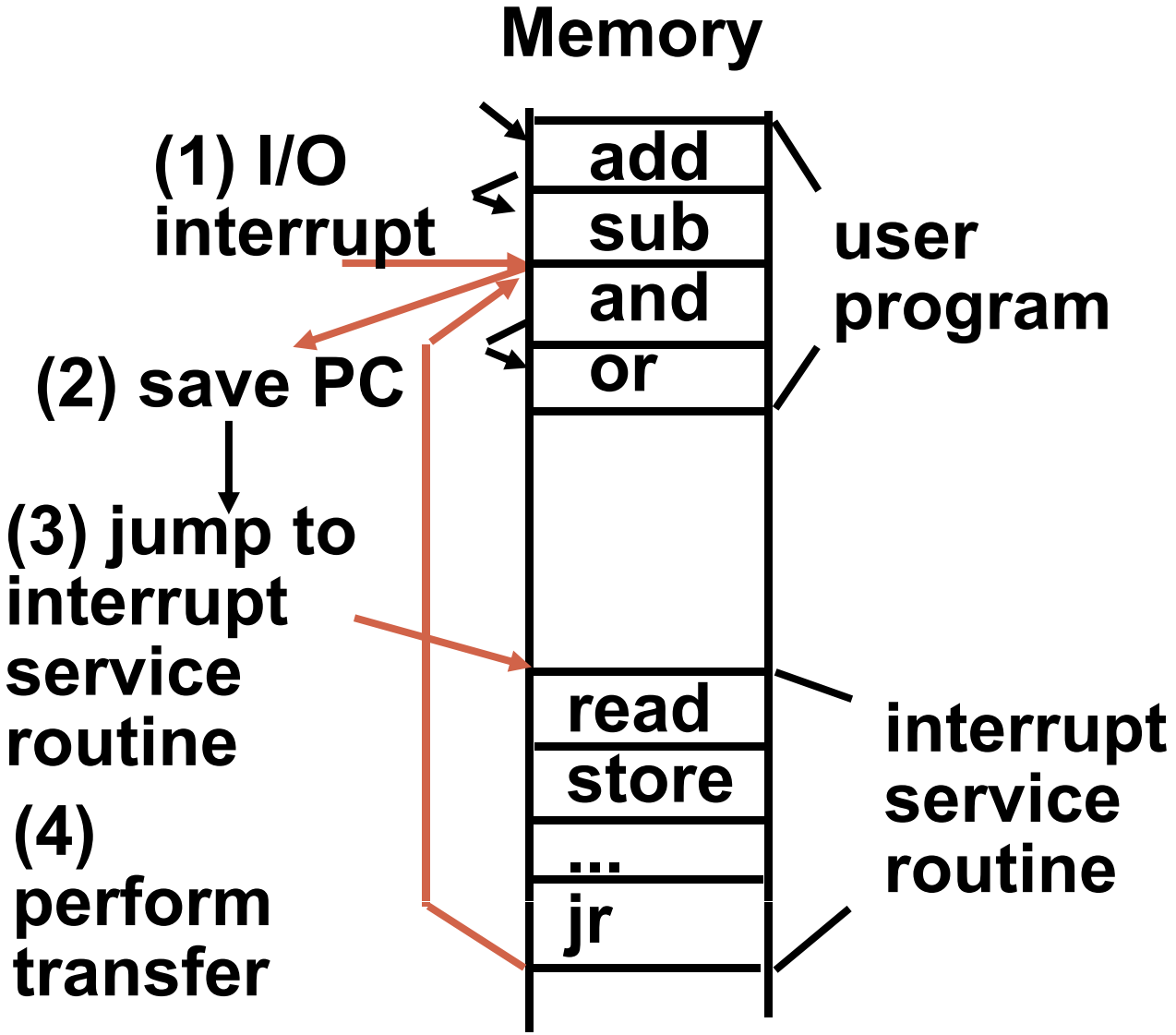
- B) Interrupção de I/O
  - Interrupção é similar a uma exceção
    - Porém, não é sincronizada à execução de instruções
    - Pode invocar o manipulador (handler) da interrupção entre instruções.
    - Muitas vezes a própria interrupção identifica o dispositivo que a disparou
  - Interrupções Prioritárias
    - Alguns dispositivos exigem atenção mais urgente que outros.
    - Nesses casos, outras interrupções em curso podem ser interrompidas.

Interrupções de I/O

- Uma interrupção de I/O é como uma excessão de overflow, porém:
  - Mais informações precisam ser obtidas sobre ela.
  - É assíncrona em relação à execução de uma instrução
    - Pode ocorrer em qualquer ponto da execução (estágio do pipeline)
    - Não impedem a conclusão da execução de nenhuma instrução.

Tratamento de Interrupções

Ex: transferência de dados de I/O

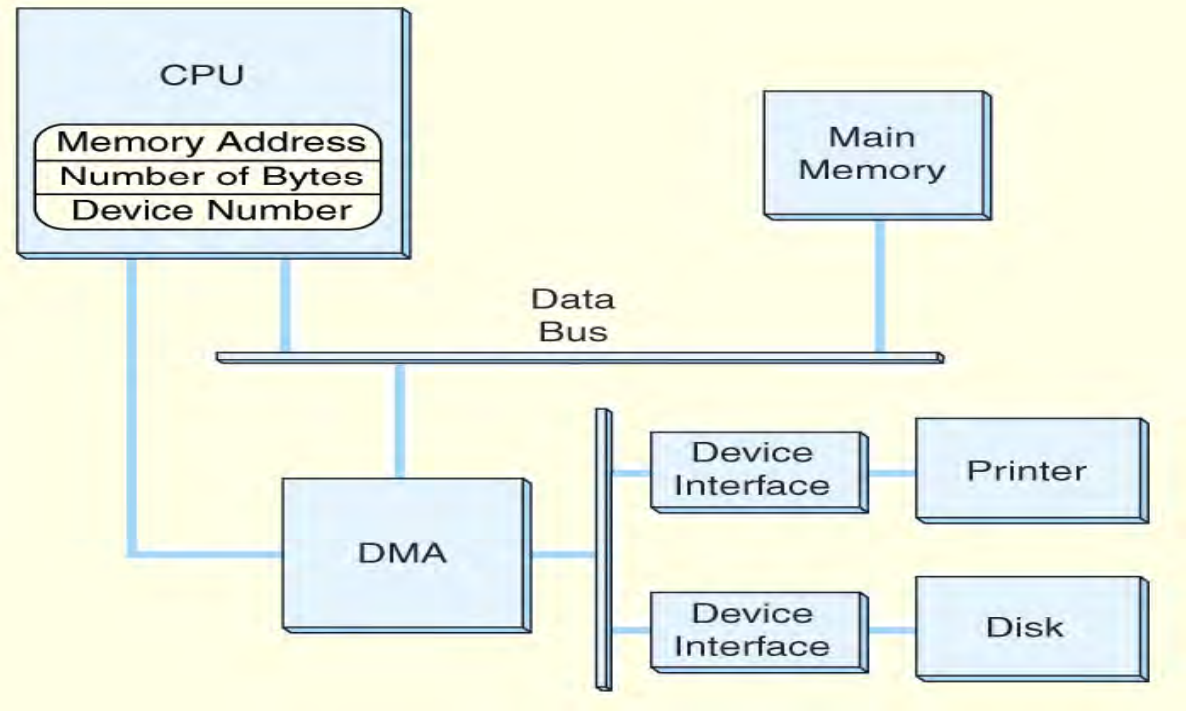


Transferências Dispositivos ←→ Memória

- I/O via Polling ou Interrupções
  - CPU transfere dados entre a memória e os registradores de I/O
  - ...muito demorado para dispositivos ou operações que exigem alta velocidade e grande volume de dados.

Transferências Dispositivos ←→ Memória

- Alternativa: Acesso Direto à Memória (DMA)
  - S.O. indica o endereço inicial na memória
  - A contraladora de I/O transfere dados de/para a memória independentemente da CPU
  - O controlador de DMA sinaliza o fim da operação (ou erro)



Interação DMA/Cache

- Se a DMA grava um bloco de memória cuja cópia está na cache:
  - A linha de cache correspondente deve ser marcada como inválida
- No caso de caches do tipo write-back, se a DMA lê um bloco marcado na cache como "dirty", dados desatualizados serão lidos.
- É preciso assegurar a coerência do cache
  - Remova (flush) os blocos da cache antes de serem usados pela DMA
  - Ou... use endereços de memória "non-cacheable" para operações de I/O

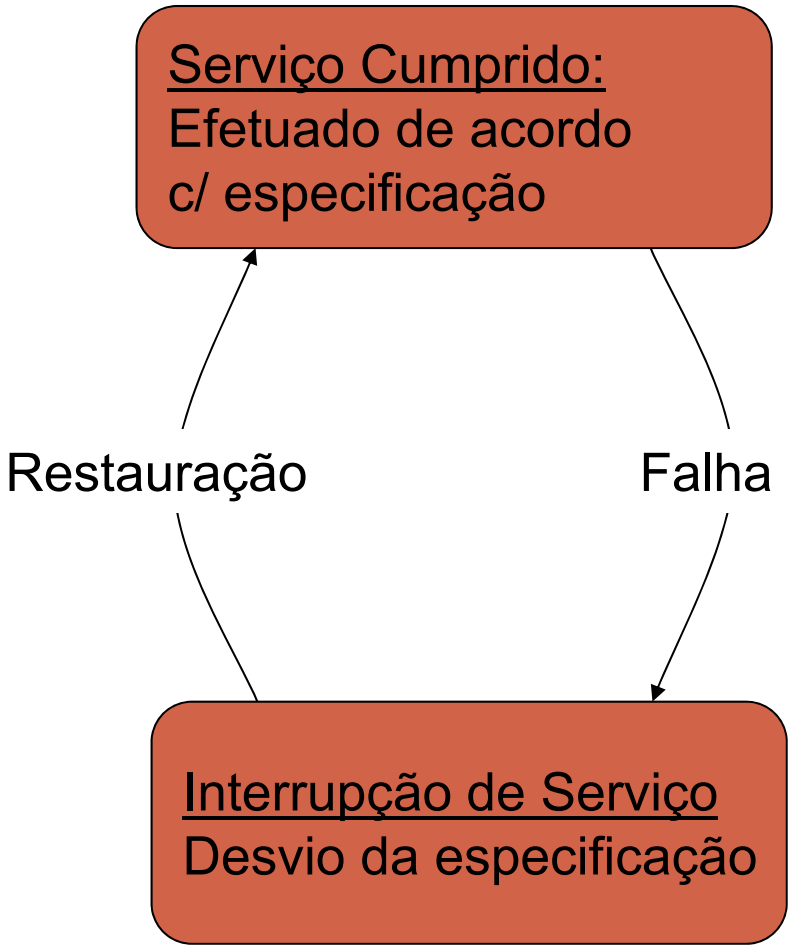
Interação DMA/Memória Virtual

- S.O. utiliza endereços de memória virtual
  - Porém, os blocos DMA podem não corresponder a endereços físicos consecutivos, o que pode reduzir seu desempenho.
- DMA usando endereços virtuais:
  - A tradução deverá ser feita pelo controlador.
- DMA usando endereços físicos
  - Pode ser necessário subdividir a transferência em grupos de páginas, ou executar múltipças transferências em sequencia, ou alocar páginas físicas contíguas p/ otimizar a operação do DMA.

Caracterizando Sistemas de I/O

- Resistência é importante!
  - Especialmente para dispositivos de armazenamento
- Desempenho
  - Latencia / Tempo de Resposta
  - Produção / Throughput (bandwidth)
- Desktops & Sistemas Embarcados
  - Interessados principalmente em tempo de resposta e diversidade de dispositivos
- Servidores
  - Interessados principalmente em throughput e capacidade de expansão.

Resistência



- Defeito: falha de um componente: pode ou não resultar em uma falha do sistema.

Medidas de Resistência

- Confiabilidade: tempo médio p/ a ocorrência de uma falha (MTTF)
- Interrupção de Serviço: tempo médio p/ reparo (MTTR)
- Tempo médio entre falhas:
  - MTBF = MTTF + MTTR
- Disponibilidade = MTTF / (MTTF + MTTR)
- Melhorando a Disponibilidade:
  - p/ aumentar MTTF: medidas p/ evitar, tolerar e prever defeitos
  - p/ reduzir MTTR: ferramentas e processos p/ diagnóstico e reparos.

Medindo o Desempenho de I/O

- Desempenho de I/O depende de:
  - Hardware: CPU, memória, controladores, barramentos
  - Software: S.O., SGBD, Aplicações.
  - Carga de Trabalho: taxa de requisições de I/O, padrões (ex: volume e tipo de transferências).

- Sistemas de I/O podem ser projetados/configurados de modo a balancear tempo de resposta e throughput

Medindo o Desempenho de I/O

- Medindo Transações
  - Acesso a pequeno volume de dados em SGBD
  - Interesse em medir taxa de I/O, e não taxa de dados.

- Medindo throughput
  - Sujeito a limites no tempo de resposta, e gerenciamento de falhas.
  - ACID (Atomicity, Consistency, Isolation, Durability)
  - Custa total por transação

Benchmarks usados p/ medir I/O

- Transaction Processing Council (TPC) benchmarks (www.tcp.org)
  - TPC-APP: B2B servidor de aplicações e web services
  - TCP-C: entrada de pedidos online
  - TCP-E: processamento online de transações em empresas de corretagem
  - TPC-H: suporte a decisões, ad-hoc queries
- SPEC System File System (SFS)
  - Carga de trabalho sintética p/ um servidor NFS server, baseda no monitoramento de sistemas reais
- SPEC Web Server benchmark
  - Três cargas de trabalho: Bancos, E-commerce, Suporte

I/O vs. CPU Performance

- Amdahl's Law
  - Na medida que o paralelismo aumenta, não esqueça do I/O
  - Exemplo:
    - Benchmark gasta 90s em tempo de CPU, e 10s em tempo de I/O
  - Duplique o número de CPUs a cada 2 anos, porém sem modificar o desempenho do sistema de I/O:

Year	CPU time	I/O time	Elapsed time	% I/O time
now	90s	10s	100s	10%
+2	45s	10s	55s	18%
+4	23s	10s	33s	31%
+6	11s	10s	21s	47%

Atenção: Desempenho Máximo (Peak Performance)

- Taxas máximas de I/O são quase inatingíveis na prática
  - Normalmente, algum outro componente do sistema limita o desempenho
  - Ex: transferências p/ memória via barramento
    - Ex: Competição pelo barramento c/outros dispositivos
  - Ex:. PCI bus: peak bandwidth ~133 MB/sec
    - Na prática, sustenta-se no máximo 80MB/sec

Conclusão

- Sistemas de I/O recebem pouca atenção em comparação com CPU e Memória.
- Porém, são fundamentais para a funcionalidade dos sistemas, a manutenção do desempenho necessário, e a escalabilidade dos mesmos.
- Sistema do tipo HPC muitas vezes tem alto custo devido à sofisticação dos subsistemas de I/O.
- Por fim, a interoperabilidade é muito importante.
  - Ver: Ethernet, WIFI, WIMAX, 3G, USB, HDMI …….. ☺