Centro de Ciência e Tecnologia Laboratório de Ciências Matemáticas ENF Ciência da Computação

Arquitetura de Computadores Aula 07

Capítulo 3 Chips de CPU Barramento de Computador

- CPUs são contidas em um único chip
- Cada chip tem um conjunto de pinos por meio dos quais deve ocorrer toda a comunicação com o mundo externo
- Pinos -> sinal da CPU → mundo externo
 - sinal do mundo externo ---> CPU
 - ambos
 - De endereço
 - De dados
 - De controle

Comunicação entre CPU com memória e dispositivos de E/S

- Ao buscar uma instrução, a CPU coloca o endereço de memória desta instrução em seus pinos de endereço
- Ativa uma ou mais linhas de controle para informar que ela quer ler uma palavra, por exemplo
- A memória responde colocando a palavra requisitadas nos pinos de dados da CPU, e ativa um sinal que informa o que acabou de fazer
- Percebido o sinal, a CPU aceita a palavra e excuta a instrução

Comunicação entre CPU com memória e dispositivos de E/S

- Ao buscar uma instrução, a CPU coloca o endereço de memória desta instrução em seus pinos de endereço
- Ativa uma ou mais linhas de controle para informar que ela quer ler uma palavra, por exemplo
- A memória responde colocando a palavra requisitadas nos pinos de dados da CPU, e ativa um sinal que informa o que acabou de fazer
- Percebido o sinal, a CPU aceita a palavra e excuta a instrução

Resumindo: a CPU se comunica com a memória e dispositivo de E/S **apresentando** sinais em seus pinos e **aceitando** sinais em seus pinos.

Desempenho

- Número de pinos de endereço
- Número de pinos de dados

Endereço:

 m pinos pode endereçar até 2^m localizações de memória valores comuns de m: 16, 20, 32, e 64

Dados:

 n pinos pode ler ou escrever uma palavra de n bits em uma única operação valores comuns de n: 8,16, 32, 36 e 64

Outros pinos

De Controle

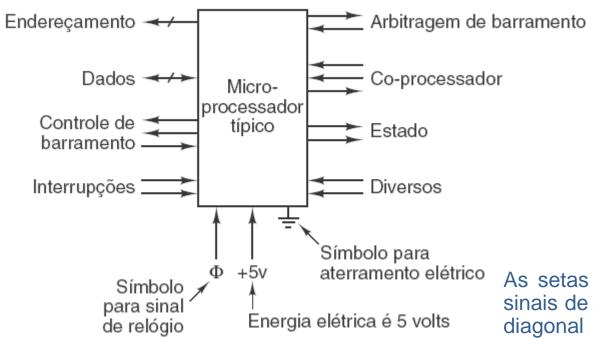
- regulam o fluxo e a temporização de dados que vem da CPU e vão para ela
 - 1. Controle de barramento
 - 2. Interrupções
 - 3. Arbitragem de co-processador
 - 4. Estado
 - 5. Diversos

Energia Elétrica

Terra

Sinal de relógio

Pinagem lógica de uma CPU genérica



As setas indicam sinais de entrada e sinais de saída. Os segmentos de reta diagonal indicam que são utilizados vários pinos. Há um número que indica quantos são os pinos para uma CPU específica.

Barramento

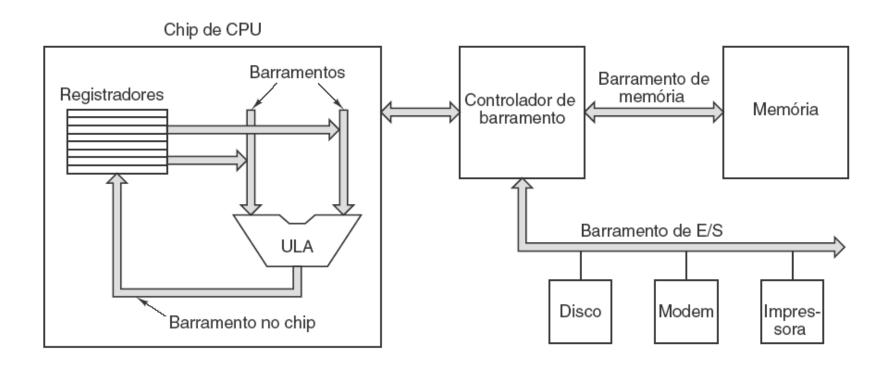
Um caminho elétrico comum entre vários dispositivos Podem ser externos ou internos a CPU

Primeiros computadores pessoais

- Somente um barramento externo: barramento de sistema
- Consistia em 50 a 100 fios de cobre paralelos gravados na placa mãe, com conectores a intervalos regulares para ligação com a memória e placas de E/S

Computadores pessoais moderno

 Tem um barramento de uso especial entre a CPU e a memória, e ao menos, um outro para dispositivos de E/S



Sistema mínimo, com um barramento de memória e um barramento de E/S

- Projetistas tem liberdade para usar qualquer tipo de barramento que quiserem dentro do chip de CPU
- Mas e para a ligação de placas projetadas por terceiros?

Protocolo de Barramento

Regras bem definidas sobre o modo de funcionamento do barramento, às quais todos os dispositivos a ele ligados tem de obedecer

Além disso, são necessárias especificações mecânicas e elétricas

Ex. de barramentos

- Omnibus (PDP-8)
- Unibus (PDP-11)
- Multibus (8086)
- Barramento VME (equipamento para laboratório de física)
- Barramento IBM PC (PC/XT)
- Barramento ISA (PC/AT)
- Barramento EISA (80386)
- Microchannel (PS/2)
- Nubus (Macintosh)
- Barramento PCI (muitos PCs)
- Barramento SCSI (muitos PCs e estações de trabalho)
- Universal Serial Bus (PCs modernos)
- FireWire (equipamentos eletrônicos e de consumo)

Funcionamento dos barramentos

- Mestres: dispositivos ativos; podem iniciar transferências no barramento
- **Escravos**: dispositivos passivos; esperam requisições

Mestre	Escravo	Exemplo		
CPU	Memória	Buscar instruções e dados		
CPU	Dispositivo de E/S	Iniciar transferência de dados		
CPU	Co-processador	CPU que passa instruções para o co-processado		
E/S	Memória	DMA (acesso direto à memória)		
Co-processador	CPU	Co-processador que busca operandos na CPU		

Funcionamento dos barramentos

- Sinais binários emitidos por dispositivos de computador muitas vezes são muito fracos para energizar um barramento (principalmente se for longo e tiver muitos dispositivos ligados a ele)
- Controlador de barramento: um amplificador digital; a maioria dos mestres se conecta ao barramento por esse chip.
- Receptor de barramento: chip onde a maioria dos escravos se conectam
- Transceptor de barramento: chip combinado, usado por dispositivos que podem agir como mestres e escravos

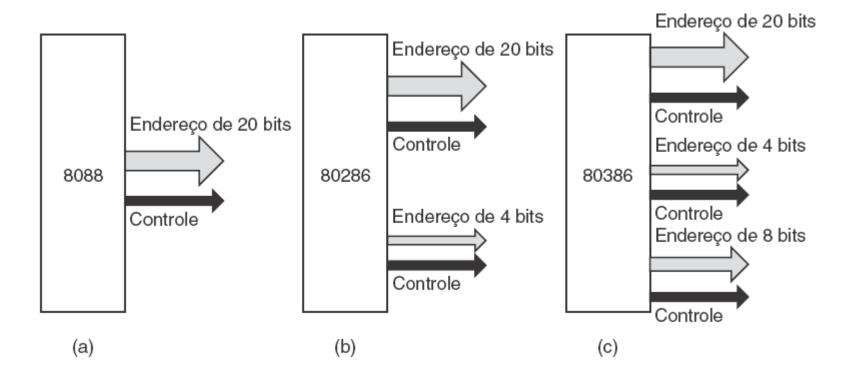
Largura do barramento

 Quanto mais linhas de endereço tiver um barramento, mais memória CPU pode endereçar diretamente

barramento com n linhas de endereço — CPU pode endereçar 2^n localizações de memória

- Memórias de grande porte precisam de muitas linhas de endereço no barramento. Problema: barramentos largos precisam de mais fios, ocupam mais espaço físico, precisam de conectores maiores, custo elevado!
- Ex.: um sistema com barramento de endereço de 64 linhas e 2³² bytes de memória custará mais do que do que um sistema com 32 linhas de endereço e 2³² bytes de memória

3.4.3 Largura do barramento



Crescimento de um barramento de endereço ao longo do tempo

3.4.3 Largura do barramento

- Não só o nº de linhas de endereços tendem a crescer, mas o número de linhas de dados
- Duas maneiras de aumentar a largura de banda de dados:
 - Reduzir o ciclo de tempo do barramento (+ transferências por seg.)
 - ✓ Problemas: atraso diferencial do barramento incompatibilidade com placas antigas
 - Aumentar a largura de dados (mais bits por transferência)
 - Não leva a um projeto limpo

3.4.3 Largura do barramento

- Para contornar o problema de barramentos muito largos, às vezes os projetistas optam por um **barramento multiplexado**.
- Em vez de as linhas de endereços e dados serem separadas, há pro exemplo, 32 linhas para endereços e dados juntos.
- Multiplexar as linhas reduz a largura (e o custo) do barramento, mas resulta em um sistema mais lento.

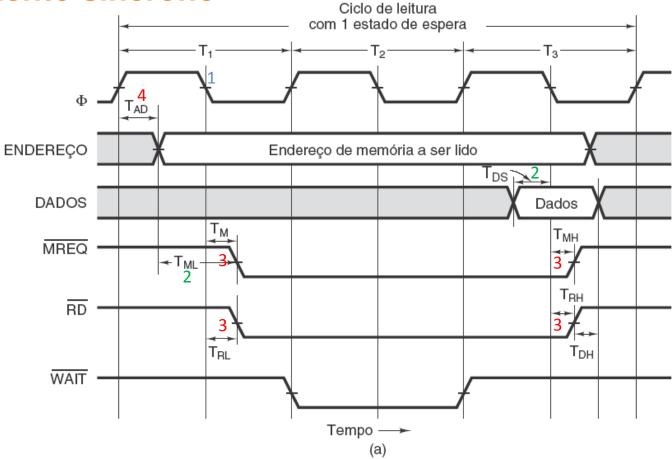
Barramento síncrono

- Tem uma linha comandada por um oscilador de cristal
- O sinal nessa linha consiste em uma onda quadrada com uma frequência em geral entre 5 MHz e 100 MHz
- Todas as atividades do barramento tomam um número inteiro desses ciclos denominados ciclos de barramento

Barramento assíncrono

- Não tem um relógio mestre
- Ciclos de barramento podem ter qualquer largura requerida e não são os mesmos entre todos os pares de dispositivos

Barramento síncrono



Barramento síncrono

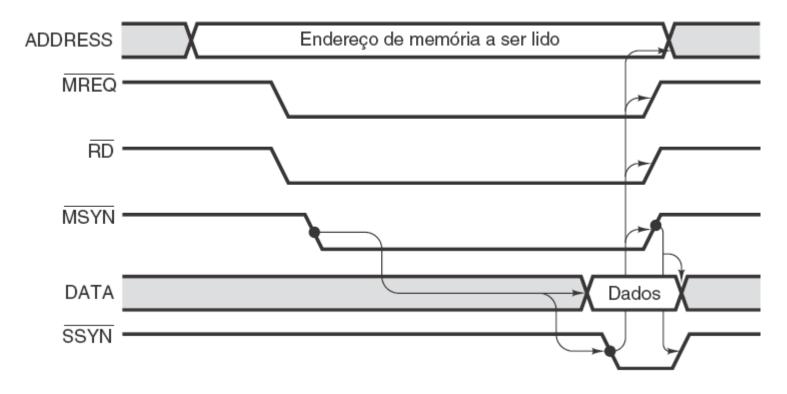
Símbolo	Parâmetro		Máx.	Unidade
T _{AD}	Atraso de saída do endereço		4	nsec
T _{ML}	Endereço estável antes de MREQ	2		nsec
T _M	Atraso de $\overline{\text{MREQ}}$ desde a borda descendente de Φ em T1		3	nsec
T _{RL}	Atraso de RD desde a borda descendente de Φ em T1		3	nsec
T _{DS}	Tempo de ajuste dos dados antes da borda descendente de Φ	2		nsec
T _{MH}	Atraso de $\overline{\text{MREQ}}$ desde a borda descendente de Φ em T3		3	nsec
T _{RH}	Atraso de $\overline{\text{RD}}$ desde a borda descendente de Φ em T3		3	nsec
T _{DH}	Tempo de sustentação dos dados desde a negação de RD	0		nsec

 Se um barramento síncrono tiver uma coleção heterogênea de dispositivos, alguns rápidos, alguns lentos, o barramento tem de ser ajustado para o mais lento, e os mais rápidos não podem usar todo o seu potencial

Barramento assíncrono

 Operações completas são independentes de temporização. Cada evento é causado por um evento anterior e não por um pulso de relógio.

Barramento assíncrono



MSYN (Master SYNchronization) SSYN (Slave SYNchronization)

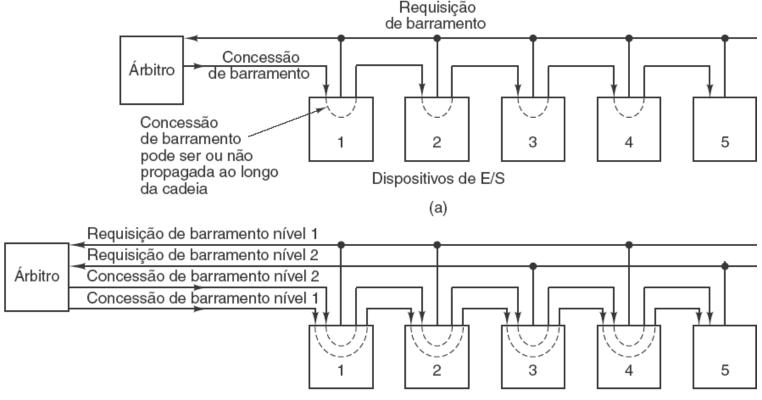
3.4.5 Arbitragem de barramento

Mecanismo de arbitragem

- Evita o caos quando dois ou mais dispositivos querem se tornar mestres de transferência de dados ao mesmo tempo
- Centralizado
- Descentralizado

3.4.5 Arbitragem de barramento

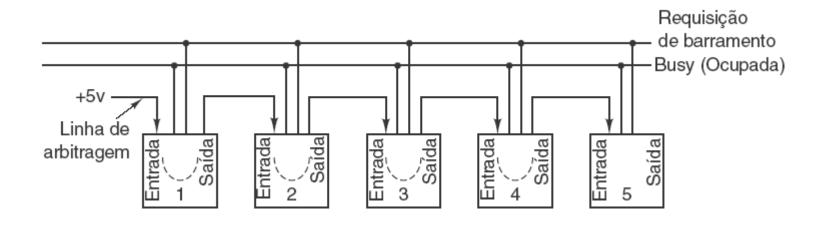
Centralizado



- (a) Árbitro de barramento centralizado de um nível usando encadeamento em série (daisy chaining).
- (b) O mesmo árbitro usando dois níveis.

3.4.5 Arbitragem de barramento

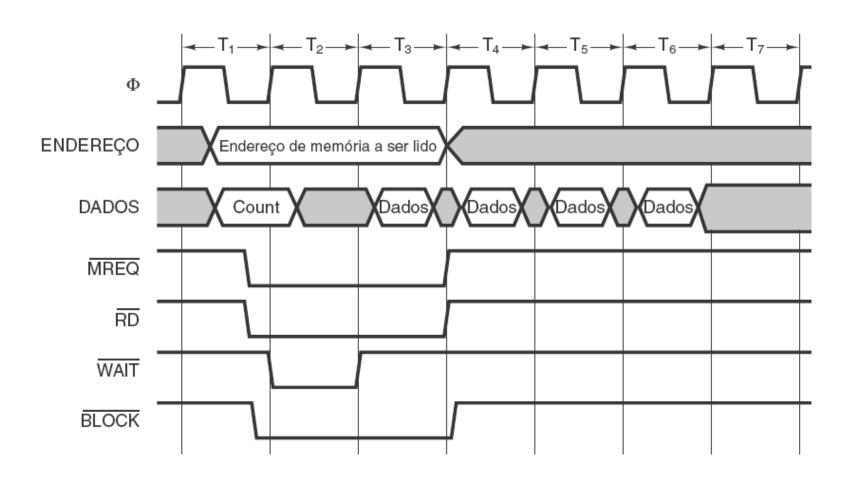
Descentralizado



 Não existe apenas ciclos de barramento comum, com um mestre lendo de um escravo ou escrevendo para ele. Existem vários outros tipos de ciclos

Transferência de bloco

- Quando é usado caching, é desejável buscar uma linha inteira de cache por vez. Ex. 16 palavras de 32 bits consecutivas.
- Quando uma leitura de bloco é iniciada, o mestre de transferência de dados informa ao escravo quantas palavras serão transferidas
- Em vez de retornar apenas uma palavra, o escravo entrega uma palavra durante cada ciclo até esgotar aquele nº de palavras
- Um sinal extra, \overline{BLOCK} , é ativado para indicar que foi requisitada uma transferência em bloco.



Ler-modificar-escrever

- Um sistema multiprocessador, com duas ou mais CPUs no mesmo barramento, muitas vezes é necessário garantir que só uma CPU por vez use algum estrutura de dados crítica na memória
- Como organizar: uma variável que é 0 quando nenhuma CPU estiver usando a estrutura de dados, e 1 quando estiver em uso
- Mas, e duas CPUs lerem a variável em ciclos de consecutivos?
 Caos!

Ler-modificar-escrever

- É um tipo de ciclo que evita que uma CPU rival possa usar o barramento e assim interferir com a operação da primeira CPU
- Permite a qualquer CPU ler uma palavra da memória, inspecionar e modificar essa palavra, e escrevê-la novamente na memória, tudo sem liberar o barramento.

Manipular interrupções

- Quando ordena que um dispositivos de E/S faça algo, a CPU normalmente espera uma interrupção quando o trabalho for concluído. A sinalização da interrupção requer o barramento
- Vários dispositivos podem querer causar uma interrupção simultaneamente: os mesmo tipos de problemas de arbitragem que há nos ciclos de barramento comum estão presentes aqui
- Solução: atribuir prioridade a dispositivos e usar um árbitro centralizado para dar prioridade aos dispositivos mais críticos em relação a tempo

Exemplo de chips de CPU

- Pentium 4
- UltraSPARCIII
- 8051

3.5

Primeiro Pentium 4

- Lançado em 4 de novembro de 2000
- CPU de 42 milhões de transistores de funcionava em 1,5 GHz com uma largura de linha de 0,18 mícron
 - Largura de linha: a largura dos fios entre transistores, bem como uma medida do tamanho dos próprios transistores
 - Quanto menor a largura de linha, mais transistores podem caber em um chip, também permite velocidades mais altas do relógio

3 anos depois

 55 milhões de transistores de funcionava em 3,2 GHz com uma largura de linha de 0,09 mícron

Do ponto de vista de software

- É uma máquina completa de 32 bits
- Totalmente compatível com 8088
- Tem todas as mesmas características ISA de nível de usuário que os chips:
 - √ 80386
 - **√** 80486
 - ✓ Pentium
 - ✓ Pentium II
 - ✓ Pentium Pro
 - ✓ Pentium III
- Tem também os mesmos registradores, as mesma instruções, e uma implementação completa no chip do padrão IEEE 754 de ponto flutuante
- Algumas novas instruções destinadas primeiramente a aplicações de multimídia

Do ponto de vista de hardware

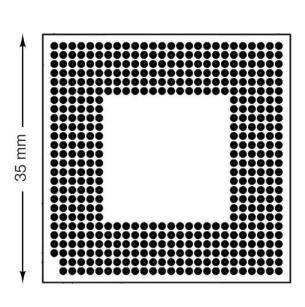
- É em parte uma máquina de 64 bits: pois, pode transferir dados de e para a memória em unidade de 64 bits
- No nível da microarquitetura (NetBurst), é radicalmente diferente de todos seus antecessores
- Tem um paralelismo mais profundo, duas ULAS: cada uma funcionando a duas vezes a frequência do relógio para permitir duas operações por ciclo
- Prevê dois conjuntos de registradores e alguns outros recursos internos que permitem um programa passe para outro com muita rapidez, como se o computador tivesse duas CPUs físicas

Barramento

- São usados dois barramentos primário externos, ambos síncronos
- O barramento de memória é usado para acessar a (S)DRAM principal
- O barramento PCI é usado para falar com dispositivos de E/S
- Às vezes um barramento **herdado** (antigo) é ligado ao barramento PCI para permitir a ligação de dispositivos periféricos antigos

Chips de CPU

- Vem em um pacote de 35mm de lado e contém 478 pinos na parte interior, 85 para energia, e 180 aterrado para reduzir ruído
- Os pinos são organizados como um quadrado de 26x26 sem nada na parte do meio (14x14)
- O chip é equipado com um suporte de montagem para um dissipador de calor que distribui o calor e um ventilador que resfria o chip



3.5.1 **Pentium 4**

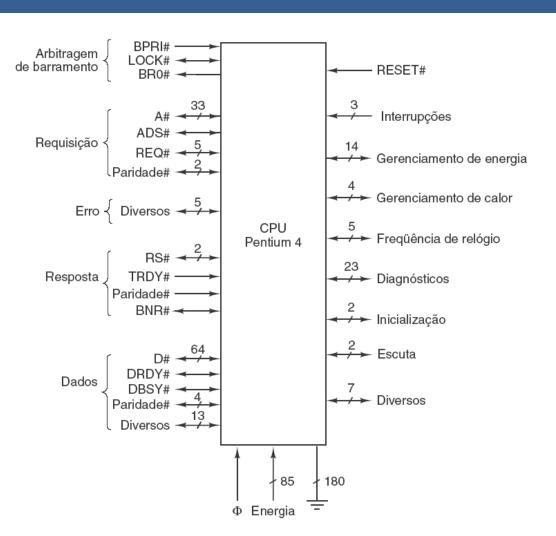
Pinagem lógica

- Dos 478 pinos, 198 são usados para sinais, 85 conexões com energia elétrica, 180 terras e 15 avulsos para uso futuro
- Alguns dos sinais lógicos usam dois ou mais pinos (tal como o endereço de memória requisitado), de modo que há somente 56 sinais diferentes

3.5.1 **Pentium 4**

Pinagem lógica

Nomes em letras maiúsculas são nomes oficiais usados pela Intel para sinais individuais. Nomes em maiúsculas e minúsculas são grupos de sinais relacionados ou descrições de sinais.



Familia Sun UltraSPARC

- Família é a linha da Sun de CPUs SPARC de 64 bits
- Essas CPUS estão totalmente de acordo com a arquitetura SPARC versão 9, que também é para CPUs de 64 bits
- São usadas em estação trabalho e servidores Sun, bem como várias outras aplicações
- Família inclui:
 - UltraSPARC I
 - UltraSPARC II
 - UltraSPARC III
- Muitos semelhantes em termos de arquitetura
- Principais diferenças: data de lançamento, velocidade de relógio e algumas instruções extras introduzidas em cada modelo

- É um máquina RISC tradicional e totalmente compatível com o código binário da arquitetura SPARC V8 de 32 bits
- Pode executar programas binários SPARC V8 sem modificação
- Só se desvia da arquitetura em um único ponto: no acréscimo do conjunto de instruções VIS 2.0 projetado para aplicações gráficas em 3D, decodificação MPEG em tempo real, compressão de dados, processamento de sinal, execução de programas Java e trabalho em rede
- Foi projetada para ser utilizada no negócio principal da Sun: grandes servidores multiprocessados de memória compartilhada usados na internet e intranets corporativas

Primeira UltraSPARC III

- Foi lançada em 2000 a 600MHz usando linhas de alumínio de 0,18 mícron
- Chips com 29 milhões de transistores
- Chips foram fabricados pela Texas Instruments

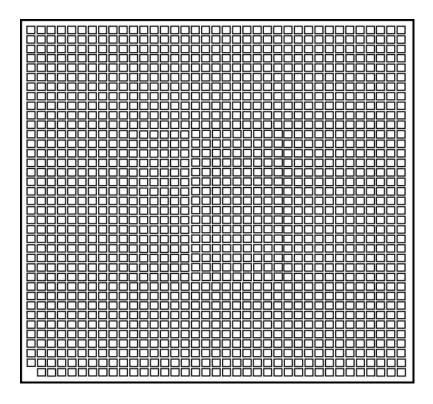
2001

• Tl melhorou sua tecnologia: começou a fabricar chips de 0,15 mícron de 900MHz usando fios de cobre em vez de alumínio

2002

Largura de linha caiu para 0,13 mícron e o relógio foi elevado até 1,2 GHz

- Tem 6 paralelismos internos, entre eles:
 - dois de 14 estágios para operações com inteiros
 - dois para operações com ponto flutuante
 - um para operações de carga/armazenagem
 - um para desvios
- Adota uma abordagem diferente em relação ao caching, barramentos mais largos e outros fatores que melhoram o desempenho
- Vem em um arranjo de grade geodésica (Land Grid Array) de 1.368 pinos. Esse pacote consiste em uma matriz quadrada de 37x37 = 1.369 pinos na parte inferior do chip, faltando o pino do canto inferior esquerdo

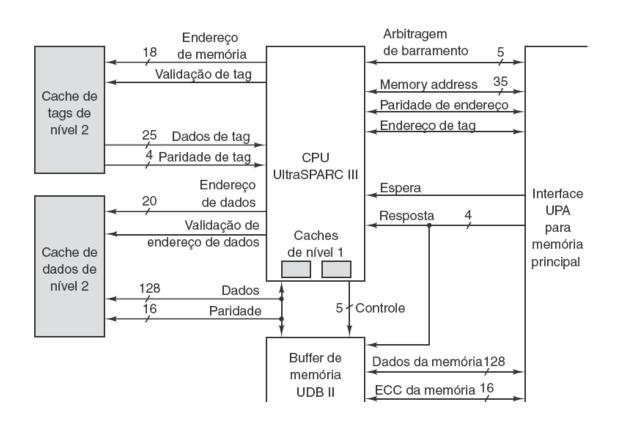


Chip de CPU UltraSPARC III

- Usa um barramento de endereços de 43 bits de largura, o que permite que ele tenha até 8TB de memória principal
- O barramento de dados tem 128 bits de largura, o que permite a transferência de 16 bytes por vez entre CPU e memória
- A velocidade do barramento é 150MHz, o que dá uma largura de banda de memória de 2,4 GB por segundo, muito mais rápida do que os 528 por segundo do PCI
- Para conectar (várias) CPUs UltraSPARC para comunicação com (várias) memórias, a Sun desenvolveu a UPA (Ultra Port Architecture – arquitetura ultraporta)

UPA

- Pode ser implementada como um barramento, um comutador ou uma combinação dos dois
- A implementação não importa para a CPU, porque a interface com ela é definida com precisão e é essa interface que o chip de CPU deve suportar (e suporta)



Núcleo de um sistema UltraSPARC

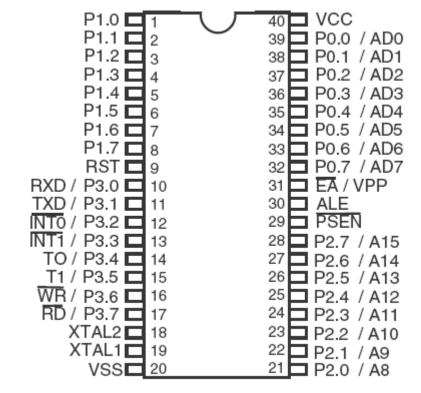
- É o microcontrolador mais popular em uso hoje, em grande parte por causa de seu custo muito baixo
- É um chip simples, fazer interface com ele é algo fácil e barato
- Normalmente vem em um pacote-padrão de 40 pinos, embora haja outros pacotes no mercado para uso especial
- Tem 16 linhas de endereços, portanto pode endereçar 64 KB de memória
- O barramento de dados tem 8 bits de largura (transferência de dados entre CPU e memória são feitos um byte por vez)

- Tem uma variedade de linhas de controle
- 32 linhas de E/S são organizadas em 4 grupos de 8 bits cada, como 4 portas. Cada uma dessas linhas é bidirecional e pode ser lida ou escrita sob controle do programa. Esse modo primário de interação com o mundo externo é o que o tona tão valioso: um único chip basta para as funções de CPU, memória e E/S
- Cada uma dessas linhas de E/S, pode ser ligada a uma tecla, comutador, LED ou outro dispositivo

Pinagem

Vem com 4KB de ROM interna (64KB de memória externa podem ser conectados por um

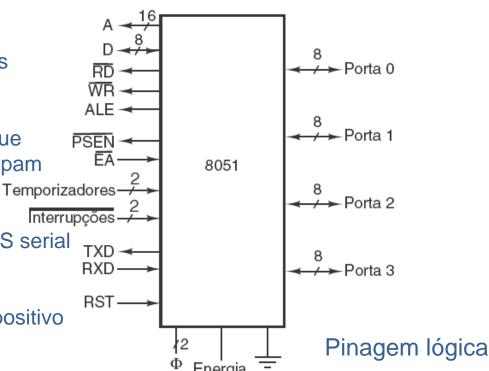
barramento)



Pinagem física

- Pinagem
- Os 7 primeiros sinais do lado esquerdo são usados para fazer interface com memórias externas, se estiverem presentes
- As duas linhas de temporizador permite que temporizadores externos sejam entradas para a CPU
- As duas linhas de interrupção permitem que dois dispositivos externos diferentes interrompam a CPU

 Tempo
- As linhas TXD e RXD são para permitir E/S serial para um terminal modem
- A linha RST permite que o usuário ou dispositivo de hadware externo reajuste o 8051



3.6 Exemplos de barramentos

- Barramento ISA (uma ligeira expansão do barramento IBM PC original)
- Barramento PCI
- PCI express
- Barramento serial universal (USB)

- Barramento IBM PC
 - Tinha 62 linhas de sinais, incluindo 20 para endereços de memória, 8 para dados e uma linha de cada para ativar leitura de memória, escrita de memória, leitura de E/S e escrita de E/S
 - Sinais para requisitar e conceder interrupções e usar DMA
 - Em termos físicos: era gravado na placa mãe do PC com cerca de meia dúzia de conectores espaçado de 2cm nos podiam ser inseridas placas. Cada placa tinha um tinha um rebordo compatível com o conector. Esse rebordo tinha 31 faixas banhadas a ouro de cada lado, que faziam contato elétrico com o conector.

- Problema
 - Lançamento do PC/AT
 - Partir do zero e projetar um barramento de 16bits inteiramente novo?
 - Ou, insistir no barramento PC e suas 20 linhas de endereço e 8 linhas de dados, deixando de aproveitar a capacidade do 80286 de acessar 16MB de memória e transferir palavras de 16 bits?
- Solução
 - Ampliar o barramento PC
 - As placas de encaixe do PC tem um conector de borda com 62 contatos que n\u00e3o ocupa todo o comprimento da borda
 - A solução PC/AT foi acrescentar um segundo conector de borda na parte inferior da placa, adjacente ao principal, e projetar o circuito AT para trabalhar com os dois tipos de placa

- Segundo conector no barramento PC/AT
 - Contém 36 linhas, sendo 31 para mais linhas de endereço, de dados, de interrupção e mais canais DMA, bem como energia e aterramento
 - Restante: trata das diferenças entre transferência de 8 bits e 16 bits

- Série PS/2
 - IBM decide que é hora de começar de novo
 - Motivos: decisão técnica (PCI já era realmente obsoleto)
 - desejo de colocar um obstáculo no caminho das empresas que fabricam clones do PC que tinham conquistado uma parte desconfortavelmente grande do mercado
- Máquinas PS/2 de faixa média e superior foram equipadas com um barramento completamente novo e protegido por uma muralha de patentes e advogados

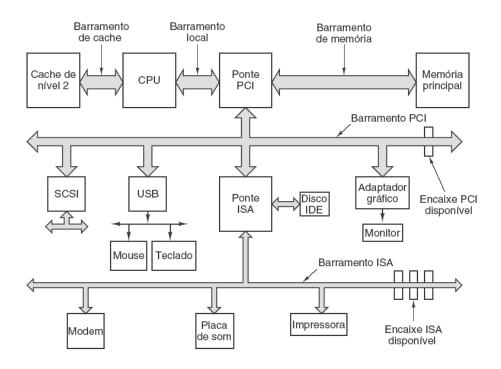
- Industria de computadores pessoais reagem
- Adotam seu próprio padrão ISA (Industry Stand Arquiteture arquitetura padrão da indústria)
- Era basicamente o barramento PC/AT, mas funcionando em 8,33 MHz
- Vantagem de conservar compatibilidade com máquinas e placas existentes
- Era baseado em um barramento que a IBM tinha licenciado literalmente para muitas empresas

 Mais tarde o barramento ISA foi ampliado para 32 bits com mais algumas características. Foi denominado: EISA (Extended ISA – ISA estendido)

- IBM original maioria das aplicações baseada em textos
- Gradualmente interfaces gráficas começaram a ser usadas
- As taxas máximas de funcionamento do barramento ISA, e também do EISA nem chegavam perto do que era necessário para apresentação de vídeo em tela
- Em 1990, a Intel percebe o que estava para acontecer, e desenvolve um novo barramento com uma largura de banda muito mais alta do que a do próprio barramento EISA: Barramento PCI

- Barramento PCI (Peripheral Component Interconnect Bus barramento de interconexão de componente periférico)
- PCI original
 - Transferia 32 bits por ciclo e funcionava em 33MHz (tempo de ciclo de 30 nanossegundos) para uma largura de banda total de 133 MB/s
- Em 1993 foi lançado o PCI 2.0, em 1995 o PCI 2.1
- PCI 2.2
 - Tem características para computadores portáteis (principalmente para economizar energia da bateria)
 - Funciona em até 66 MHz e pode manipular transferências de 64 bits para largura de banda total de 528 MB/s

- Problemas
 - Não era muito bom para um barramento de memória
 - Não era compatível com todas antigas placas ISA existentes
- Solução
 - Projetar computadores com três ou mais barramentos



- Ponte PCI
 - Conecta a CPU, a memória e o barramento PCI
- Ponte ISA
 - Conecta o barramento PCI ao barramento ISA e também suporta um ou dois discos IDE
- Existe mais de um tipo de placa PCI: á opções para tensão, largura e temporização
- As placas tem versões de 32 bits e 64 bits
 - 32 bits tem 120 pinos
 - 64 bits tem os mesmo 120 pinos mais 64 pinos adicionais

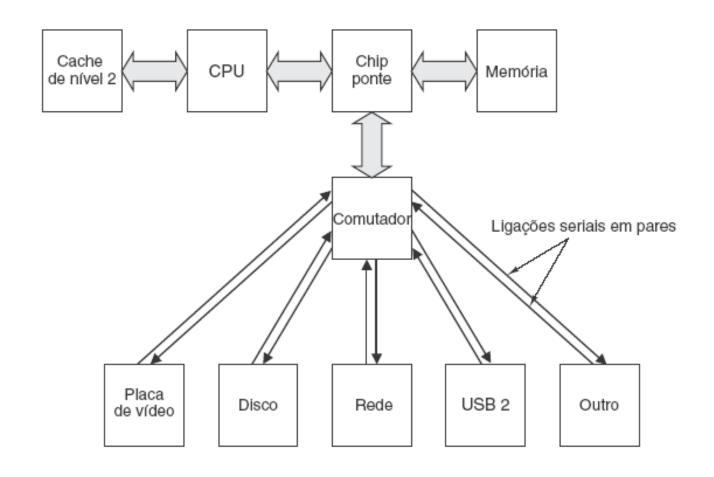
- Barramento AGP (Accelerated Graphics Port Bus barramento de porta gráfica acelerada)
- No final da década de 1990 temos um aumento:
 - Resoluções de monitores
 - Demanda por vídeo em tela inteira e movimento total (alto grau de interação, no contexto de jogos)
- Intel acrescenta mais um outro barramento: só para comandar a placa gráfica
- AGP 1.0
 - Funcionava a 264 MB/s, o que foi definido como 1x
 - Mais lento que PCI, mas dedicado a comandar a placa gráfica
 - Com o passar dos anos surgiram novas versões; AGP 3.0 funciona a 2,1 GB/s (8x)

- A necessidade de maior largura de banda de E/S causou uma confusão na antes limpa arquitetura interna do PC
- Barramento PCI não é mais o elemento central que mantém unidas as partes do PC. O chip ponte se apossou de parte desse papel
- Problema: há cada vez mais dispositivos de E/S muito rápidos para o barramento PCI
- Toda vez que um dispositivo de E/S fica muito rápido para o barramento PCI, a Intel acrescenta uma porta especial para o chip ponte para permitir que o dispositivo desvie do barramento PCI

- Problema: placas são muito grande, não cabem em laptops
- Solução: PCI Express
 - Tem pouco haver com o barramento PCI
 - Na verdade nem é um barramento!

Arquitetura do PCI express

- Coração da solução é: se livrar do barramento paralelo com seus muitos mestres e escravos e passar para o projeto baseado em conexões seriais ponto a ponto de alta velocidade
- Resumo: um PC é um conjunto de chips de CPU, memória e controladores de E/S que precisa ser interconectados
- PCI express: fornece um comutador de uso geral para conectar chips usando ligações seriais



Arquitetura do PCI express

- Três pontos de diferença

 - ✓ Utilizações de conexões seriais ponto a ponto estreitas
 X barramento paralelo largo
 - ✓ Modelo conceitual: um mestre de transferência de dados que emite um comando a um escravo para ler uma palavra ou um bloco de palavras x um dispositivo que envia um pacote de dados a um outro dispositivo.