

FACULDADE PITÁGORAS

PRONATEC

DISCIPLINA: ARQUITETURA DE COMPUTADORES

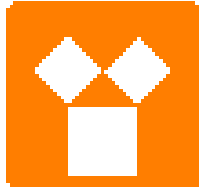
Prof. Ms. Carlos José Giudice dos Santos

carlos@oficinadapesquisa.com.br

www.oficinadapesquisa.com.br

APOSTILA IV

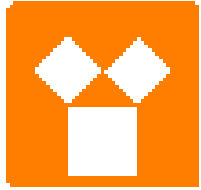
MICROPROCESSADORES



Objetivos

Ao final desta apostila, o aluno deverá ser capaz de:

1. Conhecer a arquitetura interna de uma CPU
2. Conhecer a arquitetura externa de uma CPU
3. Conhecer as principais tecnologias utilizadas em processadores.

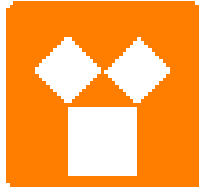


ESQUEMA DE UM COMPUTADOR

Uma Unidade Central de Processamento (CPU - *Central Processing Unit*) é a parte principal de um computador, sendo responsável pela execução de programas que fazem do computador uma máquina extremamente versátil.

É importante notar que CPU é um circuito eletrônico chamado processador. Entretanto, dentro de um computador existem várias CPU's (em placas de comunicação, gráficas e de som, em discos rígidos, etc). Além disso, os usuários e alguns técnicos costumam se referir ao computador (gabinete do PC) como CPU, e embora seja tecnicamente errado falar assim, tornou-se um termo consagrado pelo uso.

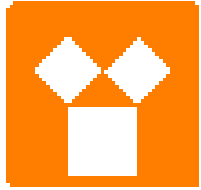
Nesta apostila, usaremos o termo **processador** para designar CPU ou microprocessador. Outro conceito que será utilizado nesta apostila é o de caixa preta, que veremos a seguir.



O QUE É CAIXA PRETA

Internamente um processador é constituído por muitos circuitos eletrônicos digitais que são resultados de várias tecnologias. Já sabemos que o processador é capaz de processar instruções (software) e que é responsável pela capacidade máxima de endereçamento da memória principal e pelo controle dos periféricos, direta ou indiretamente (via chipset).

Os livros costumam utilizar o conceito de caixa preta como uma forma de dominar a complexidade. Toda caixa preta tem entrada(s) e saída. Sabemos qual é a entrada e podemos ver o resultado, que é a saída. A caixa preta esconde o processo de transformação da entrada em saída. No caso do processador, sabemos que a saída ocorre à medida que os dados entram, e que o processamento (transformação dos dados em saída) será mais rápido se o clock (frequência de operação) for maior.

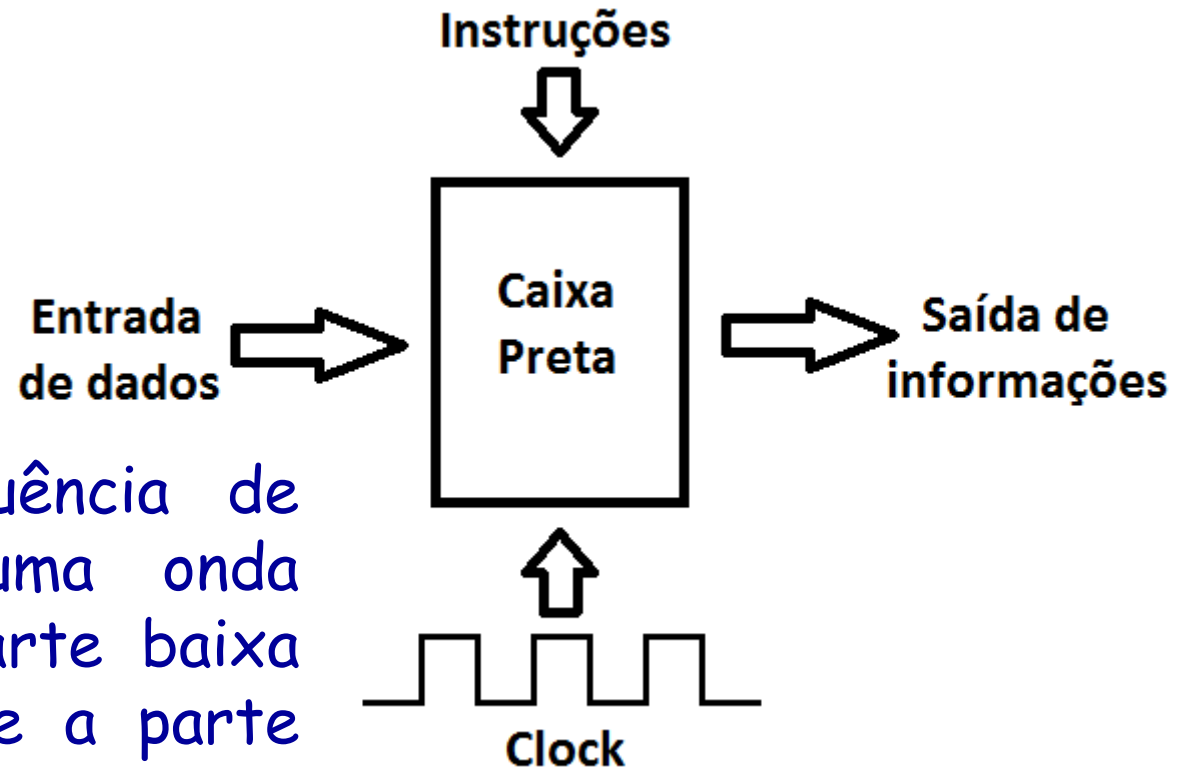


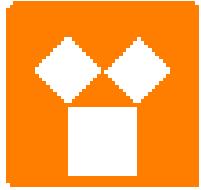
O PROCESSADOR COMO CAIXA PRETA

Para que o processador possa ser representado como uma caixa preta, temos alguns detalhes importante, a saber:

- Entrada de dados
- Entrada de instruções
- Entrada do clock
- Saída de informações

Sabemos que a frequência de clock é dada por uma onda quadrada, em que a parte baixa corresponde ao zero, e a parte alta corresponde a 1.

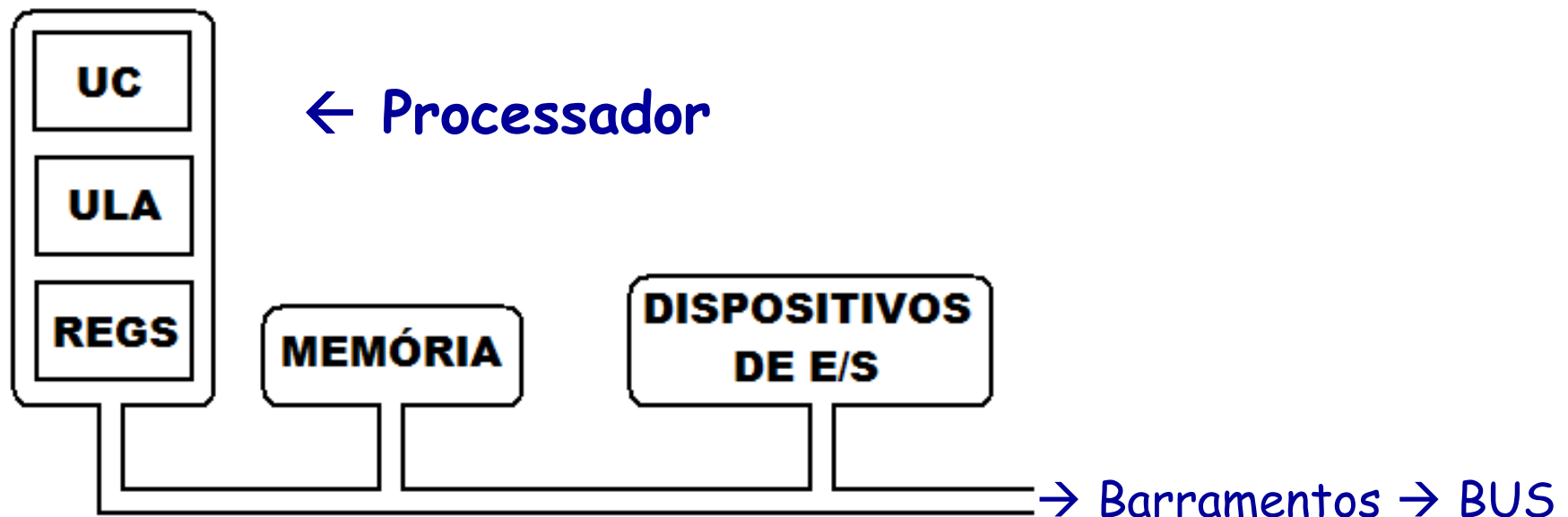


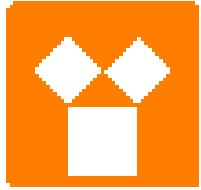


ARQUITETURA INTERNA BÁSICA

Internamente, qualquer processador básico possui uma unidade de controle (UC ou *Control Unit*), uma unidade de lógica aritmética (ULA ou *Arithmetic Logic Unit*) e os registradores (*registrators* ou simplesmente REGS). O processador se comunica com o mundo exterior por meio de barramentos.

Os barramentos do processador se comunicam com os barramentos do computador (placa-mãe).





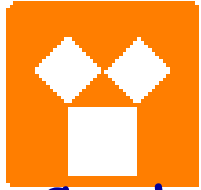
BARRAMENTOS DE UM COMPUTADOR

Uma CPU só se comunica com outros dispositivos por meio de barramentos. De acordo com Vasconcelos (2003), "barramentos são conjuntos de sinais digitais através dos quais o processador transmite e recebe dados de circuitos externos".

Em um computador existem basicamente 3 tipos de barramentos:

- **Barramento do processador**
- **Barramento de memória**
- **Barramentos de E/S**

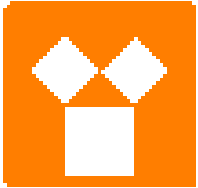
Estes dois barramentos
ficam na placa-mãe.



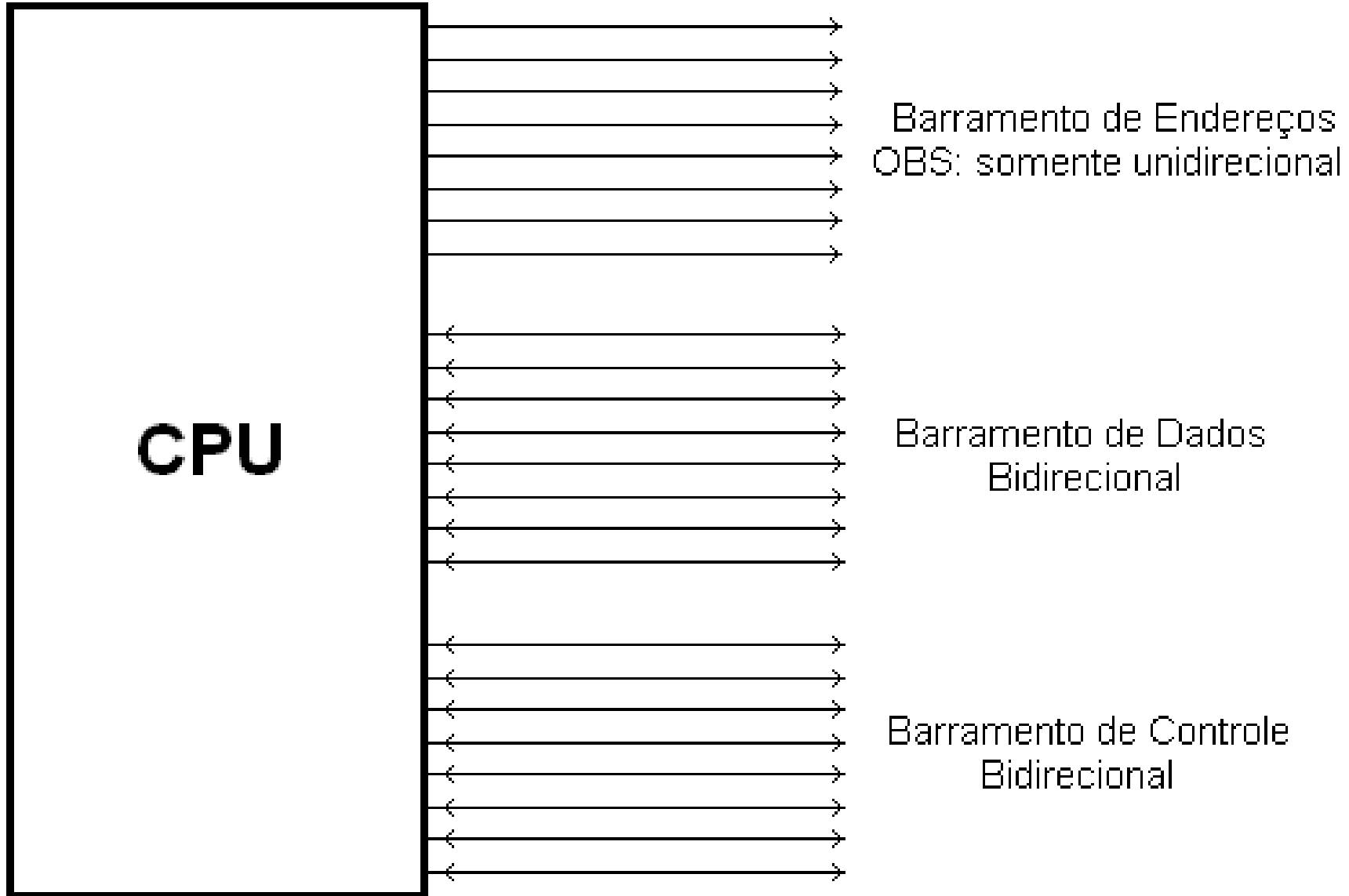
BARRAMENTO DO PROCESSADOR - I

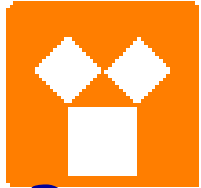
Conhecido também como barramento local (*Local Bus*) ou barramento de sistema, é aquele que faz contato direto com o Chipset (dividido em *North Bridge* e *South Bridge*), que estudaremos mais tarde. É através do Chipset que a CPU se comunica com a memória (Barramento de Memória) e com outros dispositivos (Barramentos de E/S, por exemplo, ISA, EISA, PCI, AGP, AMR, IDE, SCSI, USB, Firewire, entre outros).

O Barramento do Processador é dividido em três grupos, a saber: o Barramento de Dados, o Barramento de Endereços e o Barramento de Controle. Um esquema destes barramentos pode ser conferido no próximo slide.



BARRAMENTO DO PROCESSADOR - II

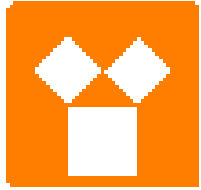




BARRAMENTO DO PROCESSADOR - III

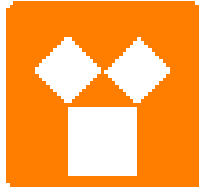
Barramento de Endereços (ou Address Bus): Este barramento tem função dupla: acessar um endereço de memória RAM ou acessar um periférico para transmitir ou receber dados. Resumindo, o **Address Bus** acessa **Memória** ou **I/O**.

Barramento de Dados (Data Bus): Este barramento é utilizado em conjunto com o Address Bus para enviar ou receber dados. A largura do barramento de dados varia de acordo com o processador. Em geral, o número de vias deste barramento corresponde à quantidade de bits que o processador opera internamente (tamanho da palavra). O processador Intel 8088 foi uma exceção a esta regra.



BARRAMENTO DO PROCESSADOR - IV

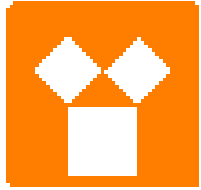
Barramento de Controle: Este barramento contém vários sinais que são necessários ao funcionamento do processador. Entre outras atribuições, tem a função de controlar o tráfego do barramento de dados. Alguns dos seus sinais são de saída, outros são de entrada, outros são de entrada e de saída (bidirecionais). Existem sinais para indicação do tipo de operação (leitura ou escrita), sinais de especificação de destino/origem de dados (memória ou E/S), sinais de sincronismo, sinais de interrupção, sinais que permitem a outro dispositivo tomar o controle do barramento, sinais de clock, sinais de programação e diversos outros.



BARRAMENTO DE MEMÓRIA

Em CPUs antigas, a memória era ligada diretamente ao processador, e funcionava em sincronia. Com o crescimento impressionante de velocidades das CPUs, a memória passou a se comunicar com a CPU por meio do Chipset, de forma assíncrona.

Em geral, o barramento de memória é utilizado também para trafegar dados, por meio de uma técnica chamada de multiplexação (que permite que um único canal seja compartilhado por mais de um sinal).



BARRAMENTOS DE E/S - I

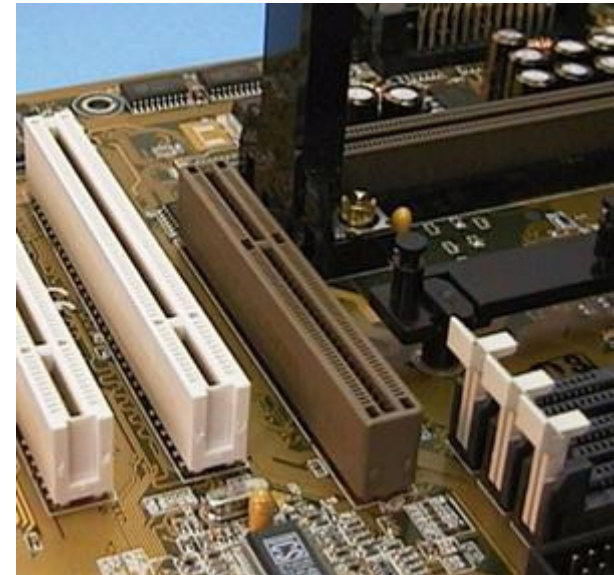
São barramentos utilizados para a CPU se comunicar com placas ou interfaces, podendo (ou não) ser acessados por meio de slots. Os mais comuns são:

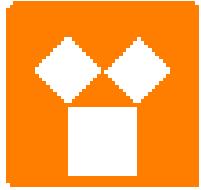
AGP: Lançado pela Intel em 1997, o *Accelerated Graphics Port* foi criado para acelerar o desempenho de placas de vídeo em PCs equipados com o recém lançado Pentium II (ver figura) →.

Este padrão evoluiu para o AGP 8x e o AGP Pro. Hoje não é mais utilizado

Fonte:

<http://www.laercio.com.br/artigos/hardware/hard-013/hard-013.htm>



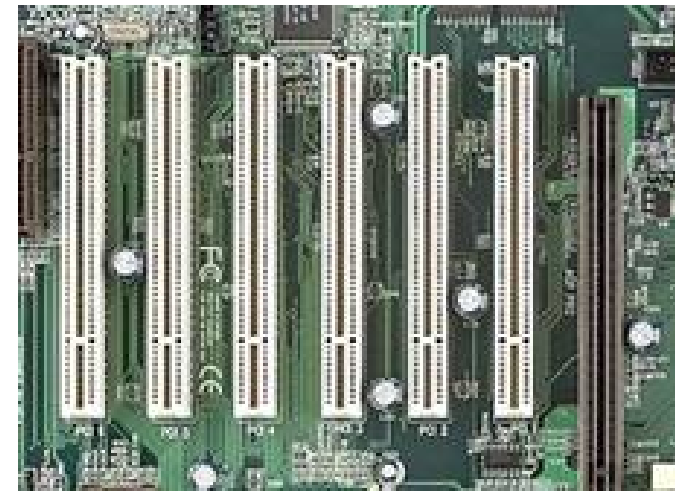


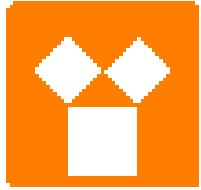
BARRAMENTOS DE E/S - II

PCI: O *Peripheral Component Interconnect* é o padrão mais utilizado para barramento de E/S e para slots de expansão. Utilizando estes slots, pode-se conectar placas de vídeo, placas aceleradoras gráficas, placas de captura de vídeo, placas de rede, placas de interface (SCSI, por exemplo), etc. Em geral, as interfaces para controle de disco rígido e para comunicação USB utilizam este tipo de barramento, apesar de não usarem slots. Este padrão evoluiu para o Fast PCI ou **PCI Express**.

Fonte:

<http://www.laercio.com.br/artigos/hardware/hard-013/hard-013b.htm>

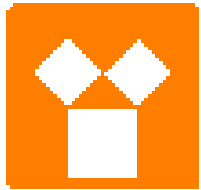




Microarquiteturas Internas de CPU's

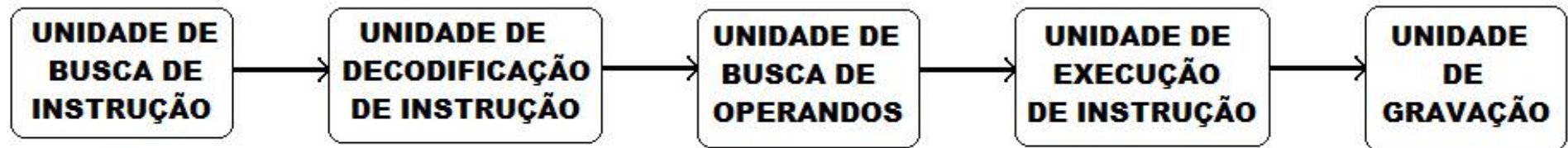
Já sabemos que a frequência de clock é uma das responsáveis pela velocidade de processamento, mas não é a única. O clock tem um limite, a partir do qual é impossível aumentar, pois o gasto de energia e o calor dissipado seriam altos demais.

Em função disso, os projetistas de processadores criaram técnicas para aumentar a velocidade de processamento sem ter que aumentar o clock. Entre estas técnicas estão o pipelining, a execução por arquitetura superescalar, a execução dinâmica de instruções, o barramento duplo independente, as instruções SIMD e SSE e outras tecnologias que serão vistas posteriormente.



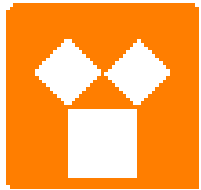
PIPELINING

Pipelining é o nome que se dá ao paralelismo ao nível de instrução. Sua implementação se dá através de uma estrutura chamada *pipeline*. Basicamente, é uma divisão de tarefas que permite a execução de mais de uma instrução:



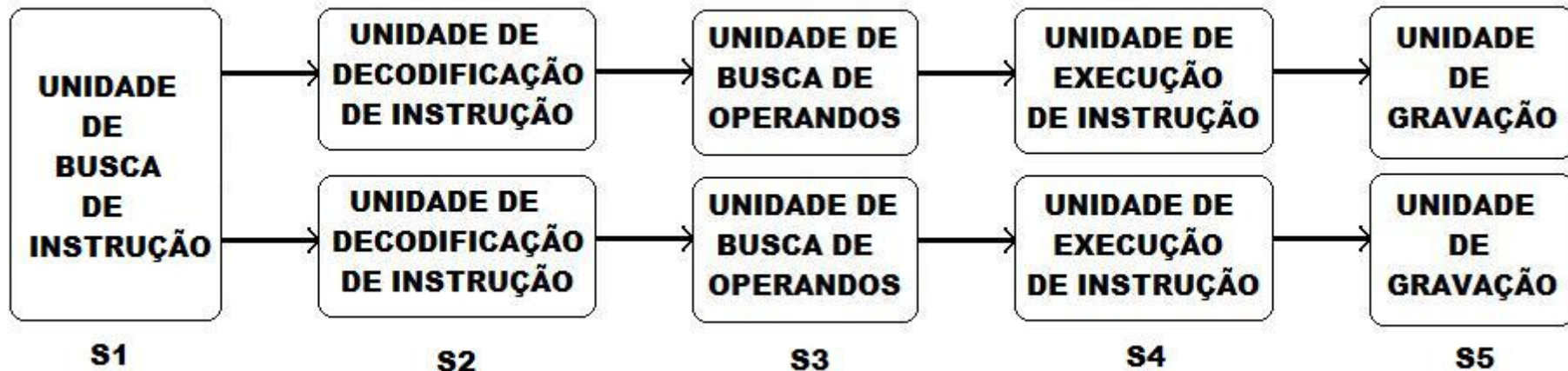
	S1	S2				S3				S4	S5
S1: Stage 1 →	1	2	3	4	5	6	7	8	9		
S2: Stage 2 →		1	2	3	4	5	6	7	8		
S3: Stage 3 →			1	2	3	4	5	6	7		
S4: Stage 4 →				1	2	3	4	5	6		
S5: Stage 5 →					1	2	3	4	5		
→ Tempo →		→ Tempo →				→ Tempo →					

Nesse pipeline (típico do 486), a cada ciclo de clock uma instrução pode ser completada (em tese), o que pode resultar em um ganho de velocidade 5 vezes maior.

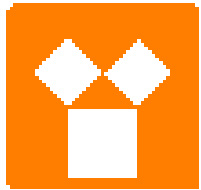


ARQUITETURA SUPERESCALAR (I)

Se um pipeline é bom, dois (ou mais) pipelines devem funcionar melhor ainda. Este é o princípio da arquitetura superescalar. A partir do processador Pentium, a Intel passou a utilizar este tipo de estrutura:

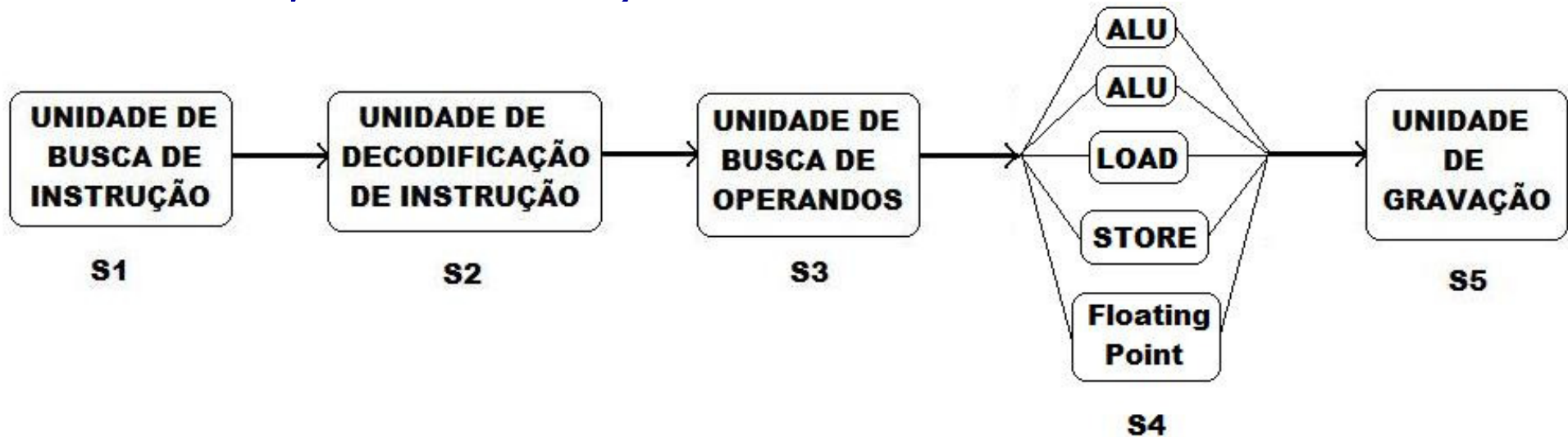


O pipeline de 5 estágios do 486 (slide anterior) às vezes demorava ao executar alguma instrução, o que atrasava toda a fila. No Pentium, os projetistas fizeram dois pipelines em paralelo, de modo que duas instruções (que não dependem uma da outra) podem ser executadas ao mesmo tempo.

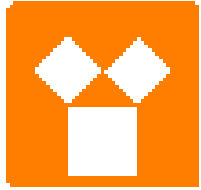


ARQUITETURA SUPERESCALAR (II)

Com o tempo percebeu-se que o estágio 4 (S4 → execução de instruções) era o gargalo de todo o processo. Quando não havia um par de instruções compatíveis, o processamento parava em S4 até que uma instrução mais complexa fosse completada. Para resolver isso, no Pentium II em diante, o pipeline foi modificado, conforme figura abaixo:



O estágio 4 (S4) passou a contar com várias ULAs para não permitir que instruções inacabadas atrasassem a fila. Variações deste tipo de pipeline são utilizadas até hoje.

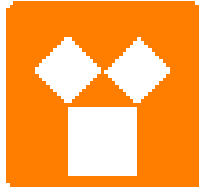


MULTITHREADING - I

A Multithreading é uma técnica que permite aumentar a velocidade de processamento de uma CPU, e o suporte a esta técnica deve estar previsto no SO.

Threads são processos que possuem a capacidade de se dividirem em dois ou mais processos concorrentes que podem ser executados simultaneamente.

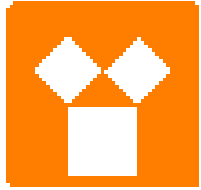
Se o processador possuir um único núcleo, essa divisão ocorre de maneira virtual (multithreading virtual), em geral por linha de execução ou por chaveamento, dando a impressão de que dois processos estão sendo executados simultaneamente.



MULTITHREADING - II

O Multithreading virtual foi uma tecnologia criada pela Intel com o nome de Hyper-Threading (HT). Esta tecnologia permite simular dois processadores virtuais em um único processador físico (de apenas um núcleo). Foi introduzida primeiramente no Xeon (mercado de servidores) e depois no Pentium 4.

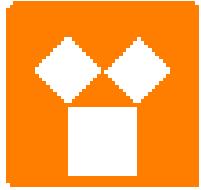
O Multithreading real acontece quando um processador possui mais de um núcleo físico. Nesse caso, dois (ou mais) processos podem realmente ser executados simultaneamente. Um bom exemplo de aplicação dessa técnica são em jogos: um processo pode atualizar a imagem enquanto outro processo atualiza o áudio do jogo.



MULTITHREADING - II

O Multithreading virtual foi uma tecnologia criada pela Intel com o nome de Hyper-Threading (HT). Esta tecnologia permite simular dois processadores virtuais em um único processador físico (de apenas um núcleo). Foi introduzida primeiramente no Xeon (mercado de servidores) e depois no Pentium 4.

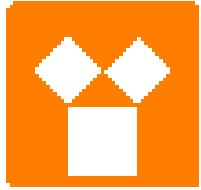
O Multithreading real acontece quando um processador possui mais de um núcleo físico. Nesse caso, dois (ou mais) processos podem realmente ser executados simultaneamente. Um bom exemplo de aplicação dessa técnica são em jogos: um processo pode atualizar a imagem enquanto outro processo atualiza o áudio do jogo.



ARQUITETURA DE INSTRUÇÕES - I

A motivação para o uso do poder de processamento paralelo é a demanda cada vez maior de aplicações que necessitam de grande poder de processamento. Utilização militar (projeto de aeronaves, submarinos e armas nucleares), pesquisas científicas (por exemplo, sequenciamento genético), simulações aerodinâmicas, prospecção mineral e petrolífera, modelagem climática, sistemas especialistas e de inteligência artificial são algumas das aplicações que necessitam de grande poder de processamento.

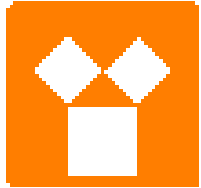
Os processadores funcionavam (até o Pentium) geralmente com uma arquitetura de instruções denominada SISD (Single Instruction Single Data), ou seja, uma única instrução atua sobre um único dado.



ARQUITETURA DE INSTRUÇÕES - II

Logo, na arquitetura SISD (*Single Instruction Single Data*), um único fluxo de instruções opera sobre um único fluxo de dados. Entretanto, os projetistas de processadores perceberam que existiam casos em que uma mesma instrução era aplicada a um conjunto de dados.

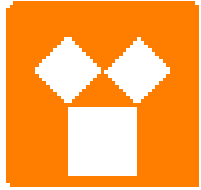
Este era o caso, por exemplo, de instruções de vídeo: para mudar a cor de todos os pontos de uma tela, um processador de arquitetura SISD tem que executar a mesma instrução, uma por uma, em cada ponto da tela. Nos primeiros processadores, esta mudança de cor era tão lenta que era possível ver a tela mudando de cor, linha após linha. Foi então que surgiu a ideia de se criar uma arquitetura que permitisse aplicar uma única instrução a vários dados, de uma única vez. Surgiu assim a arquitetura SIMD (*Single Instruction Multiple Data*), utilizada a partir do Pentium MMX.



ARQUITETURA DE INSTRUÇÕES - III

SIMD (*Single Instruction Multiple Data*): nesta classe, uma única instrução pode ser aplicada simultaneamente sobre um conjunto de dados. Os aplicativos ainda seguem uma organização sequencial, mas o acesso a múltiplos dados é facilitado pela organização da memória em diversos módulos.

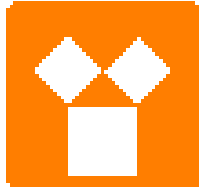
Este tipo de arquitetura de instruções é ideal para trabalhar com aplicações gráficas, em que um grande volume de dados deve ser modificado por uma única instrução. Ainda há apenas uma Unidade de Controle, mas existem diversas unidades funcionais. Corresponde aos processadores vetoriais e matriciais. O Intel Pentium MMX e o AMD K6 são os primeiros representantes desta classe.



ARQUITETURA DE INSTRUÇÕES - IV

A AMD criou uma tecnologia chamada 3D Now!, que utiliza instruções SIMD com dados em ponto flutuante para permitir um aumento significativo do processamento de gráficos. Esta tecnologia foi implementada inicialmente do AMD K6-2, e foi a partir deste processador que a AMD começou a rivalizar com a Intel, fabricando processadores de qualidade a um preço mais competitivo.

A Intel deu o troco criando uma extensão de instruções SIMD conhecidas como SSE (Streaming SIMD Extensions), que permitiu novas implementações de aplicações multimídia. O Intel Pentium III foi o primeiro processador a ter esta tecnologia integrada.



Outras tecnologias de processadores

Houve uma imensa evolução no projeto de processadores, resultando em inúmeras tecnologias.

Para poder entender estas tecnologias, torna-se necessário entender primeiro o que é cache, chipset e as tecnologias de construção de memória RAM, razão pela qual vamos estudar estas tecnologias de construção de processadores posteriormente.

Por enquanto, foquem nas tecnologias já estudadas.