ARQUITETURA BÁSICA DE UM MICROCOMPUTADOR

O computador é um processador de informações. As informações fornecidas pelo usuário são processadas e o resultado desse processamento é retornado ao usuário, normalmente através do monitor de vídeo ou da impressora. O "PROGRAMA" é que define o que o computador vai fazer exatamente com os dados que são fornecidos pelo usuário. Existem programas que efetuam uma simples soma entre dois números, assim como existem programas que são utilizados para projeto e simulação de funcionamento de carros, aviões, naves espaciais etc.

O matemático inglês Charles Babbage apresentou em 1833, o projeto de sua "MÁQUINA ANALÍTICA", que foi considerado por muitos o precursor dos atuais computadores eletrônicos. A máquina analítica foi projetada para ser capaz de realizar operações matemáticas a partir de programas que ficavam armazenados em cartões perfurados. Era um projeto inteiramente mecânico, baseado em engrenagens de diferentes tamanhos.

Em 1914, Jhon Von Neumann deu o passo decisivo para a construção de um computador eletrônico, valendo dos grandes avanços que a eletrônica realizava através da válvula. Jhon propôs a idéia de "PROGRAMA INTERNO", estabelecendo os fundamentos para que em 1946, entrasse em funcionamento o primeiro computador eletrônico, o ENIAC, desenvolvido na Universidade da Pensilvânia, nos Estados Unidos.

A partir de então, a tecnologia utilizada no desenvolvimento de computadores evolui cada vez mais, possibilitando a construção de computadores cada vez menores e com maior capacidade de processamento.

CPU – é quem gerencia todo o sistema e executa os programas. A CPU normalmente é chamada de processador ou microprocessador. Existem muitos tipos de processadores e o seu nome é utilizado para identificar o próprio computador. Um computador Pentium tem este nome porque ele utiliza como CPU, um Processador Pentium.

MEMÓRIA – Armazena os dados que estão sendo ou que serão processados, assim como os programas que são executados pela CPU.

DISPOSITIVOS DE I/O (Entrada/Saída) – é por onde os dados entram e saem do computador.

Exemplos:
TECLADO (entrada)
MONITOR (saída)
MOUSE (entrada)
MODEM (entrada e saída)

LOCALIDADES DE MEMÓRIA

A memória é utilizada para armazenar informações (dados ou programas). É composta por várias localidades onde são armazenadas as informações. Cada localidade possui um endereço e é capaz

armazenar o equivalente a um caractere. Serão usadas então seis localidades para armazenar o nome CARLOS.

Fisicamente falando, cada localidade de memória é capaz de armazenar um conjunto de 8bits (1BYTE). Qualquer informação, então, para estar armazenada numa localidade de memória, deve estar codificada no sistema binário. O sistema binário é o sistema de numeração utilizado pelos computadores, ele é composto por dois algarismos 0 e 1. Praticamente, qualquer informação pode ser representada no sistema binário, números , letras, palavras, imagens etc., onde serão utilizados apenas os algarismos 0 e 1.

Exemplo:

A – 01000001(2)

Na localidade onde está armazenado a letra "A", existe na verdade um conjunto de bits que representam a letra "A". Isto é necessário porque a CPU só processa informações ou executa programas que estejam codificados no sistema binário.

VIAS DE DADOS, ENDEREÇO E CONTROLE

Para gerenciar o sistema, a CPU precisa se comunicar com a memória, dispositivos de I/O. Esta comunicação será viabilizada a partir das VIAS (Barramentos) que ligam a CPU à memória, dispositivos de I/O e DMA. A CPU pode se comunicar com a memória ou dispositivo de I/O de duas formas:

ESCRITA – é quando a CPU envia uma informação para ser armazenada numa localidade de memória ou para um dispositivo de I/O. Por exemplo, quando a CPU envia uma informação para o vídeo mostrar na tela, dizemos que a CPU escreveu no vídeo.

LEITURA – é quando a CPU busca uma informação na memória ou num dispositivo de I/O. Quando uma tecla é pressionada no teclado e aparece no vídeo, é por que a CPU efetuou uma leitura no teclado e escreveu o dado lido no vídeo.

VIA DE ENDEREÇOS – é utilizada quando a CPU quer se comunicar com a memória ou um dispositivo de I/O. Cada localidade de memória, assim como cada dispositivo de I/O possui um endereço, que é um valor numérico representado por nós no sistema hexadecimal.

Quando a CPU vai efetuar uma leitura no teclado, ela coloca na via de endereços o valor 060H que é o endereço do teclado. O endereço chegará a todos que estiverem ligados a via de endereços, mas só será identificado pelo teclado. Da mesma forma, quando a CPU vai efetuar uma escrita numa localidade de memória, ela colocará na via de endereços o endereço da mesma. A CPU reconhece cada um dos dispositivos de I/O, assim como cada localidade de memória, pelo seu respectivo endereço.

VIA DE CONTROLE - é por ela que a CPU envia e recebe os sinais de controle. Estes sinais informam ao sistema, qual o tipo de comunicação será efetuada e como será efetuada. Para fazer uma leitura no teclado, a CPU coloca na via de endereços o endereço do mesmo e pela via de controle, mandará sinais que irão informar ao teclado que a comunicação será de leitura. O mesmo vale para comunicação de escrita.

VIA DE DADOS – é por onde trafegam os dados. É por ela que os dados entram e saem da CPU e também do sistema. Quando a CPU, por exemplo, efetua uma escrita na memória, o dado sai da CPU e chega até a memória pela via de dados.

BARRAMENTOS

De uma Maneira bem simples, poderíamos dizer que um barramento é uma via de comunicação. Em um micro, temos vários barramentos. Os motivos são muitos, a começar pelos aspectos históricos do próprio PC.

O principal barramento do micro, é o barramento local, a via de comunicação que conecta o processador aos circuitos primordiais da placa-mãe: a memória RAM, a memória cache e o chipset. Damos o nome de chipset ao conjunto de circuitos integrados de apoio existentes na placa-mãe. Dentre os diversos circuitos presentes no chipset, destacamos o controlador de memória (RAM) e o controlador de cache além de controlador de barramento, no caso de existirem outros barramentos no micro o que certamente ocorre.

O barramento local é o mais rápido, pois os circuitos se comunicarão com o processador com seu desempenho máximo. Entretanto, o barramento não é padronizado: cada processador deverá utilizar o seu próprio modelo, de acordo com as suas características. Aliás, é por esse motivo que cada processador necessita de um modelo de placa-mãe diferente.

Para que uma simples placa de vídeo ou um disco rígido possam ser utilizados em qualquer micro, independentemente do processador instalado (ou seja, independentemente do modelo de barramento local empregado), utilizamos diversos modelos de barramentos de expansão. Dentre eles, podemos destacar:

- ISA (Industry) Standard Architecture)
- EISA (Extended Industry Standard Architecture)
- VLB (VESA local Bus)
- PCI (Peripheral Component Interconnect)
- AGP (Accelerated Graphics Port)
- USB (Universal Seial Bus)
- Firewire (também, chamado IEEE1394)
- IrDA (Infrared Develops Association)

Todos esses modelos de barramentos são disponibilizados na placa-mãe do micro, através de conectores, chamados slots. A exceção fica por conta dos barramentos USB, Firewire, IrDA, que são externos ao micro, como veremos.

O maior problema em relação aos barramentos de expansão é a velocidade. Embora a maioria dos periféricos utilizados no micro seja lenta (como a impressora, o mouse, a unidade de disquete, a unidade de CD-ROM e o teclado), há basicamente três classes de periféricos bastante prejudicadas pela baixa velocidade:

- Vídeo.
- Discos Rígidos.
- Interfaces para rede local.

A princípio, pode parecer que, se o micro tiver esses periféricos integrados à própria placa-mãe, eles trabalharão à mesma velocidade da placa-mãe (isto é, à mesma velocidade do barramento local). Porém, isso não é verdade. No projeto de placa-mãe integradas, como em micros de arquitetura fechada, normalmente os periféricos integrados comunicam-se com o processador através de uma extensão do barramento de expansão, chamado de barramento X (eXtension bus) – ou seja, mesmo o periférico estando integrado na placa-mãe, ele é tratado como se estivesse conectado a um dos slots de expansão.

BARRAMENTO LOCAL

Como o barramento local é utilizado na comunicação do processador com os circuitos básicos e que demandam velocidade (especialmente a memória RAM e o cachê de memória), esse barramento é totalmente transparente ao usuário, mesmo quando temos em mãos uma placa-mãe. Simploriamente poderíamos dizer que o barramento local está na placa-mãe e nada mais.

Podemos dividir o barramento local (e a maioria dos barramentos de expansão) em três grupos:

- Barramento de dados
- Barramento de endereços
- Barramento de controle

Ao comentarmos do processador Pentium tem um barramento de dados de 64 bits, queríamos dizer que o barramento local terá o seu barramento de dados de 64 bits. Como conseqüência, o acesso à memória será feito a 64 bits por vez. Da mesma forma, quando dizemos que o processador trabalha externamente a 66 MHz, isso significa que é esta a freqüência de operação do barramento local.

BARRAMENTO ISA

Dos barramentos existentes é o mais antigo e já obsoleto, embora ainda utilizado. Sua freqüência é de 8,33 MHz. O barramento ISA (Industrie Standard Architeture) foi concebido originalmente com largura de oito bits, sendo posteriormente modificado para trabalharem16bits.

Ainda encontramos à venda placas de som, rede e modems padrão ISA, todos de 16 bits. Isto se explica pelo fato destas placas não transferirem tão elevado número de bits/s (bits por segundo) quanto uma placa de vídeo, por exemplo, necessita atualmente.

Sua taxa de transferência é de oito bits/s, para os barramentos de oito bits de largura e 16 bits/s para o barramento ISA com largura de 16 bits. Como diferenciar uma placa ISA 8 bits de outra de 16 bits? Basta observar seus contatos. As placas de 8 bits têm apenas uma linha de contatos, enquanto as placas ISA de 16 bits têm duas linhas separadas por um chanfro que permite o encaixe correto nos slots ISA fabricados para aceitar tanto placas de 8 quanto de 16 bits. Se você observar um slot ISA numa placa-mãe de um PC padrão XT (Extended Technology) irá observar que ele é menor, aceitando somente placas ISA de 8 bits.

BARRAMENTO EISA x BARRAMENTO MCA (IBM)

A utilização de um barramento proprietário foi um balde de água fria nos concorrentes da IBM. Outros fabricantes que quisessem construir microcomputadores compatíveis teriam de formular seu próprio padrão. E como não havia qualquer tipo de padronização e muito menos entendimento entre os diversos fabricantes independentes, alguma coisa a respeito devia ser feita.

A empresa Compaq foi uma das primeiras empresas a levantar sua bandeira em defesa da arquitetura aberta e compatibilidade e padronização entre os diversos fabricantes, liderando um grupo formado pelos nove maiores fabricantes de micros "de marca" do mundo na época na tentativa de se criar um novo padrão de barramento de expansão mais rápido e com arquitetura aberta ou seja, quem quisesse utilizá-lo em seu projeto.

O novo padrão criado pelos fabricantes chamava-se EISA, totalmente compatível com o antigo ISA. O barramento EISA tem as seguintes características:

- Barramento de dados de 32 bits
- Barramento de endereços de 32 bits
- Freqüência de operação de 8 MHz.

O slot EISA é muito parecido com o slot ISA, pois ambos têm o mesmo tamanho. No slot EISA, as linhas adicionais de dados, controles e endereços que não existiam no ISA foram colocadas entre os contatos convencionais, fazendo com que o slot EISA fosse compatível tanto com interfaces ISA quanto EISA.

Quando inserimos uma interface ISA em um slot EISA, o percurso dela é limitado por travas no conector, impedindo que ela faça contato com sinais EISA. Já quando inserimos uma interface EISA, essa limitação não ocorre e os seus contatos ajustam-se plenamente com todos os sinais EISA.

Mas havia um problema: para manter total compatibilidade com o barramento ISA, o barramento EISA teve de usar a mesma freqüência de sinais do barramento ISA. Mesmo tendo a capacidade de trabalhar com dados de 32 bits, endereçar até 4 GB de memória e ser uma arquitetura aberta (além de ter modos de transferência de dados ausentes no barramento ISA), o EISA não se tornou tão popular, pois ainda apresentava um gargalo para interfaces que exigiam alto desempenho.

BARRAMENTO VLB

As arquiteturas ISA e EISA tinham muitas diferenças, mas a forma como a memória e a CPU eram acessadas era praticamente a mesma. Em sistemas que experimentam níveis muito altos de tráfego em seus barramentos, a latência envolvida em operações de E/S e processamento de dados pode se tornar muito pronunciada, podendo ocorrer "time out". Essas situações podem ocorrer com adaptadores de rede ou controladoras SCSI, podendo haver perda ou corrupção dos dados. Percebendo-se isso, alguns fabricantes apresentaram soluções bem originais para este problema, simplesmente colocando na própria placa de CPU todos os circuitos da placa de vídeo. A comunicação entre a CPU e a memória de vídeo podia ser feita de forma direta, sem encontrar pelo caminho o lento barramento ISA. Esta técnica era chamada de Local Bus, e resultava em um considerável aumento de desempenho.

No final de 1992 criou-se a VESA "A" Local Bus (Video Eletronics Standards Association) para desviar o tráfego mais intenso, como vídeo, com um barramento local conectado diretamente ao barramento da CPU. Este barramento foi criado tendo em vista aumentar a velocidade de transferência de dados entre a placa de CPU e a placa SVGA, mas outras placas de expansão também poderiam utilizá-lo. Desta forma, o barramento de dados não teria problemas com um tráfego tolerável entre os dispositivos periféricos. Não necessitava de chips especiais como era o caso do EISA, era uma arquitetura aberta ao contrário do MCA, e tratava-se de um padrão industrial, uma grande vantagem sobre os barramentos proprietários. É fisicamente representada por um conector especial de expansão.

Havia um problema inerente à conexão de dispositivos diretamente ao barramento da CPU: a interface entre o barramento da CPU e o barramento VESA era dependente da CPU. Isso implicava em que, ao fazer o upgrade da CPU, havia necessidade de fazer o upgrade da placa VESA em conjunto.

Com a evolução dos processadores, o padrão VESA "A" foi se tornando relativamente cada vez mais lento. Para resolver o problema de latência do barramento inerente a essa situação, foi criado um virtual local bus e conectado ao barramento da CPU via buffer. Essa solução ficou conhecida como VESA "B".

Os buffers permitiam que os sinais fossem armazenados por breves períodos, enquanto o barramento estivesse ocupado. Mesmo assim, havia outros problemas: o que aconteceria quando mais de um dispositivo necessitava utilizar o barramento da CPU ao mesmo tempo? Aconteceria uma arbitragem e apenas um dispositivo utilizaria o barramento enquanto os demais aguardariam a sua liberação. Com isso, poderia ocorrer uma latência e o sistema poderia operar ineficientemente barramento PCI (Peripheral Component Interconnect).

BARRAMENTO PCI

A arquitetura PCI é semelhante à VESA quanto à conexão ao local bus da CPU, mas é muito mais elegante e completa ao utilizar uma ponte PCI-HOST, um dispositivo de cache que provê uma única interface entre a CPU, memória e o PCI local bus. A arquitetura PCI permite que a CPU continue a buscar informação do cache enquanto o controlador de cache possibilita a um dispositivo de expansão acessar a memória do sistema, ou seja, operações concorrentes no mesmo barramento.

Outra grande vantagem do barramento PCI é que até 256 dispositivos podem ser atacados a um único PCI local bus, e 256 barramentos PCI podem existir em um único sistema.

BUS MATERING

A técnica de Bus Mastering introduzida com o barramento PCI consiste em liberar o processador da tarefa de controlar diretamente todas as transferências de dados entre os dispositivos. O periférico que utilizar o Bus Mastering pode utilizar o barramento PCI da forma mais racional possível, sem

que o processador precise interferir no processo diretamente. Isto irá conferir ao sistema maior rapidez, pois o processador fica disponível para outras tarefas.

Quando for necessário escolher entre uma placa de som, rede ou modem padrão ISA ou PCI, lembre-se do recurso de Bus Mastering que irá permitir sensível melhoria no desempenho do sistema. Por isso o barramento PCI é ideal para placas de som com efeitos em 3D. Uma placa de som sem este recurso pode ser ISA. No caso das placas de som, rede e modem deve-se preferir o padrão PCI também por que a tendência é que as placas-mãe deixem de oferecer slots ISA, como já vem ocorrendo. Em caso de upgrade o usuário correria o risco de não poder aproveitar sua placa padrão ISA. Placas de vídeo exigem atualmente pelo menos barramento PCI.

TECNOLOGIA PLUG & PLAY

Mas havia outras inovações como a tecnologia Plug and Play (também identificada como PnP e P&P). Esta tecnologia permite ao BIOS identificar os dispositivos instalados na placa-mãe por intermédio de uma BIOS existente em cada dispositivo Plug and Play, o que não ocorria antes quando era necessário configurar os dispositivos manualmente pelo Setup. Para que tudo funcione é necessário que o Sistema Operacional seja Plug and Play o que ocorreu a partir do Windows 95, quando ao instalar uma placa de modem, por exemplo, esta é reconhecida durante a inicialização do Sistema Operacional. Neste momento basta informar ao S.O. onde está(ão) o(s) driver(s) da placa a ser instalada.

BARRAMENTO AGP

Este barramento (Acelerator Graphics Port) foi projetado pela Intel, novamente para valorizar seus processadores que possuem um co-processador aritmético bastante evoluído em relação à concorrência.

Foi projetado somente para placas de vídeo. Sua freqüência base é de 66 MHz combinada com a largura do barramento de 32 bits. Assim sua taxa de transferência é de 264 MB/s que pode aumentar com o recurso da multiplicação do clock, ou seja, uma placa AGP pode ser Modo 1x (66 MHz), Modo 2x 132 MHz) e Modo 4x (264 MHz).

Repare que as taxas de transferências aumentam significativamente chegando a 1056 MB/s no modo 4x. Com o grande avanço dos processadores e dos jogos para computador, as placas de vídeo padrão AGP estão se tornando bastante requisitadas pelos usuários.

BARRAMENTO USB

O USB é a tentativa de criar um novo padrão para a conexão de periféricos externos. Suas principais armas são a facilidade de uso e a possibilidade de se conectar vários periféricos em uma única porta USB.

Apesar do "boom" ainda não ter acontecido, já existem no mercado vários periféricos USB, que vão de mouses e teclados à placas de rede, passando por scanners, impressoras, Zip drives, modems, câmeras de videoconferência e muitos outros.

Podemos conectar até 127 periféricos a uma única saída USB em fila, ou seja, conectando o primeiro periférico à saída USB da placa mãe e conectando os demais a ele.

A saída USB do micro é o nó raiz do barramento. A este nó principal podemos conectar outros nós chamados de hubs. Um hub nada mais é do que um benjamim que disponibiliza mais encaixes, sendo 7 o limite por hub. O hub possui permissão para fornecer mais níveis de conexões, o que permite conectar mais hubs ao primeiro, até alcançar o limite de 127 periféricos permitidos pela porta USB. A idéia é que periféricos maiores, como monitores e impressoras possam servir como hubs, disponibilizando várias saídas cada um. Os "monitores USB" nada mais são do que monitores comuns com um hub USB integrado.

BARRAMENTO FIREWIRE

A idéia do barramento Firewire é bastante parecida com a do USB. A grande diferença é o seu foco. Enquanto o USB é voltado para periféricos normais que todo PC apresenta externamente, o Firewire vai mais além: pretende simplesmente substituir o padrão SCSI.

Dentre os periféricos-alvo do Firewire, encontram-se, além dos "comuns" câmeras de vídeo, scanners de mesa, videocassete, fitas DAT, etc.

O Firewire apresenta as demais idéias e características do barramento USB. Podemos conectar até 63 periféricos ao barramento Firewire. Seu cabo poderá até 4,5 m em cada trecho (ou seja, entre dois periféricos).

O Windows 98 reconhece o barramento Firewire. Em outros sistemas operacionais (inclusive no Windows 95), é necessário que seja instalado um driver apropriado para que esse barramento possa ser acessado.

O barramento Firewire utiliza a especificação IEEE 1394 (IEEE é o Instituto de Engenheiros Eletricistas e Eletrônicos dos Estados Unidos). Como americanos simplesmente adoram siglas e acrônimos, muito provavelmente o Firewire será chamado de IEEE 1394 por aí.

BARRAMENTO IrDA

O IrDA é um barramento sem fios: a comunicação é feita através de luz infravermelha, na mesma forma que ocorre na comunicação do controle remoto da televisão. Você pode ter 126 periféricos IrDA; "conversando" com uma mesma porta. É muito comum notebooks com uma porta IrDA ; podemos, assim, transferir arquivos de um notebook para outro (ou mesmo para um micro desktop) sem a necessidade de cabos ou imprimir em uma impressora com porta IrDA sem a necessidade de cabos.

O barramento IrDA pode se utilizado para conectar vários tipos de periféricos sem fio ao micro, tais como teclado, mouse e impressora. O barramento pode estar conectado diretamente à placa-mãe do micro ou então disponível através de um adaptador IrDA conectado à porta serial do micro.

Existem dois tipos padrões IrDa:

- IrDA 1.0: Comunicações a até 115.200 bps.
- IrDA 1.1: Comunicações a até 4.194.304 bps (4 Mbps).

Tanto o Windows 98 quanto o Windows 95 OSR2 reconhecem automaticamente portas IrDA instaladas no micro. No caso das primeiras versões do Windows 95, é preciso instalar um driver.

No caso de placas-mãe com porta IrDA, deverá habilitá-la no setup do micro. Você pode configurar seu funcionamento em dois modos:

- Full-duplex: em que os periféricos podem trocar daos simultaneamente.
- Hall-duplex: Em que somente um periférico pode transmitir dados por vez.

INTRODUÇÃO DMA

O DMA (Direct Memory Access - acesso direto a memória) é um controlador existente integrado na placa-mãe desde a época do primeiro PC. Ele permite que periféricos façam transferências de dados para a memória RAM sem a intervenção do microprocessador. Isto economiza um tempo enorme.

O DMA é usado em transferências de grupos de dados em casos nos quais o microprocessador não pode ser sobrecarregado. Em uma transferência de DMA, seqüências de dados presentes em uma área de memória são enviados diretamente para um dispositivo de E/S, sem intervenção direta da CPU. Quem faz o trabalho de transferência é o circuito controlador de DMA (antigamente era o chip 8237A, agora está embutido no chipset). Desta forma, em um jogo o microprocessador pode ficar dedicado a processar o teclado e movimentar gráficos na tela, ao mesmo tempo em que o controlador de DMA envia para a placa de som, seqüências de áudio digitalizado, em intervalos de tempo constante. O microprocessador só precisa indicar ao controlador de DMA, qual é o número de bytes a serem transferidos, qual é o endereço inicial de memória, e qual é o intervalo de tempo entre bytes (ou words) consecutivos. Ao terminar a transferência, o controlador de DMA avisa à placa de som, que por sua vez interrompe a CPU para indicar que o trecho já foi transferido.

Nem todos os dispositivos de I\O utilizam o DMA para transferência de dados para a memória, apenas aqueles que manipulam grandes quantidades de informações e precisam de velocidade na transferência. Este é o caso dos periféricos (disps. de I\O): scanner, drive de disquete, placa de som, winchester, entre outros.

DMA - O DMA é um microprocessador com uma única finalidade, transferir dados de um dispositivo de I\O para a memória sem a interferência da CPU, como já havíamos dito anteriormente. Quando o programa solicita ao dispositivo a transferência de dados (quando vamos por exemplo, scannear uma foto), o scanner envia ao controlador de DMA, um sinal solicitando a transferência. O controlador de DMA recebe o pedido e envia um sinal pela via de controle à CPU, pedindo autorização para assumir o controle dos barramentos. A CPU entrega o controle dos barramentos ao DMA, que se encarrega de fazer a transferência dos dados do scanner para a memória (scannear a foto). Ao término da transferência, o DMA devolve o controle dos barramentos à CPU. Portanto, a CPU gerencia todo o sistema, exceto o DMA. Um computador pode funcionar sem o DMA, mas

praticamente todos os computadores comerciais fazem uso do DMA. Por isso este bloco foi incluído na arquitetura básica do microcomputador.

Vamos dar um exemplo simples. Imagine um arquivo de 50 KB gravado em disquete. Se não existisse o recurso de DMA, a transferência seria feita byte-a-byte, ou seja, seriam necessárias mais de 50.000 instruções por parte do processador para que esta transferência fosse executada.

No mundo real, porém, a transferência seria controlada pelo controlador de DMA e com um detalhe importantíssimo: o processador não interage no processo, ficando disponível para executar outra tarefa. Bastaria uma única instrução para o controlador de DMA iniciar o processo.

CONCLUSÃO DMA

Quando a CPU vai efetuar uma escrita, ela primeiramente coloca na via de endereços o endereço do dispositivo ou da localidade de memória com o qual irá se comunicar. Todos que estiverem conectados a via de endereços irão receber o endereço, mas o mesmo só será reconhecido pelo dispositivo em questão. Depois a CPU envia pela via de controle, sinais que irão informar ao dispositivo que tipo de comunicação será efetuada, neste caso de escrita. Após receber os sinais de controle, o dispositivo já sabe que a CPU quer falar com ele e o que. O dispositivo então se prepara para a transferência de dados. Quando o dispositivo está pronto para a transferência, a CPU envia o dado ao dispositivo pela via de dados. O mesmo acontece para a leitura, primeiro a CPU coloca na via de endereços o endereço do dispositivo, depois ela envia sinais de controle informando ao dispositivo que a comunicação será de leitura e quando o dispositivo estiver pronto, a CPU busca o dado pela via de dados.

Ultra DMA / Ultra ATA :

O uso desta tecnologia permite que se alcance performances superiores que os sistemas convencionais.

Ultra DMA-33:

O modo Ultra DMA, também chamado de Ultra-ATA, permite a transferência de dados entre o HD e sua interface na velocidade de 33.3MB por segundo.

Como comparação, o uso da interface convencional, a velocidade de transferência é de 16.6MB por segundo.

Para o uso desta tecnologia os seguintes parâmetros são requeridos:

- Sistema Operacional compatível com a tecnologia Ultra DMA
- Bios da placa CPU compatível com a tecnologia Ultra DMA
- Disco rígido compatível com Ultra DMA

Sistemas Operacionais x Suporte Ultra DMA:

- DOS: não suporta Ultra DMA
- Windows 95: suporta se realizada sua atualização.O driver de software existente suporta apenas o modo PIO. Para suporte Ultra DMA, é necessária a atualização do driver "Intel Bus Master DMA" .

- Windows 95 OSR2: suporta se realizada sua atualização. O driver de software existente suporta apenas o modo DMA, não suportando o CRC. Para suporte Ultra DMA, é necessária a atualização do driver "Intel Bus Master DMA".
- Windows 98: suporta Ultra DMA em todas suas funções.

Benefícios da Tecnologia Ultra DMA:

- Permite a utilização dos discos tipo Ultra DMA em placas de CPU que não suportam este modo. Neste caso, apenas os benefícios do Ultra DMA serão perdidos.
- Aumento da performance do sistema dobrando a velocidade de transferência de dados entre o sistema / HD. Permite também a fabricação de discos de alta capacidade de armazenamento que podem ser conectados nas interfaces das placas de CPU atuais.
- Ultra DMA incorpora mecanismo de correção de erro tipo CRC (cyclical redundancy checking) aumentando a confiabilidade do sistema e garantindo a integridade dos dados armazenados . Se uma condição de erro é verificada , a operação de transferência de dados é repetida de modo a assegurar sua integridade. NoUltra DMA, o sistema CRC protege os dados lidos e escritos.
- A tecnologia Ultra DMA é 100% compatível com a Fast ATA-2 (EIDE) e IDE utilizadas atualmente. Isto permite a ligação do disco Ultras DMA na interface atuais da placa de CPU, com o mesmo cabo são de 40 pinos. Entretanto, para o aproveitamento dos benefícios da tecnologia Ultra DMA, o sistema deverá ser também compatível com Ultra DMA conforme descrevemos acima.
- Suporte a múltiplos drives conectados a um mesmo cabo de sinal.

FIM DA PARTE I .....

Arquitetura Pipeline ou Arquitetura Superescalar.

O Pentium têm como se fossem dois processadores 486 trabalhando em paralelo em seu interior. Cada "processador" desses é chamado de canalização e essa arquitetura é conhecida por arquitetura pipeline. No caso do Pentium, a primeira canalização é chamada "U" e a segunda é chamada de "V". A idéia fundamental do processamento em pipeline é dividir um bloco de lógica combinatória em vários blocos separados por registros, a que se chamam andares do pipeline. Vamos dizer que temos um programa com duas instruções seguidas: A + B e C + D. Nesse caso, haveria a possibilidade de executá-los em qualquer ordem, pois o resultado de um não interferiria no outro. O Pentium é capaz de processar as 2 informações independentes, uma em cada canalização, obtendo o dobro de desempenho sobre qualquer microprocessador convencional.

O problema é quando, no entanto, ocorre quando temos um programa com instruções dependentes, como por exemplo, A+B e Resultado + C. Resultado+C não pode ser resolvido antes de A + B. Nesse caso, o processamento ficaria igual a de um microprocessador convencional. Nesse caso, a segunda canalização ficaria vazia. Programas convencionais para 80386/80486 funcionarão perfeitamente e muito mais rápido no Pentium, porém o Pentium "enxerga" apenas a próxima instrução do programa, deixando várias vezes a canalização "V" vazia, fazendo com que o processador seja subutilizado.

Isso indica que os programas devem ser lançados especificamente para arquitetura Pentium a fim de utilizar esse mecanismo de pipeline.

Podemos observar claramente que programas, como o Windows 9x, são compilados para "Pentium". Na verdade, o que o compilador faz é organizar as instruções do programa de modo que a canalização "V" do processador fique cheia a maior parte do tempo. Mas é claro que os programas modelados para Pentium funcionam para processadores mais antigos, pois isso trata-se apenas de um "re-arranjo" do programa.

Por isso podemos observar o ganho de desempenho do Windows 95 é muito maior do que em um micro com Windows 3x quando fazemos a troca de processador para um Pentium.

MEMÓRIAS RAM

A RAM (Random Access Memory – Memória de acesso aleatório) é a memória usada em alta escala e cada vez em maior quantidade nos computadores e tem como principais características:

- Permite a leitura e a gravação de dados, enquanto as ROMs só servem para leitura.
- É volátil, isto é, a memória RAM perde todos os seus dados assim que é desligada. Isto não é nenhum problema, pois quando o computador é ligado, o sistema operacional novamente é transferido do disco rígido para a RAM.

Existem vários tipos de RAM com diversas características e para diversas aplicações. A mais conhecida é a DRAM (dinâmica) e a SRAM (estática).

Conceito DRAM

A memória DRAM (Dynamic RAM) é a memória mais conhecida no computador. Muitas vezes, quando dizemos que o nosso computador tem 16 ou 32 MB de memória ou de RAM, na verdade estamos nos referindo à DRAM. A DRAM é uma memória relativamente rápida e que tem o objetivo de armazenar o maior volume de dados na troca dinâmica CPU-Memória.

DRAM é mais lenta que a SRAM. É comum hoje encontrar nas DRAMs 60ns de tempo de acesso, enquanto que nas SRAM é de 10 a 15ns. E, por ter uma qualidade superior, a SRAM é mais cara que a DRAM. Pode-se identificar a velocidade da memória observando o chip. Para uma memória de 60ns, os fabricantes geralmente colocam -6, 60, -60, 6, 06 para sua identificação." (observe a figura abaixo).

A constituição da SRAM e DRAM tem uma variação bem simples: a DRAM necessita de pulsos de 15ns para manter seu conteúdo, de forma que a energia não fique o tempo todo abastecendo os chip. Esse pulso periódico é o refresh. Cada bit da DRAM necessita de um transistor e de um capacitor.

Os chips de DRAM diferenciam nos seguintes aspectos:

- número de células na memória;
- tamanho de cada célula na memória;
- tempo de acesso;
- encapsulamento;
- O número de células é relacionado com a capacidade de armazenamento (ou posições de memória), existindo chips com 8KB a 16MB de células de memória. O tamanho das células de memória é o número de bits que cada célula armazena.

Existem chips de memória com 1, 4, 8, 9, 32 ou 36 bits. O tempo de acesso, como falamos anteriormente, é medido em bilionésimos de segundo, conhecidos como nano-segundos.

Tipos DRAM

FPM DRAM (Fast Page Mode DRAM)

Esta é a tecnologia com que os circuitos da memória RAM são tradicionalmente construídos. Este tipo de memória DRAM foi utilizada bastante durante os anos de 80, e também até aproximadamente 1995. A FPM existe em módulo de SIMM de 72 terminais e de 32 bits, utiliza uma matriz de

capacitores (lugar onde se armazenam dados internamente), e os dados são lidos ou armazenados por vez, freqüência de barramento é de 66 MHz, os dados são transferidos utilizando ciclos 5-3-3-3 (5 ciclos para a primeira leitura e três ciclos para cada uma dos três ciclos seguintes, sendo que cada ciclo dura 15ns em um Pentium com clock externo de 66 MHz), tempo de acesso 80ns, 70ns e 60ns Podem ser usadas em qualquer placa de CPU Pentium, mas já é considerada obsoleta para os processadores Pentium II.

EDO DRAM (Extended Data Out DRAM)

Criada em 1995, a EDO DRAM é obtida a partir de um melhoramento de engenharia nas memórias FPM DRAM. Isto foi possível graças a uma pequena modificação em sua estrutura interna, permitindo que o processador acesse um endereço de memória ao mesmo tempo em que ela ainda está entregando um dado pedido anteriormente. O resultado é uma economia de tempo, o que equivale a um aumento de velocidade. Podemos encontrá-lo em módulos SIMM de 72 e DIMM de 168 de 32 bits, utilizando uma matriz de capacitores, freqüência de barramento de 66 MHz, os dados são transferidos utilizando ciclos 5-2-2-2, tempo de acesso de 70, 60 ou 50ns. É suportada por todas as placas de CPU Pentium, a partir das que apresentam o chipset i430FX. Placas de CPU Pentium II também as suportam.

Desvantagens

As DRAMs utilizam complicados processos de multiplexação de linhas de endereçamento, que reduz a quantidade de vias elétricas do acesso, porém isso aumenta a complexidade dos circuitos de controle. É daí que surgem os termos RAS (Row Address Strobe) e CAS (Column Address Strobe), ou seja, sinalizadores de quando a matriz de memória está recebendo um apontamento de linha (row) ou coluna (column). Mesmo com tanta complexidade de implementação, acha-se que o custo ainda é compensador em relação às SRAMs.

Como opera a DRAM

Uma memória DRAM pode ser pensada como um arranjo de células, como uma tabela ou planilha. Estas células são feitas de capacitores e contém um ou mais bits de dados, a depender da configuração do chip. Esta tabela é endereçada através de decodificadores de linha e coluna, que por sua vez recebem seus sinais de geradores de clock, denominados geradores CAS (Column Address Strobe) e RAS (Row Address Strobe). De modo a minimizar o tamanho do pacote de dados, os endereços de linha e coluna são multiplexados em buffers. Por exemplo, se há 11 endereços, então haverá 11 linhas e 11 colunas de buffers. Transistores de acesso chamados "sense amps" ou amplificadores de sinal (numa tradução livre) são conectados a cada coluna de modo a possibilitar as operações de leitura e recuperação do chip. Uma vez que as células são capacitores que se descarregam para cada operação de leitura, o sense amp precisa recuperar ou restaurar o dado ali armazenado antes do fim do ciclo de acesso. Abaixo um gráfico ilustra uma célula de memória sendo acessada pela linha e coluna.

Os capacitores usados nas células de dados tendem a "perder" sua carga com o tempo, desta maneira requerem uma renovação constante e periódica dos dados, sob pena de estes se perderem. Este ciclo de renovação chama-se refresh cycle. Um controlador determina o tempo entre os ciclos de refresh, e um contador assegura que toda a matriz (todas as linhas) sofrem refresh. Logicamente isso significa que alguns ciclos da máquina são usados para a operação de refresh, e isto impacta na performance.

Um acesso típico de memória ocorreria da seguinte maneira. Primeiro, os bits da linha de endereço são colocados nos pinos de endereçamento. Após um período de tempo o sinal RAS cai (a voltagem diminui), o que ativa os sense amps e provoca o travamento da linha de endereço no buffer. Quando o sinal RAS se estabiliza, a linha selecionada é transferida para os sense amps ou transistores. Logo após, os bits da coluna de endereço são preparados, e então travados no buffer quando o sinal CAS cai, ao mesmo tempo em que o buffer de saída (output buffer) é ativado. Quando o sinal CAS se estabiliza, os transistores selecionados alimentam seus dados para o buffer de saída.

Encapsulamento

Até o final dos anos 80, a memória DRAM era feita com o encapsulamento DIP, que tinha que ser encaixada na placa-mãe. Logo depois surgiu o encapsulamento SIPP, que deu lugar, por sua vez, ao encapsulamento SIMM. Veja cada um dos chips abaixo:

Um chip de memória

Um módulo de memória SIPP

Um módulo de memória SIMM

O SIMM surgiu por volta de1992 e, até hoje, os chips de memória que compõem as placas adaptadoras são do tipo DIP (Dual In-Line Package).

Com o SIPP (Single In-Line Pin Package), surgiu o que é chamado módulos de memória, que eram vários chips de DRAM numa fileira de terminais que se encaixavam num soquete. Esse tipo de encapsulamento foi bastante usado até o início dos anos 90. Visualmente, pode ser uma mistura do que é o DIP e o SIMM.

Mas logo que o SIPP tornou-se popular, surgiu o SIMM (Single In-Line Memory Module), que é eletricamente igual ao SIPP, possuindo de diferente apenas a forma de seus contatos para afixação na placa-mãe. Podemos dizer que o SIPP possui perninhas e o SIMM, contatos na borda inferior.

Módulo de memória de 30 vias

Mais tarde, surgiram os módulos SIMM de memória de 72 vias, que são um pouco maiores do que os de 30, operando a 32 bits, que os últimos 486 fabricados usavam muito, também, algumas vezes, em conjunto com os de 30 vias. Esses módulos de memória de 72 vias podem ter até 32MB e um único módulo.

Módulo de memória de 72 vias

Em 1997 surgiram as memórias no encapsulamento DIMM (Dual In-Line Memory Module), que é uma módulo de memória com um encaixe igual ao do SIMM, mas que é de 168 pinos, praticamente o dobro do tamanho de um SIMM. Essa memória é de 64 bits. Assim, para um Pentium, basta um desses módulos de memória para funcionar.

Módulo de memória DIMM de 32MB

Armazenamento de dados

- O dispositivo/CPU envia os dados para a SRAM, que os absorve rapidamente.
- A SRAM envia os dados para a DRAM.
- Caso a DRAM não seja suficientemente grande para armazenar os dados, envia-os para o HD, que possui um espaço reservado para servir de memória temporária.
- A informação retorna, quando necessário, realizando o caminho inverso.

Cabe ressaltar que a SRAM somente serve como passagem rápida de dados. Ela agiliza o sistema, mas não armazena por grandes períodos, como a DRAM e o HD.

Nesse processo o HD é utilizado para auxiliar a memória DRAM.

SRAM e comparações

A memória cache é uma das grandes obras de engenharia dos PCs. Apesar de não ser essencial, a sua presença costuma balancear a morosidade das memórias DRAM (Dynamic Random Access Memory) ou mesmo SDRAM (Syncronous DRAM) frente à voracidade dos processadores.

Os circuitos de memória SRAM (Static RAM) são os constituintes da memória cache. Os primeiros tipos de memória utilizadas na indústria (quando ainda nem existiam os módulos SIMM de 30 vias) eram similares às SRAMs. Dadas as exigências cada vez maiores por quantidade de memória e o elevado custo das SRAM, ficou claro para a indústria e para os engenheiros que um tipo novo de memória deveria ser empregada, afinal os custos das SRAMs ficaria proibitivo para quantidades cada vez mais elevadas.

Alterando drasticamente a concepção das células de memória, surgiram os modelos DRAM, que, como a nomenclatura indica, são radicalmente diferentes dos SRAM. O produto era tão barato e também tão funcional, que a estrutura básica permanece até hoje.

Enquanto as DRAM foram evoluindo com novas idéias e aplicações práticas, sempre tendo-se em vista a manutenção de custos baixos em detrimento de performance, as SRAM mantiveram-se exatamente iguais ao que eram no princípio, porém, com a descoberta de novas técnicas em microeletrônica e as almejadas reduções das dimensões dos dispositivos, as SRAM puderam ganhar muita velocidade e até reduções de custo.

Em microeletrônica, dispositivos pequenos implicam em coisas boas e ruins. Entre as boas estão menores tempos de acesso, maior exigência de potência, maior quantidade de células elementares e obviamente menor custo relativo de produção. Entre as coisas ruins estão a maior suscetibilidade a ruídos, limitação de potência, efeitos eletromagnéticos entre circuitos internos até então inexistentes, limitações do emprego de certos materiais e processos de produção mais complexos.

Como as SRAMs são de 8 a 10 vezes mais rápidas do que as DRAMs, bolou-se uma arquitetura que utilize uma mínima quantidade de SRAM para tentar promover uma melhora na performance. Assim surgiu a memória cache.

A memória cache é uma pequena quantidade de SRAM que por meio de algoritmos refinados consegue manter boa parte dos dados requisitados pelo processador quase sempre em seus domínios, ou seja, dados da SDRAM ou DRAM são transferidos para suas células e daí o processador consulta simultaneamente o cache e a DRAM em busca dos dados. Obviamente, sempre que o cache possuir os dados, o processador o extrairá mais rapidamente dele. A preocupação é justamente com aquele quase sempre. Nenhum algoritmo pode prever com 100% de acurácia quais dados serão requisitados. Além disso, num primeiro momento, esses algoritmos nem tem idéia das regiões da memória que serão necessárias. Somando a essas limitações físicas do cache, ou seja, a quantidade de SRAM disponível e mais importante ainda, o alcance do cache.

O algoritmo, a quantidade de SRAM e o alcance são três fatores extensivamente pesquisados para que a memória cache continue sendo vantajosa. O alcance do cache indica qual a região ou quantidade de memória DRAM ficará na cobertura do cache, isto é, que região possui probabilidade não nula de ser encontrada no cache. Os resultados de testes mostram exatamente o que ocorre quando há uma certa quantidade de DRAM fora de alcance. Nada muito dramático, mas definitivamente limitador.

Como se sabe, a memória cache vem sofrendo algumas transformações, especialmente a conhecida como cache nível 2 (L2). Gradualmente ela está sendo incorporada junto do processador. Isto vem ocorrendo porque com o cache fisicamente mais próximo e transferindo dados por um barramento especial, usualmente chamado de backside bus (BSB), freqüências mais elevadas de troca de dados podem ser empregadas. Seria complexo, mas não impossível, para os fabricantes de placas-mãe implementarem vias elétricas operando com freqüências elevadas como por exemplo 800 MHz. O problema é suas singelas e longas trilhas de condutoras em meio a dezenas de circuitos. Se assim fosse, as placas-mãe ficariam totalmente dependentes da freqüência do processador utilizado, limitando grandemente a compatibilidade. O Pentium III Coppermine (sérieE) e os Celeron com L2, por exemplo, possuem um L2 interno e operando na mesma freqüência de processamento. Os primeiros Athlon, o Pentium II e os III não Cu-mine fazem o cache L2 operar na metade da freqüência de processamento, porém os caches são implantados externamente nos próprios cartuchos, dispensando auxilio da placa-mãe.

E qual a quantidade de cache ideal? Note que aqui fala-se sempre do cache L2. O cache L3 (placa-mãe), no caso dos K6-III é de pouca significância frente ao interno de 256 KB rodando na mesma freqüência de processamento. O L1, existente em todo processador, é de suma importância e é indispensável da arquitetura dos processadores, por isso, não há como discutir a sua quantidade. Percebe-se no entanto, que a quantidade geral de cache vem aumentando gradativamente. O tamanho do cache deve ser tal que a maioria da porção mais ativa dos programas desenvolvidos durante a existência de uma dada geração de processadores consiga caber nele. Nos sistemas multitarefa, ou seja, todos os atuais, o cenário é mais complexo, afinal há vários programas operando simultaneamente e correndo pela ocupação do cache. Em qualquer caso, é consenso que a quantidade de cache e principalmente seu alcance sejam os maiores possíveis.

Bibliografia:

http://www.pr.gov.br/celepar/celepar/batebyte/edicoes/2000/bb100/estagiario.htm

http://users.hotlink.com.br/rmenezes/informa/memoria/memoria.htm

http://www.geocities.com/ResearchTriangle/4480/academic-files/dram.html

http://www.kingston.com/king/mg0.htm

http://www.tomshardware.com/guides/ram/index.html

http://www.whatis.com

http://www.lpc.ufrj.br/linhasdepesquisa/arquitetura.html

http://mega.rnl.ist.utl.pt/~ic-ac/aulas/26-aula/Aula26.html

http://www.douglastorres.hpg.ig.com.br/Ciencia_e_Educacao/9/pipeline.htm

http://www.azoresdive.com/jsim/pipeline.htm

http://www.ic.unicamp.br/~981612/ea960/desafio_4/mips.html

http://www.salvidicas.hpg.ig.com.br/faq1.htm

http://www.pcfacil.com.br/glossario.asp?termo=296

http://www.wagnerzanco.hpg.ig.com.br/cursos/arquiteturadopc/capitulo1/arquiteturadopc.

http://www.net-rosas.com.br/~hncbq/Hardware.htm