

Avanços na Arquitetura de Memória Cache

Layla Chris Rodrigues Ferreira
UNESP - IBILCE

1. INTRODUÇÃO

A comunicação entre o processador e a memória, bem como entre o processador e os dispositivos de entrada e saída acabaram sendo afetadas, devido ao grande aumento de velocidade nos processadores durante o decorrer dos anos. As memórias também tiveram avanços significativos, porém, mesmo com seu aperfeiçoamento, não conseguiram acompanhar o ritmo de processamento [5].

Como consequência, o desempenho da CPU foi reduzido, devido à baixa velocidade de transferência de dados e alto tempo para acesso dos mesmos. Dessa forma, com a finalidade de minimizar o desequilíbrio de velocidade entre o processador e a memória principal, foram criadas memórias individuais para o processador, sendo conhecidas como memórias *cache*, além de registradores [9]. Ambos localizam-se no topo da hierarquia de memória e portanto, possuem alta velocidade, alta volatilidade (perde as informações armazenadas no caso de ausência de energia), custo muito alto e baixa capacidade de armazenamento.

Um processador que não possui memória *cache* irá, em cada ciclo de instrução, acessar pelo menos uma vez a memória principal. Considerando que o acesso à mesma demora mais do que a própria execução da instrução, o processador permanecerá muito tempo ocioso, acarretando uma queda significativa no desempenho do sistema [13].

Assim, as memórias *cache* são dispositivos intermediários na transferência de dados entre o processador e a memória principal e que visam melhorar o desempenho do computador [13]. Para isto, armazenam os dados e instruções que são mais frequentemente usados pelo processador, permitindo que os mesmos sejam acessados rapidamente.

2. MEMÓRIAS CACHE

Com a contínua demanda por desempenho computacional, as memórias *cache* vêm sendo largamente adotadas nas últimas décadas, nos mais diversos tipos de projetos arquiteturais de processadores [10].

Uma vez incorporada, ao haver a necessidade por um dado, este será procurado primeiramente nas memórias *cache* presentes [3]. Caso o dado não seja encontrado, será preciso acessar a memória principal.

Ao observar a hierarquia de memória representada pela figura 1, nota-se que o tempo de acesso à memória é variável, já que se a informação for encontrada na *cache*, o tempo de acesso será curto pelo fato da mesma se localizar no topo da hierarquia, mas caso seja necessário ir para as camadas mais

inferiores na busca pela informação, o tempo de acesso total será a soma de todos os tempos de acesso de cada uma das camadas percorridas.



Figure 1. Hierarquia de Memória [12]

Ao longo de décadas, diversas técnicas foram utilizadas a fim de obter os melhores desempenhos com a utilização de memórias *cache* [10].

2.1. ARQUITETURA

A memória *cache* é dividida em níveis com o objetivo de otimizar a transferência de dados:

- **Nível 1:** denominado *cache* L1, proveniente do inglês level 1. Localiza-se no interior do processador, sendo a menor porção da memória presente dentro do mesmo (entre 16 KB e 256 KB).
- **Nível 2:** conhecido como *cache* L2, este nível pode localizar-se tanto no interior quanto no exterior do processador. Foi desenvolvido com o objetivo de suprir a baixa capacidade de armazenamento da *cache* L1 e para melhorar o desempenho do sistema, já que é mais uma chance para que a informação desejada não precise ser procurada na memória principal.

- **Nível 3:** denominado *cache* L3, é um tipo não tão comum devido à complexidade dos processadores atuais.

A capacidade de armazenamento dos níveis aumenta do topo para as camadas inferiores, porém a velocidade e custo possuem o sentido contrário, sendo a *cache* L1 a mais veloz e mais cara de todas.

2.2. FUNCIONAMENTO

O processador inicia a busca pela instrução ou dados desejados, procurando primeiramente na *cache* L1. Caso a informação necessária não esteja presente, a busca passará para a *cache* L2. Estando aí presente, o processador a transferirá para a L1 junto de um bloco de dados. Se a informação também estiver ausente em L2, a busca passará a ser feita em L3 (se houver) e caso não a encontre, o alvo será a memória principal seguida da memória secundária, caso nada tenha sido encontrado.

Ao ser localizada, a informação é transferida para o processador. Caso ela não esteja presente na memória, a informação será transferida da memória principal para a *cache* e para o processador que permaneceu ocioso até então. Durante a busca dos dados ausentes, é copiado um bloco (ou linha) inteiro da memória principal. Isto ocorre para diminuir a taxa de falha dos acessos futuros. Dois conceitos surgem com esse funcionamento:

- **Cache Hit:** instrução ou dados procurados estão presentes na memória *cache*.
- **Cache Miss:** instrução ou dados procurados estão ausentes na memória *cache*.

Ao transferir blocos, é preciso armazenar os endereços dos mesmos ou parte deles, para que possam ser identificados. No momento em que for feita uma leitura de dados ou instrução pelo processador, os endereços armazenados na *cache* são comparados com o endereço fornecido pelo processador, para saber se houve acerto ou falha [12]. Em caso de acerto, a informação armazenada na *cache* é transferida para o processador, evitando o acesso à memória principal. A figura ??fig:memory ilustra o funcionamento detalhado acima.

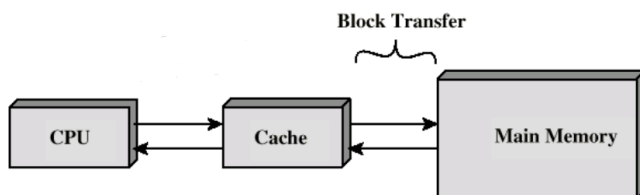


Figure 2. Funcionamento de um Sistema com Memória *Cache* [11]

Assim, já que os programas costumam executar um grupo de instruções várias vezes em diversos momentos, a maioria das tecnologias *cache* desenvolvidas baseia-se em dois princípios de localidade, os quais são uma tentativa de prever qual será a demanda futura da CPU por instruções que estão na memória principal e que são definidos abaixo de acordo com [3]:

- **Princípio de localidade espacial:** se um item é referenciado, provavelmente seus vizinhos também o sejam. Nesta situação, encaixam-se funções, métodos, laços de repetições, contadores ou variáveis que são constantemente modificadas.
- **Princípio de localidade temporal:** se um item é referenciado, provavelmente ele será referenciado novamente em um curto espaço de tempo. Neste caso, encaixam-se partes procedurais sequenciais de um programa, tais como vetores, matrizes, *arrays*, entre outros.

Dessa forma, por meio desses dois princípios, diversos avanços foram feitos nas memórias *cache* no decorrer dos anos, como por exemplo, os modos de endereçamento de blocos da memória, as políticas de alocação e substituição de dados da memória *cache*, além das estratégias de escritas [3], entre outros.

Ao unir os projetos de arquitetura de processadores com as memórias *cache*, os esforços passaram a ser a redução das faltas de dados na memória e, além disso, a redução das penalidades causadas por essas faltas [3].

3. AVANÇOS NA ARQUITETURA DE MEMÓRIA CACHE

As memórias *cache* nem sempre estiveram presentes nos computadores. Sua primeira utilização ocorreu em 1989 com o lançamento do processador 486 da Intel.

3.1. EVOLUÇÃO DA MEMÓRIA CACHE NOS PROCESSADORES INTEL

Com a implementação de uma *cache* L1 de 8 KB de dados e instruções, o processador 486 obteve um desempenho significativamente maior do que seu antecessor, o 386.

Ao perceber a melhoria no desempenho do computador, foram desenvolvidas placas com memória *cache* embutida para os 386. A capacidade de armazenamento variava de acordo com a placa-mãe, mas os valores típicos nessa época eram de 64 KB e 128 KB [9]. Com isso, surgiram as primeiras *cache* L2, sendo assim denominadas somente após o surgimento do Pentium Pro. Até então elas eram chamadas de *cache* externa. Dessa maneira, foi lançada uma nova versão do 486 com o adicional de *cache* L2 externa.

3.1.1. 386DX

Como dito anteriormente, esse processador não possuía uma *cache* de memória embutido, porém havia um controlador de *cache* externo ao processador, sendo a mesma opcional, podendo incluí-la ou não no computador. Aqueles que não a continham, eram muito mais lentos [1].

Nessa época, o controlador de *cache* utilizava a arquitetura *write-through*, a qual permitia, quando um dado precisava ser armazenado na *cache*, que o controlador atualizasse a memória RAM imediatamente.

O problema com esses processadores foram que, apesar de uma tecnologia de memória mais rápida, esta era mais lenta do que o barramento do sistema [11].

3.1.2. 486DX

Lançado em 1989, a Intel incluiu neste processador 8 KB de *cache* L1 ou interna, como era denominada, além de uma unidade de interface de barramento aprimorada, taxa de execução de uma instrução por ciclo de *clock* e uma unidade de ponto flutuante adicional. Todos esses avanços praticamente dobraram o desempenho do processador em relação ao 386 [1]. O problema com essa arquitetura foi que o aumento da velocidade do processador tornou-se um gargalo para o acesso à *cache* [11].

A tecnologia foi tão avançada para a época que novas versões do 486 foram lançadas em 1991 com tecnologia de memória *cache* mais rápida que a memória principal. As versões que surgiram são detalhadas de acordo com [16]:

- **486DX2:** processador com *clock* interno duplicado.
- **486DX4** processador com *clock* interno triplicado, que aceitava *cache* L1 e L2.
- **OverDrive:** um processador em cima do outro.

Os novos 486 passaram a usar a arquitetura *write-back*, fazendo com que a CPU escrevesse dados diretamente na *cache*, deixando ao sistema a tarefa de atualizar a memória RAM. Esta técnica vem sendo utilizada até os dias atuais. O problema com as novas versões é que a partir do 486DX2 os processadores passaram a usar *clock* interno e externo diferentes um do outro. Dessa forma, a *cache* L2 era acessada com um *clock* menor do que o processador.

3.1.3. PENTIUM

Em 1993, a Intel lançou o processador Pentium com dois circuitos de *cache* L1 de 16 KB, sendo um circuito de 8 KB para dados e outro também de 8 KB para instruções [16]. Na época, a *cache* L2 ainda localizava-se na placa-mãe e seus valores eram entre 256 KB e 512 KB.

O problema nesse processador ocorre quando a busca por instruções e a unidade de execução solicitam simultaneamente acesso à *cache*. Neste caso, a busca por instruções é interrompida enquanto a unidade de execução ganha o acesso.

3.1.4. PENTIUM PRO

Em 1995, a Intel investiu no lançamento do Pentium Pro, o qual introduziu a ideia da *cache* L2 interna ao processador, permitindo-o acessá-la com o mesmo *clock* interno. Dessa forma, ambos trabalhavam na mesma velocidade.

Neste processador, a arquitetura da memória *cache* era L1 com 8 KB para dados e 8 KB para instruções e L2 interna com capacidade de armazenamento entre 256 KB e 1 MB. Assim como nos 486DX, o aumento na velocidade do processador tornou-se um gargalo para o acesso à *cache* L2.

3.1.5. PENTIUM MMX

Em 1996, a Intel lançou o processador Pentium MMX com 57 novas instruções criadas para manipular e processar dados de vídeo, áudio e gráficos. As instruções existentes até então manipulavam dados de 8 ou 16 bits, enquanto que as novas criadas passaram a ser capazes de manipular dados agrupados em pacotes de 64 bits [16].

Para ser capaz de armazenar as novas instruções, a *cache* L1 teve sua capacidade dobrada para 32 KB, sendo 16 KB para dados e 16 KB para instruções. Dessa forma, reduz-se a chance do processador acessar áreas mais lentas da memória.

3.1.6. PENTIUM II

O processador Pentium II, lançado em 1997, combinava as avançadas tecnologias do Pentium Pro com os recursos da tecnologia de aperfeiçoamento de meios eletrônicos do Pentium MMX [16]. Dessa forma, esperava-se a transformação de dados brutos em detalhados modelos em 3D em questão de segundos.

O tamanho da *cache* L1 neste processador era de 16 KB para dados e 16 KB para instruções. E o tamanho da *cache* L2 era de 512 KB. Foram criados Pentium II com *clocks* de 233, 266 e 300 MHz. A velocidade da *cache* L1 é igual à do *clock* em cada caso, ou seja, para 233 MHz, a L1 possuía 233 de velocidade. Para 266 Mhz, L1 possuía 266 e assim por diante. A *cache* L2 trabalhava na mesma velocidade do processador. Então, para 233 Mhz, L2 possuía 117. Para 266 Mhz, L2 tinha 133 e para 300 Mhz, L2 trabalhava com 150 de velocidade.

O problema com essa arquitetura é o mesmo do Pentium Pro, isto é, o aumento na velocidade do processador continua sendo um gargalo para o acesso à *cache* L2 [11].

3.1.7. PENTIUM II XEON

O processador Pentium II Xeon, lançado em junho de 1998, possuía características inovadoras e técnicas especificamente concebidas para estações de trabalho e servidores que utilizavam aplicações profissionais exigentes, tais como serviços de Internet, armazenamento de dados corporativos, criação de conteúdos digitais e automação eletrônica e mecânica. Sistemas computacionais baseados nesse processador permitiam multiprocessamento com quatro, oito, ou mais processadores [1].

Quanto à arquitetura de memória *cache*, possuía a *cache* L1 e 512 KB, 1 MB ou 2 MB de *cache* L2, variando de acordo com o *clock* (400 ou 450 MHz).

3.1.8. CELERON

Lançado em agosto de 1998, era um Pentium II sem memória *cache* L2 integrada, porém foi lançado o Celeron-A com 32 KB de *cache* L1 e 128 KB de *cache* L2. Era a linha econômica da Intel para concorrer com o Duron da AMD [12].

3.1.9. PENTIUM III

Lançado em fevereiro de 1999, este processador veio com 70 novas instruções, as quais aumentaram visivelmente o desempenho de gráficos avançados, 3D, *streaming* de áudio, vídeo e aplicações de reconhecimento de voz [1].

Neste processador a *cache* L1 continuou com 32 KB de armazenagem, sendo 16 KB para dados e 16 KB para instruções. A *cache* L2 possuía 512 KB a 2 MB para armazenagem. Também foi adicionada uma *cache* externa L3, porém o primeiro uso desse nível foi no processador Alpha 21164, sendo opcional nos processadores seguintes.

O problema com essa arquitetura era que algumas aplicações lidavam com grandes bases de dados e precisam acessar rapidamente enormes quantidades de informações, o que acaba sendo afetado pelas *caches* internas por ainda serem muito pequenas.

3.1.10. PENTIUM III XEON

Lançado em outubro de 1999, este processador possuía as mesmas características do Pentium II Xeon, porém aceitava *clocks* de 600, 667, 733, 800, 866, 933 e 1 Ghz. Também opera com 512 KB, 1 MB ou 2 MB de *cache* L2.

3.1.11. PENTIUM 4

Lançado em novembro de 2000, utilizava a mesma estrutura interna do seu antecessor Pentium III, mas com algumas modificações para torná-lo mais rápido, tais como: sua unidade lógica aritmética trabalhava com o dobro do *clock*

interno do processador, continha 144 novas instruções em relação ao Pentium III, entre outros novos aperfeiçoamentos [12].

A memória *cache* L2 era integrada trabalhando na mesma frequência interna do processador, assim como ocorre no Celeron e nos últimos modelos do Pentium III [12]. Também contém 2 MB de *cache* L3. O caminho de dados entre a *cache* L1 e a L2 passou a ser de 256 bits. Nos processadores anteriores, esse valor era apenas 64 bits [9]. Dessa maneira, a comunicação nesses processadores pode ser quatro vezes maior, operando com o mesmo *clock*.

3.1.12. ITANIUM

Lançado em junho de 2001, este processador possuía melhor performance para trabalhar com dados criptografados e possuía um sistema de detecção de e correção de erros, além de diversas outras melhorias.

A arquitetura da sua memória *cache* contava com 36 KB de *cache* L1, sendo 16 KB para dados e 16 KB para instruções, 256 KB de *cache* L2 e de 2 a 4 MB de *cache* L3, operando a 800 MHz.

No período de 2001 a 2006, surgem diversas novas tecnologias e processadores. Contudo, a partir de 2001, após o lançamento de novas versões do Pentium 4, a indústria já podia construir processadores com transistores tão pequenos que tornou-se muito difícil aumentar o *clock* por limitações físicas, principalmente porque o calor gerado era tão grande que não podia ser dissipado pelos resfriadores convencionais [1].

Em 2007, as barreiras térmicas que atrasavam o avanço dos processadores levaram os fabricantes a criar novas saídas para continuar desenvolvendo novos produtos com maior poder de processamento que os anteriores. Uma das saídas mais palpáveis foi colocar vários núcleos em um mesmo *chip*. Esses novos processadores ficaram conhecidos como *multi-core*, sendo o Pentium D o primeiro lançamento nesta categoria.

3.1.13. PENTIUM DUAL CORE

Lançado em 2006 e existente no mercado até hoje, os processadores Pentium Dual Core diferenciam-se dos Pentium D principalmente quanto à dissipação térmica ou aquecimento, além de possuírem arquiteturas baseadas na tecnologia *Core* [1], que é muito mais rápida do que a *Netburst*, na qual o Pentium D é baseado.

Os Pentium Dual Core possuem versões para computadores de mesa e portáteis, sendo que as *cache* L2 podem atingir até 4 MB, dependendo da versão. Possuem 64 Kb de *cache* L1 (32 Kb para dados e 32 KB para instruções). É importante salientar o fato de que, com o uso de dois núcleos de processamento, os valores das *cache* devem ser duplicados.

3.1.14. INTEL CORE 2

O lançamento do *Core 2* substituiu a marca Pentium. Os modelos mais comuns são o *Core 2 Duo*, *Core 2 Quad*, *Core 2 Extreme* e *Core 2 Solo*.

Com o *Core 2*, a Intel preferiu investir na melhoria do barramento, da memória *cache* e em técnicas que trouxessem aumento de desempenho sem elevar a frequência, o consumo elétrico e o calor dissipado. Dessa maneira, esses processadores possuem baixas frequências, mas melhor performance em relação aos anteriores.

Os processadores *Core 2* possuem 64 KB de *cache* L1, sendo 32 KB para dados e 32 KB para instruções. Aceitam 2 MB, 3 MB, 4 MB ou 6 MB de *cache* L2 compartilhada. No *Core 2 Quad*, esta é dividida em duas outras *cache*, sendo uma compartilhada entre os núcleos 1 e 2 e a outra entre os núcleos 3 e 4 citeoito.

3.1.15. INTEL CORE i3

Lançado em janeiro de 2010, o Intel *Core i3* é a nova família da Intel que surgiu para substituir a antiga linha *Core 2 Duo*, sendo baseada na tecnologia *Nehalem*. Apesar de possuir dois núcleos de processamento, houveram melhorias em diversas funcionalidades, tais como utilização de canal duplo para memória RAM (memória trabalha aos pares), tecnologia *Hyper-Threading* (permite a realização de várias tarefas simultaneamente) [2], entre outras.

A arquitetura de memória *cache* neste processador opera com 64 KB de *cache* L1, sendo 32 KB para dados e 32 KB para instruções e 256 KB de *cache* L2 para cada núcleo. Também permite até 4 MB de *cache* L3 [2].

3.1.16. INTEL CORE i5

Lançado em 2009, o Intel *Core i5* executa tarefas mais pesadas que o i3. Está disponível em modelos de dois ou quatro núcleos. Possui tecnologia Intel *Hyper-Threading* e *Turbo Boost*. Esta trabalha verificando frequência, voltagem e temperatura do processador. Ao notar queda em um desses valores, esta ferramenta aumenta a frequência, conseguindo um melhor desempenho nas aplicações [2].

A memória *cache* L1 continua com 64 KB (32 KB para dados e 32 KB para instruções) e 256 KB de *cache* L2 [2] para cada núcleo. Quanto à *cache* L3, opera com 4 MB nos modelos com dois núcleos e com 8 MB nos modelos com 4 núcleos.

3.1.17. INTEL CORE i7

O último modelo em tecnologia de processamento é o Intel *Core i7*. Todos esses processadores possuem quatro núcleos, exceto o i7-980X que possui seis núcleos. Esta arquitetura possui tecnologia Intel *Hyper-Threading*, *Turbo Boost* e ainda o recurso Intel QPI (*QuickPath Interconnect* [2]. Este

funciona aumentando a largura de banda, permitindo mais transmissão de dados e diminuindo a latência, ao invés de aumentar a frequência, como o *Turbo Boost*.

Possui memória *cache* L1 de 64 KB (32 KB para dados e 32 KB para instruções), 256 Kb de *cache* L2 e 8 MB de *cache* L3, sendo que o i7-980X possui 12 MB de *cache* L3.

É importante ressaltar que a Intel, na tecnologia *Nehalem* dos processadores i3, i5 e i7, utiliza um sistema onde as *cache* L1 e L2 armazenam cópias de dados também armazenados na *cache* L3. Isto acarreta um aumento de desempenho, pois cada núcleo não precisa verificar o conteúdo das demais na busca por dados, bastando checar L1 e L2 e em seguida, a L3 compartilhada.

Outro fator importante decorrente da família *Core* atual é a redução no tamanho das *cache* L2, o qual é compensado nas *cache* L3 compartilhada entre todos os núcleos. Os processadores i3, i5 e i7 ainda estão passando por melhorias, evoluindo de geração em geração.

3.2. EVOLUÇÃO DA MEMÓRIA CACHE NOS PROCESSADORES AMD

Como o Intel 486, o AM486 da AMD, lançado em 1993 (4 anos após o 486 da Intel), foi construído com um coprocessador matemático integrado. Porém, a frequência do seu barramento interno era de 40MHz, fazendo ele ser mais rápido que as primeiras versões do Intel 486 [1], proporcionando o início da popularidade da AMD.

Após o surgimento do Pentium, a AMD lançou uma versão melhorada do AM486 que, mesmo com sua frequência máxima de 150MHz e 1.600.000 transistores, não era competitivo ao rival e não foi bem aceito pelos consumidores [1].

O AM486 possuía 8 KB de *cache* L1 e a *cache* L2 era externa, sendo opcional aos computadores com esse processador.

3.2.1. K5

Introduzido em março de 1996, seu principal concorrente era a família de processadores Intel Pentium. O processador K5 possuía *pipelines* duplos com menos restrições para a execução de instruções simultâneas, além de aceitarem previsão de desvio [3].

Sua arquitetura de memória *cache* era composta por *cache* L1 de 8 KB de dados e 16 KB de instruções e *cache* L2 externa ao processador. O grande problema com esse processador, é que se tornou complexo o desenvolvimento de versões mais rápidas.

3.2.2. K6

Lançado em 1997, o k6 oferecia um desempenho competitivo em aplicativos comerciais e *desktop* sem perder desem-

penho com o cálculo de ponto flutuante, que é uma funcionalidade essencial para jogos e algumas tarefas multimídia [1].

O K6 era um processador de sexta geração, com uma arquitetura muito parecida com a dos processadores Pentium II e Pentium III. Sua memória *cache* L1 tinha capacidade de 64 KB, sendo 32 KB para dados e 32 KB para instruções, além de suportar a tecnologia MMX. Ao saber do sucesso do K6, a Intel decide antecipar o lançamento do Pentium II.

3.2.3. K6 II

Lançado em 1998, o K6-II foi o primeiro processador a incorporar um conjunto de 21 novas instruções para processamento 3D [1]. Esta funcionalidade é chamada 3DNow! A maior diferença entre o K6 II e os demais processadores é a sua frequência de operação externa. Ele utilizava 100 MHz do barramento, enquanto os outros processadores operavam com 66 MHz externos com multiplicador de 4.5x [1].

O K6-II utilizava *cache* L2 de 512 KB ou 1 MB externa ao processador, assim como os outros processadores da época. Seu tamanho variava de acordo com a placa-mãe. Possuía *cache* L1 de 64 KB (32 KB para dados e 32 KB para instruções).

3.2.4. K6 III

Lançado em fevereiro de 1999, K6 III trouxe como diferencial uma memória *cache* em 3 níveis, onde L1 e L2 localizavam-se internamente e L3 na placa-mãe. Os tamanhos das *cache* neste processador eram 64 KB de *cache* L1 (32 KB para dados e 32 KB para instruções) e 256 KB de *cache* L2.

O uso do *cache* L3 aumentou de 3 a 8% o desempenho do K6 III, dependendo da quantidade usada.

3.2.5. AMD ATHLON (K7)

Lançado em junho de 1999, o AMD Athlon foi o primeiro processador com frequência acima de 1 GHz [1]. Com esta criação, a AMD desvinculou a produção de *chips* compatíveis com os Intel. Este processador foi projetado especificamente do zero para executar sistemas *Windows* com performance excepcional [1].

Os AMD Athlon ou K7, representaram a primeira grande vitória da AMD contra a Intel. A arquitetura da memória *cache* nesses processadores era composta por 128 KB de *cache* L1 (64 KB para dados e 64 KB para instruções) e 512 KB de *cache* L2 externo ao processador.

3.2.6. AMD ATHLON XP

Lançado em 2000, o Athlon XP processava mais instruções por *clock* do que o Pentium 4 e portanto, era mais eficiente. Sua memória *cache* operava com 128 KB de *cache* L1 (64 KB para dados e 64 KB para instruções) e 512 KB de *cache*

L2. De 2000 a 2003, a AMD lançou diversas versões dos processadores Athlon, as quais são citadas a seguir:

- **ATHLON THUNDERBIRD (T-BIRD):** Lançado em junho de 2000, a AMD substituiu os 512 KB de *cache* L2 externa por 256 KB de L2 interna, que operava na mesma frequência do processador. Sua *cache* L1 continuou com 128 KB, sendo 64 KB para dados e 64 KB para instruções.
- **ATHLON XP PALOMINO:** Lançado em outubro de 2001, este processador sofreu alterações no *design* do núcleo com o objetivo de elevar em aproximadamente 10% seu desempenho se comparado com um *Thunderbird* de mesma frequência. Essa nova arquitetura consumia em torno de 20% a menos de energia que seu antecessor. Sua memória *cache* continuou com os mesmos tamanhos, sendo 128 KB para L1 e 256 KB para L2.
- **ATHLON XP THOROUGHbred:** Esta versão de processadores ainda possui duas outras versões denominadas Athlon XP *Thoroughbred* A e B. A inovação da primeira, foi o refinamento do circuito do chip, que passou a ter 130 nanômetros. A segunda versão permitiu que os Athlon XP conseguissem obter valores bem superiores de *clock*, chegando a 2.25 GHz [1]. As arquiteturas de memórias *cache* de ambas as versões, continuam idênticas aos seus antecessores.
- **ATHLON XP BARTON:** Lançado no início de 2003, este processador obteve aumento no tamanho do *cache* L2, que passou a operar com 512 KB. O *cache* L1 continuou com 128 KB.
- **ATHLON XP THORTON:** lançado em fevereiro de 2003, é uma versão do Athlon XP idêntica ao Barton, porém com redução no tamanho do *cache* L2 pela metade, voltando a ter 256 KB.

3.2.7. DURON

Lançado em 2000, este processador foi uma alternativa de baixo custo ao Athlon, do mesmo modo que a Intel fez com o Celeron. A grande diferença entre o Duron e o Athlon foi na arquitetura da memória *cache*, que teve o L2 reduzido para 64 KB, contrastando os 256 KB e 512 KB dos processadores Athlon [1].

As versões provenientes do Duron, tais como *Spitfire*, *Morgan* e *Applebred* possuem 192 KB de memória *cache*, sendo 128 KB destinados para o L1 (64 KB para dados e 64 KB para instruções) e 64 KB para o L2.

3.2.8. ATHLON 64

Lançado em setembro de 2003, este é o primeiro processador da oitava geração, além de ser o primeiro de 64 bits [16] e que manteve compatibilidade com programas de 32 bits.

A arquitetura da sua memória *cache* é composta por 64 KB de *cache* L1 para dados e 64 KB de *cache* L1 para instruções, totalizando 128 KB. O *cache* L2 aceita 512 KB ou 1MB, dependendo da versão [1]. Surgiram diversas versões mais econômicas e mais estáveis deste processador.

3.2.9. AMD ATHLON 64 X2

Lançado em 2006, o Athlon 64 X2 foi o primeiro processador de dois núcleos da AMD, competindo com o Pentium D. A maior vantagem que este processador possui sobre o da Intel é que seu controlador de memória está embutido dentro do processador, fazendo com que este controle as memórias e o multiprocessamento [1]. Essa tecnologia já era utilizada no Athlon 64.

A arquitetura de memória *cache* neste processador operava com 2 X 64 KB de *cache* L1 para e 2 x 64 KB de *cache* L1 para instruções. Já o L2, dependendo da versão, aceitava 2 x 256 KB, 2 x 512 Kb ou 2 x 1 MB. Lembrando que os valores são sempre duplicados devido à existência de dois núcleos de processamento.

3.2.10. PHENOM

Lançado em 2008, o processador Phenom opera com diversos núcleos, dependendo da sua versão. Em qualquer um deles, a *cache* de memória L1 opera com 128 KB (64 KB para dados e 64 KB para instruções) e 512 KB de *cache* L2 por núcleo, além de 2 MB de *cache* L3 compartilhada entre todos os núcleos [4]. Algumas versões dos processadores Phenom são:

- **PHENOM X4 AGENA:** Lançado em 2007, o Phenom Agena foi o primeiro processador da 10ª geração e o primeiro *quad-core* [4]. Este processador foi criticado por ser lançado após os *Core 2* da Intel e ainda assim ser menos eficiente por não suportar altas frequências [1].
- **PHENOM X3 TOLIMAN:** Lançado em 2008, era a versão *tri-core* dos processadores Phenom. Ele foi baseado no Angena, porém possuía um núcleo a menos. Dessa forma, também foi criticado por não suportar altas frequências.

3.2.11. PHENOM II

Lançado em fevereiro de 2009, O Phenom II possuía versões com dois, três e quatro núcleos, tais como:

- **PHENOM II X2 CALLISTO:** Processador com dois núcleos, sofreu avanço na arquitetura de memória *cache* em relação ao seu antecessor Phenom, ao ter sua L3

com capacidade triplicada, passando a operar com 6 MB compartilhados entre os núcleos. As *cache* L1 e L2 continuaram com 128 KB e 512 KB, respectivamente.

- **PHENOM II X3 HEKA:** Este processador operava com três núcleos e também possuía 128 KB de *cache* L1 e 512 KB de L2 por núcleo, além de 6 MB de L3 compartilhada entre os mesmos.
- **PHENOM II X4:** Possuiu as versões Zosma e Deneb, ambas operando com quatro núcleos e com as mesmas capacidades que os anteriores em sua arquitetura de memória *cache*. **PHENOM II X6 THUBAN:** Com a mesma arquitetura de memória *cache* que seus antecessores, este processador possuía seis núcleos de processamento.

Diversos outros processadores *multi-core* da AMD surgiram nos anos seguintes, se fazendo presentes até os dias atuais. Entre eles, os mais conhecidos são:

- **AMD FX 4-CORE BLACK EDITION:** Também conhecido como família FX-4000, surgiu para ser usado em servidores por serem capazes de lidar com grandes cargas de trabalhos *multithreaded* geradas nos *data centers*, porém acabaram sendo utilizados em computadores de uso geral também. Composto por quatro núcleos, sua memória *cache* opera com 4 x 16 KB de *cache* L1 para dados e 2 x 64 KB para instruções [4], totalizando 192 KB de *cache* L1. Também possui 2 x 2 MB de *cache* L2 e 8 MB de L3 compartilhada.
- **AMD FX 6-CORE BLACK EDITION:** Representante da família Fx-6000, opera com seis núcleos e é voltado para aplicações que exigem um pouco mais de desempenho. Sua arquitetura de memória *cache* é composta por 6 x 16 KB de L1 para dados e 3 x 64 KB para instruções, totalizando 288 KB de L1. Além disso, conta com 3 x 2 MB de *cache* L2 e 8 MB de *cache* L3 [4].
- **AMD FX 8-CORE BLACK EDITION:** Também conhecido como família FX-8000, este processador é composto por oito núcleos de processamento e voltado para aplicações profissionais altamente paralelizadas e para computadores equipados com uma ou mais placas de vídeo de última geração [4]. Sua memória *cache* opera com 8 x 16 KB de L1 para dados e 4 x 64 KB para instruções, além de 4 x 2 MB de L2 e 8 MB de *cache* L3.

É possível notar que a disputa entre Intel e AMD vem motivando a criação de processadores cada vez mais rápidos e com altas performances.

3.3. EVOLUÇÃO DA MEMÓRIA CACHE NOS PROCESSADORES CYRIX

Cyrix foi uma empresa fundada em 1988 localizada em Richardson no Texas e que fabricava microprocessadores. Lançou o *Cyrix DX4* para disputar com o 486 DX2 da Intel e o *Cyrix 5x86* para disputar com o K5 da AMD. Em seguida, lançou o *Cyrix 6x86*, *6x86MX* e o *M2*. A tabela a seguir mostra os processadores desenvolvidos pela *Cyrix*, bem como suas arquiteturas de memórias *cache*.

Table 1. Processadores *Cyrix*

Processadores	Cache L1
Cx486SLC	1 KB
2Cx486DLC	1 KB
Cx486SLC2	1 KB
Cx486DRX2	1 KB
Cx486SRX2	1 KB
Cx486S	2 KB
Cx486S2	2 KB
Cx486DX	8 KB
Cx486DX2	8 KB
6x86 (M1)	16 KB
6x86 MX	64 KB
MII	64 KB
MIIe	64 KB
Media GX	16 KB

É possível notar, que os processadores *Cyrix* possuíam em suas arquiteturas, apenas a *cache* L1. O *Media GX* possuía *cache* L2, porém, além de ser opcional, localizava-se na placa-mãe. Outro fator importante, é que estes processadores utilizavam a operação *write-back* na *cache*, aumentando em até 15% o desempenho das aplicações em relação à configuração *write-through*.

Em 1999, a *Cyrix* foi comprada pela *VIA Technologies*, uma tradicional fabricante de *chipsets* [15]. Juntas, produziram o *Cyrix-III*, também denominado C3.

3.3.1. *Cyrix-III* (C3)

Lançado em 2001, o C3 foi projetado com o objetivo de ter um baixo custo e baixo consumo. Este processador possuía um desempenho por ciclo de *clock* muito lento, assim como sua frequência de operação. Dessa maneira, sua única vantagem era consumir pouco e ter um preço relativamente acessível.

Existiram diversas versões do C3. A versão original, ainda chamada *Cyrix-III* não possuía *cache* L2. A atualização dessa versão, baseada no núcleo Samuel 2, incorporou 64 KB de L2, enquanto a *cache* L2 operava com 128 KB, sendo 64 KB para dados e 64 KB para instruções [14]. Aqui, a *cache* opera

com mapeamento associativo de quatro vias, isto é, ela é dividida em vários blocos, contendo quatro linhas cada. Uma vantagem desta configuração, é que é fácil de ser implementada, porém, quando todas as linhas forem ocupadas, o controlador de *cache* terá que liberar uma delas para armazenar a próxima instrução carregada do mesmo bloco de memória.

A *VIA Technologies* também lançou o C7, que embora tenha recebido algumas melhorias em relação ao C3, ainda era um processador muito simples, já que possuía apenas duas unidades de execução de inteiros e uma única unidade de execução para instruções de ponto flutuante [14]. Apesar de fabricar o C7 desde 2005 com diversas tecnologias novas, a *VIA Technologies* não obteve muito sucesso comercial.

3.4. EVOLUÇÃO DA MEMÓRIA CACHE NOS PROCESSADORES DEC

Digital Equipment Corporation (DEC), foi uma companhia estadunidense pioneira na indústria de computadores. Apesar de ser conhecida como DEC, seu nome original era DIGITAL. Nos anos 70 e 80, seus minicomputadores PDP e VAX era muito populares. A DIGITAL acabou sendo comprada pela Compaq, a qual se fundiu com a *Hewlett-Packard* (HP). Em 2006, seus produtos ainda eram produzidos, porém sob o nome HP. A tabela a seguir lista os principais modelos da DIGITAL e suas respectivas capacidades de memória *cache*.

Table 2. Processadores DEC

Modelo/Número	L1	L2	L3
EV4/21064	8KB+8KB	512KB-2MB	-
EV4S/21064	8KB+8KB	512kb-2MB	-
EV45/21064A	16KB+16KB	512KB-2MB	-
LCA4/21066	8KB+8KB	512KB-2MB	-
LCA4/21068	8KB+8KB	512KB-2MB	-
LCA45/21066A	8KB+8KB	512KB-2MB	-
LCA45/21068A	8KB+8KB	512KB-2MB	-
EV5/21164	8KB+8KB	96KB	Até 64MB
EV56/21164A	8KB+8KB	96KB	Até 64MB
PCA56/21164PC	8KB+16KB	96KB	Até 4MB
PCA57/21164PC	16KB+32 KB	96KB	Até 4 MB
EV6/21264	64KB+64KB	1MB-16MB	-
EV67/21264A	64KB+64KB	1MB-16MB	-
EV68AL/21264B	64KB+64KB	1MB-16MB	-
EV68CB/21264C	64KB+64KB	1MB-16MB	-
EV68CX/21264D	64KB+64KB	1MB-16MB	-
EV7/21364	64KB+64KB	1.75MB	-
EV7Z/21364	64KB+64KB	1.75MB	-
EV78/EV79/21364A	64KB+64KB	1.75MB	-
EV8	21464	64KB+64KB	3MB

A versão EV8 e a EV78/EV79 não chegaram a ser pro-

duzidas, pois a DEC foi comprada pela Compaq nesta época, porém, seus lançamentos estavam previstos para 2003 e 2004, respectivamente.

Na grande maioria dos processadores listados, a *cache* L1 operava com mapeamento associativo de duas vias com linhas de 64 bytes. A *cache* de dados e a L2 utilizavam *write-back* como técnicas para manutenção de coerência. As *caches* L3 eram externas ao processador e eram mapeadas diretamente.

3.5. EVOLUÇÃO DA MEMÓRIA CACHE NOS PROCESSADORES IBM

A IBM fez acordos com a Intel para ter acesso aos *designs* 386 e 486, porém sem permissão para vendê-los. Dessa maneira, a IBM desenvolveu, além do 386SLC, 486SLC2 e do *Blue Lighting*, os modelos do *PowerPC* [6]. A tabela a seguir apresenta alguns dos modelos produzidos pela IBM e as respectivas capacidades de memória *cache*.

Table 3. Processadores IBM

Modelo	L1	L2	L3
386SLC	8KB	-	-
486SLC	16KB	-	-
486SLC2	16KB	-	-
486SLC3	16KB	-	-
5x86	16KB	-	-
6x86	16KB	-	-
6x86L	16KB	-	-
6x86MX	64KB	-	-
PowerPC 601	32KB	-	-
PowerPC 603	8KB+8KB	-	-
PowerPC 604	16KB+16KB	-	-
PowerPC 620	32KB+32KB	-	-
PowerPC G3	32KB+32KB	512KB	-
PowerPC G4	64 KB	256KB - 1MB	-
PowerPC G5	64 KB	512KB - 1MB	-
Power 4	64KB+32KB	1.41MB	32MB
Power 5	64KB+32KB	1.88MB	36MB
Power 6	64KB+64KB	4MB	32MB
Power 7	32KB+32KB	256KB	32MB
Power 8	64KB+32KB	512KB	8 MB

Até a produção do *Blue Lighting*, as memórias *cache* possuíam técnicas para manutenção de coerência do tipo *write-through*. A partir dele, a IBM substituiu esta técnica para a *write-back*, melhorando a performance dos processadores futuros em torno de 30%.

O processador *Power 4* possui dois núcleos de processamento, onde cada um contém uma *cache* L1 privada conectada à uma *cache* L2 compartilhada entre eles. Esta é ligada a um interconector para o uso de mais um processador e o mesmo é ligado à L3. O último nível da memória *cache* é

ligada à outro interconector de algum outro processador, se disponível [7]. A figura 3 ilustra esta arquitetura.

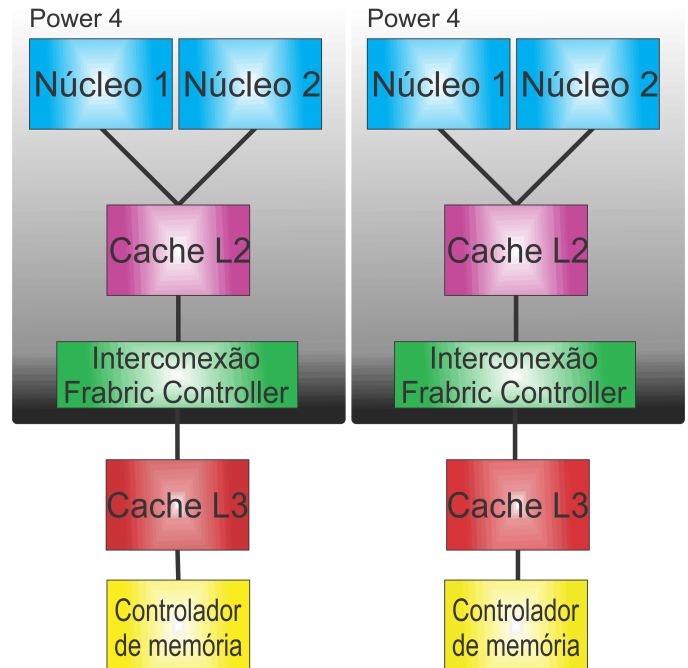


Figure 3. Arquitetura do Power 4

O processador *Power 5* também possui dois núcleos que compartilham a mesma memória *cache* L2. Esta é conectada diretamente à memória *cache* L3. Esses núcleos suportam *Simultaneous Multithreading*, podendo trabalhar com duas *threads* simultaneamente [7]. A figura 4 representa esta arquitetura.

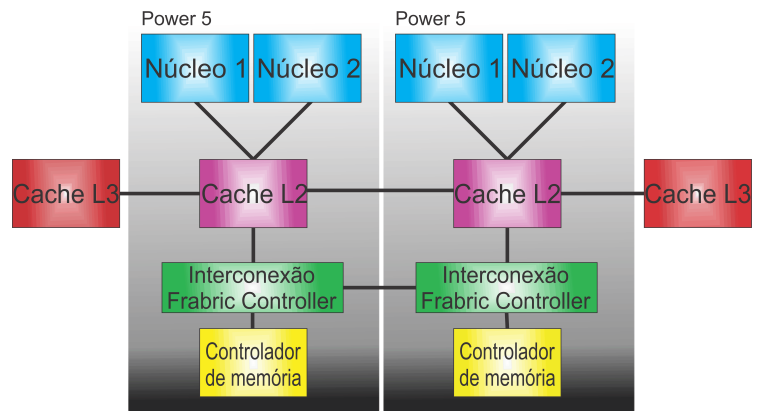


Figure 4. Arquitetura do Power 5

O processador *Power 6* também opera com dois núcleos,

nos quais existem memórias *cache* L1 e L2 privadas para cada um. Cada L2 possui uma conexão exclusiva para o controlador de memória *cache* L3 [7]. A figura 5 ilustra esta arquitetura.

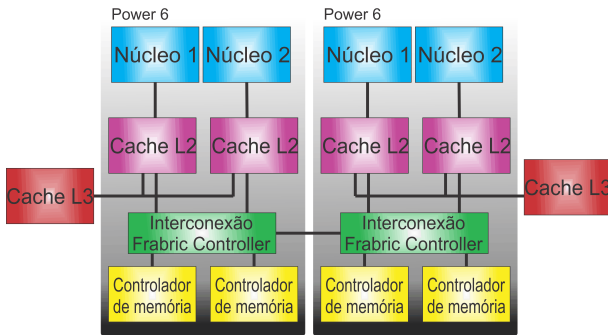


Figure 5. Arquitetura do Power 6

O processador *Power 7* está disponível nas versões com quatro, seis ou outro núcleos de processamento. O *Power 8* possui controladores de memória que são especificados para utilizar memórias DDR3 ou DDR4. Estes são controladores genéricos emparelhados a um componente externo denominado Centauro, o qual atua como uma textitcache L4. A figura 6 representa a arquitetura deste processador.

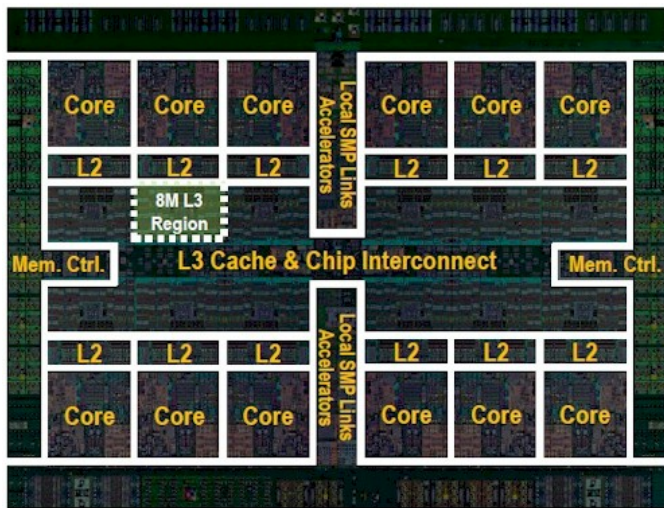


Figure 6. Arquitetura do Power 8 [8]

Houveram várias outras empresas que produziram processadores, tais como *Transmeta Crusoe*, *Texas Instruments*, *NexGen*, entre muitas outras, porém, não tiveram muito sucesso, já que o mercado estava muito competitivo com grandes marcas, como a Intel.

4. CONCLUSÃO

Este artigo apresentou um estudo sobre os conceitos fundamentais de memórias *cache*, além dos principais processadores disponíveis no mercado desde 1989, quando surgiu o primeiro computador com memória *cache*.

Como é possível notar, o avanço ocorrido nas *cache* durante o decorrer das décadas foi suficiente para colaborar com um grande aumento no desempenho dos computadores. Sabendo o limite físico dos *chips*, a dúvida que permanece para os projetistas de *hardware* é o que pode ser mudado para atingir desempenhos ainda mais altos nos processadores. Futuramente, a memória *cache* pode sofrer alterações em sua arquitetura para que esse objetivo seja atingido.

REFERENCES

- [1] R. B. Almeida. Evolução dos processadores - comparação das famílias de processadores intel e amd. Disponível em: <<http://www.ic.unicamp.br/ducatte/mo401/1s2009/T2/089065-t2.pdf>> Acessado em 23 de maio, 2009.
- [2] C. A. Alves. Microprocessadores 2. Disponível em: <http://www.feis.unesp.br/Home/departamentos/engenhariaeletrica/cap_00_micro2_programa_2014.pdf> Acessado em 25 de maio.
- [3] M. A. Z. Alves. Avaliação do compartilhamento das memórias cache no desempenho de arquiteturas multi-core. Master's thesis, Universidade Federal do Rio Grande do Sul, 2009.
- [4] AMD. Um histórico de inovação: 2000 – 2009. Disponível em: <<http://www.amd.com/pt-br/innovations/2000-2009>> Acessado em 26 de maio.
- [5] D. L. O. de Azevedo and N. M. de Oliveira. Memórias cache: Arquitetura, evolução e conceitos. Disponível em: <http://www.deinfo.ufrpe.br/sites/deinfo.ufrpe.br/files/artigos_aoc/Artigo%20-%20Di%C3%B3genes%20e%20Nicolas%20-%20cache%20> de maio.
- [6] IBM. A ibm no brasil. Disponível em: <http://www.ibm.com/br/ibm/history/ibm_brasil.phtml> Acessado em 28 de maio.
- [7] IBM. *The POWER4 Processor Introduction and Tuning Guide*. IBM Redbooks, 2001.
- [8] IBM. *Performance Optimization and Tuning Techniques for IBM Processors, including IBM Power 8*. IBM Redbooks, 2014.

- [9] M. Khan and T. Bandy. A study of recent advances in cache memories. *IEEE Conference Publications*, pages 398–403, 2014.
- [10] D. Patterson and J. Hennessy. *Organização e Projeto de Computadores*. Brasil: Editora Campus, 2005.
- [11] E. Sha. Cache memory. Disponível em: <<https://www.utdallas.edu/edsha/parallel/2010S/Cache-Memory.pdf>> Acessado em 23 de maio, 2010.
- [12] G. Silva. Microarquiteturas de alto desempenho. Disponível em: <<http://www.dcc.ufrj.br/gabriel/microarq/Hierarquia.pdf>> Acessado em 22 de maio, 2010.
- [13] A. S. Tanenbaum. *Structured Computer Organization*. Prentice Hall, 1999.
- [14] V. Technologies. Via c3 in ebga datasheet. Disponível em: <http://www.via.com.tw/download/mainboards/6/13/VIA_C3_EBGA%20datasheet110.pdf> Acessado em 27 de maio.
- [15] V. Technologies. Via corporate history. Disponível em: <<http://www.via.com.tw/en/company/history.jsp>> Acessado em 27 de maio.
- [16] F. Tuset. Evolução dos processadores. Disponível em: ><http://equipe.nce.ufrj.br/gabriel/arqcomp2/Hist%F3rico.pdf>< Acessado em 24 de maio, 2006.