

Memória cache grupo-associativa interna ao processador:

- Um determinado bloco da memória principal somente pode ser em um único conjunto (ou grupo) da cache;

Smartcache:

- É uma metodologia que se aplica somente para vários núcleos, mas que apresenta uma cache de nível 1 (L1) exclusiva para cada núcleo e uma cache de nível 2 (L2) compartilhada entre os vários núcleos;

Memória virtual paginada:

- Necessita de suporte do sistema operacional para tratamento de “page fault” (falta de espaço para uma página na memória principal);
- Necessita de uma a quatro tabelas de descrição de páginas, localizadas na própria memória principal;
- É gerenciada pelo microprocessador, na tradução de um endereço virtual em físico;

Memória virtual segmentada:

- Gerenciada pelo microprocessador;
- Necessita de suporte do sistema operacional para determinação onde um segmento de memória virtual vai ser carregado na memória física;
- Opera sobre blocos de tamanho variável, denominados de segmentos;
- Necessita de suporte do sistema operacional para tratamento de “segment fault”;

Interrupções:

- Interrupções de hardware e software compartilham o mesmo esquema de tratamento de interrupções;
- O vetor de interrupções é usado para localizar a rotina do tratamento de interrupções;
- Os pedidos de interrupção chegam pelo pino INTR e podem ser desabilitados por software; se chegar pelo pino NMI, não pode ser desabilitado por software;
- Interrupções de hardware e software não são síncronas em relação a execução de instruções;
- O vetor de interrupções não determina a prioridade das interrupções;

Memória:

- Conforme aumenta a geração das memórias DDR, as frequências ficam maiores e as temporizações diferentes;
- DDR sempre mantém duas transferências por ciclo de relógio;
- DDR2 permite o dobro da taxa que DDR, para uma mesma frequência de operação;
- 2-3-4-7-1 -> latência do CAS = 2 (primeiro número);

Pentium MMX - Aritmética de saturação:

- As operações são executadas na aritmética tradicional, porém alguns códigos são suprimidos;
- Resultados que implicariam num estouro de representação são devolvidos como o maior representável, no caso de overflow, ou menor representável, no caso de underflow;
- Para operandos de 16 bits com sinal, a faixa é de $[-32768, +32767]$;
- Novo tratamento aritmético para casos de overflow;
- Novos tipos de operandos para instruções de memória;
- Instruções extras para processamento vetorial;
- As instruções MMX ajudam a melhorar o desempenho em aplicações 2D;
- Instruções Streaming SIMD ou 3DNow ajudam a melhorar o desempenho em aplicações 3D;

Ponto flutuante:

- Instruções da Intel manipulavam 4 números em ponto flutuante, enquanto as da AMD manipulavam 2 números;
- As extensões aos conjuntos de instruções trabalham sobre número em ponto flutuante de 32 bits;

Memória virtual vs. Cache:

- Ambas contêm dados ou instruções frequentemente utilizados, para que esses possam ser acessados de forma mais rápida;
- Ambas necessitam de suporte de hardware no processar. Não podem ser implementadas apenas somente por software;
- Ambas são suportadas por hardware na conversão de endereços;
- Ambas são suportadas por hardware na determinação da ausência de um bloco;
- A técnica de write-through torna desnecessário salvar um bloco alterado antes da sua substituição;
- A ausência de um bloco não é indicada por uma interrupção do microprocessador;
- Não são suportadas pelo sistema operacional na busca de um novo bloco e substituição de um antigo;

Superescalaridade:

- Várias unidades de execução, de forma a possibilitar que várias instruções possam terminar no mesmo ciclo de relógio;
- Renomeação de registradores -> Eliminar falsas dependências -> WAR e WAW;

Turbo Boost:

- É uma tecnologia que permite que os núcleos do processador aumentem sua frequência de operação, desde que operem nas condições adequadas;

Compilador:

- Unidades com instruções MMX;
- Unidades com instruções SSE;
- Arquiteturas VLIW;

Barramentos:

- PCI-Express:
 - o Serial, mas bidirecional;
 - o 2500MHz;
- ISA:
 - o Primeiro barramento utilizado nos IBM-PC;
 - o 8,33 MHz;
- PCI:
 - o 33MHz;

Exercício:

- Cache nível 2 de 512 KBytes, com tamanho de linha de 64 bytes, 8-way set associative:
 - Tamanho cache: $512\text{KB} = 2^{19}$;
 - Tamanho do bloco: $64\text{ Bytes/bloco} = 2^6 \rightarrow \text{OFFSET} = 6\text{ bits}$;
 - Número de blocos = 2^{13} ;
 - 8-Way set associative;
 - $2^{13}/2^3 = 2^{10} = 1024$ conjuntos $\rightarrow 10\text{ bits}$ para determinar o conjunto;
 - Número de blocos / conjuntos = $8192 / 1024 = 8\text{ blocos/conjunto}$;

Exercício:

- Barramento hipotético, 50 e 100 MHz, largura de dados de 128 bits, 14 transferências de dados por ciclo de relógio. Qual taxa máxima?
 - o $128\text{ bits} = 16\text{ bytes}$;
 - o $\text{Taxa} = 16 * 14 * 100 = 22400\text{ MB/s}$;

Exercício:

- Os endereços de memória gerados pela CPU têm 28 bits;
- Os 12 bits menos significativos identificam o byte dentro da página (OFFSET);
- Os 16 bits mais significativos servem para acessar uma tabela de descritores de página;
- Os 14 bits formam o endereço posicional da página;
 - o $\text{Tamanho do endereço} = 2^{28} = 268.435.456 = 256\text{ MBytes}$;
 - o $\text{Tamanho do bloco} = 2^{12} = 4.096 = 4\text{ KBytes}$;
 - o $\text{Tamanho memória física} = 268.435.456 / 4.096 = 65.536 = 64\text{ MBytes}$;

Exercício:

- Cache de blocos de 16 bytes;
- Tamanho da cache = 16 Kbytes;

- 256 grupos (conjuntos);
- Memória de 4 GB;
 - Endereço de acesso à memória composto de $2^{32} = 4.294.967.296 = 32$ bits;
 - OFFSET = $2^4 = 4$ bits;
 - Conjuntos = $2^8 = 8$ bits;
 - Tag = $32 - 4 - 8 = 20$ bits;

Exercício:

- Mapeamento direto;
- Cache de blocos de 16 bytes;
- Tamanho da cache = 16 Kbytes;
- Memória de 4 GB;
 - Endereço de acesso à memória composto de $2^{32} = 4.294.967.296 = 32$ bits;
 - OFFSET = $2^4 = 4$ bits;
 - Número de blocos = $2^{14} / 2^4 = 2^{10} = 10$ bits;
 - Tag = $32 - 10 - 4 = 18$ bits