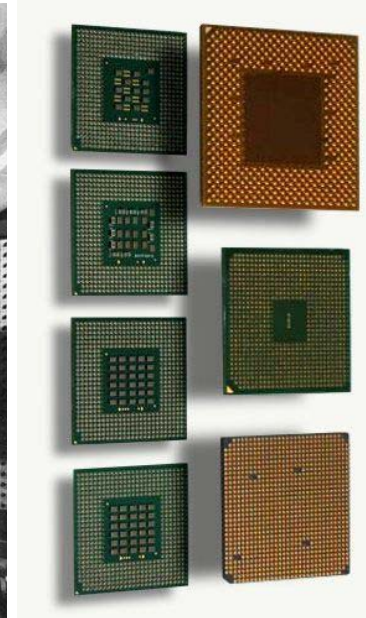
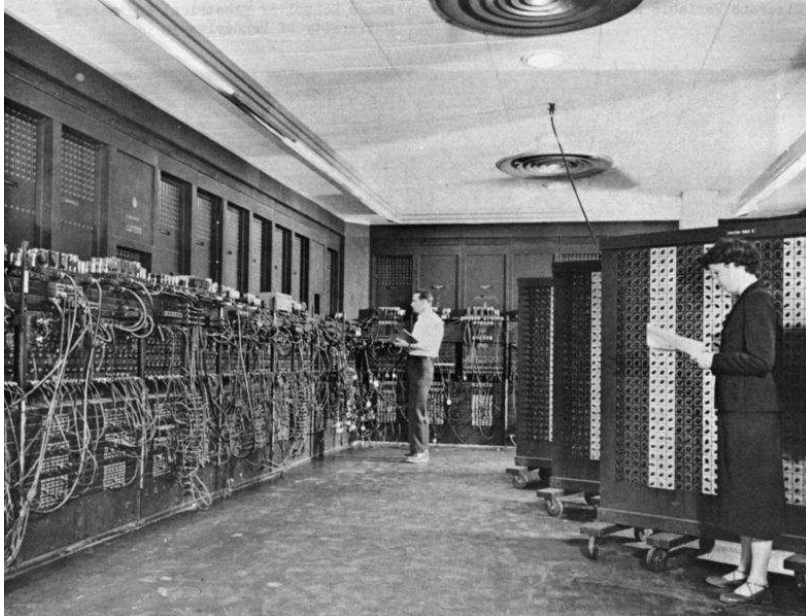


# Curso Básico de Hardware

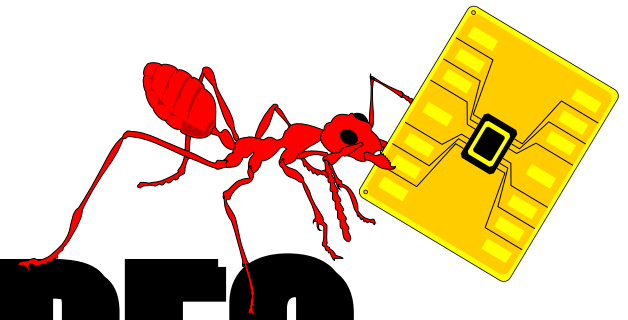


## História da Informática – das Válvulas até o Athlon 64

Revista GdH Nº01 - Jan/2007  
Revista GdH Nº03 - Mar/2007  
Revista GdH Nº06 - Jun/2007

**Carlos E. Morimoto**  
<http://www.guiadohardware.net>

# A evolução dos COMPUTADORES



**H**oje em dia, quando ouvimos falar em processadores de 2 ou 3 GHz, dá até sono, de tão comuns que eles já se tornaram. Pouca gente já ouviu falar no 8088, que foi o processador usado no PC XT, em 1981 e muito menos no Intel 4004, o primeiro microprocessador, lançado em 71.

Nas próximas páginas falarei sobre os processadores e computadores que fizeram parte da história, começando não a partir da década de 70 ou 80, mas no século XIX. Sim, na época dos nossos bisavós os computadores já existiam, apesar de extremamente rudimentares. Eram os computadores mecânicos, que realizavam cálculos através de um sistema de engrenagens, acionado por uma manivela ou outro sistema mecânico qualquer. Este tipo de sistema, comum na forma de caixas registradoras predominou até o início da década de 70, quando as calculadoras portáteis se popularizaram.

No final do século XIX surgiu o relê, um dispositivo eletromecânico, formado por um magneto móvel, que se deslocava unindo dois contatos metálicos. O relê foi muito usado no sistema telefônico, no tempo das centrais analógicas. Nas localidades mais remotas, algumas continuam em atividade até os dias de hoje.

Os relês podem ser considerados uma espécie de antepassados dos transístores. Suas limitações eram o fato de serem relativamente caros, grandes demais e ao mesmo tempo muito lentos: um relê demora mais de um milésimo de segundo para fechar um circuito.

Também no final do século XIX, surgiram as primeiras válvulas. As válvulas foram usa-

das para criar os primeiros computadores eletrônicos, na década de 40.

As válvulas tem seu funcionamento baseado no fluxo de elétrons no vácuo. Tudo começou numa certa tarde quando Thomas Edison, inventor da lâmpada elétrica estava brincando com a sua invenção. Ele percebeu que ao ligar a lâmpada ao polo positivo de uma bateria e uma placa metálica ao polo negativo, era possível medir uma certa corrente fluindo do filamento da lâmpada até chapa metálica, mesmo que não existisse contato entre eles. Havia sido descoberto o efeito termoiônico, o princípio de funcionamento das válvulas.

As válvulas já eram bem mais rápidas que os relês, atingiam frequências de alguns megahertz, o problema é que esquentavam demais, consumiam muita eletricidade e se queimavam com facilidade. Era fácil usar válvulas em rádios, que usavam poucas, mas construir um computador, que usava milhares delas era extremamente complicado, e caro.

Apesar de tudo isso, os primeiros computadores come-

çaram a surgir durante a década de 40, naturalmente com propósitos militares. Os principais usos eram a codificação e decodificação de mensagens e cálculos de artilharia.

Sem dúvida, o computador mais famoso daquela época foi o ENIAC (Electronic Numerical Integrator Analyzer and Computer), construído em 1945. O ENIAC era composto por nada menos do que 17.468 válvulas, além de 1.500 relês e um grande número de capacitores, resistores e outros componentes.

No total, ele pesava 30 toneladas e era tão volumoso que ocupava um grande galpão. Outro grave problema era o consumo elétrico: um PC típico atual, com um monitor LCD, consome cerca de 100 watts de energia, enquanto o ENIAC consumia incríveis 200 kilowatts.

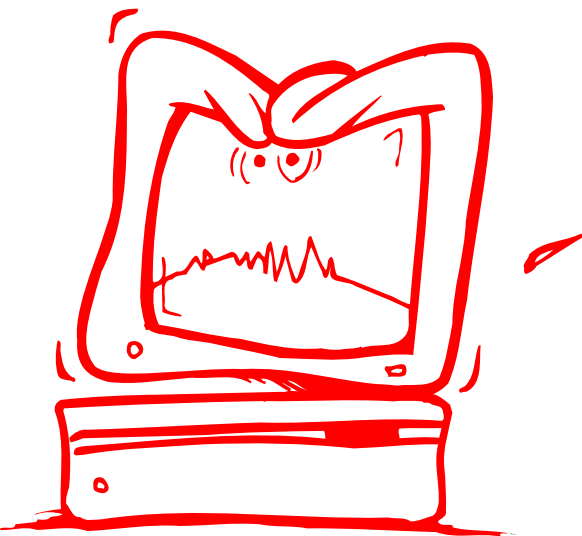
Construir este monstro, custou ao exército Americano 468.000 dólares da época, que correspondem a pouco mais de US\$ 10 milhões em valores corrigidos.

Porém, apesar do tamanho, o poder de processamento do ENIAC é ridículo para os padrões atuais, suficiente para processar apenas 5.000 adições, 357 multiplicações ou 38 divisões por segundo. O volume de processamento do ENIAC foi superado pelas calculadoras portáteis ainda na década de 70 e, hoje em dia, mesmo as calculadoras de bolso, das mais baratas, são bem mais poderosas do que ele.

A idéia era construir um computador para quebrar códigos de comunicação e realizar vários tipos de cálculos de artilharia para ajudar as tropas aliadas durante a segunda Guerra mundial. Porém, o ENIAC acabou sendo terminado exatos 3 meses depois do final da Guerra e acabou sendo usado durante a guerra fria, contribuindo por exemplo no projeto da bomba de hidrogênio.



Relê



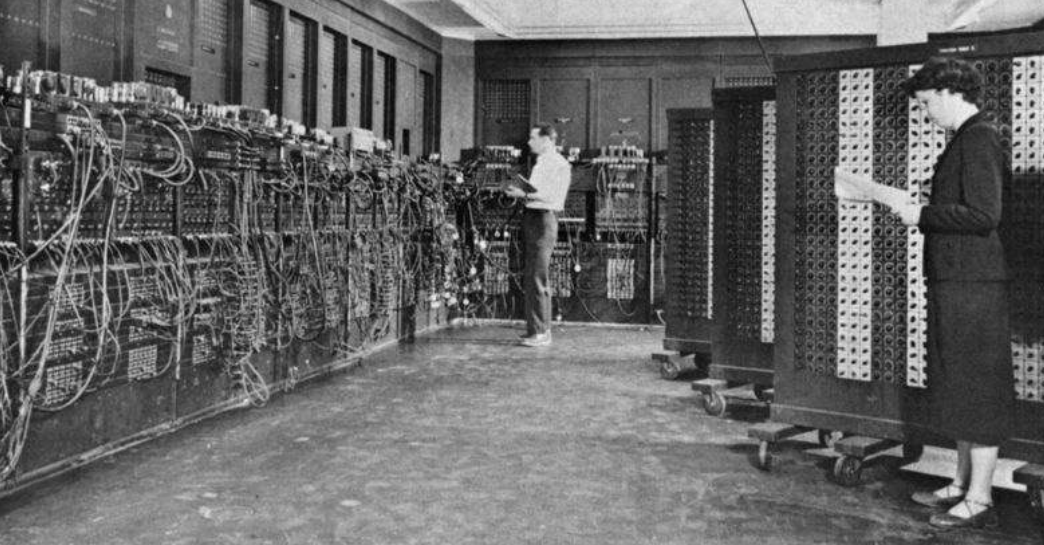


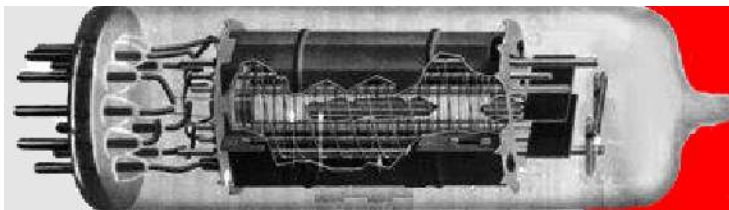
Foto do acervo do Exército dos EUA

Se você acha que programar em C ou em Assembly é complicado, imagine como era a vida dos programadores daquela época. A programação do ENIAC era feita através de 6.000 chaves manuais e, ao invés de teclas, toda a entrada de dados era feita através de cartões de cartolina perfurados, que armazenavam algumas poucas operações cada um.

Uma equipe preparava os cartões, incluindo as operações a serem realizadas, formando uma pilha, outra ia trocando os cartões no leitor do ENIAC e uma terceira "traduzia" os resultados, também impressos em cartões, para o padrão decimal.

O ENIAC também possuía sérios problemas de manutenção. Em média, a cada 5 minutos alguma das válvulas se queimava, tornando necessárias manutenções frequentes.

Abaixo está a foto de uma válvula muito usada na década de 40:



Vendo essa foto é fácil imaginar por que as válvulas eram tão problemáticas e caras: elas eram simplesmente complexas demais.

Mesmo assim, na época, as válvulas eram o que existia de mais avançado, permitindo que computadores como o ENIAC executassem em poucos segundos cálculos que um matemático equipado com uma calculadora mecânica demorava horas para executar.

Durante a década de 1940 e início da de 1950, a maior parte da indústria continuou trabalhando no aperfeiçoamento das válvulas, obtendo modelos menores e mais confiáveis. Porém, vários pesquisadores, começaram a procurar alternativas menos problemáticas.

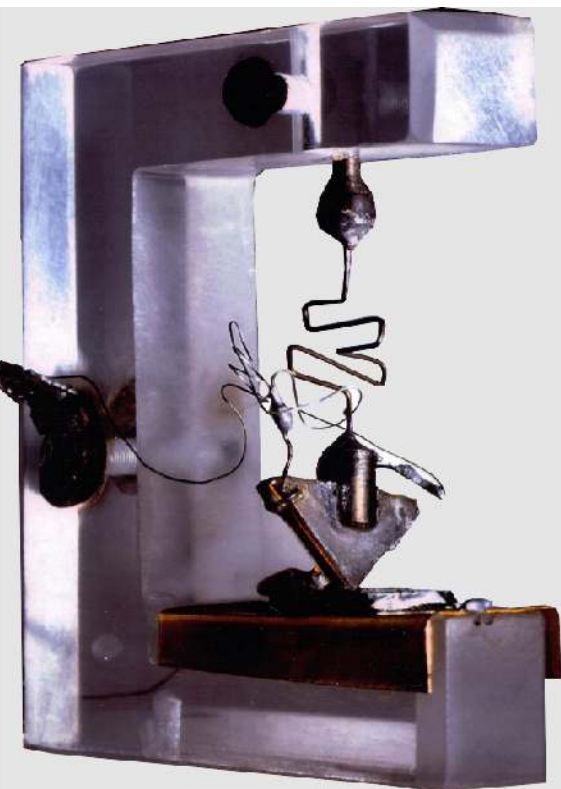
Várias destas pesquisas tinham como objetivo a pesquisa de novos materiais, tanto condutores, quanto isolantes. Os pesquisadores começaram então a descobrir que alguns materiais não se enquadravam nem em um grupo nem no outro, pois de acordo com a circunstância, podiam atuar tanto quando isolantes quanto como condutores, formando uma espécie de grupo intermediário que foi logo apelidado de grupo dos semicondutores.

Haviam encontrado a chave para desenvolver o transistor. O primeiro protótipo surgiu em 16 de dezembro de 47 (veja a foto abaixo), onde era usado um pequeno bloco de germânio (que na época era junto com

o silício o semicondutor mais pesquisado) e três filamentos de ouro. Um filamento era o polo positivo, o outro o polo negativo, enquanto o terceiro tinha a função de controle. Tendo apenas uma carga elétrica no polo positivo, nada acontecia, o germânio atuava como um isolante, bloqueando a corrente. Porém, quando uma

certa tensão elétrica era aplicada usando o filamento de controle, um fenômeno acontecia e a carga elétrica passava a fluir para o polo negativo. Haviam criado um dispositivo que substituíra a válvula, que não possuía partes móveis, gastava uma fração da eletricidade gasta por uma e, ao mesmo tempo, era muito mais rápido.





Este primeiro transistor era muito grande, mas não demorou muito para que este modelo inicial fosse aperfeiçoado. Durante a década de 50, o transistor foi aperfeiçoado e passou a gradualmente dominar a indústria, substituindo rapidamente as problemáticas válvulas. Os modelos foram diminuindo de tamanho, caindo de preço e tornando-se mais rápidos. Alguns transistores da época podiam operar a até 100 MHz. Claro que esta era a frequência

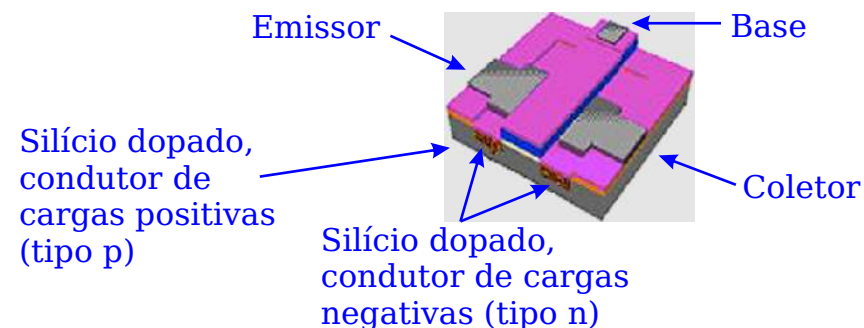
que podia ser alcançada por um transistor sozinho, nos computadores da época, a frequência de operação era muito menor, já que em cada ciclo de processamento o sinal precisa passar por vários transistores.

Mas, o grande salto foi a substituição do germânio pelo silício. Isto permitiu miniaturizar ainda mais os transistores e baixar seu custo de produção. Os primeiros transistores de junção comercial (já similares aos atuais) foram produzidos partir de 1960 pela Crystalonics, decretando o final da era das válvulas.

A idéia do uso do silício para construir transistores é que adicionando certas substâncias em pequenas quantidades é possível alterar as propriedades elétricas do silício. As primeiras experiências usavam fósforo e boro, que transformavam o silício em condutor por cargas negativas ou condutor por cargas positivas, dependendo de qual dos dois materiais fosse usado. Estas substâncias adicionadas ao silício são chamadas de impurezas, e o silício “contaminado” por elas é chamado de silício dopado.

O funcionamento de um transistor é bastante simples, quase elementar. É como naquele velho ditado “as melhores invenções são as mais simples”. As válvulas eram muito mais complexas que os transistores e mesmo assim foram rapidamente substituídas por eles.

Um transistor é composto basicamente de três filamentos, chamados de base, emissor e coletor. O emissor é o polo positivo, o coletor o polo negativo, enquanto a base é quem controla o estado do transistor, que como vimos, pode estar ligado ou desligado. Veja como estes três componentes são agrupados num transistor moderno:



Quando o transistor está desligado, não existe carga elétrica na base, por isso, não existe corrente elétrica entre o emissor e o coletor. Quanto é aplicada uma certa tensão na base, o circuito é fechado e é estabelecida a corrente entre o emissor e o receptor.

Cada transistor funciona como uma espécie de interruptor, que pode estar ligado ou desligado, como uma torneira que pode estar aberta ou fechada, ou mesmo como uma válvula. A diferença é que o transistor não tem partes móveis como uma torneira e é muito menor, mais barato, mais durável e muito mais rápido que uma válvula.

A mudança de estado de um transistor é feito através de uma corrente elétrica. Esta mudança de estado por sua vez pode comandar a mudança de estado de vários outros transistores

ligados ao primeiro, permitindo ao processador dados. Num transístor esta mudança de estado pode ser feita bilhões de vezes por segundo, porém, a cada mudança de estado é consumida uma certa quantidade de eletricidade, que é transformada em calor. É por isso que quanto mais rápidos tornam-se os processadores, mais eles se aquecem e mais energia consomem.

Um 386, por exemplo, consumia pouco mais de 1 watt de energia e podia funcionar sem nenhum tipo de resfriamento. Um 486DX-4 100 consumia cerca de 5 Watts e precisava de um cooler simples, enquanto Athlon 64 chega a consumir 80 Watts de energia e precisa de no mínimo um bom cooler para funcionar bem. Em compensação, a versão mais rápida do 386 operava a apenas 40 MHz,



Intel 4004

enquanto os processadores atuais já superaram a barreira dos 3.0 GHz.

O grande salto veio quando descobriu-se que era possível construir vários transístores sobre o mesmo wafer de silício. Isso permitiu diminuir de forma gritante o custo e tamanho dos computadores. Entramos então na era do microchip.

O primeiro microchip comercial foi lançado pela Intel em 1971 e chamava-se 4004. Como o nome sugere, ele era um processador que manipulava palavras de apenas 4 bits (embora já trabalhasse com instruções de 8 bits). Ele era composto por pouco mais de 2000 transístores e operava a apenas 740 kHz. Embora fosse muito limitado, ele foi muito usado em calculadoras, área em que representou uma pequena revolução. Mais do que isso, o sucesso do 4004 mostrou a outras empresas que os microchips eram viáveis, criando uma verdadeira corrida evolucionária, em busca de processadores mais rápidos e avançados.

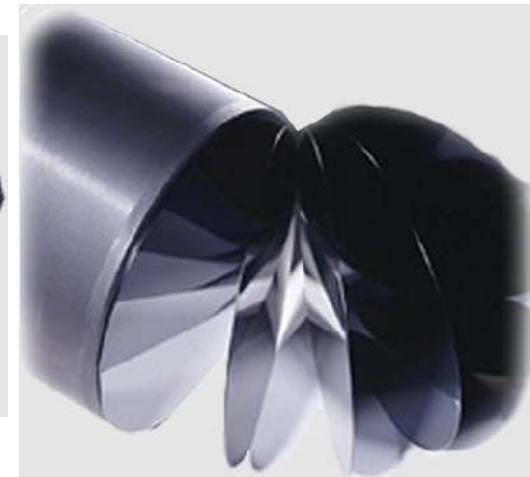
Em 1972 surgiu o Intel 8008, o primeiro processador de 8 bits e, em 1974, foi lançado o Intel 8080, antecessor do 8088, que foi o processador usado nos primeiros PCs. Em 1977 a AMD passou a vender um clone do 8080, inaugurando a disputa Intel x AMD, que continua até os dias de hoje.



8080, da AMD

## Como são fabricados os processadores

O componente básico para qualquer chip é o wafer de silício que é obtido através da fusão do silício junto com os materiais que permitirão sua dopagem posteriormente. Inicialmente são produzidos cilindros, com de 20 a 30 centímetros de diâmetro, que posteriormente são cortados em fatias bastante finas:



Estas “fatias” por sua vez são polidas e tratadas, obtendo os wafers de silício. A qualidade do wafer determinará o tipo de chip que poderá ser construído com base nele.

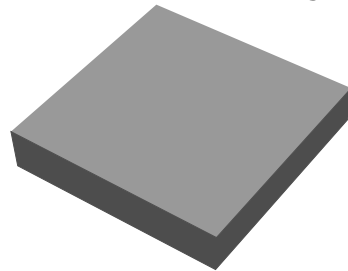
Wafers de baixa qualidade, usados para para construir circuitos rudimentares, com poucos milhares de transistores podem ser comprados a preços bastante baixos, a partir de milhares de fornecedores diferentes. Entretanto, para produzir um processador moderno, é preciso utilizar wafers de altíssima qualidade, que são extremamente caros.

Embora o silício seja um material extremamente barato e abundante, toda a tecnologia necessária para produzir os wafers faz com que eles estejam entre os produtos mais caros produzidos pelo homem. Cada wafer de 30 centímetros custa de mais de 20 mil dólares para um fabricante como a Intel, mesmo quando comprados em grande quantidade.

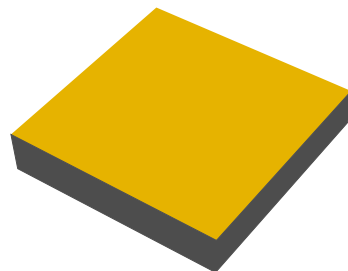
Cada wafer é usado para produzir vários processadores, que no final da produção são separados e encapsulados individualmente. Não seria possível mostrar todos os

processos usados na fabricação de um processador, mas para lhe dar uma boa idéia de como eles são produzidos, vou mostrar passo a passo a construção de um único transistor. Imagine que um Core 2 Duo possui 291 milhões de transistores e cada wafer permite produzir algumas centenas de processadores.

Tudo começa com o wafer de silício em seu estado original:

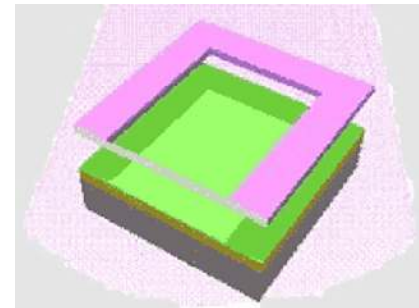


A primeira etapa do processo é oxidar a parte superior do wafer, transformando-a em dióxido de silício. Isto é obtido expondo o wafer a gases corrosivos e altas temperaturas. A fina camada de dióxido de silício que se forma é que será usada como base para a construção do transistor.



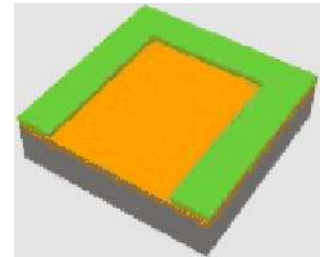
Em seguida é aplicada uma camada bastante fina de um material fotosensível sobre a camada de dióxido de silício.

Usando uma máscara especial, é jogada luz ultravioleta apenas em algumas áreas da superfície. Esta máscara tem um padrão diferente para cada área do processador, de acordo com o desenho que se pretende obter. A técnica usada aqui é chamada de litografia óptica. Existem várias variações da tecnologia, como a EUVL (Extreme Ultra Violet Lithography), usada nos processadores atuais. Quanto mais avançada a técnica usada, menores são os transistores, permitindo o desenvolvimento de processadores mais complexos e rápidos.

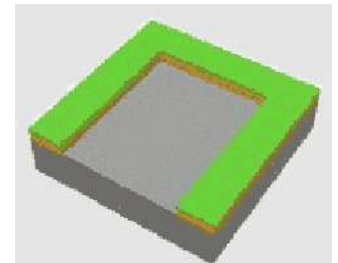


A camada fotosensível é originalmente sólida, mas ao ser atingida pela luz ultravioleta transforma-se numa substância gelatinosa, que pode ser facilmente removida.

Depois de remover as partes moles da camada fotosensível, temos algumas áreas do dióxido de silício expostas, e outras que continuam cobertas pelo que restou da camada:

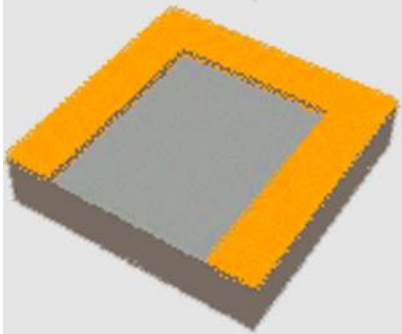


O wafer é banhado com um produto especial que remove as partes do dióxido de silício que não estão protegidas pela camada fotosensível. O restante continua intacto.

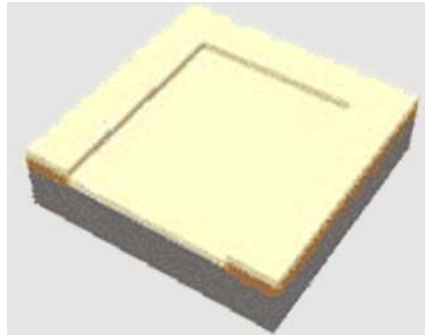




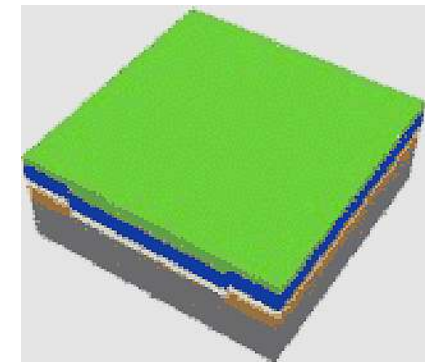
Finalmente, é removida a parte que restou da camada fotosensível. Note que como temos substâncias diferentes é possível remover uma camada de cada vez, ora o dióxido de silício, ora a própria camada fotosensível. Com isto é possível “desenhar” as estruturas necessárias para formar os transistores. Temos aqui pronta a primeira camada. Cada transistor é formado para várias camadas, dependendo do



projeto do processador. Neste exemplo, temos um transistor simples, de apenas quatro camadas, mas os processadores atuais utilizam um número muito maior de camadas, mais de vinte em alguns casos, dependendo da densidade que o fabricante pretende alcançar.

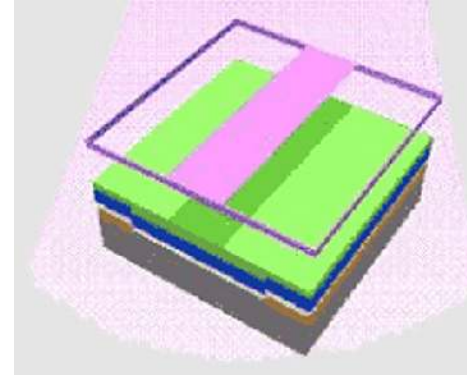


Começa então a construção da segunda camada do transistor. Inicialmente o wafer passa novamente pelo processo de oxidação inicial, sendo coberto por uma nova camada (desta vez bem mais fina) de dióxido de silício. Note que apesar da nova camada de dióxido, o desenho conseguido anteriormente é mantido.

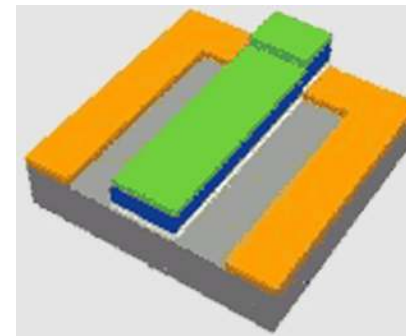


Em seguida é aplicada sobre a estrutura uma camada de cristal de silício. Sobre esta é aplicada uma nova camada de material fotosensível.

Novamente, o wafer passa pelo processo de litografia, desta vez utilizando uma máscara diferente.

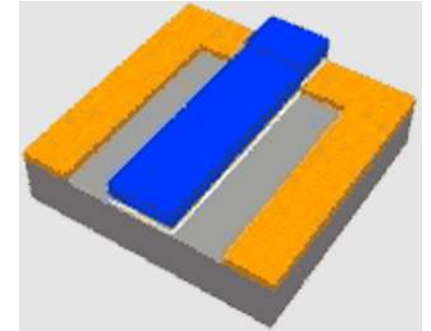


Novamente, a parte da camada fotosensível que foi exposta à luz é removida, deixando expostas partes das camadas de cristal de silício e dióxido de silício, que são removidas em seguida.

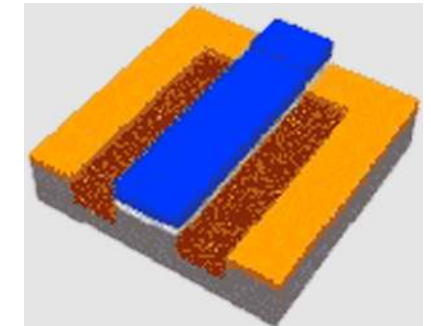


Como na etapa anterior, o que restou da camada fotosensível é removida. Terminamos a

construção da segunda camada do transistor.

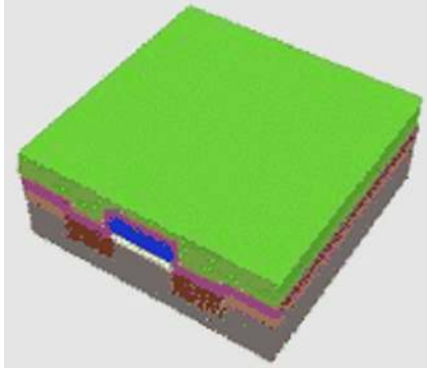


Chegamos a uma das principais etapas do processo de fabricação, que é a aplicação das impurezas, que transformarão partes do wafer de silício num material condutor. Estas impurezas também são chamadas de íons. Note que os íons aderem apenas à camada de silício que foi exposta no processo anterior e não nas camadas de dióxido de silício ou na camada de cristal de silício.

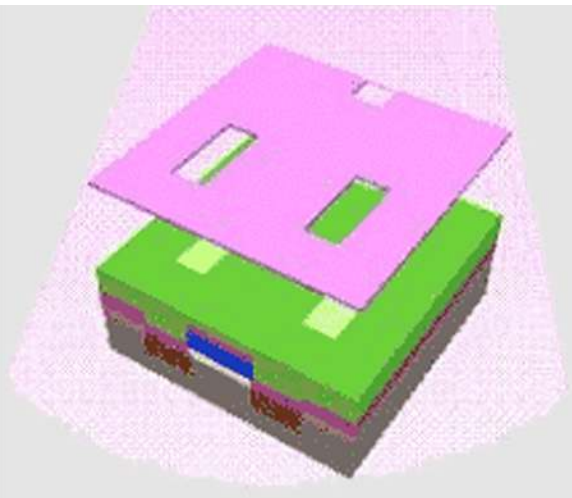




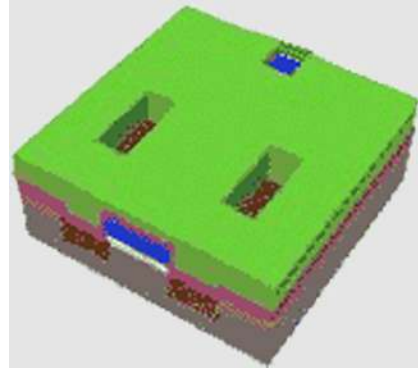
É adicionada então uma terceira camada, composta de um tipo diferente de cristal de silício e novamente é aplicada a camada fotosensível sobre tudo.



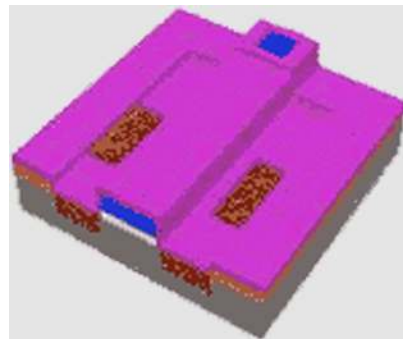
O wafer passa novamente pelo processo de litografia, usando mais uma vez uma máscara diferente.



As partes do material fotosensível expostas à luz são removidas, expondo partes das camadas inferiores, que são removidas em seguida.

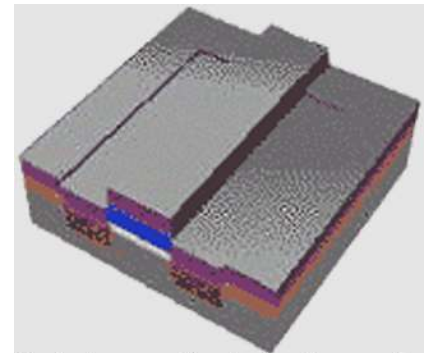


Temos agora pronta a terceira camada do transístor. Veja que a estrutura do transístor já está quase pronta, faltando apenas os três filamentos condutores.

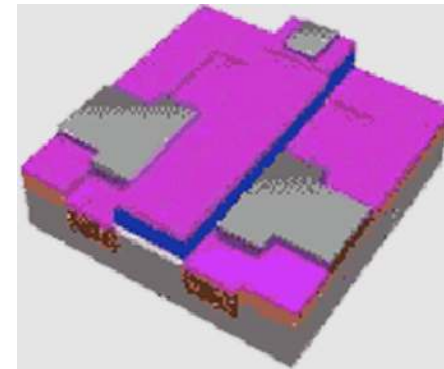


Uma finíssima camada de metal é aplicada sobre a estrutura anterior. Nos processadores atuais, que são produ-

zidos através de uma técnica de produção de 0.065 micron, esta camada metálica tem o equivalente a apenas 3 átomos de espessura.



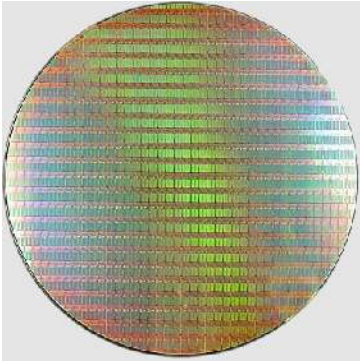
O processo de aplicação da camada fotosensível, de litografia e de remoção das camadas é aplicado mais uma vez, com o objetivo de remover as partes indesejadas da camada de metal. Finalmente temos o transístor pronto.



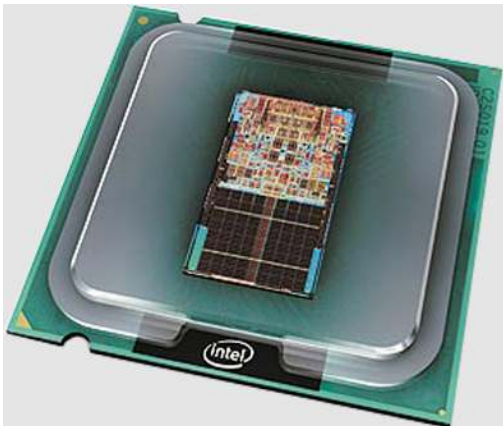
Cada processador é constituído por vários milhões de transístores, divididos em diversos grupos de componentes, entre eles as unidades de execução (onde as instruções são realmente processadas) e os caches. Como todo processador atual processa várias instruções por ciclo, são incluídos diversos circuitos adicionais, que organizam e ordenam as instruções, de forma a aproveitar da melhor maneira possível os recursos disponíveis.

No final do processo, toda a área do wafer é coberta por processadores. Destes, muitos acabam sendo descartados, pois qualquer imperfeição na superfície do wafer, partícula de poeira, ou anomalia durante o processo de litografia acaba resultando numa área defeituosa. Temos também os

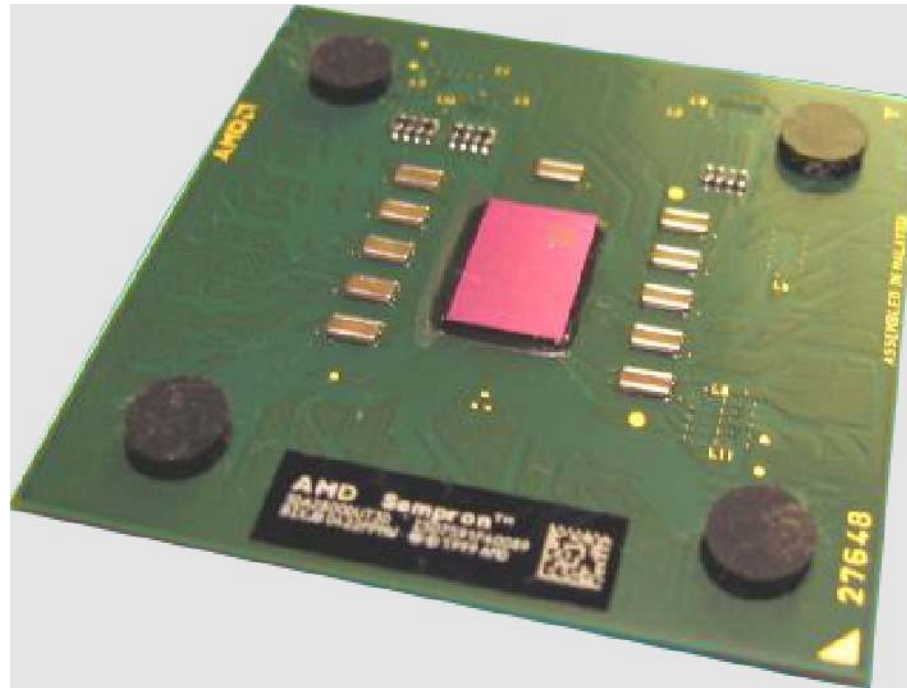
processadores "incompletos", que ocupam as bordas do wafer; que também são descartados:



Cada processador é testado individualmente, através de um processo automático. O wafer é finalmente cortado e os processadores "bons" são finalmente encapsulados, ou seja, instalados dentro da estrutura que os protege e facilita o manuseio e instalação:



O formato do encapsulamento varia de processador para processador. Geralmente temos um spreader, ou seja, uma proteção de metal sobre o die do processador, que fica entre ele e o cooler. Entretanto em muitos processadores, como os Athlons, Durons e Semprons antigos, é usado um encapsulamento mais simples, onde a parte central é a própria parte inferior do wafer de silício, exposta para melhorar a dissipação de calor. Nestes casos, é preciso redobrar os cuidados na hora de instalar e remover o cooler, pois qualquer dano ao núcleo será suficiente para inutilizar o processador:



Só a título de curiosidade, o Intel 4004 era produzido numa técnica de 10 micra, onde cada transistor mede o equivalente a 1/100 de milímetro. Parece pouco, mas estes transistores parecem pirâmides se comparados aos atuais.

O 486 já foi produzido numa técnica de 1 micron, onde cada transistor ocupa uma área 100 vezes menor. Enquanto o 4004 tinha apenas 2.000 transistores, o 486 tinha um milhão deles.

Como a velocidade de operação do transistor está diretamente relacionada a seu tamanho, o 486 é também brutalmente mais rápido. Enquanto o 4004 opera a 740 kHz, o 486 atingiu 100 MHz (nas versões fabricados pela Intel).

Mas, isso não é nada se comparado com os processadores atuais. Um Core 2 Duo X6800 é fabricado numa técnica de 0.065 micron (237 vezes menores que os do 486!), possui 291 milhões de transistores e opera a 2.93 GHz.

Estão previstos processadores fabricados numa técnica de 0.045 micron em 2008 e 0.032 micron em 2010. Depois disso não se sabe até onde a tecnologia poderá evoluir, pois os fabricantes estão se aproximando dos limites da matéria. A 0.032 micron já temos transistores ocupando uma área equivalente a poucas centenas de átomos de silício.



# Os supercomputadores

Nas décadas de 1940 e 1950, todos os computadores do mundo eram gigantescos e caros, agregando tudo o que havia mais avançado em termos de conhecimento humano. Pois bem, vindo de hoje, pode parecer ridículo que qualquer calculadora de mão de 3 reais possa ter um poder de processamento muito superior ao de um ENIAC, que só de manutenção consumia o equivalente a quase 200.000 dólares por dia (em valores corrigidos). Mas, os supercomputadores continuam existindo, tão grandes e caros quanto um ENIAC, porém incomparavelmente mais rápidos do que os micros de mesa, como o que você está usando neste exato momento.

Estes mastodontes que estão por trás de muitos dos avanços da humanidade, que apesar de estarem escondidos em grandes salas refrigeradas são alvo de grande curiosidade.

Enquanto escrevo, o supercomputador mais rápido do planeta (segundo o <http://www.top500.org/>) é o IBM Blue Gene/L, desenvolvido pela IBM. Ele é composto por nada menos do que 131.072 processadores PowerPC e possui 32 terabytes de memória RAM.

Para chegar a estes números, a IBM desenvolveu módulos relativamente simples, cada um contendo 2 processadores, 512 MB de RAM e uma interface de rede gigabit Ethernet, similares a um PC doméstico. Estes módulos foram agrupados em racks (chamados de nós), cada um com 128 deles. No final, chegaram a 512 racks, interligados por uma complexa malha de cabos de rede, rodando um software próprio de gerenciamento. Esta gigantesca estrutura funciona como um cluster, onde o processamento é dividido em pequenos pedaços e dividido entre os módulos.

Veja uma foto mostrando parte das instalações, publicada com autorização da IBM:



Os primeiros supercomputadores começaram a surgir na década de 60, alias uma década de muitos avanços, já que no final da década de 50 foi feita a transição das válvulas para os transistores. Cada transistor era centenas de vezes menor que uma válvula, era muito mais durável e tinha a vantagem de gerar pouco calor.

Todos os computadores da década de 60 já utilizavam transistores, o que permitiu o

desenvolvimento dos primeiros minicomputadores. Naquela época, minicomputador era qualquer coisa do tamanho de um armário, com uma capacidade de processamento inferior ao de uma agenda eletrônica atual, das mais baratas.

Os computadores de grande porte, porém, continuaram a ser desenvolvidos, passando a ser chamados de supercomputadores. O primeiro supercomputador para fins comer-



ciais foi o CDC 6600, que foi seguido pelos IBM 360/95 e 370/195.

Na década de 70 surgiu uma nova revolução: o microchip. Um microchip sozinho oferecia uma capacidade de processamento equivalente à de um minicomputador, mas em compensação era escandalosamente menor e mais barato. Surgiram então os primeiros microcomputadores.

Os supercomputadores da década de 70 já eram centenas de vezes mais poderosos do que os produzidos uma década antes. Os principais modelos foram o CDC 7600, o BSP, produzido pela Burroughs e o ASC da Texas Instruments.

Estes sistemas atingiram a marca de 100 megaflops, ou seja, 100 milhões de cálculos de ponto flutuante por segundo. Esta é a mesma capacidade de processamento de um Pentium 60, porém atin- gida 20 anos antes :).

No final da década de 70 sugiram os supercomputa- dores Cray, produzidos pela Seymour. O primeiro da linha, chamado de Cray 1, também processava 100 megaflops, porém o Cray-XMP atingiu a

incrível marca de 1 gigaflop, ainda no início da década de 80. Esta é uma capacidade de processamento próxima à de um Pentium II 350.

Só para efeito de comparação, o Blue Gene/L, que citei a pouco, possui 360 teraflops de poder de processamento, ou seja, é 360 mil vezes mais rápido.

Apesar de mesmo um "PC de baixo custo" atualmente possuir um poder de processamento superior ao de um supercom- putador que a 15 anos atrás custava 5 milhões de dólares, a demanda por sistemas cada vez mais rápidos continua.

As aplicações são várias, englobando principalmente pesquisas científicas, aplicações militares diversas e vários tipos de aplicativos financeiros e relacionados à Internet, aplicativos que envolvem uma quantidade absurda de processamento, e claro, envolvem instituições que podem pagar muito mais do que 5 ou 10 mil dólares por um computador o mais rápido possível. Existindo demanda... aparecem os fornecedores.

Atualmente, todos os supercomputadores são construídos com base em praticamente os mesmos componentes que temos em micros de mesa, memória, HDs, e processadores, Intel, IBM ou AMD.

Ao invés de usar apenas um disco rígido IDE, como num micro de mesa, um supercomputador utiliza um array de centenas de HDs, sistemas semelhantes ao RAID, mas numa escala maior, que permitem gravar dados de forma fragmentada em vários discos e ler os pedaços simultaneamente a partir de vários HDs, obtendo taxas de transferência muito altas.

Processadores e memória RAM geralmente são agrupados em nós, cada nó engloba de um a quatro processadores e uma certa quantidade de memória RAM e cache. Isso garante que os processadores tenham um acesso à memória tão rápido quanto um PC de mesa. Os nós por sua vez são interligados através de interfaces de rede, o que os torna partes do mesmo sistema de processamento. Como

neurônios interligados para formar um cérebro. Um nó sozinho não tem uma capacidade de processamento tão surpreendente assim, mas ao interligar algumas centenas, ou milhares de nós a coisa muda de figura.

Uma opção mais barata para instituições que precisam de um supercomputador, mas não possuem muito dinheiro disponível, é usar um sistema de processamento distribuído, ou cluster. Um cluster formado por vários PCs comuns ligados em rede.

O exemplo mais famoso de processamento distribuído foi o projeto Seti@Home, onde cada voluntário instalava um pequeno programa que utili- zava os ciclos de processa- mento ociosos da máquina



para processar as informações relacionadas ao projeto. Os pacotes de dados de 300 KB cada chegavam pela Internet e demoravam várias horas para serem processados. Isso permitiu que mais de 2 milhões de pessoas, muitas com cone-xão via modem participassem do projeto. O sistema montado pela Seti@Home foi considerado por muitos o supercomputador mais poderoso do mundo, na época.

Este tipo de sistema pode ser construído usando por exemplo a rede interna de uma empresa. Rodando o software adequado, todos os micros podem fazer parte do sistema, alcançando juntos um poder de processamento equivalente ao de um supercomputador. O mais interessante é que estes PCs poderiam ser usados normalmente pelos funcionários, já que o programa rodaria utilizando apenas os ciclos ociosos do processador.

A tecnologia de cluster mais usada atualmente são clusters Beowulf, formados por vários computadores interligados em rede. Não é necessário nenhum hardware muito sofisticado: um grupo de PCs parrudos, ligados através de uma rede gigabit já é o suficiente para montar um cluster Beowulf capaz de rivalizar com muitos supercomputadores em poder de processamento. A idéia é criar um sistema de baixo custo, que possa ser utilizado por universidades e pesquisadores com poucos recursos.

O primeiro cluster Beowulf foi criado em 1994 na CESDIS, uma subsidiária da NASA e era formado por 16 PCs 486 DX-100 ligados em rede. Para manter a independência do sistema e baixar os custos, os desenvolvedores optaram por utilizar o Linux.

Estes clusters não servem para processar dados em tempo real (um game qualquer por exemplo), mas apenas para processar grandes quantidades de dados, que podem ser quebrados em pequenas partes e divididos entre os vários computadores. Uma área onde são populares é na aplicação de efeitos especiais e renderização de imagens para filmes de cinema. Há inclusive casos de filmes como Shrek e Final

Fantasy que foram feitos inteiramente em clusters Beowulf.

## A evolução dos computadores pessoais

Até aqui, falei sobre os supercomputadores e sobre a evolução dos processadores, que evoluíram das válvulas para o transistor e depois para o circuito integrado. Vou agora falar um pouco sobre os primeiros computadores pessoais, que começaram a fazer sua história a partir da década de 70. Tempos difíceis aqueles :).

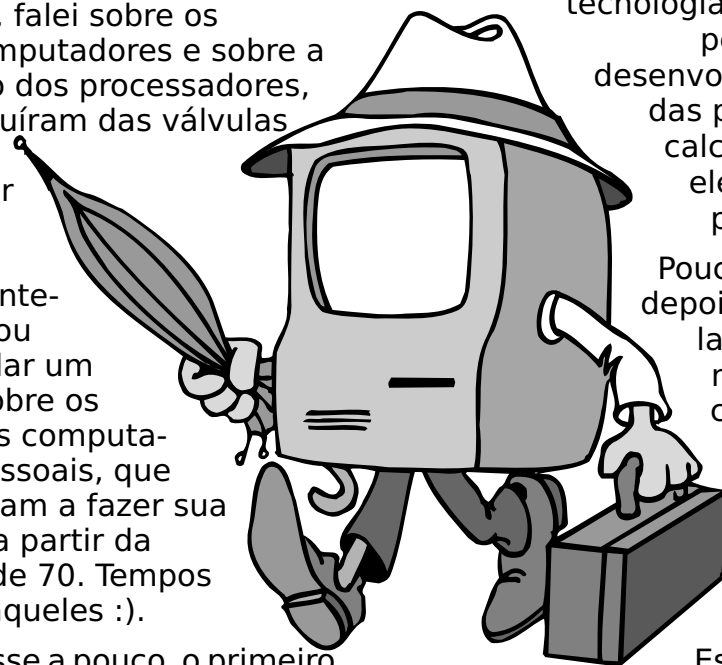
Como disse a pouco, o primeiro microchip, o 4004, foi lançado pela Intel em 71. Era um projeto bastante primitivo, que processava instruções de 8 bits, através de um barramento rudimentar, que permitia transferir apenas 4 bits por

ciclo, e operava a meros 740 kHz. Na verdade, o 4004 era tão lento que demorava 10 ciclos para processar cada instrução, ou seja, ele processava apenas 74 mil instruções por segundo (cerca de 15 vezes mais rápido que o ENIAC). Hoje em dia esses números parecem piada, mas na época era a última palavra em

tecnologia. O 4004 permitiu o desenvolvimento das primeiras calculadoras eletrônicas portáteis.

Pouco tempo depois, a Intel lançou um novo processador, que fez sucesso durante muitos anos, o 8080.

Este já era um processador de 8 bits e operava a incríveis 2 MHz: "Ele é capaz de endereçar até 64 KB de memória e é rápido, muito rápido!" como dito num anúncio publicitário do Altair 8800 que, lançado em 1974, é



considerado por muitos o primeiro computador pessoal da história.

O Altair era baseado no 8080 da Intel e vinha com apenas 256 bytes de memória, realmente bem pouco, mesmo para os padrões da época. Estava disponível também uma placa de expansão para 4 KB. Em teoria, seria possível instalar até 64 KB, mas o custo tornava o upgrade inviável.

No modelo básico, o Altair custava apenas 439 dólares, na forma de kit (onde você precisava soldar manualmente todos os componentes). Em valores corrigidos, isso equivale a quase 4.000 dólares, mas na época esse valor foi considerado uma pechincha, tanto que foram vendidas 4.000 unidades em 3 meses, depois de uma matéria da revista Popular Eletronics.

Esse “modelo básico” consistia nas placas, luzes, chips, gabinete, chaves e a fonte de alimentação, junto claro com um manual que ensinava como montar o aparelho. Existia a opção de comprá-lo já montado, mas custava 182 dólares (da época) a mais.

Em sua versão básica, o Altair não tinha muita utilidade prática, a não ser a de servir como fonte de aprendizado de eletrônica e programação. Entretanto, pouco tempo depois, começaram a surgir vários acessórios para o Altair: um teclado que substituíam o conjunto de chaves que serviam para programar o aparelho, um terminal de vídeo (bem melhor que ver os resultados na forma de luzes :), um drive de disquetes (naquela época ainda se usavam disquetes de 8 polegadas), placas de expansão de memória e até um modelo de impressora. Até mesmo Bill Gates (antes mesmo da fundação da Microsoft) participou, desenvolvendo uma versão do Basic para o Altair.

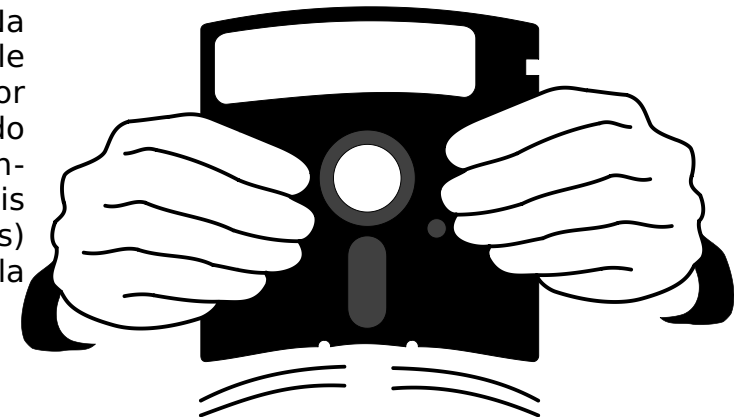
Se você tivesse muito dinheiro, era possível chegar a algo que se parecia com um computador moderno, capaz de editar textos e criar planilhas rudimentares. Algumas empresas perceberam o nicho e passaram a vender versões "completas" do Altair, destinadas ao uso em empresas, como neste anúncio, publicado na revista Popular Eletronics, onde temos

um Altair "turbinado", com o terminal de vídeo, impressora, dois drives de disquete e 4 KB de memória:



O Altair serviu para demonstrar a grande paixão que a informática podia exercer e que, ao contrário do que diziam muitos analistas da época, existia sim um grande mercado para computadores pessoais.

Pouco depois, em 1976, foi fundada a Apple, tendo como sócios Steve Jobs (que continua ativo até os dias de hoje) e Steve Wozniak. Na verdade, a Apple só foi fundada por que o projeto do Apple I (desenvolvido pelos dois nas horas vagas) foi recusado pela





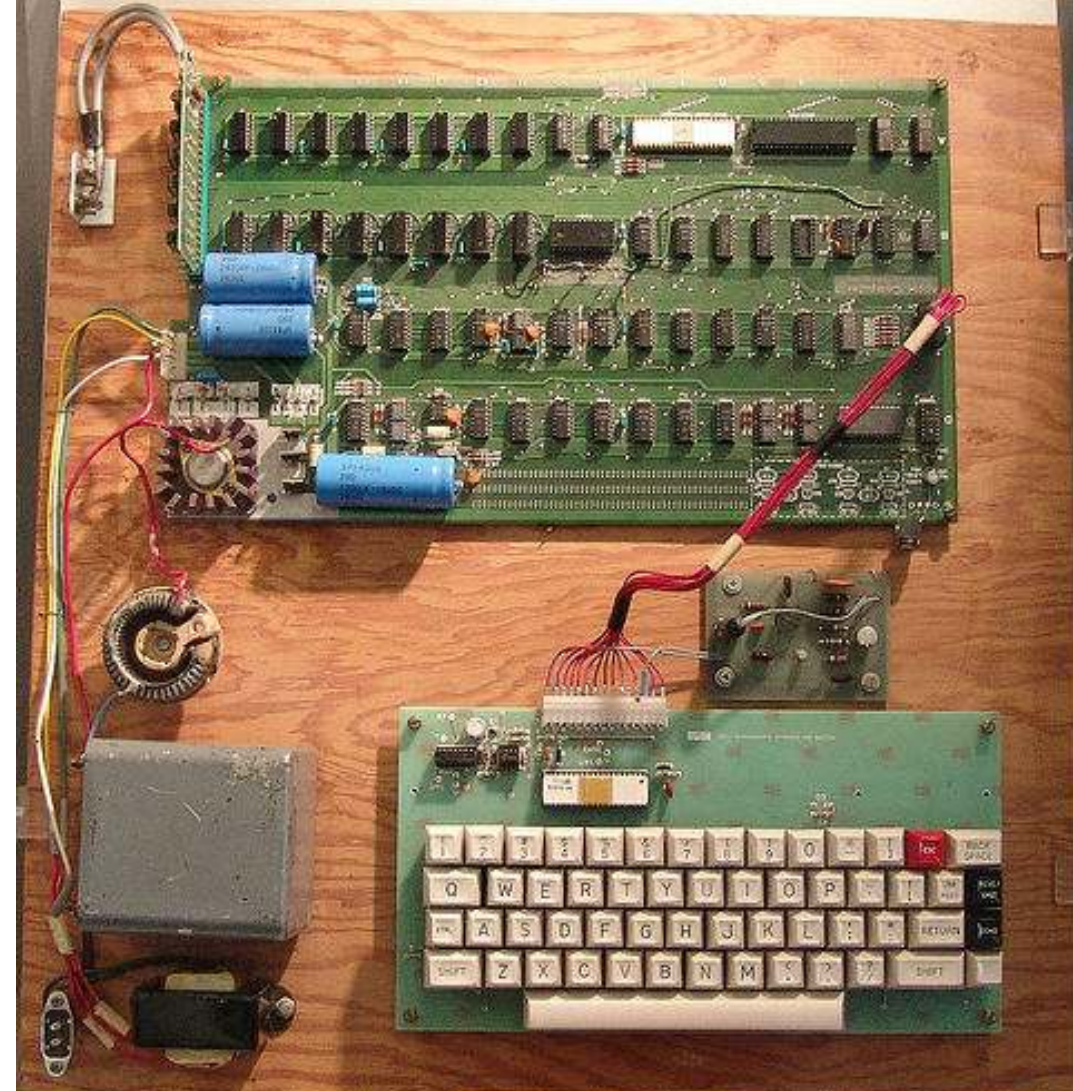
Atari e pela HP. Uma frase de Steve Jobs descreve bem a história: “Então fomos à Atari e dissemos “Ei, nós desenvolvemos essa coisa incrível, pode ser construído com alguns dos seus componentes, o que acham de nos financiar?” Podemos até mesmo dar a vocês, nós só queremos ter a oportunidade de desenvolvê-lo, paguem-nos um salário e podemos trabalhar para vocês. Eles disseram não, fomos então à Hewlett-Packard e eles disseram “Nós não precisamos de vocês, vocês ainda nem terminaram a faculdade ainda”.

O Apple I não foi lá um grande sucesso de vendas, vendeu pouco mais de 200 unidades a 666 dólares (pouco mais de US\$ 5000 em valores corrigidos) cada uma. Mesmo assim, os lucros sustentaram a Apple durante o primeiro ano, abrindo caminho para o lançamento de versões mais poderosas. Quem comprou um, acabou fazendo um bom negócio, pois hoje em dia um Apple I em bom estado chega a valer US\$ 50.000.

Diferente do Altair, o Apple I era vendido já montado. A placa era vendida "pelada" dentro de uma caixa de papelão, sem nenhum tipo de gabinete, por isso era comum que os Apple I fossem instalados dentro de caixas de madeira.

O Apple I era baseado no processador 6502, um clone do Motorola 6800, fabricado pela MOS Technology. Ele era um processador de 8 bits, que operava a apenas 1 MHz. Em termos de poder de processamento, o 6502 perdia para o 8080, mas isso era compensado pelos "espaçosos" 8 KB de memória, que eram suficientes para carregar o interpretador BASIC (que ocupava 4 KB), deixando os outros 4 KB livres para escrever e rodar programas.

Uma das vantagens é que o Apple I podia ser ligado diretamente à uma TV, dispensando a compra de um terminal de vídeo. Ele possuía também um conector unidade de fita (o controlador era vendido separadamente por 75 dólares) e um conector proprietário reservado para expansões futuras:



Naquela época, as fitas K7 eram o meio mais usado para guardar dados e programas. Os disquetes já existiam, mas eram muito caros.

Os grandes problemas das fitas K7 eram a lentidão e a

baixa confiabilidade. No Apple I, os programas eram lidos a meros 1500 bits por segundo e em outros computadores o acesso era ainda mais lento, com de 250 a 300 bits. Era preciso ajustar cuidadosamente o volume no aparelho

de som antes de carregar a fita e, conforme a fita se desgastava, era preciso tentar cada vez mais vezes antes de conseguir uma leitura sem erros.

Na época, existiam até programas de rádio que transmitiam softwares como parte da programação. O locutor avisava e em seguida "tocava" a fita com o programa. Os interessados precisavam ficar com o aparelho de som à mão para gravar a cópia. Estes programas de rádio foram a primeira rede de pirataria de softwares de que se tem notícia, décadas antes da popularização da internet ;).



O Apple I foi logo aperfeiçoado, surgindo então o Apple II, lançado em 1977. Este sim fez sucesso, apesar do preço salgado para a época: US\$

1298, que equivalem a quase 10.000 dólares em valores corrigidos.

O Apple II vinha com apenas com 4 KB de memória, mas incluía mais 12 KB de memória ROM, que armazenava um interpretador BASIC e o software de bootstrap, lido no início do boot. Isso foi uma grande evolução, pois você ligava e já podia começar a programar ou carregar programas. No Apple I, era preciso primeiro carregar a fita com o BASIC, para depois começar a fazer qualquer coisa.

O BASIC era a linguagem mais popular na época (e serviu como base para diversas linguagens modernas), pois tem uma sintaxe simples se comparado com o C ou Assembly, utilizando comandos derivados de palavras do Inglês.

Este é um exemplo de programa em BASIC simples, que pede dois números e escreve o produto da multiplicação dos dois:

```
10 PRINT "MULTIPLICANDO"  
20 PRINT "DIGITE O PRIMEIRO NUMERO:"  
30 INPUT A  
40 PRINT "DIGITE O SEGUNDO NUMERO:"  
50 INPUT B  
60 LETC=A*B  
70 PRINT "RESPOSTA:", C
```

Este pequeno programa precisaria de 121 bytes de memória para rodar (os espaços depois dos comandos são ignorados, por isso não contam). Ao desenvolver programas mais complexos você esbarrava rapidamente na barreira da memória disponível (principalmente se usasse um ZX80, que tinha apenas 1 KB ;), o que obrigava os programadores a otimizarem o código ao máximo. Aplicativos comerciais (e o próprio interpretador BASIC) eram escritos diretamente em linguagem de máquina, utilizando diretamente as instruções do processador e endereços de memória, de forma a extraírem o máximo do equipamento.

Voltando ao Apple II, a memória RAM podia ser expandida até 52 KB, pois o processador Motorola 6502 era capaz de

endereçar apenas 64 KB de memória, e 12 KB já correspondiam à ROM embutida. Um dos "macetes" naquela época era uma placa de expansão, fabricada pela recém formada Microsoft, que permitia desabilitar a ROM e usar 64 KB completos de memória.

Além dos jogos, um dos programas mais populares para o Apple II foi o Visual Calc, ancestral dos programas de planilha atuais:



O Apple II já era bem mais parecido com um computador atual. Vinha num gabinete plástico e tinha um teclado incorporado. A versão mais básica era ligada na TV e usava o famigerado controlador de fita K7, ligado um aparelho de som para carregar programas. Gastando um pouco mais era



possível adquirir separadamente uma unidade de disquetes.



A linha Apple II se tornou tão popular que sobreviveu até o início dos anos 90, quase uma década depois do lançamento do Macintosh. O último lançamento foi o Apple IIc Plus, que utilizava um processador de 4 MHz (ainda de 8 bits) e vinha com um drive de disquetes de 3.5", já similar aos drives atuais.

Outra inovação do Apple I e Apple II em relação ao Altair e outros computadores anteriores é o tipo de memória usada. O Apple I foi o primeiro a utilizar memórias DRAM, que é essencialmente a mesma tecnologia utilizada até hoje em pentes de memória.

Ao longo das primeiras décadas, a memória RAM passou por duas grandes evoluções. No ENIAC, não existia uma unidade de memória dedicada. Parte das válvulas eram reservadas para armazenar as informações que estavam sendo processadas. Não existia unidade de armazenamento, além

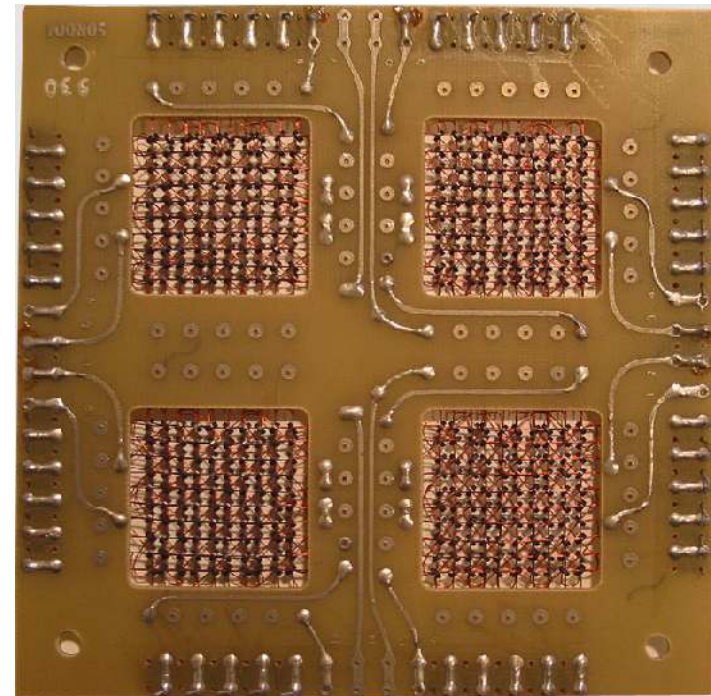
dos cartões perfurados e as anotações feitas manualmente pelos operadores.

Na década de 50 surgiram as memórias core, um tipo antiquado de memória onde são usados anéis de ferrite, um material que pode ter seu campo magnético alterado através de impulsos elétricos, armazenando o equivalente a um bit 1 ou 0). Estes anéis de ferrite eram carinhosamente chamados de "donuts" (rosquinhas) e eram montados dentro de uma complexa rede de fios, que transportavam os

impulsos elétricos usados para ler e escrever dados.

Cada anel armazenava apenas um bit, de forma que você precisava de 8.192 deles para cada KB de memória. Inicialmente a malha de fios era "tecida" manualmente, mas logo começaram a ser usadas máquinas, que permitiram miniaturizar bastante as estruturas.

Este é um exemplo de placa de memória core. Ela mede 11 x 11 cm (um pouco menor que um CD), mas armazena apenas 50 bytes.





Estas placas eram ligadas entre si, formando "pilhas" organizadas dentro de estruturas maiores. Imagine que, para atingir 1 MB de memória no início da década de 1960, você precisaria de quase 21 mil destas plaquinhas.

Este é um exemplo de unidade de memória, construída usando placas de memória core, que está em exposição no museu no MIT. Apesar do tamanho, ela possui apenas 64 KB:



Por serem muito caras e precisares de um grande número de circuitos de apoio, as memórias core ficaram restritas aos computadores de grande porte. O Altair já utilizava memórias "modernas" na forma de chips, para ser exato, dois chips de 1024 bits (ou 128 bytes) cada um.

O Altair utilizava chips de memória SRAM (static RAM), que eram rápidos e confiáveis, porém muito caros. Na memória SRAM, são usados de 4 a 6 transístores para cada bit de dados (as do Altair usavam 4 transistores), o que multiplica o custo dos chips. Atualmente, as memórias SRAM são usadas nos caches L1 e L2 dos processadores, o tipo mais rápido e caro de memória que existe.

O Apple I inovou utilizando um "novo" tipo de memória, as DRAM (dynamic RAM), onde é usado um único transistor para cada bit de dados. Embora à primeira vista pareçam mais simples, os chips de memória DRAM são muito mais complicados de se trabalhar (principalmente se considerarmos as limitações da época), pois eles são capazes de armazenar os dados por

apenas uma pequena fração de segundo. Para conservar os dados, eles precisam de um circuito de refresh, que lê e regravava os dados a cada 64 miléssegundos (ou menos, de acordo com o projeto).

Apesar de todas as dificuldades, foi o uso de memórias DRAM no Apple I que permitiu que ele viesse com 8 KB de memória, custando pouco mais que um Altair, que vinha com meros 256 bytes. A partir daí, as memórias DRAM se tornaram norma, o que continua até os dias de hoje.

Voltando à história, em 1979 surgiu um outro modelo interessante, desta vez da Sinclair, o ZX-80. Ele não era tão poderoso quanto o Apple II, mas tinha a vantagem de custar apenas 99 dólares (pouco mais de 400 em valores corrigidos) Foi o computador mais popular até então, com 100.000 unidades vendidas (entre 1979 e 1981), sem contar uma grande quantidade de clones, produzidos em diversos países ao longo da década de 80.

O ZX-80 era baseado no NEC-780C, um clone do Z80, que operava a 3.25 MHz. Ele era relativamente poderoso para

os padrões da época, mas aquecia bastante. Segundo as más línguas, ele foi o primeiro processador overclockado da história ;).

Para cortar custos ele vinha de fábrica com apenas 1 KB de memória RAM, combinados com 4 KB de memória ROM que armazenavam o interpretador BASIC, usado pelo aparelho. Como em qualquer sistema popular da época, os programas eram armazenados em fitas K7 e ele era ligado diretamente na TV:



Considerando preço, o Z80 foi uma máquina surpreendente, mas claro, tinha pesadas limitações, mesmo se comparado com outras máquinas da época. Apesar dele já vir com uma saída de vídeo, a resolução gráfica era de apenas 64x48, mesmo em modo monocromático, já que o adaptador de

vídeo tinha apenas 386 bytes de memória. Existia também uma opção de modo texto (usada para programas em BASIC, por exemplo), com 32x24 caracteres.

O processador Z80 se tornou incrivelmente popular, superando de longe as vendas de qualquer outro processador da história. Versões modernizadas do Z80 (que conservam o mesmo projeto básico, mas são produzidos com técnicas modernas de fabricação e trabalham a frequências mais altas) fazem sucesso até hoje, sendo utilizados em todo tipo de eletrônicos, incluindo impressoras, aparelhos de fax, controladores diversos, robôs de uso industrial, brinquedos, diversos tipos de calculadoras, video-games (incluindo o Game Boy e Game Boy color), diversos modelos populares de MP3Players, entre inúmeros outros exemplos. Apesar de não ser nenhum campeão de velocidade, o Z80 é um chip extremamente barato e fácil de programar, já que todos os seus truques são bem conhecidos e documentados.

Aqui no Brasil tivemos os TK80 e TK82 da Microdigital e o NE-

Z80 da Prológica, produzidos na época da reserva de mercado. Eles concorriam com os computadores compatíveis com os Apple, como o AP II, Exato, Craft II e Magnex M10. A linha CP (200, 300, 400 e 500) da Prológica era baseada em chips Z80 e tínhamos também clones da linha MSX, como os Expert 1.0 e Expert Plus.

A reserva de mercado estagnou o desenvolvimento tecnológico do país, de forma que clones de computadores de 8 bits, lançados a uma década atrás era tudo que nossa indústria conseguia produzir. Isso perdurou até 1992, quando a reserva de mercado foi abolida, permitindo a entrada de computadores importados. Em pouco tempo, todos estes computadores de 8 bits foram substituídos por PCs 386 e 486.

Concluindo nosso passeio pela década de 70, outro que não poderia deixar de ser citado é o Atari 800. Sim, apesar de ser mais vendido como um video-game, o Atari 800 também podia ser usado com um computador relativamente poderoso, chegando a ser adotado nos laboratórios de



informática de algumas universidades. Ele foi o antecessor do Atari 2600, o video-game conhecido por aqui.

Ele vinha de fábrica com 16 KB de memória RAM, que podiam ser expandidos para até 48 KB, com mais 10 KB de memória ROM. O sistema operacional era o Atari-OS, uma versão do BASIC.



Originalmente, o sistema vinha apenas com a entrada para os cartuchos, com o sistema operacional ou jogos, mas era possível adquirir separadamente uma unidade de disquetes e um teclado, que o transformavam num computador completo. Não existiram muitos programas para o Atari, já que o foco foram sempre os jogos. O principal uso do Atari como computador era desenvolver programas em BASIC, por isso seu uso em escolas.

## A década de 80

Como profetizado por Gordon Moore, os processadores vem, em média dobrando de desempenho a cada 18 meses desde o início da década de 70. Uma década é uma verdadeira

eternidade dentro do mercado de informática, o suficiente para revoluções acontecerem e serem esquecidas.

Depois dos dinossauros da primeira metade da década de 70, os computadores pessoais finalmente começaram a atingir um nível de desenvolvimento suficiente para permitir o uso de aplicativos sérios. Surgiram então os primeiros aplicativos de

processamento de texto, planilhas, e até mesmo programas de editoração e desenho.

Depois dos Apple I e Apple II, ZX80, Ataris e outros computadores de 8 bits, chegamos finalmente à era PC.

A IBM de 1980 era uma gigantesca empresa, especializada em mainframes e terminais burros para eles. Entretanto, percebendo a crescente demanda por computadores pessoais, decidiram criar um pequeno grupo (que originalmente possuía apenas 12 desenvolvedores) para desenvolver um computador pessoal de baixo custo.

Este era considerado um projeto menor dentro da IBM, apenas uma experiência para testar a demanda do mercado. O projeto chegou a ser marginalizado dentro da empresa, pois muitos executivos acreditavam que o IBM PC poderia concorrer com outros produtos do portfólio da IBM.

Depois de quase um ano de desenvolvimento, o primeiro PC foi lançado em 12 de agosto de 1981.





Para cortar custos e acelerar o desenvolvimento, a equipe decidiu que usaria apenas componentes padrão, que pudessem ser encontrados facilmente no mercado. O processador escolhido foi o Intel 8088, uma versão econômica do processador 8086, que havia sido lançado pela Intel em 78. Quando a IBM estava desenvolvendo seu computador pessoal, chegou a ser cogitado o uso do

competir na mesma faixa de preço dos computadores de 8 bits mais populares e, ao mesmo tempo, possuir um desempenho bem superior devido ao seu processador de 16 bits. O 8088 é capaz de acessar até 1 MB de memória RAM (embora o PC original suportasse apenas 64 KB, devido a limitações da placa mãe) e funciona a 4.77 MHz, recursos incríveis para a época, já que estamos falando de um processador lançado no final de 1979.

Lembre-se de que o principal concorrente do IBM PC era o Apple II que, embora fosse mais barato e contasse mais softwares disponíveis, usava um processador de 8 bits, de apenas 1 MHz e meros 4 KB de memória RAM.

8086, mas acabou sendo escolhido o 8088 devido ao seu baixo custo.

Tanto o 8086 quanto o 8088 são processadores de 16 bits e eram considerados bastante avançados para a época. Um processador de 16 bits é capaz de endereçar mais memória (até 64 KB de memória de cada vez) e processar instruções muito mais complexas que os processadores de 8 bits usados até então.

A grande diferença entre os dois é que o 8086 é um processador de 16 bits "puro", enquanto o 8088 se comunica com os demais periféricos usando um barramento de 8 bits. Isso naturalmente prejudicava o desempenho, mas trouxe uma vantagem importante: a possibilidade de usar os componentes de 8 bits usados em outros computadores da época, que eram muito mais populares e baratos.

Esta arquitetura permitiu ao primeiro PC

Entretanto, o aspecto técnico não foi o determinante para o sucesso do PC. Ele era um bom computador para a época, mas era caro e não tinha nada que os concorrentes não pudessem usar em seus produtos. Ele tinha tudo para ser apenas mais um no mercado, se não fosse um diferencial importante: a arquitetura aberta.

Diferente dos Apples e outros computadores da época, qualquer fabricante podia desenvolver e vender acessórios para o PC, sem pagar royalties ou fazer acordos de licenciamento. Como todos os componentes eram abertos, era possível também desenvolver clones, computadores compatíveis com o PC, fabricados por outras empresas. Isso lentamente fez com que toda a indústria passasse a orbitar em torno do PC, fazendo com que a plataforma crescesse assustadoramente.

Voltando ao tema original, o PC original tinha, em sua versão mais simples, apenas 16 KB de memória RAM, com direito apenas ao gabinete e teclado. A partir daí, tudo era opcional, incluindo o monitor (você



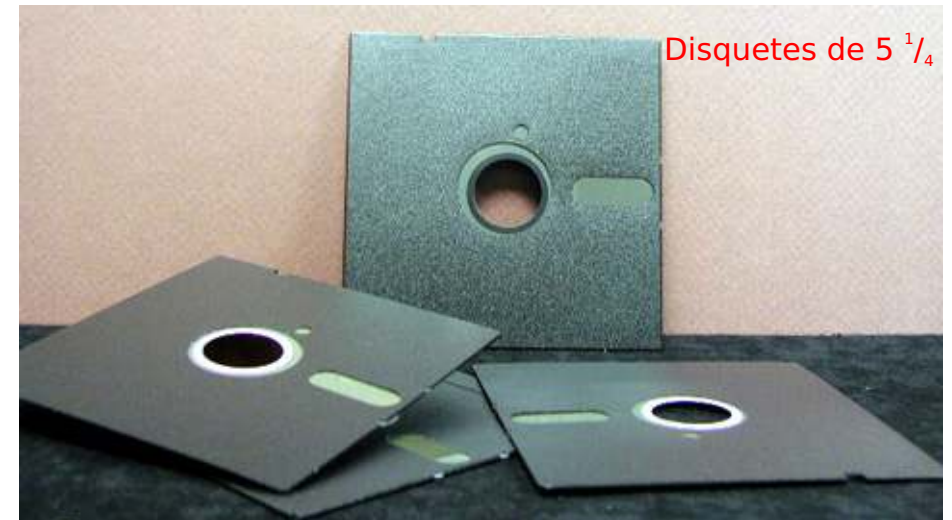
podia usar uma TV, embora a qualidade da imagem ficasse ruim), os drives de disquete e o HD.

Na configuração básica, o PC custava "apenas" 1564 dólares da época, mas incluindo mais 48 KB de memória, dois drives de disquete e um monitor mono de 12", o preço chegava facilmente a 2500 dólares, que equivalem a mais de 7000 dólares em valores atuais.

Na época, os HDs ainda eram um componente caro e

exótico. Em 1981, um Seagate ST-506 (o modelo mais popular até então) custava mais de 1000 dólares (da época) e tinha apenas 5 MB de capacidade. Este da foto é um ST-225 (também da Seagate), um modelo de 20 MB, lançado em 1984, que foi muito usado nos micros 286. Estes primeiros modelos ainda utilizavam motores de passo para mover as cabeças de leitura (como nos drives de disquete), por isso os problemas eram comuns.

Ao usar um PC sem HD, o sistema operacional e todos os programas era carregados a partir de disquetes de 5¼. Inicialmente eram usados disquetes de 180 KB, mas eles foram logo substituídos por disquetes de 360 KB (onde eram usadas as duas faces do disco) e, alguns anos mais tarde, por disquetes "alta densidade", com 1.2 MB. Os disquetes de 3.5", com 1.44 MB que usamos hoje em dia passaram a ser usados nos PCs apenas em 1987, com o lançamento do IBM PS/2. Existiu ainda um padrão de disquetes de 2.8 MB, lançado nos anos 90, que acabou não pegando.



O PC era monotarefa, de forma que para carregar outro programa, você precisava primeiro encerrar o primeiro e trocar o disquete dentro do drive. O segundo drive de disquetes era um item extremamente popular (e necessário), pois os disquetes de 5¼ eram extremamente frágeis e a mídia se degradava com o tempo, de forma que você precisava copiar os discos frequentemente.

Conforme foram sendo lançados PCs com mais memória RAM, surgiu o "macete" de criar um ramdisk (um pedaço da memória RAM usado como se fosse um HD) e usá-lo para

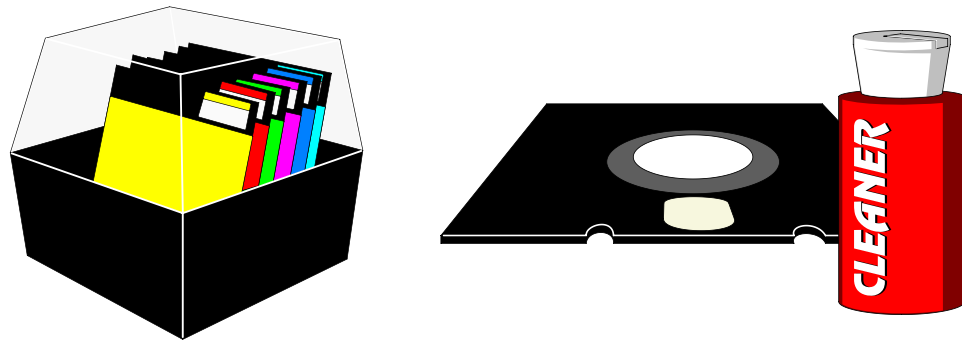
copiar disquetes sem precisar de um segundo drive :). Também era comum aparecerem versões "capadas" dos principais programas, com componentes e bibliotecas desativados ou removidos, de forma a rodar nos PCs com menos memória RAM. Naquela época, ainda não existia memória swap, de forma que se você não tivesse memória suficiente, os programas simplesmente não rodavam.

O sistema operacional usado no PC original era o MS-DOS 1.0 (na época ainda chamado de PC-DOS), que foi desenvolvido às pressas pela Microsoft com base num sistema operacional mais simples, chamado QDOS, comprado da Seattle Computers, uma pequena empresa desenvolvedora de sistemas. Na verdade, a Microsoft foi a segunda opção da IBM, depois de ter sua proposta de licença recusada pela Digital Research, que na época desenvolvia versões do seu CP/M para várias arquiteturas diferentes.

Na época, a IBM acreditava que ganharia dinheiro vendendo as máquinas e não vendendo sistemas operacionais e softwares, o que era considerado um negócio menor, dado de bandeja para a Microsoft.

Com o passar do tempo, os executivos da IBM se arrependeram amargamente da decisão, pois a concorrência entre os diversos fabricantes derrubou os preços e as margens de lucro dos PCs, enquanto a Microsoft conseguiu atingir um quase monopólio do sistema operacional para eles e, sem concorrentes de peso, passou a trabalhar com margens de lucro cada vez maiores.

Um fabricante de memórias, como a Micron, trabalha normalmente com margens de lucro abaixo de 1%.



Conseguem ganhar dinheiro apenas por venderem quantidades muito grandes. Um integrador como a Dell trabalha com margens de 3 a 5% (e leva prejuízo às vezes, nas unidades que ficam muito tempo em estoque ou não vendem), enquanto a Microsoft (mesmo com toda a pirataria)

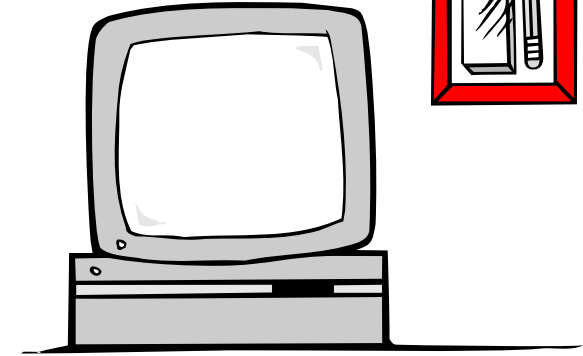
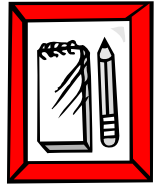
trabalha com margens superiores a 80% vendendo o Windows e Office; um negócio da China :).

Hoje em dia, a IBM sequer fabrica PCs. Mesmo os famosos notebooks IBM Thinkpad são agora fabricados e vendidos pela Lenovo, uma empresa chinesa que comprou os direitos sobre a marca em 2000.

Voltando à história, dois anos depois foi lançado o PC XT, que apesar de continuar usando o 8088 de 4.77 MHz, vinha bem mais incrementado, com 256 KB de RAM, disco rígido de 10 MB, monitor CGA e o MS-DOS 2.0. O XT se tornou um computador bastante popular e chegou a ser fabricado no Brasil, durante a reserva de mercado. Enquanto os americanos já usavam muitos 386, os clones tupiniquins do XT eram a última palavra em tecnologia aqui no Brasil...

Depois do XT, o próximo passo foi o PC AT (lançado em 1984), já baseado no Intel 286. Na verdade, o processador 286 foi lançado em Fevereiro de 1982, apenas 6 meses após a IBM ter lançado o seu primeiro PC, porém, demorou até que a IBM conseguisse desenvolver um computador baseado nele, pois foi preciso desenvolver toda uma nova arquitetura. Da placa de vídeo ao gabinete, praticamente tudo foi mudado, o que

Em caso de emergência,  
quebre o vidro





somado à burocracia e a longos períodos de testes antes do lançamento, levou mais de 2 anos.

Atualmente, o período de desenvolvimentos dos periféricos é muito mais curto. Quase sempre quando um novo processador é lançado, já temos placas mãe para ele disponíveis quase que imediatamente, pois o desenvolvimento é feito de forma simultânea.



O PC AT vinha com um processador 286 de 6 MHz (depois surgiram versões mais rápidas, de 8, 12 e até 16 MHz),

HD de 10 MB, monitor EGA (640x350, com 64 cores) e já usava disquetes de 5¼ de 1.2 MB. Como a memória RAM ainda era um item muito caro, existiam versões com de 256 a 2 MB de RAM. Embora fosse extremamente raro usar mais de 2 MB, existia a possibilidade de instalar até 16 MB.

O processador 286 trouxe vários avanços sobre o 8088. Ele utilizava palavras binárias de 16 bits tanto interna quanto externamente, o que permitia o uso de periféricos de 16 bits, muito mais avançados do que os usados no PC original e no XT. O custo dos periféricos desta vez não chegou a ser um grande obstáculo, pois enquanto o PC AT estava sendo desenvolvido, eles já podiam ser encontrados com preços mais acessíveis.

Para manter compatibilidade com os periféricos de 8 bits usados no PC original e no XT, a IBM desenvolveu os slots ISA de 16 bits, que permitem usar tanto placas de 8 bits, quanto de 16 bits. As placas de 8 bits são menores, e usam apenas a primeira série de pinos do slot, enquanto as placas de 16 bits

usam o slot completo. Devido à sua popularidade, o barramento ISA continuou sendo usado por muito tempo. Em 2004 (10 anos depois do lançamento do PC AT) ainda era possível encontrar placas mãe novas com slots ISA, embora atualmente eles estejam extintos.



Slots ISA

O principal avanço trazido pelo 286 são seus dois modos de operação, batizados de “Modo Real” e “Modo Protegido”. No modo real, o 286 se comporta exatamente como um 8086 (apesar de mais rápido), oferecendo total compatibilidade com os programas já existentes. Já no modo protegido, ele manifesta todo o seu potencial, incorporando funções mais avançadas, como a capacidade de acessar até 16 MB de memória RAM (apesar de ser um processador de 16 bits, o 286 usa um sistema de endereçamento de memória de 24 bits), multitarefa, memória virtual em disco e proteção de memória.

Assim que ligado, o processador opera em modo real, e com uma instrução especial, passa para o modo protegido. O problema é que trabalhando em modo protegido, o 286 deixava de ser compatível com os programas escritos para o modo real, inclusive com o próprio MS-DOS. Para piorar, o 286 não possuía nenhuma instrução que fizesse o processador voltar ao modo real, o que era possível apenas resetando o micro. Isso significa que um programa escrito para rodar em modo protegido, não poderia usar nenhuma das rotinas de acesso a dispositivos do MS-DOS, tornando inacessíveis o disco rígido, placa

de vídeo, drive de disquetes memória, etc., a menos que fossem desenvolvidas e incorporadas ao programa todas as rotinas de acesso a dispositivos necessárias.

Isso era completamente inviável para os desenvolvedores, pois para projetar um simples jogo, seria praticamente preciso desenvolver todo um novo sistema operacional. Além disso, o programa desenvolvido rodaria apenas em micros equipados com processadores 286, que ainda eram minoria na época, tendo um público alvo muito menor. De fato, apenas algumas versões do UNIX e uma versão do OS/2 foram desenvolvidas para utilizar o modo protegido do 286.

Basicamente, os micros baseados no 286 eram usados para rodar aplicativos de modo real, que também podiam ser executados em um XT, aproveitando apenas a maior velocidade do 286.

Devido às várias mudanças na arquitetura, destacando o acesso mais rápido à memória e alterações no conjunto de instruções do processador, que permitiam realizar muitas operações de maneira mais rápida e eficiente, um 286 consegue ser quase 4 vezes mais rápido que um 8088 do mesmo clock.

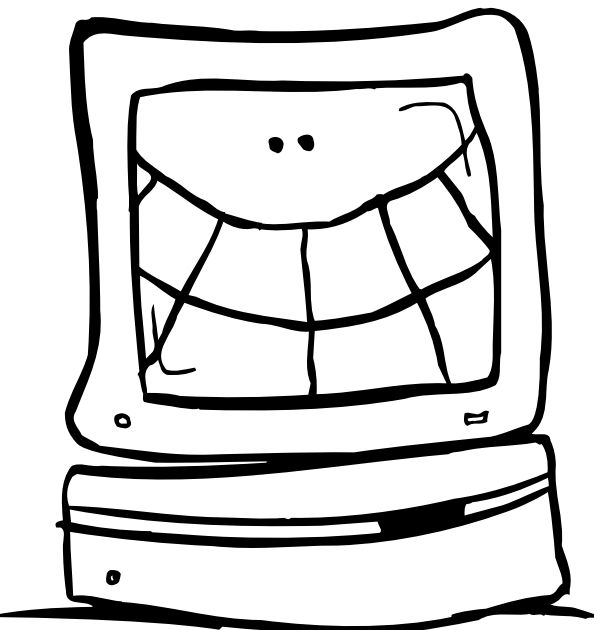
Em outubro de 1985 a Intel lançou o 386, que marcou o início dos tempos modernos. Ele trouxe vários recursos novos. Para começar, o 386 trabalha tanto interna quanto externamente com palavras de 32 bits e é capaz de acessar a memória usando um barramento de 32 bits, permitindo uma transferência de dados duas vezes maior. Como o 386 pode trabalhar com palavras binárias de 32 bits, é possível acessar até 4 GB de memória ( $2^32$  elevado à 32ª potência), mesmo sem usar a segmentação de endereços, como no 8088.

Assim como o 286, o 386 continua possuindo os dois modos de operação. A diferença é que no 386 já é possível alternar entre o modo real e o modo protegido livremente. Os programas que rodavam sobre DOS, podiam chavear o

processador para o modo protegido, para beneficiar-se de suas vantagens e voltar ao modo real sempre que precisavam usar alguma sub-rotina do DOS, de maneira transparente ao usuário. Neste caso, era usado um programa de DPML ("DOS Protected Mode Interface", ou "interface DOS de modo protegido") para fazer o chaveamento entre os dois modos.

Toda vez que o programa precisava usar alguma sub-rotina do DOS, ele passava o comando ao chaveador e ficava esperando. O chaveador por sua vez, colocava o processador em modo real, executa o comando, chaveava o processador para o modo protegido e entregava o resultado ao aplicativo, que continuava trabalhando como se nada tivesse acontecido. Um bom exemplo de programa de DPML é o DOS4GW, que é usado por muitos jogos antigos que rodam sobre o MS-DOS, como o DOOM, Sim City 2000 e vários emuladores de video-games.

O esquema de chaveamento também era utilizado pelo Windows 3.x, que incluía todas



as rotinas necessárias, dispensando qualquer programa de DPMI. O Windows 95/98 também pode chavear para o modo real caso precise carregar algum driver de dispositivo de modo real. No Windows XP os programas DOS passaram a ser executados dentro de um emulador (ao invés de nativamente), o que reduziu a compatibilidade do sistema.

Ter um processador 386 é o requisito mínimo para rodar qualquer sistema operacional moderno. Com um 386, memória RAM e espaço em disco suficiente, você pode rodar o Windows 95 e aplicativos, embora bem lentamente devido à pouca potência do processador. Você pode também instalar distribuições Linux antigas e (usando algum truque para burlar a detecção da configuração mínima para rodar o sistema), até mesmo instalar o Windows 98.

Com um simples 286, no máximo você poderia rodar o DOS e aplicativos mais simples, que trabalhem somente com o modo real. Também era possível rodar o Windows 3.0, porém em modo "Standard", onde é

possível acessar todos os 16 MB de memória permitidos pelo 286, mas sem memória virtual nem multitarefa.

Apenas o Athlon 64 e os processadores Intel com o EM64 (o conjunto de instruções compatíveis com os processadores de 64 bits da AMD), vieram a quebrar esta compatibilidade histórica. Os processadores de 64 bits atuais são perfeitamente compatíveis com os aplicativos de 32 bits, mas programas otimizados para eles não rodam mais nas máquinas antigas. Embora mais suave e gradual, estamos assistindo a uma migração similar à que ocorreu na transição do 286 para o 386.

Voltando ao lançamento do 386, embora o processador tenha sido lançado em 1985, a IBM só foi capaz de lançar um PC baseado nele em 1987, dando tempo para a Compaq sair na frente. Este foi um verdadeiro marco pois, de repente, as companhias perceberam que não eram mais obrigadas a seguir a IBM. Qualquer um que tivesse tecnologia suficiente poderia sair na frente, como fez a Compaq. A partir daí, a

IBM começou a gradualmente perder a liderança do mercado, tornando-se apenas mais um entre inúmeros fabricantes de PCs.

Naturalmente, a Apple não ficou sentada durante este tempo todo. Além de continuar aperfeiçoando a linha Apple II, a empresa começou a investir pesadamente no desenvolvimento de computadores com interface gráfica e mouse. A "inspiração" surgiu numa visita de Steve Jobs ao laboratório da Xerox, onde computadores com interface gráfica eram desenvolvidos desde a década de 70 (embora sem sucesso comercial, devido ao custo proibitivo).

Em 1983, eles apareceram com uma grande novidade, o Lisa. Em sua configuração original, o Lisa vinha equipado com um processador Motorola 68000 de 5 MHz, 1 MB de memória RAM, dois drives de disquete de 5¼ de alta densidade (eram usados discos de 871 KB), HD de 5 MB e um monitor de 12 polegadas, com resolução de 720 x 360. Era uma configuração muito melhor do que os PCs da época, sem falar que o Lisa já usava uma interface gráfica bastante elaborada e contava com uma suíte de aplicativos de escritório à lá Office. O problema era o preço: 10.000 dólares da época (suficiente para comprar 5 PCs).





Embora não houvesse nada melhor no mercado, o Lisa acabou não atingindo o sucesso esperado. No total, foram produzidas cerca de 100.000 unidades em dois anos, porém a maior parte delas vendidas com grandes descontos, muitas vezes abaixo do preço de custo (como um lote de 5.000 unidades vendido para a Sun em 1987, depois que o Lisa já havia sido descontinuado). Como a Apple investiu aproximadamente US\$ 150 milhões no desenvolvimento do Lisa, a conta acabou ficando no vermelho.

Apesar disso, o desenvolvimento do Lisa serviu de base para o Macintosh, um computador mais simples, lançado em 1984. Este sim fez um

grande sucesso, chegando a ameaçar o império dos PCs. A configuração era similar à dos PCs da época, com um processador de 8 MHz, 128 KB de memória e um monitor de 9 polegadas. A grande arma do Macintosh era o MacOS 1.0 (derivado do sistema operacional do Lisa, porém otimizado para consumir muito menos memória), um sistema inovador de vários pontos de vista.

Ao contrário do MS-DOS ele era inteiramente baseado no uso da interface gráfica e mouse, o que o tornava muito mais fácil de ser operado. O MacOS continuou evoluindo e incorporando novos recursos, mas sempre mantendo a mesma ideia de interface amigável

Depois do Mac original, chamado apenas de "Macintosh", a Apple lançou um modelo atualizado, com 512 KB de memória RAM. Para diferenciá-lo do primeiro, a Apple passou a chamá-lo de Macintosh 512k. O modelo antigo continuou sendo vendido até outubro de 1985 como uma opção de baixo custo, passando a ser chamado de Macintosh 128k.

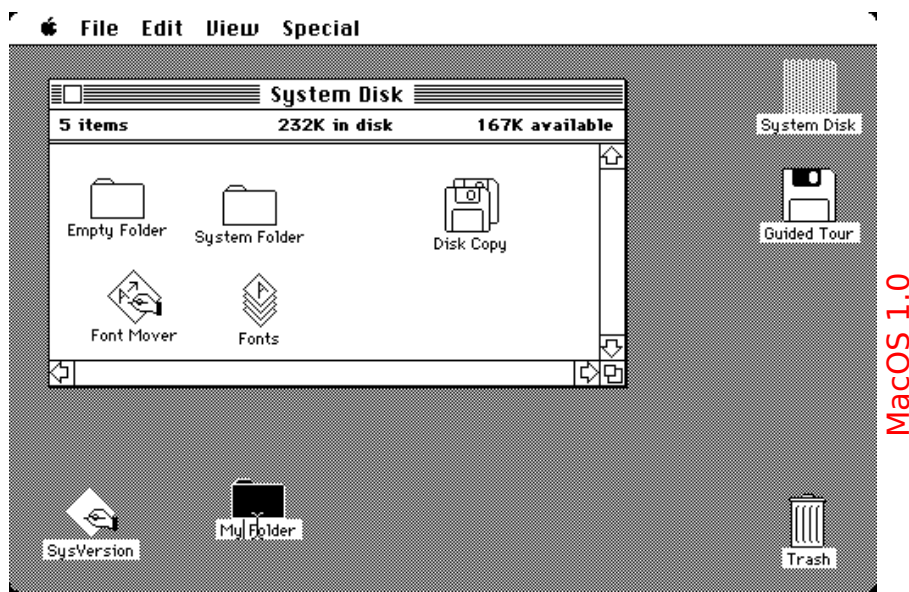
Pouco tempo depois foi lançado o Mac Rescue, uma placa de expansão que ampliava os 128 ou 512k de memória para 4 MB (algo assustador para a época) e dava "de brinde" um ramdisk de 2 MB para armazenamento de arquivos e programas (as primeiras versões do Mac não possuíam HD). O Mac já utilizava um recurso de hibernação, de forma que muita gente nunca desligava o aparelho, apenas o colocava pra dormir, preservando os dados do ramdisk. Embora fosse um upgrade muito bem vindo, o Mac Rescue não foi muito popular, pois era caro demais.

Neste meio tempo, a Microsoft teve tempo de desenvolver a primeira versão do Windows, anunciada em novembro de 1983. Ao contrário do MacOS, o Windows 1.0 era uma interface bastante primitiva, que fez pouco sucesso.

Ele rodava sobre o MS-DOS e podia executar tanto aplicativos for Windows quanto os programas para MS-DOS. O problema era a memória.

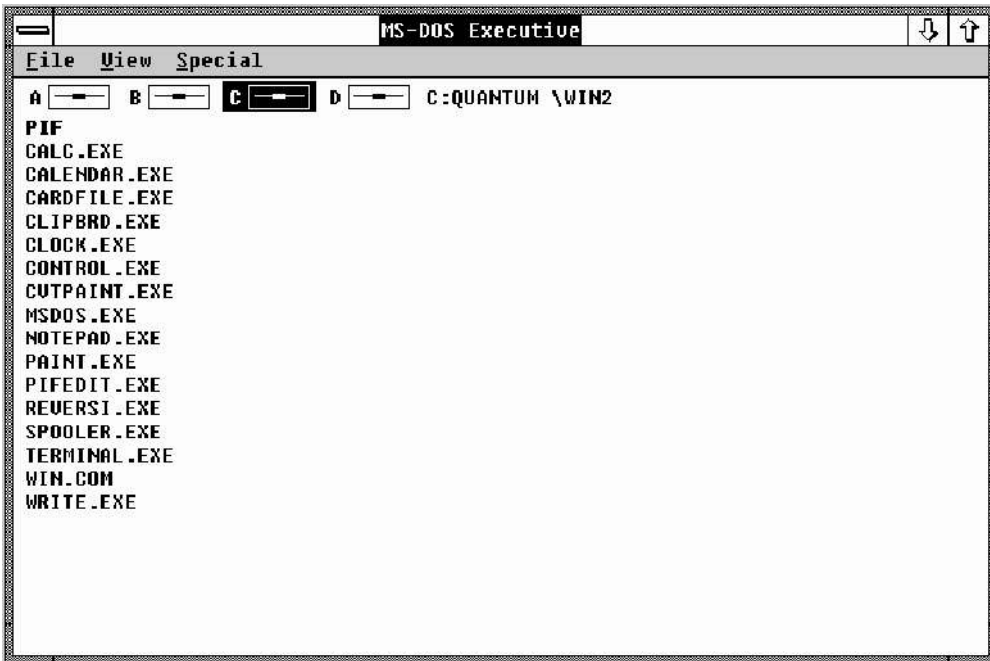
Os PCs da época vinham com quantidades muito pequenas de memória RAM e na época ainda não existia a possibilidade de usar memória virtual (que viria a ser suportada apenas a partir do 386). Para rodar o Windows, era preciso primeiro carregar o MS-DOS. Os dois juntos já consumiam praticamente toda a memória de um PC básico da época. Mesmo nos PCs mais parrudos não era possível rodar muitos aplicativos ao mesmo tempo, novamente por falta de memória.

Como os aplicativos for Windows eram muito raros na época, poucos usuários viram necessidade de utilizar o Windows para rodar os mesmos aplicativos que rodavam (com muito



mais memória disponível...) no MS-DOS. Sem contar que a versão inicial do Windows era bastante lenta e tinha vários bugs.

O Windows começou a fazer algum sucesso na versão 2.1, quando os micros 286 com 1 MB ou mais de memória já eram comuns. Com uma configuração mais poderosa, mais memória RAM e mais aplicativos, finalmente começava a fazer sentido rodar o Windows. O sistema ainda tinha vários bugs e travava com frequência, mas alguns usuários começaram a migrar para ele.



## Windows 2.0

O Windows emplacou mesmo a partir da versão 3.11. Ele era relativamente leve, mesmo para os PCs da época e já suportava o uso de memória virtual, que permitia abrir vários programas, mesmo que a memória RAM se esgotasse. Já existiam também

vários aplicativos for Windows e os usuários tinham a opção de voltar para o MS-DOS quando desejassem.

Foi nesta época que os PCs começaram a recuperar o terreno perdido para os Macintoshs da Apple. Convenhamos, o Windows 3.11 travava com muita frequência, mas tinha muitos aplicativos e os PCs eram mais baratos que os Macs.

Na época começaram a surgir os primeiros concorrentes para o Windows, como o OS/2 da IBM.

Desde o início da era PC, a Microsoft e a IBM vinham trabalhando juntas no desenvolvimento do MS-DOS e outros programas para a plataforma PC. Mas, em 1990 a IBM e a Microsoft se desentenderam e cada uma ficou com uma parte do trabalho feito, com o qual tentaram tomar a liderança do mercado de sistemas operacionais.

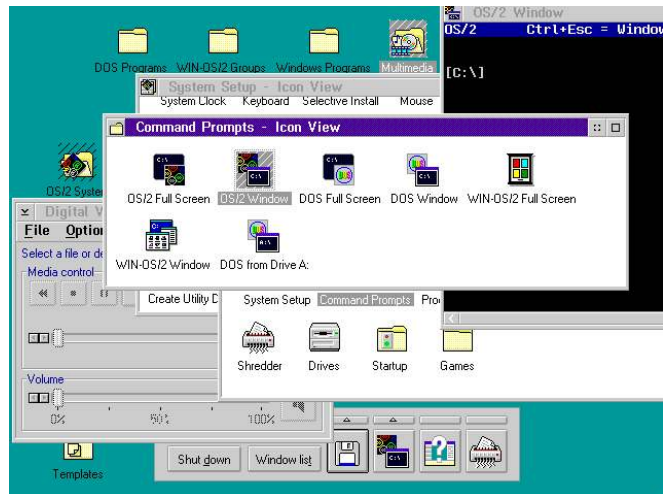
Alguns brincam que a IBM ficou com a parte que funciona e a Microsoft com o resto, mas a verdade é que apesar do OS/2 da IBM ser tecnicamente superior ao Windows 95, foi o

sistema das janelas quem levou a melhor, pois era mais fácil de usar e contava com a familiaridade dos usuários com o Windows 3.1, enquanto a IBM derrapava numa combinação de falta de investimento, falta de marketing e falta de apoio aos desenvolvedores.

Inicialmente, o OS/2 era incompatível com os softwares desenvolvidos para o Windows, o que era um grande empecilho, já que o Windows era muito mais popular entre os desenvolvedores. Depois de muita negociação, a IBM conseguiu um acordo com a Microsoft, que permitia que o OS/2 executasse o Windows 3.11 dentro de uma máquina virtual, oferecendo compatibilidade com seus programas. Entretanto, o tiro acabou saindo pela culatra, pois desestimulou ainda mais o desenvolvimento de aplicativos nativos para o OS/2, fazendo com que ele acabasse concorrendo (em desvantagem) com o Windows em seu próprio território. Rodar programas windows dentro do OS/2 era muito mais problemático e o desempenho era inferior,

fazendo com que mais e mais usuários preferissem usar o Windows diretamente.

Embora esteja oficialmente morto, o OS/2 ainda é utilizado por algumas empresas e alguns grupos de entusiastas. Em 2005 a Serenity comprou os direitos sobre o sistema, dando origem ao eComStation, um sistema comercial disponível no <http://www.ecomstation.com/>.



OS/2 Warp 3

Um sistema muito mais bem sucedido, que começou a ser desenvolvido no início da década de 90 é o Linux, que todos já conhecemos. O Linux tem a vantagem de ser um sistema aberto, que atualmente conta com a colaboração de milhares de desenvolvedores voluntários espalhados pelo globo, além do apoio de empresas de peso, como a IBM. Mas, no começo o sistema era muito mais complicado que as distribuições atuais e não contava com as interfaces gráficas exuberantes que temos hoje em dia. Embora o Linux seja forte em servidores desde o final da década de 1990, usar Linux em desktops é algo relativamente recente.

Voltando ao lado negro da força, a Microsoft continuou melhorando seu sistema ao longo da década de 1990. Foram

lançados o Windows 95, depois o 98 e finalmente ME, com todos os problemas que conhecemos mas com a boa e velha interface fácil de usar e uma grande safra de aplicativos que garantiram a manutenção e crescimento da popularidade do sistema.

Paralelamente, a Microsoft desenvolvia uma família de sistemas Windows destinadas a servidores, o Windows NT, que chegou até a versão 4, antes de ser transformado no Windows 2000.

Em seguida, as duas famílias Windows fundiram-se no Windows XP, um sistema destinado tanto ao uso doméstico quanto em estações de trabalho e servidores, que é bem mais estável que o 98 e ME, por ser baseado no NT. O XP foi aperfeiçoado ao longo dos anos seguintes, dando origem ao Vista.

A Apple por sua vez, passou por duas grandes revoluções. A primeira foi a migração do MacOS antigo para o OS X, que por baixo da interface polida, é um sistema Unix, derivado do BSD. A segunda aconteceu em 2005, quando a Apple anunciou a migração de toda a sua

linha de desktops e notebooks para processadores Intel.



Mac OS X

Do ponto de vista do hardware, os Macs atuais não são muito diferentes dos PCs, você pode inclusive rodar Windows e Linux através do boot camp. Entretanto, só os Macs são capazes de rodar o Mac OS X, devido ao uso do EFI, um firmware especial, que substitui o BIOS da placa mãe.

Esta combinação permitiu que a Apple se beneficiasse da redução de custo nos processadores e outros componentes para micros PCs, mas ao mesmo tempo conservasse seu principal diferencial, que é o software.

E a história continua... ;)



# do 486 ao Athlon

por Carlos E. Morimoto



Com o final da reserva de mercado, em 1992, a importação de computadores voltou a ser permitida no Brasil, sem maiores restrições. Rapidamente, os computadores de 8 bits e os XT's de fabricação nacional foram substituídos por micros 386 e 486 e muita gente teve a chance de comprar o primeiro micro. Isto coincidiu justamente com o início da era 486, fazendo com que eles fossem vendidos em enorme quantidade aqui no Brasil. É por isso que ainda é relativamente comum ver micros 486 por aí, muitos ainda funcionando, mas é muito raro encontrar um 386 ou anterior.

Assim como outros processadores da época, o 486 era vendido encapsulado dentro de uma pastilha de cerâmica, que ao mesmo tempo protegia o processador e facilitava a dissipação do calor. Os contatos ficavam nas beiradas internas do processador e eram ligados aos pinos externos através de filamentos de ouro. Atualmente, são utilizados compostos plásticos resistentes ao invés de cerâmica e os contatos do processador são feitos através de minúsculos pontos de solda disponíveis na parte inferior do próprio waffer de silício.

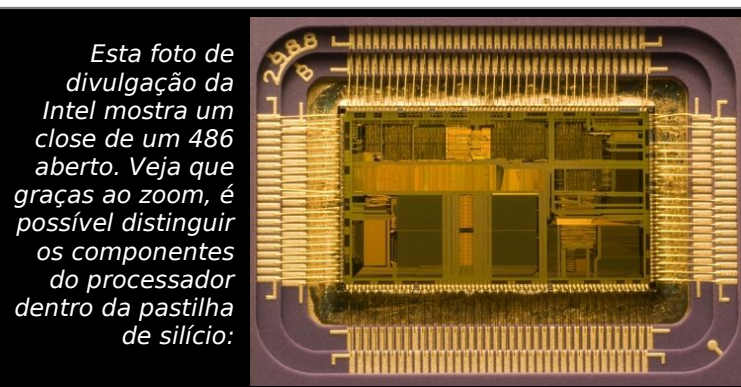
O 486 possuía 1.2 milhões de transistores e era fabricado numa técnica de 1 micron. Isto significa que cada transistor media um milionésimo de centímetro. Como tínhamos 1.2 milhões deles, o die do processador tinha cerca de 120 milímetros quadrados. Para efeito de comparação, o 386 tinha apenas 275.000 transistores, quase 5 vezes menos.

Este brutal aumento de complexidade pode ser justificado por três inovações introduzidas pelo 486. Em primeiro lugar, ele usa um co-processador aritmético integrado, ao invés de um chip separado, como no 386. Em segundo, ele incorpora 8 KB de cache ultra-rápido diretamente no processador, complementando o cache mais lento disponível na placa mãe. O cache interno passou a ser chamado de cache L1 (level 1, ou nível 1) e o cache da placa mãe de cache L2.

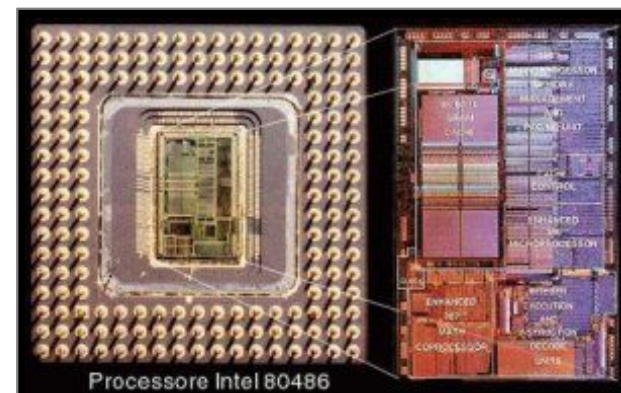
O cache L1 integrado se tornou um item de série em todos os processadores a partir do 486, pois melhora de forma considerável o desempenho do processador, oferecendo acesso instantâneo os dados que estão sendo processados. Mesmo em casos onde o cache L2 opera à mesma frequência no processador (como no 486 DX-33, onde a placa mãe também trabalhava a 33 MHz, ou nos processadores atuais, onde o cache L2 também é integrado ao processador) existe uma grande diferença nos tempos de acesso dos caches L1 e L2. Enquanto (nos processadores atuais) o cache L1 trabalha sempre com tempos de latência de 2 ou 3 ciclos, o L2 trabalha com latência de 7 ciclos ou mais e é acessado através de um barramento mais estreito.

Voltando ao 486, foi introduzido também o processamento de instruções em etapas, recurso que é utilizado até os dias de hoje. A unidade de execução do 486 é composta por um pipeline de 5 estágios, que possuem funções distintas. Cada instrução passa sucessivamente por cada um dos 5 estágios, sendo que cada um deles faz seu trabalho em apenas um ciclo, passa a instrução adiante e recebe outra. Poderíamos fazer uma analogia com uma linha de produção, que passa a ter 5 trabalhadores ao invés de um. O uso do pipeline trouxe duas vantagens. A primeira é que muitas instruções complexas, que consumiam vários ciclos do 386, passaram a ser executadas numa única passagem. Em segundo, a redução do trabalho feito em cada estágio permitiu que o processador fosse capaz de atingir frequências mais altas.

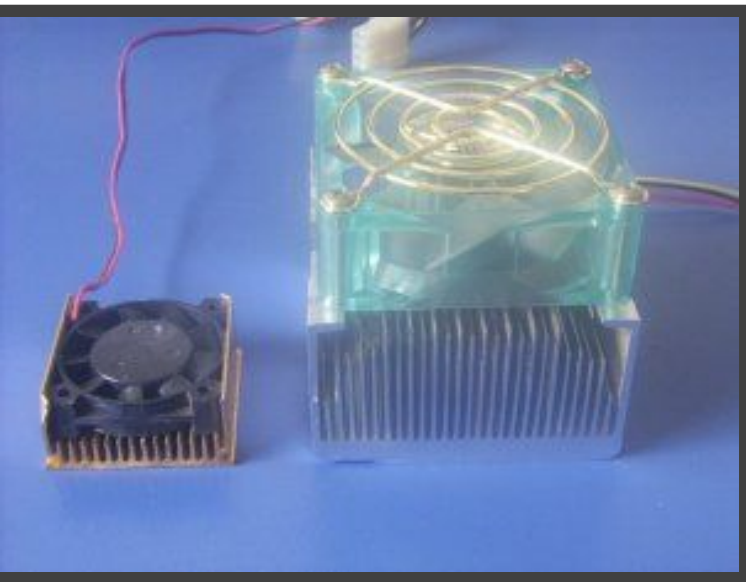
O 486 foi vendido em três versões: o 486DX (a versão completa completa), o 486SX (uma versão de baixo custo, sem o co-processador aritmético) e o 486SL (uma versão de baixo consumo, destinada a notebooks).



*Esta foto de divulgação da Intel mostra um close de um 486 aberto. Veja que graças ao zoom, é possível distinguir os componentes do processador dentro da pastilha de silício:*



Os 486 (a partir do DX-33) foram também os primeiros processadores Intel a utilizarem coolers. Naquela época, eram usados dissipadores com menos de um centímetro de altura e exaustores minúsculos. Conforme os processadores passaram a dissipar cada vez mais calor, os coolers foram crescendo na mesma proporção, até chegar às monstruosidades que vemos atualmente ;). Aqui temos um cooler para 486 e um cooler low-end atual:



Como disse, este é apenas um cooler low-end, que pode ser usado em processadores AMD de até 1.8 GHz. Coolers para processadores dual core, ou voltados para quem deseja fazer overclock são muito maiores e mais pesados, com exaustores maiores, bases de cobre e hot-pipes, como este Arctic Freezer:



O cobre é capaz de absorver quantidades maiores de calor que o alumínio, mas em compensação é mais caro e mais difícil de se trabalhar.

O alumínio, por sua vez, permite criar lâminas mais

finas, que facilitam a dissipação do calor. Por isso, é comum que os coolers atuais utilizem bases de cobre e dissipadores de alumínio. Os hot-pipes são barras de cobre, maciças, ou contendo fluído, que permitem "transportar" o calor da base do cooler para o dissipador.

Os hot-pipes são muito comuns em notebooks, onde o cooler é montado "na horizontal", com a base de uma lado, o exaustor do outro e dois ou mais hot-pipes interligando as duas peças, como neste Toshiba A45:



Outra inovação introduzida pelo 486 foi a multiplicação de clock, onde o processador trabalha numa frequência mais alta que a placa-mãe. Isto parece natural hoje em dia, mas na época causou espanto.

Graças a todas as inovações, as versões iniciais do 486 eram capazes de operar a 50 MHz ou mais. Por questões estratégicas, a Intel optou por lançar versões iniciais de 25 e 33 MHz, que funcionavam bem nas placas da época. Quando, pressionada pelos concorrentes, a Intel decidiu gastar os cartuchos e lançar um 486 operando a 50 MHz, se viu num beco sem saída, pois na época não existia tecnologia para produzir uma placa mãe capaz de trabalhar estavelmente a 50 MHz.

A solução foi incluir um circuito simples, que fazia com que o processador trabalhasse ao dobro da frequência da placa mãe. Surgiu então o 486 DX-2 50, seguido pelo DX-2 66. Pouco depois, a Intel introduziu uma nova técnica de fabricação, com transistores de 0.6 micron, e foi capaz de lançar o 486 DX-4 100, que trabalha ao triplo da frequência da placa mãe:

Processador	Placa mãe	Multiplicador
486 25 MHz	25 MHz	1x
486 33 MHz	33 MHz	1x
486DX-2 50 MHz	25 MHz	2x
486DX-2 66 MHz	33 MHz	2x
486DX-2 80 MHz	40 MHz	2x
486DX-4 75 MHz	25 MHz	3x
486DX-4 100 MHz	33 MHz	3x
486DX-4 120 MHz	40 MHz	3x



#### JP5~7: CPU Speed Jumpers

CPU Clock	JP5	JP6	JP7
50Mhz			
60Mhz			
66Mhz			
75Mhz			

#### JP10,11: CPU Internal Clock Speed Jumpers

IDT	Intel	Cyrix	AMD	JP10	JP11
Reserved	1.5X/ 3.5X	Reserved	K5 1.5X/ K6 3.5x		
Reserved	2.0X	2.0X	Reserved		
Reserved	2.5X	M2 2.5X	2.5X		
C6 3.0X	3.0X	M2 3.0X	K6 3.0X		

Com isso, surgiu também a possibilidade de fazer overclock do processador, usando uma frequência ou multiplicador maior que o nominal. Era muito comum usar um 486DX-2 66 a 80 MHz ou um 486DX-4 100 a 120 MHz, aumentando a frequência da placa mãe de 33 para 40 MHz. Obviamente, fazer overclock aumenta o consumo e aquecimento do processador (exigindo muitas vezes o uso de um cooler mais parrudo), além da possibilidade de redução da vida útil. Mesmo assim, ele se tornou incrivelmente popular, por permitir aumentar o desempenho do equipamento a custo zero.

A partir das placas para Pentium II, os ajustes passaram a ser feitos através do setup, ao invés de jumpers. A placa detecta automaticamente as configurações usadas pelo processador, mas você pode alterar as configurações manualmente para fazer overclock.

Se, por acaso, você exagerar na dose e o micro não der mais boot, use o jumper do CMOS (o único ainda encontrado em todas as placas atuais) para limpar o CMOS, restaurando as configurações originais.

Aqui temos um exemplo de overclock (numa placa moderna) feito através das configurações do setup, onde a placa foi configurada para operar a 148 MHz (ao invés de 133), resultando num aumento proporcional da frequência do processador.



Mesmo depois do lançamento do Pentium 1, a plataforma 486 ainda teve uma sobrevida surpreendente graças aos processadores AMD 5x86, Cyrix Cx5x86 e ao Pentium overdrive da Intel, que serviam como opções de upgrade para quem tinha um 486DX-2 ou 486DX-4, prometendo um desempenho similar aos processadores Pentium low-end.

A grosso modo, um 486 possui um desempenho duas vezes superior ao de um 386 do mesmo clock. Isto significa que mesmo um 468SX de 25 MHz ganha fácil de um 386 de 40 MHz. Por outro lado, um 486 apresenta apenas 50 a 60% do desempenho de um Pentium 1 do mesmo clock. Isto significa que mesmo um Pentium 60 consegue superar um 486 DX4-100

O AMD 5x68 é um 486 "modernizado", lançado pela AMD no final de 1995. Ele vinha com 16 KB de cache L1, que operava em modo write-back (onde o cache trabalha cacheando tanto as operações de leitura, quanto as de escrita) e era fabricado numa técnica de produção de 0.35 micron, o que permitia que ele trabalhasse a 133 MHz (4x 33 MHz) e também o tornava menor e mais barato de se produzir. Na época, ele era o processador mais barato do mercado e permitia montar micros usando placas para 486, que também eram bem mais baratas. Isso o tornou especialmente popular aqui no Brasil.

Apesar do clock de 133 MHz, o 5x86 concorria com o Pentium 75 em termos de desempenho. Era comum overclocar estes processadores para 160 MHz (aumentando a frequência da placa mãe de 33 para 40 MHz), resultando num desempenho próximo ao de um Pentium 90.

Concorrendo com o 5x86, a Cyrix lançou o Cx5x86, um processador que utilizava uma arquitetura bastante similar ao Pentium da Intel, mas usava placas para 486. A versão de 100 MHz rivalizava com o 5x86 de 133 MHz, enquanto a de 120 MHz conseguia superá-lo com uma certa margem. O maior problema é que o Cx5x86 era bem mais caro, por isso acabou não sendo uma opção muito popular.

Completando o time, a Intel lançou o Pentium overdrive, um Pentium adaptado para utilizar as placas para 486, que existia em versões de 63 e 83 MHz. Apesar da diferença de clock, o overdrive de 83 MHz

conseguia competir de igual para igual com o 5x86 de 133 MHz, com o vencedor variando de acordo com a aplicação usada. Apesar disso, o AMD 5x86 acabou sendo mais popular devido à questão do custo.

Muitas placas-mãe desta época vinham sem memória cache, trazendo no lugar um encaixe marrom, ao lado do processador, que permitia encaixar um módulo COAST (cache on a stick), com 128, 256 ou 512 KB de cache. Estas placas são bem diferentes da primeira safra de placas para 486, pois já possuem slots PCI e utilizam pentes de memória de 72 vias, assim como as placas para Pentium 1. Aqui temos um 5x86, com o módulo COAST encaixado:



Com o Pentium, a Intel tentou solucionar os principais gargalos de desempenho da plataforma anterior, mudando vários aspectos não apenas do processador, mas também nas placas-mãe para ele.

A primeira mudança trazida pelo Pentium foi o aumento da quantidade de cache L1, que passou a ser de 16 KB, dividido em dois blocos de 8 KB. Isso melhorou a eficiência do cache (em relação ao cache unificado do 486), permitindo que o processador consiga acessar instruções e dados necessários para executá-las simultaneamente, ao invés de precisar fazer duas operações separadas. Além do aumento no tamanho, o cache passou a ser acessado através de um barramento interno de 512 bits (256 bits para o cache de dados e mais 256 para o de instruções), contra apenas 128 bits no 486. Assim como no 5x86 da AMD, passou também a ser utilizado cache write-back, que também cacheava as operações de escrita. O cache usado no 486, cacheia apenas as operações de leitura, o que permite ao processador ganhar tempo ao ler dados, mas não ajuda na hora de gravar, onde o processador tinha que esperar pela memória RAM.

Com a popularização dos games e aplicativos 3D de uma forma geral, o desempenho do processador em ponto flutuante passou a ser cada vez mais importante. Combinado com as melhorias no cache, foi desenvolvido um co-processador aritmético cerca de 5 vezes mais rápido que o usado no 486. O co-processador do Pentium era tão rápido (comparado com outros processadores da época), que mesmo um K6-2 perdia para um Pentium da mesma frequência em jogos e aplicativos muito dependentes de cálculos de ponto flutuante.

O Pentium também foi o primeiro processador Intel a usar uma arquitetura superescalar. Internamente, o Pentium possui duas unidades de execução, com a inclusão de um circuito de branch prediction, encarregado de dividir as instruções entre as duas unidades e antecipar o processamento de instruções, de forma a manter ambas ocupadas na maior parte do tempo.

Esta tendência de "executar mais por ciclo de clock" continua até os dias de hoje, com os processadores incluindo um número cada vez maior de unidades de execução, mais memória cache, circuitos aprimorados de branch prediction, sem falar nos processadores dual-core e quad-core. Mesmo que existisse um 486 ou Pentium capaz de operar a 2 ou 3 GHz, o desempenho seria ridiculamente inferior ao de um processador atual operando na mesma frequência.



Acompanhando as melhorias no processador, foram feitas mudanças nas placas-mãe, de forma a melhorar o desempenho de acesso à memória. Em primeiro lugar, as placas para Pentium operam a frequências mais altas: 60 ou 66 MHz, de acordo com o processador

mbora as memórias FPM e EDO usadas na época não fossem capazes de acompanhar a mudança (elas passaram a trabalhar usando tempos de espera), o aumento da frequência permitiu usar chips de cache L2 mais rápidos (já que eles operavam na frequência da placa mãe). Além disso, o Pentium passou a acessar a memória a 64 bits, ao invés de 32 bits como no 486, o que efetivamente dobrou a velocidade do barramento com a memória.

Como os pentes de 72 vias usados na época eram módulos de 32 bits, era necessário utilizá-los aos pares, sempre 2 ou 4 módulos, sendo que os módulos dentro de cada par deveriam ser idênticos. Estes pentes de 72 vias são antecessores diretos dos módulos DIMM usados atualmente.

Aqui temos uma comparação entre um módulo DIMM, um módulo de 72 vias e um antigo módulo de 30 vias, dos usados nos micros 386 e nos primeiros 486:



Os processadores Pentium existiram em versões de 60 a 200 MHz, sempre com a placa-mãe operando a 60 ou 66 MHz e usando um multiplicador de 1x, 1.5x, 2x, 2.5x ou 3x para atingir a frequência anunciada. Era comum fazer overclock aumentando a frequência da placa mãe para 75 MHz, opção oferecida pela maioria das placas soquete 7 da época.

Assim como na época dos micros 486, as placas mãe para processadores Pentium (com exceção de placas muito antigas) suportam várias frequências de barramento e vários multiplicadores distintos. Na maioria dos casos é possível configurar a placa mãe para utilizar qualquer processador da família.

Em 1996 a Intel lançou o **Pentium MMX**, que foi o integrante final da família Pentium 1. Ele chegou ao mercado acompanhado de uma forte campanha de marketing, que divulgavam as novas instruções MMX como a maior invenção da humanidade depois da batata frita ;). As instruções MMX permitiam que até 4 instruções simples fossem combinadas numa única instrução complexa e fossem assim processadas num único ciclo de clock. A Intel esperava que elas fossem capazes de multiplicar o desempenho do processador em aplicativos multimídia, principalmente em compressão e descompressão de áudio, mas no final acabou dando com os burros n'agua.

As instruções MMX continuam disponíveis nos processadores atuais, mas nunca foram muito utilizadas devido a uma combinação de fatores.

Para não depender apenas das novas instruções, a Intel aumentou o cache L1 do processador, de 16 para 32 KB. Com isto, o MMX passou a ser um pouco mais rápido do que um Pentium 1 da mesma frequência, mesmo nos aplicativos sem otimização. Lembre-se de que naquela época o cache L2 do processador ainda fazia parte da placa mãe e operava a apenas 66 MHz. Um cache L1 competente era essencial.

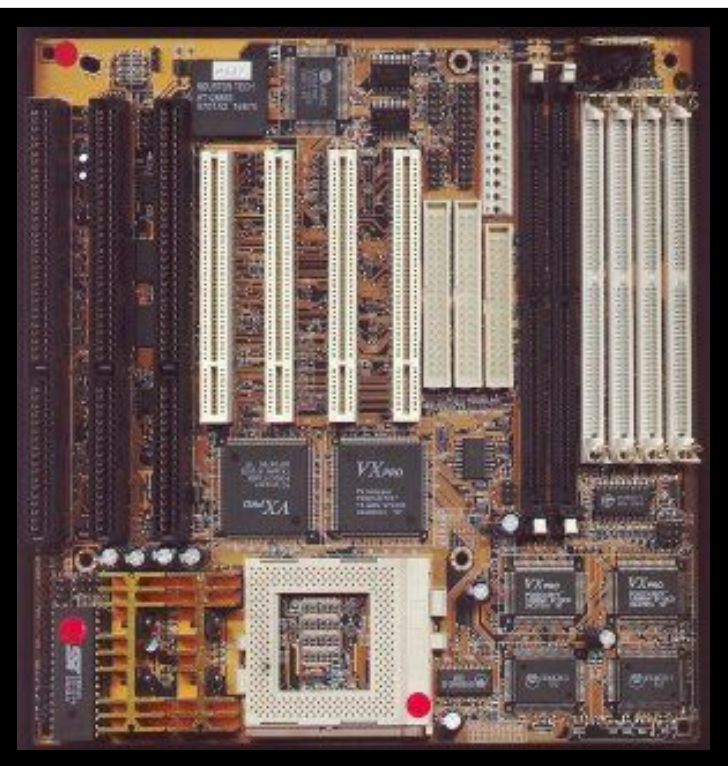
O MMX foi lançado em versões de 200 e 233 MHz, ambas compatíveis com a grande maioria das placas soquete 7 existentes. Ele também foi o primeiro processador Intel a usar um encapsulamento plástico com um dissipador metálico, ao contrário da cerâmica usada nos anteriores. Esta mudança foi na verdade bastante benéfica, pois o dissipador metálico é muito mais eficiente na dissipação do calor do que a cerâmica, o que melhora a eficiência do cooler.





Aqui temos uma VX Pro, uma placa bem comum no final da era Pentium. Ela suporta todos os processadores da família, até o 233 MMX. Note que ela inclui três slots ISA, pois os modems e placas de som ISA ainda eram bastante comuns.

Além dos 4 soquetes para pentes de 72 vias, ela possui também dois soquetes para módulos DIMM, mas eles não servem para muita coisa, pois o chipset só é capaz de endereçar chips de memória de até 2 MB, permitindo que fossem utilizados pentes de no máximo 32 MB. Mesmo que você tentasse instalar um pente de maior capacidade, a placa só reconhecia 16 ou 32 MB, de acordo com o número de chips no pente.



Durante a era Pentium, tivemos também os chips K5 e K6 (da AMD), o Cyrix 6x86 e o IDT C6, que eram compatíveis com as mesmas placas, mas que possuíam uma fatia relativamente pequena do mercado. A Cyrix foi comprada pela Via em 1999, que usou sua tecnologia para desenvolver os processadores Via C6 e Via C7, que são vendidos atualmente. A IDT era uma fabricante menor, que desapareceu sem deixar rastros. Das antigas concorrentes, sobrou a AMD, que com o lançamento do Athlon 64 e derivados, possui hoje mais de 20% do mercado.

Na época, a AMD ainda era uma empresa menor, que enfrentava problemas financeiros depois das fracas vendas dos processadores K5 e K6. Para completar, a Intel passou a utilizar um barramento proprietário no Pentium II, impedindo que a AMD desenvolvesse processadores capazes de utilizar as placas-mãe para ele.

Sem escolha, a AMD apostou tudo no K6-2, um processador relativamente poderoso, que mantinha compatibilidade com as placas soquete 7. Em termos de processamento de inteiros, o K6-2 competia diretamente com um



Pentium II do mesmo clock, o maior problema continuava sendo o co-processador aritmético, que era até duas vezes mais lento.

Para reduzir a diferença, a AMD investiu no desenvolvimento de um novo conjunto de instruções, o 3D-Now, formado por 27 novas instruções que tem o objetivo de agilizar o processamento 3D, sobretudo em jogos. O 3D-Now só ajudava em aplicativos otimizados, mas ao contrário do MMX ele era realmente eficiente, o que levou muitas empresas a otimizarem seus títulos para ele. O K6-2 também incorporou, sem muito alarde, o suporte às instruções MMX.

Junto com a AMD, fabricantes de chipsets, como a Via, SiS e Ali foram prejudicados pela decisão da Intel, pois não podiam desenvolver chipsets para o Pentium II sem o pagamento de licenças, o que continua até o dias de hoje. Isto causou uma certa "revolta" entre os fabricantes, que passaram a apoiar a AMD, desenvolvendo placas soquete 7 aprimoradas, que suportavam barramento de 100 MHz, utilizavam pentes de memória DIMM e possuíam slots AGP. Estas placas passaram a ser chamadas de placas super 7 e, junto com os processadores K6-2, se tornaram uma opção de baixo custo para quem não queria gastar com um Pentium II.

*Esta é uma Asus P5A-B, uma placa super 7 bastante comum na época. Veja que ela possui slot AGP e três soquetes para pentes de memória DIMM, que permitem o uso de módulos até 256 MB. Ela oferecia suporte a toda a família K6-2, até o K6-2 550, além de manter suporte aos processadores antigos:*



Umas das prioridades da AMD foi manter a compatibilidade com as placas soquete 7 antigas. Por isso, optaram por vender o K6-2 com o multiplicador destravado. Isso permitia instalar processadores K6-2 em placas antigas, que trabalhavam a apenas 66 MHz, desde que a placa suportasse a tensão de 2.2v (ou 2.4v nas versões mais recentes) utilizada pelo processador. Um K6-2 de 300 MHz podia ser utilizado tanto numa placa mãe configurada para operar a 100 MHz com multiplicador de 3x, quanto em uma placa configurada para operar a 66 MHz com multiplicador de 4.5x. Naturalmente, o desempenho será melhor na placa de 100 MHz, pela diferença na frequência da memória e do cache L2.

Em 1999 a AMD lançou uma última atualização para a plataforma K6, na forma do K6-3, o primeiro processador AMD a trazer cache L2 integrado ao processador. Além de manter os mesmos 64 KB de cache L1 do K6-2, o K6-3 incorporou 256 KB de cache L2 full-speed (operando na mesma frequência do processador), assim como os processadores atuais. Ele também aproveitava o cache disponível nas placas soquete 7, que passava a funcionar como um cache L3, resultando em mais um pequeno ganho.

Embora fosse perceptivelmente mais rápido que um K6-2 do mesmo clock, o K6-3 era mais caro e foi lançado no finalzinho da era soquete 7, quando a plataforma já caminhava para a obsolência. Por causa disso, ele acabou não fazendo muito sucesso, sendo substituído rapidamente pelos Athlons e Durons.

Como disse, ao desenvolver o Pentium II, a Intel optou por desenvolver um barramento proprietário (o GTL+), de forma a dificultar a vida dos concorrentes. Inicialmente a AMD continuou usando as placas soquete 7, mas em seguida respondeu com o EV6, um barramento próprio, utilizado pelos processadores Athlon e Duron. A partir daí, nunca mais tivemos um barramento padrão, que permitisse a criação de placas-mãe capazes com suporte a processadores dos dois fabricantes, como na época das placas soquete 7.

As primeiras versões do Pentium II utilizavam o encapsulamento SEPP (Singled Edge Processor Package), um formato dispendioso, onde ao invés de um pequeno

encapsulamento de cerâmica, temos uma placa de circuito, que traz o processador e o cache L2 integrado. Protegendo esta placa, temos uma capa plástica, formando um cartucho muito parecido com um cartucho de video-game, onde o cooler era instalado na parte de trás. O encaixe para ele foi batizado de slot 1:

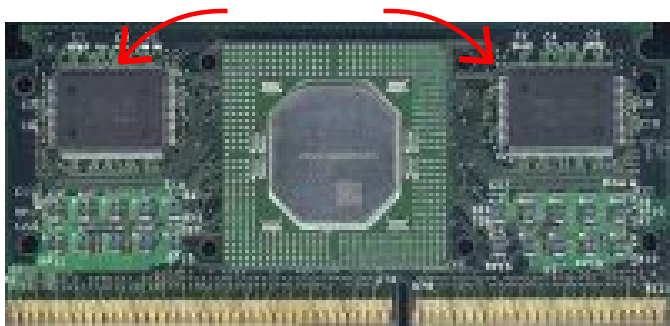


O Pentium II inclui 512 KB de cache L2, que opera à metade da frequência do processador (num Pentium II 400, por exemplo, o cache L2 opera a 200 MHz) e por isso é bem mais rápido que o cache usado nas placas soquete 7, que opera a apenas 66 ou 100 MHz. Com o cache movido para dentro do processador, as placas deixaram de trazer cache externo, o que continua até os dias de hoje. O único processador para micros PC a utilizar cache L3, foi o K6-3.

Além do cache L2, o Pentium II manteve os 32 KB de cache L1 (dividido em dois blocos de 16 KB para dados e instruções) do MMX. Abrindo o cartucho, é possível ver os dois grandes chips de cache L2 instalados próximos ao die do processador:



*Observe na imagem como é possível ver os dois grandes chips de cache L2 instalados próximos ao die do processador:*



O Pentium II foi produzido em duas arquiteturas diferentes. As versões de até 300 MHz utilizam a arquitetura Klamath, que consiste numa técnica de fabricação de 0.35 micrón, muito parecida com a utilizada nos processadores Pentium MMX. Nas versões a partir de 333 MHz já é utilizada a arquitetura Deschutes de 0.25 micrón, que garante uma dissipação de calor muito menor, o que possibilitou o desenvolvimento de processadores mais rápidos. As versões do Pentium II de até 333 MHz usam bus de 66 MHz, enquanto que as versões a partir de 350 MHz usam bus de 100 MHz, quebrando a compatibilidade com as placas da geração anterior.

Com o lançamento do Pentium II, a Intel abandonou a fabricação do Pentium MMX, passando a vender apenas processadores Pentium II que eram muito mais caros. O problema com esta estratégia foi que a Intel passou a perder terreno rapidamente no mercado de PCs de baixo custo, principalmente para o K6-2.

Para preencher a lacuna, a Intel lançou o Celeron, que inicialmente era uma simples versão castrada do Pentium II, sem os chips de cache e o invólucro plástico. Este Celeron original era muito lento, pois não possuía cache L2 algum, contando apenas com os 32 KB de cache L1.

O cache L2 é um componente extremamente importante nos processadores atuais, pois apesar da potência dos processadores ter aumentado quase 10,000 vezes nas últimas duas décadas, a memória RAM pouco evoluiu em velocidade. Pouco adianta um processador veloz, se ao todo instante ele tem que parar o que está fazendo para esperar dados provenientes da memória RAM. É justamente aí que entra o cache secundário, reunindo os dados mais importantes da memória para que o processador não precise ficar esperando. Retirando o cache L2, a performance do equipamento cai em 40% ou mais. Justamente por isso, além de perder feio para o Pentium II, o Celeron sem cache perdia até mesmo para processadores mais antigos.

Esta primeira safra foi rapidamente substituída pelo Celeron Mendocino, que trazia 128 KB de cache L1 interno (full-speed), que resolveu o problema da performance. O Mendocino foi produzido em versões de 300 a 533 MHz, sempre utilizando barramento de 66 MHz. Além de possírem um desempenho próximo ao de um Pentium II do mesmo clock (o cache do Pentium II é maior, porém mais lento), as versões de 300, 333 e 366 MHz permitiam overlocks de 50%, atingindo

respectivamente 450, 500 e 550 MHz com boa estabilidade. Não poderia ser mais simples: bastava investir num cooler de boa qualidade e instalar o Celeron Mendocino numa placa mãe configurada para operar a 100 MHz.

O Celeron Mendocino foi também o primeiro processador a utilizar o soquete 370, que mais tarde seria utilizado pelo Pentium III Coppermine e demais processadores da família. Foram produzidos também adaptadores, que permitiam usar processadores soquete 370 em placas slot 1 compatíveis:



Em fevereiro de 1999 foi lançado o **Pentium III**. A versão inicial (que utilizava o core Katmai) não trazia grandes mudanças: era basicamente um Pentium II, que utilizava o encapsulamento SEPP, usava cache externo e operava a 500 MHz. A novidade foi a inclusão das instruções SSE que, assim como as instruções 3D-Now da AMD, eram voltadas para aplicativos multimídia e 3D.



Pouco depois, em outubro de 1999, foi lançado o Pentium III **Coppermine**, uma versão aprimorada, produzida numa técnica de 0.18 micron, que trazia 256 KB de cache L2 de cache integrado (operando à mesma frequência do processador) e abandonou o formato SEPP em favor do FC-PGA, que utilizava placas mãe soquete 370. A mudança decretou a morte do slot 1, que não voltou a ser utilizado por outros processadores Intel. Apesar disso, as versões do Pentium III PC-PGA que utilizavam bus de 100 MHz ainda podiam ser usadas na maioria das placas slot 1 antigas, com a ajuda do adaptador (muitas placas precisavam de uma atualização de BIOS).



Não demorou para que a Intel lançasse também uma nova versão do Celeron, baseada na mesma arquitetura, dando continuidade à tradição de overlocks gigantescos.

O Celeron Coppermine nada mais era do que um Pentium III com metade do cache L2 desativado (128 KB), que utilizava bus de 66 MHz (ao invés de 100 ou 133, como as diferentes versões do Pentium III). Embora fosse originalmente mais lento que um Pentium III do mesmo clock, o Celeron Coppermine de 600 MHz podia tranquilamente operar a 900 MHz, utilizando bus de 100 MHz, oferecendo um desempenho similar ao de um Pentium III 800 a uma fração do custo. Eu mesmo tenho um, funcionando até hoje ;).

O Celeron acabou se revelando um bom negócio para a Intel, pois permitia aproveitar processadores Pentium III com defeitos na memória cache, que de outra forma iriam para o lixo.

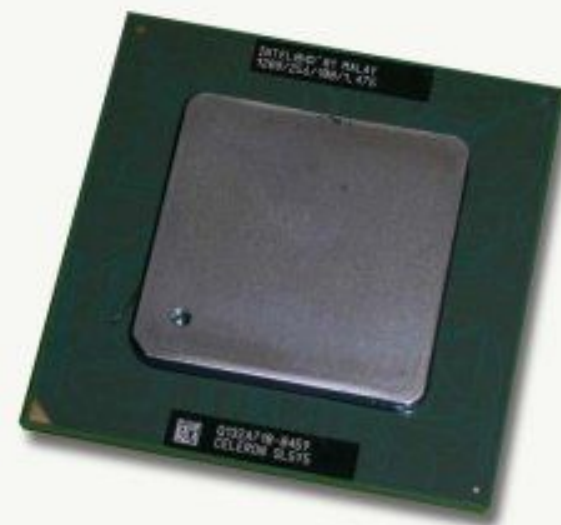
Quando ocorre um defeito no cache, em geral apenas alguns poucos bits são afetados, geralmente bits fisicamente próximos. Antes de saírem de fábrica, todos os processadores são rigorosamente testados, e os que apresentam defeitos no cache são separados. O Pentium III foi projetado de tal maneira que o cache L2 era dividido em duas seções de 128 KB, que podiam ser desabilitadas individualmente (um processo irreversível). Como é usada apenas a metade "boa" do cache, o processador funciona perfeitamente e temos mais um consumidor satisfeito.

O Celeron Coppermine foi lançado em versões de até 950 MHz, sendo que a partir do Celeron 800 passou a ser utilizado bus de 100 MHz (que melhorou um pouco o desempenho do processador em relação aos antigos).

O Pentium III passou a utilizar barramento de 133 MHz e foi lançado em versões de até 1.0 GHz, até ser substituído pelo Pentium 4.

Entre 2001 e 2002, já depois de descontinuar o Pentium III, a Intel produziu pequenas quantidades do Celeron Tualatin, uma versão aprimorada, produzida numa técnica de 0.13 micron e equipada com 256 KB de cache L2. O **Tualatin** existiu em versões de 1.0 e 1.4 GHz e era compatível com a maioria das placas soquete 7 para Pentium III (embora muitas precisassem de uma atualização de BIOS). Ele possuía um bom desempenho em relação ao Pentium 4 e era bastante econômico com relação ao consumo elétrico e aquecimento, mas a Intel optou por não levar o projeto adiante, com medo de prejudicar as vendas do Pentium 4.

***Este é um exemplar raro do Tualatin de 1.2 GHz:***



Uma questão interessante é que os processadores Pentium II e Pentium III são derivados do **Pentium Pro**, um processador destinado a workstations e servidores, lançado pela Intel em 1995, em versões de 150 a 200 MHz. Embora seja contemporâneo do Pentium I, o Pentium Pro era baseado numa arquitetura completamente diferente (a plataforma P6), inspirada nos processadores RISC.

Ao invés de processar diretamente todas as instruções complexas incluído no conjunto x86, o Pentium Pro inclui unidades de execução capazes de processar apenas instruções simples, porém a uma grande velocidade, quase sempre uma instrução em cada unidade de execução, por ciclo de clock. Antes de chegarem às unidades de execução, as instruções complexas utilizadas pelos programas são convertidas em instruções simples pelas unidades decodificadoras, incluídas no processador. O Pentium Pro inclui três unidades de execução e uma unidade decodificadora capaz de mantê-las ocupadas na maior parte do tempo.

Para completar, o processador inclui 256 KB de cache L2 full speed, incluído no mesmo encapsulamento do processador. Isto soa familiar não é? O Pentium Pro é muito similar a um Pentium III Coppermine, só que lançado 4 anos antes. Na época a Intel não possuía tecnologia para incluir o cache L2 diretamente no die do processador, por isso era utilizado um waffer separado, porém incluído no mesmo encapsulamento do processador. Isso fazia com que o Pentium Pro tivesse um formato retangular:



Naturalmente, o Pentium Pro não possuía as instruções SSE, mas os ingredientes básicos já estavam presentes. Podemos dizer que tanto o Pentium II, quanto o Pentium III são versões melhoradas do Pentium Pro, adaptadas ao mercado doméstico.

Voltando à AMD, embora o K6-2 tenha dado combate ao Pentium II e às primeiras versões do Pentium III, as limitações da antiga plataforma soquete 7 limitavam o desempenho do processador. Outro problema é que o K6-2 era uma plataforma de baixo custo, que levava ao aparecimento de placas mãe cada vez mais baratas e de qualidade cada vez pior. Os constantes problemas com as placas acabaram fazendo com que o K6-2 ficasse com má fama no mercado, embora o processador em si apresentasse um bom custo benefício.

**Em 1999 a AMD finalmente conseguiu lançar o Athlon (na época também chamado de K7),** uma plataforma completamente nova, que conseguiu solucionar os principais problemas associados ao K6-2. Apesar de toda a evolução, todos os processadores AMD lançados daí em diante, incluindo os Athlon 64 dual-core e quad-core continuam sendo baseados nesta mesma arquitetura.

As primeiras versões do Athlon utilizavam um formato de cartucho, muito similar ao usado pelo Pentium II, com chips de memória cache externos, operando à metade da frequência do processador. As placas mãe utilizavam o slot A, um conector similar ao usado pelo Pentium II, porém incompatível e com uma pinagem diferente.

O uso de cache externo atrapalhou o desempenho desta versão inicial, pois na época a Intel já vendia o Pentium III Coppermine, com cache full-speed, que era mais rápido e mais barato de se produzir. Para piorar, a AMD não conseguiu obter chips de memória cache capazes de operar a mais de 350 MHz, de forma que o divisor da frequência do cache foi aumentando conforme lançava processadores mais rápidos.

As versões de até 700 MHz do Athlon slot A trazem cache L2 operando à metade da frequência do processador. As versões de 750, 800 e 850 MHz trazem cache operando a apenas 2/5 da frequência, enquanto nas versões de 900, 950 e 1 GHz o cache opera a apenas 1/3 da frequência.



Esta cara e desajeitada versão inicial do Athlon foi rapidamente substituída pelo **Athlon Thunderbird**, que incorporou 256 KB de cache L2 full-speed e voltou a utilizar o formato soquete, dando início à era soquete A (ou soquete 462). Este mesmo formato continuou sendo usado pelos Durons, Athlons XP e Semprons, até a introdução do Athlon 64, que passou a utilizar placas-mãe baseadas no soquete 754 ou 939.



O **Duron** substituiu o K6-2 como processador de baixo custo da AMD e prestou bons serviços, concorrendo com as diferentes versões do Celeron. Em todas as suas encarnações, o Duron possui apenas 64 KB de cache L2. Entretanto, ele conserva o massivo cache de 128 KB do Athlon, o que cria uma configuração interessante, onde temos mais cache L1 do que L2. Assim como o Athlon, o Duron utiliza um sistema de cache "exclusivo", onde os dados armazenados no cache L1 são sempre diferentes dos armazenados no cache L2. O cache L1 do Celeron, por sua vez, armazena sempre cópias de informações já armazenadas no cache L2 de 128 KB. Graças a esta característica, o Duron acaba levando vantagem com relação ao cache, pois é capaz de armazenar um total de 196 KB de informações nos caches, contra apenas 128 KB no Celeron.

A versão inicial do Duron utilizava o core Spitfire (baseado no Athlon Thunderbird) e existia em versões de 600 a 950 MHz. Em seguida foi lançado o Duron Morgan (baseado no Athlon XP), que existiu em versões de até 1.3 GHz, quando foi substituído pelo Sempron.

As versões iniciais do Athlon Thunderbird utilizavam barramento de 100 ou 133 MHz, assim como o Pentium III. Entretanto, as versões seguintes (com core Thoroughbred e Barton) passaram a utilizar bus de 166 ou 200 MHz, memórias DDR e tensões mais baixas, quebrando a compatibilidade com as placas antigas.

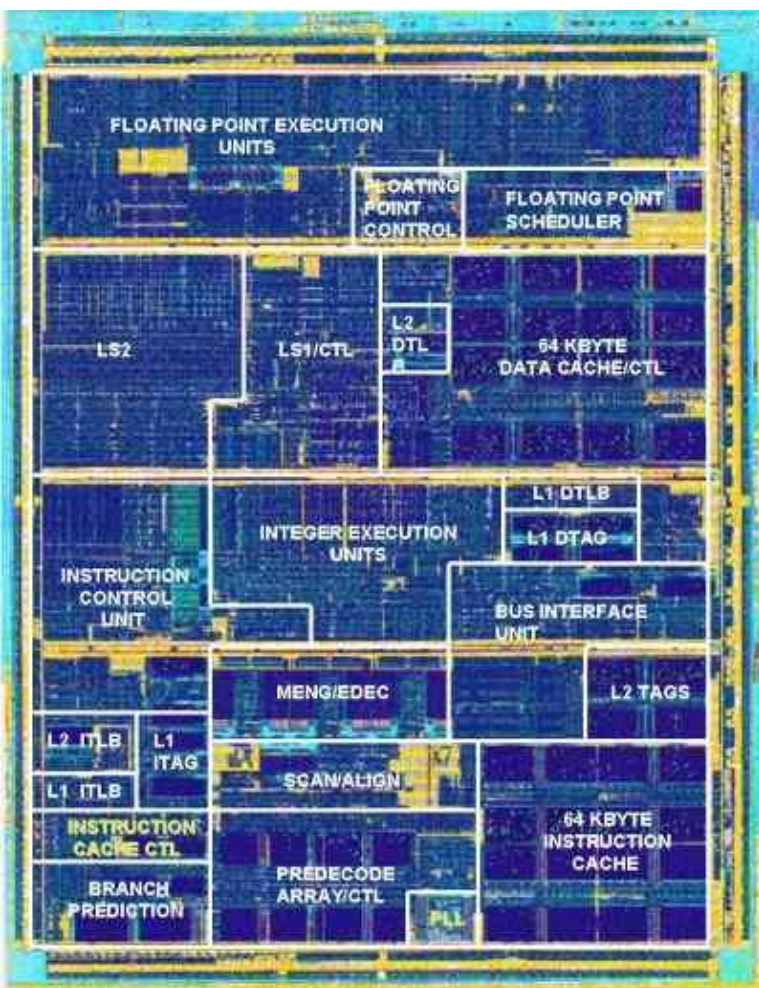
Com o uso de memória DDR, o barramento passa a realizar duas transferências por ciclo, por isso é comum que os fabricantes dobrem a frequência na hora de divulgar as especificações, chegando aos 333 (166 x 2) ou 400 MHz (200 x 2).

Como citei anteriormente, o Intel patenteou o barramento GTL+ usados pelos processadores Pentium II em diante, assim como o slot 1 e o soquete 370, de forma que a AMD não podia utilizá-los em seus processadores. A AMD optou então por licenciar o barramento EV6, desenvolvido pela Alpha Digital. O EV6 possui uma vantagem importante sobre o barramento da Intel, que é o fato de ser um barramento ponto a ponto. Nele, cada processador tem seu barramento exclusivo de comunicação com o chipset, permitindo o desenvolvimento de placas para dois ou quatro processadores Athlon, trabalhando em SMP, cada um com o seu barramento exclusivo com a memória e outros componentes, resultando em um ganho considerável de desempenho em relação ao Pentium III. Embora a AMD não tenha conseguido tirar muito proveito deste recurso nos desktops, ele ajudou o Athlon a ganhar espaço nos servidores, onde máquinas com dois processadores são comuns.

*Na próxima página temos uma foto que mostra como o Athlon é fisicamente. Esta foto foi cedida pela própria AMD e é bem interessante, pois ajuda a entender como os dados trafegam dentro do processador.*



*Veja que a foto está dividida em pequenos retângulos, cada um mostrando a área ocupada por cada componente do processador. Infelizmente tudo está em inglês, mas aqui vai uma breve legenda dos componentes mais importantes (na ordem em que cada componente aparece na ilustração, começando de cima):*



**Floating Point Execution Units:** São as unidades de execução de ponto flutuante. Nada mais é do que a parte fundamental do coprocessador aritmético, fundamental em vários aplicativos.

**Floating Point Control:** Este componente controla a ordem em que as instruções serão executadas pelo coprocessador, permitindo que as unidades de ponto flutuante fiquem ocupadas na maior parte do tempo.

**Floating Point Scheduler:** Armazena as próximas instruções que serão processadas pelo coprocessador. Este componente é essencial para o trabalho do Floating Point Control.

**64 Kbyte Data Cache:** Os 128 KB de cache L1 do Athlon são divididos em dois blocos, 64 KB para dados e 64 KB para instruções, esta divisão meio a meio é utilizada na maioria dos processadores atuais e melhora a velocidade de acesso, pois permite que os dois blocos sejam acessados simultaneamente. O Data Cache é a metade que armazena dados.

**Integer Execution Unit:** Estas são as unidades de execução de inteiros. Este é o componente básico de qualquer processador. É aqui que são processadas as operações envolvendo números inteiros.

**Instruction Control Unit:** Circuito que controla o envio de instruções para as unidades de execução de inteiros. Também ordena as instruções, de forma que possam ser processadas mais rápido.

**Bus Interface Units:** É por aqui que os dados entram e saem do processador. Controla a comunicação do processador com o chipset e com os demais componentes do micro.

**64 Kbyte Instruction Cache:** É o segundo bloco do cache L1, o bloco que armazena instruções.

**Branch Prediction:** O circuito de Branch Prediction é um dos componentes mais importantes dos processadores atuais, responsável por organizar as instruções de forma a manter as unidades de execução do processador ocupadas. Além de procurar adiante no código por instruções que podem ser "adiantadas", ele "adivinha" o resultado de operações de tomada de decisão (levando em conta fatores como o resultado de operações similares executadas anteriormente), permitindo que o processador vá "adiantado o serviço" enquanto o resultado da primeira operação ainda não é conhecido. Como todo bom adivinho, ele às vezes erra, fazendo com que o processador tenha que descartar todo o trabalho feito. Apesar disso, o ganho é muito grande, pois nos processadores atuais o circuito de branch prediction acerta em mais de 90% das vezes.

**Predecode Array:** Esta é a porta de entrada do processador. Estes circuitos convertem as instruções x86 enviadas pelos programas nas instruções simples que o processador executa internamente.

Até certo ponto, tanto o Pentium III quanto o Athlon e outros processadores x86 atuais trabalham da mesma maneira. Internamente, o processador é capaz de executar apenas instruções simples, para ser mais exato apenas quatro instruções: adição, atribuição, leitura e gravação.

Se você já estudou alguma linguagem de programação, aprendeu a importância das variáveis, que são pequenos espaços de memória reservados para guardar algum tipo de dado. Existem vários tipos de variáveis, de 8, 16, 32 bits, etc. que mudam de nome dependendo da linguagem de programação usada. A instrução de atribuição do processador é usada sempre que é necessário criar ou alterar o valor de uma variável.

Por exemplo, imagine que um programa qualquer criou uma variável de 8 bits com o número 5. A próxima instrução manda que o programa compare o valor da variável com o número 6 e, caso o número seja menor, altere o valor para 9. Como 5 é menor que 6, o programa decide fazer a alteração, feita utilizando a operação de atribuição do processador, que lê o valor 9 e grava-o no espaço de memória da variável, que passa a ter o valor 9 ao invés de 5.

A instrução de soma é a operação básica que permite fazer todo tipo de processamento, enquanto as instruções de leitura e gravação permitem mover os dados. Basicamente é só isso que um processador atual sabe fazer. Operações mais complexas, são executadas através da combinação de várias instruções simples.

Para calcular uma multiplicação, por exemplo, o processador utilizará seqüencialmente várias operações de soma. Na verdade, dentro do processador todas as operações, mesmo as mais complexas, são calculadas com base em várias operações de soma, feitas entre os valores binários processados pelo processador. Uma operação de subtração é conseguida através de uma operação de atribuição, que transforma um dos valores em negativo, seguida por uma operação de soma.

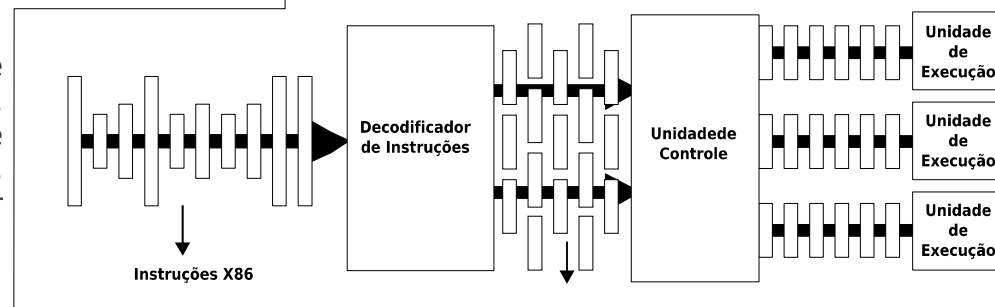
Uma operação de divisão é conseguida executando-se uma seqüência de operações de subtração e todas os demais cálculos, mesmo os cálculos mais complexos, executados pelo co-processador aritmético, são resolvidos usando apenas as quatro operações, obtidas a partir das simples instruções de soma e atribuição.

Pois bem, o conjunto de instruções x86, utilizadas pelos programas e com as quais qualquer processador destinado a micros PC deve ser compatível, é composto tanto por instruções simples (soma, subtração, leitura, gravação, comparação, etc.) quanto por instruções muito complexas, que devem ser quebradas em várias instruções simples para que possam ser executadas pelo processador.

Excluindo-se componentes de apoio, como o cache L1, deixando apenas a parte “funcional” do processador, podemos dividir o processador em três partes.

A primeira parte é o **decodificador de instruções**. Este componente tem a função de converter as instruções x86 usadas pelos programas nas instruções simples que podem ser executadas pelo processador. As instruções simples vão então para uma **unidade de controle** (o circuito de branch-prediction), que organiza as instruções da forma que possam ser executadas mais rapidamente. As instruções formam então uma fila, (scheduler) a fim de permitir que a unidade de controle tenha tempo de fazer ser trabalho. Lembre-se que os processadores atuais são superescalares, executam várias instruções por ciclo, simultaneamente, o que torna essencial a existência de algum circuito que as coloque numa ordem em que a execução de uma não dependa do resultado da execução de outra.

*Finalmente, temos as **unidades de execução**, onde as instruções preparadas e organizadas são finalmente processadas. Veja que todos os circuitos trabalham simultaneamente, visando que as unidades de execução sempre tenham algo para processar.*





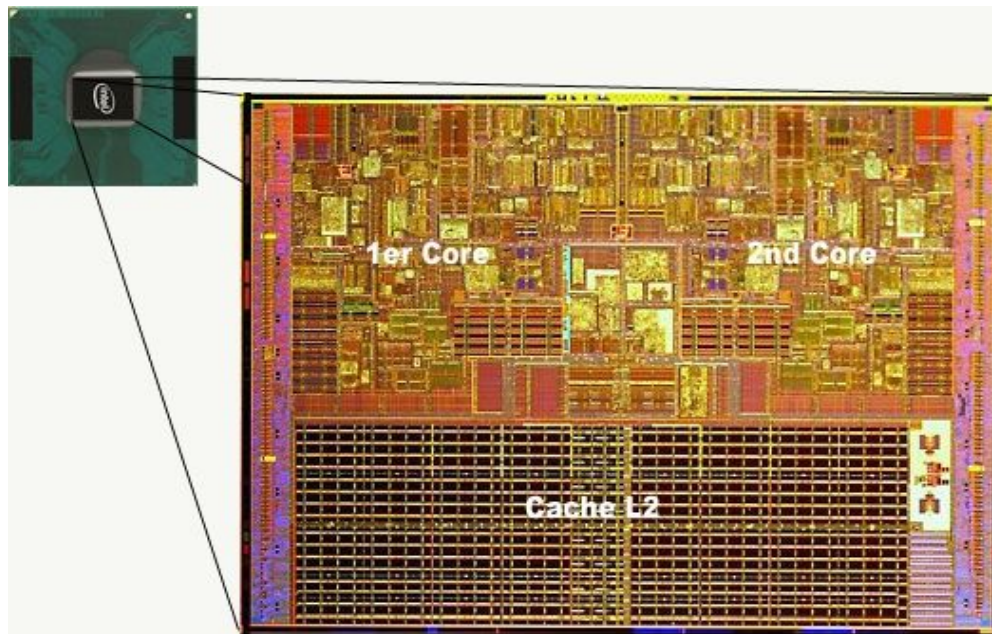
A lógica é que quanto mais unidades de execução tivermos trabalhando ao mesmo tempo, mais instruções todas juntas serão capazes de processar e quanto mais circuitos de decodificação e controle tivermos, mais eficiente será a decodificação das instruções, resultando em um processador mais rápido.

O maior limitante é que a maioria dos programas são desenvolvidos esperando que o processador processe uma instrução de cada vez. Temos também um grande número de operações de tomada de decisão, onde o processador precisa resolver uma determinada operação para então

poder decidir o que vai fazer depois. Graças aos circuitos de branch-prediction, os processadores são capazes de ir "adiantando o serviço", processando outras instruções mais adiante, enquanto a operação de tomada de decisão é solucionada. De qualquer forma, existe um limite para quanto trabalho o processador é capaz de executar por ciclo.

Com mais do que três ou quatro unidades de execução, o processador acaba ficando ocioso grande parte do tempo, de forma que processadores muito mais complexos do que isso acabariam sendo um desperdício de recursos.

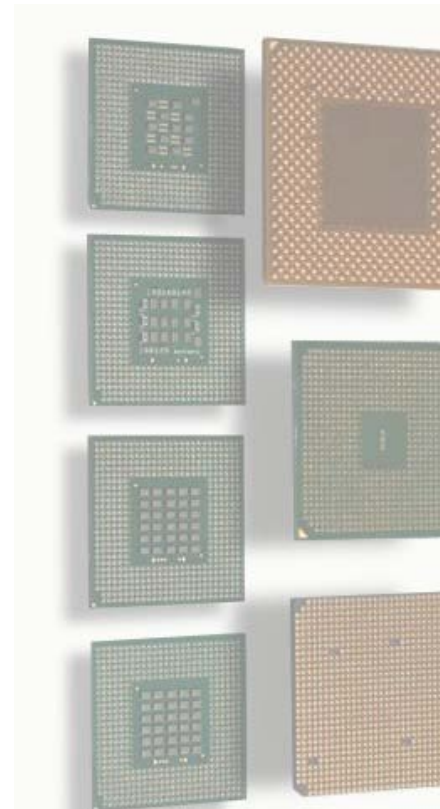
Ao invés de adicionar mais e mais unidades de execução aos processadores, os fabricantes passaram, a partir de um certo ponto, a desenvolver processadores dual-core e quad-core, onde temos dois ou quatro processadores no mesmo encapsulamento, trabalhando como se fossem um único processador. Isto resulta num melhor desempenho ao rodar vários aplicativos simultaneamente (você pode jogar e ripar um DVD ao mesmo tempo, por exemplo) e é muito mais simples e barato para os fabricantes do que desenvolver um único e processo super-processador.



*Aqui temos um diagrama da Intel, que mostra como isso funciona no Core Duo. Veja que temos um grande bloco de cache L2, que é compartilhado por dois processadores. Tudo isso dentro da mesma pastilha de silício:*

Carlos E. Morimoto.

É editor do site <http://www.guiadohardware.net>, autor de mais de 12 livros sobre Linux, Hardware e Redes, entre eles os títulos: "Redes e Servidores Linux", "Linux Entendendo o Sistema", "Linux Ferramentas Técnicas", "Entendendo e Dominando o Linux", "Kurumin, desvendando seus segredos", "Hardware, Manual Completo" e "Dicionário de termos técnicos de informática". Desde 2003 desenvolve o Kurumin Linux, uma das distribuições Linux mais usadas no país.







*Todos os processadores Intel e AMD, do 386 ao Athlon são chamados genericamente de "processadores x86", deixando claro que apesar de todas as diferenças de arquitetura, eles são compatíveis com o mesmo conjunto básico de instruções. É graças a isso que temos um número tão grande de softwares, acumulados ao longo de mais de duas décadas, que continuam funcionando nos PCs atuais. Com exceção da questão dos drivers, você poderia perfeitamente rodar o DOS e o Windows 3.11 em um Pentium 4, por exemplo.*

Duas grandes limitações da plataforma x86 é o pequeno número de registradores e o suporte nativo a apenas 4 GB de memória RAM, limitação compartilhada por todos os processadores de 32 bits.

Os registradores são pequenos espaços de memória utilizados pelo processador para armazenar informações que serão usadas para processar a instrução seguinte. Você pode pensar nos registradores como uma espécie de "cache L0". Eles estão muito próximos das unidades de execução do processador e por isso as informações podem ser acessadas imediatamente, mas em troca eles são incrivelmente limitados em tamanho. A arquitetura x86 prevê o uso de apenas 8 registradores, com 32 bits cada um. Ou seja somados os registradores armazenam apenas 8 bytes de dados.



*Processadores da família x86*

Não é possível adicionar mais registradores facilmente, pois softwares escritos para utilizarem os registradores adicionais não seriam mais compatíveis com os processadores antigos.

A solução encontrada foi utilizar o recurso chamado "register renaming" onde o processador possui um número maior de registradores, mas apenas 8 são visíveis. Os demais são gerenciados de forma dinâmica pelo processador, que vai chaveando entre eles conforme novas informações são necessárias. Esta técnica ameniza os problemas de desempenho causados pelo pequeno número de registradores, mas não soluciona a questão completamente.

Com relação à memória, o limite de 4 GB se tornou uma limitação séria em algumas áreas a partir da época do Pentium 1. A solução veio na forma do PAE (Physical Address Extension), um hack implementado pela Intel em alguns processadores a partir do Pentium Pro. O PAE consiste numa segunda tabela de endereços, com 4 bits adicionais, que permitem endereçar 16 páginas de memória, cada uma com 4 GB.

Com o PAE, passa a ser possível endereçar até 64 GB de memória. A desvantagem é que o processador continua sendo capaz de acessar apenas 4 GB por vez e o chaveamento entre diferentes páginas de memória toma um certo tempo, que acaba prejudicando bastante o desempenho. Ou seja, assim como no caso do register renaming, o PAE ameniza o problema, mas não o soluciona completamente.

A única forma de solucionar de vez as duas questões, assim como um conjunto de outros problemas comuns a todos os processadores de 32 bits, era desenvolver uma nova arquitetura, composta agora por processadores de 64 bits.

## Desenvolvendo um sucessor

Em primeiro lugar, precisamos definir o que são processadores de 64 bits e quais são as vantagens sobre os de 32. Em primeiro lugar, processadores de 64 bits não são duas vezes mais rápidos nem processam (necessariamente) o dobro de dados por ciclo de clock. A grande vantagem dos processadores de 64 bits é que eles são capazes de trabalhar com endereços de memória de 64 bits, o que permite endereçar muito mais do que 4 GB de memória RAM. Temos também um aumento no tamanho dos registradores, que passam a armazenar 64 bits de informações, ao invés de 32, o que representa um pequeno ganho de desempenho.

Outro benefício (embora menos significativo) é que eles são capazes de processar números inteiros de até 64 bits, ao invés de 32. Isto oferece ganhos de desempenho em algumas áreas específicas (como, por exemplo, softwares de criptação e em alguns aplicativos científicos) mas não ajuda muito nos aplicativos do dia a dia. Processadores de 32 bits podem processar números inteiros de 64 bits, mas eles precisam ser divididos em duas instruções separadas, o que naturalmente toma mais tempo.

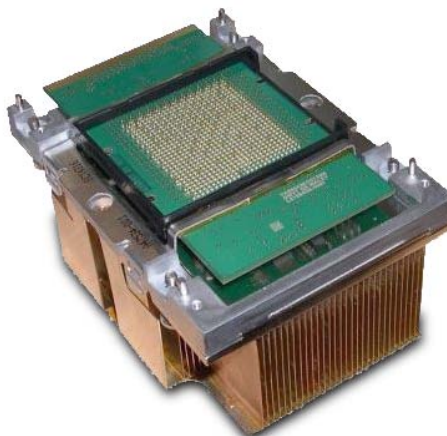
Com relação ao processamento de números de ponto flutuante, não existe ganho, pois os co-processadores aritméticos

utilizados nos processadores atuais já são capazes de processar números de ponto flutuante de 64 bits e vetores de 128 bits. Nestas duas áreas não existe um ganho direto, já que a transição já foi feita a muito tempo.

## Itanium

A primeira tentativa de criar um sucessor de 64 bits para a plataforma x86 veio da Intel, que em 2001 lançou o **Itanium**, um processador de 64 bits destinado a servidores, que quebra a compatibilidade com a família x86, passando a utilizar o IA-64, um novo conjunto de instruções. O plano inicial era popularizar a arquitetura primeiro nos servidores, onde os benefícios de um processador de 64 bits são mais evidentes, e em seguida lançar também versões destinadas a desktops. O problema com o Itanium (além do processador ser muito caro) era que não existiam softwares capazes de se beneficiar da nova plataforma. Ele incluía um sistema de emulação, que permitia rodar softwares de 32 bits, mas neste caso o desempenho era muito ruim.

As duas versões iniciais do Itanium operavam a 733 e 800 MHz. Elas foram seguidas pelas diferentes versões do Itanium 2, vendidas em versões de 900 MHz a 1.67 GHz, incluindo versões dual-core. A Intel continua investindo no desenvolvimento do Itanium, lançando novas versões, mas o futuro da plataforma é incerto.

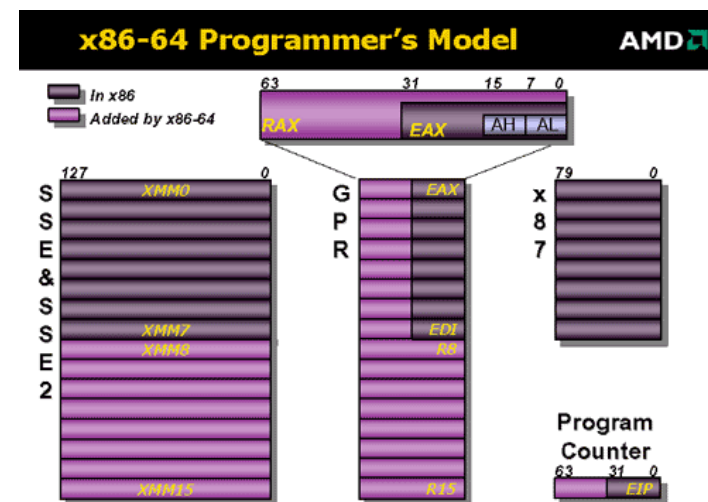


*Itanium 2*

## x86-84 (AMD64)

Ao invés de repetir o mesmo erro da Intel, lançando uma plataforma completamente nova, incompatível com os softwares atuais, a AMD decidiu desenvolver um projeto mais simples, que adicionasse suporte a instruções de 64 bits no Athlon (incluindo novos registradores, suporte a endereços de memória de 64 bits, etc.) sem entretanto quebrar a compatibilidade com os softwares de 32 bits, nem prejudicar o desempenho do processador ao executá-los. Com isto chegaram a um processador de 64 bits, com um número maior de registradores e maior capacidade de endereçamento de memória, mas que é capaz também de rodar aplicativos de 32 bits nativamente, sem perda de desempenho.

Este slide de apresentação da AMD mostra os novos registradores adicionados. Os 8 registradores x86 foram expandidos de 32 para 64 bits e foram adicionados 8 novos registradores de 64 bits, o que resultou em um espaço de armazenamento 4 vezes maior. Foram adicionados ainda 8 novos registradores para instruções SSE ou SSE2, mas neste caso não houve expansão, já que o SSE já utiliza registradores de 128 bits:



*Novos registradores x86-64*

No modo "legacy", o processador funciona como um processador x86 comum, executando instruções de 32 bits e utilizando apenas os registradores padrão. Ao mudar para o modo "long", o processador tem acesso a 16 registradores de 64 bits cada um e passa a suportar as instruções e endereços de memória de 64 bits. O chaveamento entre os dois modos é feito de forma muito rápida,



de forma que o processador pode inclusive rodar aplicativos de 32 bits dentro de um sistema operacional de 64 bits, da mesma forma que era possível rodar aplicativos DOS ou Windows 3.x (de 16 bits) dentro do Windows 98 (que já era um sistema de 32 bits).

O conjunto de instruções da AMD foi batizado de **x86-64** (e posteriormente renomeado para **AMD64**) e acabou sendo adotado também pela Intel, na forma do **EM64T**, um conjunto compatível, incluído sem muito alarde a partir do Pentium 4 com core Prescott. Pela primeira vez na história, a AMD ditou o novo padrão e a Intel se viu obrigada a segui-lo.

---

*Ao utilizar um processador Athlon 64 ou um Intel equipado com o EM64T, existem 3 possibilidades.*

---

A primeira é continuar usando as versões de 32 bits do Windows XP, Vista ou Linux, utilizando apenas aplicativos de 32 bits. Neste modo o processador se beneficia do controlador de memória integrado (no caso dos AMD) e outras melhorias na arquitetura, mas não utiliza os novos registradores, nem é capaz de endereçar mais do que 4 GB de memória nativamente.

A segunda possibilidade é utilizar um sistema operacional de 64 bits, como as versões de 64 bits do Vista e de diversas distribuições Linux. Neste caso você tem um pequeno ganho de desempenho devido ao uso dos novos registradores e o processador passa a suportar volumes maiores de memória RAM. A maioria dos aplicativos não exibe grandes ganhos de desempenho ao serem recompilados para rodarem em modo 64 bits, mas alguns (sobretudo bancos de dados) podem obter 15 ou mesmo 20% de ganho de desempenho em alguns casos.

A grande dificuldade em utilizar um sistema de 64 bits é encontrar versões nativas de todos os aplicativos. Chegamos então à terceira possibilidade, que é rodar um sistema de 64 bits, mas utilizar o modo de compatibilidade para executar aplicativos de 32 bits quando necessário. Neste modo, o sistema operacional precisa de mais memória, pois acaba tendo que manter carregadas muitas bibliotecas e componentes duplicados, mas o desempenho do processador não é prejudicado.

Como vimos, os aplicativos de 32 bits podem utilizar apenas 4 GB de memória (menos na prática, devido à limitações por parte dos sistemas operacionais). Um efeito colateral interessante é que, em um PC com 8 GB de memória, por exemplo, os aplicativos de 32 bits enxergarão e utilizarão apenas os primeiros 4 GB.

A memória adicional poderá ser usada pelo sistema operacional e aplicativos de 64 bits, mas não ajudará muito em casos onde os aplicativos de 32 bits sejam os aplicativos principais.

Com relação à memória, os processadores AMD64 são capazes de endereçar até 1 terabyte de memória física o que corresponde a 40 bits de endereços. Os 64 bits completos não são usados por questões relacionadas ao desempenho, já que existem módulos e placas que permitam utilizar tanta memória atualmente.

Apesar no AMD64 de reservar "apenas" 40 bits para o endereçamento da memória física, estão disponíveis 48 bits de endereçamento para o virtual address space, o que permite endereçar até 256 terabytes. Este limite não tem a ver com o uso de memória swap, mas sim com o suporte a arquivos grandes, como bancos de dados. Você poderia imaginar estes dois limites em ação no caso de um grande servidor, com muita memória RAM e diversos HDs em RAID, que armazenam um grande banco de dados.

Quando criamos um array RAID, o sistema passa a enxergar um único e grande disco, ao invés de vários HDs separados, de forma que podemos usar todo o espaço somado para armazenar um único e gigantesco arquivo.

Com o sistema de endereçamento do AMD64, o nosso hipotético servidor poderia

ter até 1 TB de memória RAM e um sistema operacional seria capaz de gerenciar um banco de dados de até 256 TB espalhado pelos diversos HDs, sem que fosse necessário recorrer a truques ou gambiarras para aumentar o espaço de endereçamento, como é necessário ao acessar mais do que 4 GB de RAM, ou gerenciar arquivos maiores do que 4 GB em um processador de 32 bits.

Quando este limite se tornar um empecilho, daqui a talvez mais uma década, novas revisões dos processadores podem estender os limites de endereçamento 56 ou mesmo 64 bits completos, permitindo gerenciar volumes virtualmente ilimitados de endereços.

Esta tabela, divulgada pela AMD, mostra os diferentes modos de operação suportados. Note que, além dos três modos que citei (64-Bit mode, que corresponde ao modo de 64 bits nativo; Compatibility Mode, onde são usados aplicativos de 32 bits, sob um sistema operacional de 64 bits e Protected Mode, onde são executados um sistema operacional e aplicativos de 32 bits) existe suporte também ao modo Virtual-8086 e ao Real Mode, que são os modos de legado, também suportados pelos processadores atuais, onde ele executa aplicativos de 16 bits, escritos para o DOS ou Windows 3.x:

Table 1-1. Operating Modes

Operating Mode		Operating System Required	Application Recompile Required	Defaults		Register Extensions	Typical
				Address Size (bits)	Operand Size (bits)		GPR Width (bits)
Long Mode	64-Bit Mode	New 64-bit OS	yes	64	32	yes	64
	Compatibility Mode		no	32		no	32
				16	16		16
Legacy Mode	Protected Mode	Legacy 32-bit OS	no	32	32	no	32
				16	16		
	Virtual-8086 Mode			16	16		16
	Real Mode	Legacy 16-bit OS					

Modos de operação dos processadores x86-64

Tenha em mente que a migração para os aplicativos de 64 bits é similar à migração dos sistemas operacionais de 16 bits (DOS e Windows 3.x) para os sistemas de 32 bits que ocorreu a partir de 1995. Ela acontecerá de forma gradual, de forma que daqui a 10 anos ainda estaremos rodando alguns aplicativos de 32, ou mesmo de 16 bits, utilizando o modo de compatibilidade.

## A arquitetura K8

Além do x86-64, a arquitetura K8, compartilhada por todos os processadores de 64 bits da AMD, inclui um circuito de branch-prediction sensivelmente melhorado, com um global history counter 4 vezes maior que o do Athlon K7. O global history counter é a área reservada a armazenar os resultados de operações resultadas anteriormente. Estas informações são utilizadas como base de consulta pelo circuito de branch prediction na hora de decidir qual caminho deve tomar dentro de uma operação de tomada de decisão.

Outro recurso, possivelmente o mais famoso é o controlador de memória integrado, chamado de **IMC** (Integrated Memory Controller). Tradicionalmente, o controlador de memória fazia parte do chipset e operava à mesma frequência que ele. O processador se comunicava com o chipset através do front side bus (o barramento principal) e ele era utilizado para todo o tráfego, incluindo a leitura e gravação de dados na memória.

Ao mover o controlador de memória para dentro do processador, a AMD foi capaz de reduzir sensivelmente o tempo de latência nos acessos à memória, aumentando assim o desempenho do processador em alguns pontos. No processadores baseados na arquitetura K8, o processador é ligado diretamente às trilhas da placa mãe que levam aos pentes de memória e a comunicação com o chipset é feita através de um barramento HyperTransport. No caso de sistemas com dois processadores em SMP (como no caso de servidores equipados com o Opteron), barramentos HyperTransport adicionais fazem a ligação entre os processadores.

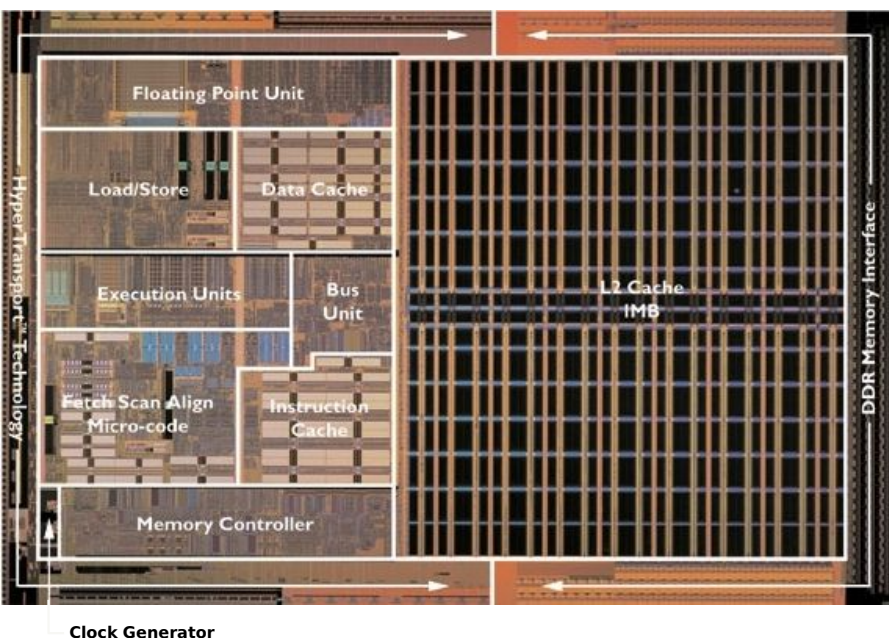
O Athlon 64 não utiliza mais um barramento frontal, como nos processadores antigos, mas sim um link HyperTransport que liga o processador ao chipset. Existe um clock de referência de 200 MHz, que substitui o FSB, gerando a frequência do processador, da memória e do próprio barramento HyperTransport. Tecnicamente, seria incorreto dizer que um Athlon 64 ou Sempron utiliza "200 MHz de FSB", já que ele não possui FSB e os 200 MHz são apenas a frequência de referência, mas para manter as coisas simples, vou continuar usando o termo "FSB" para descrever o clock da frequência de referência daqui em diante.

transferência por ciclo; isso faz com que a velocidade de transmissão seja muito baixa para os padrões atuais (apenas 133 MB/s), mesmo com o barramento transmitindo 32 bits de dados por ciclo.

O padrão HyperTransport é desenvolvido por um conjunto de fabricantes, por isso é utilizado em diversos dispositivos, indo do Xbox da Microsoft a roteadores da Cisco. Os chipsets nForce, para placas soquete A, por exemplo, já utilizavam o HyperTransport para interligar a ponte norte e ponte sul do chipset bem antes do Athlon 64 ser lançado.

No caso dos processadores AMD64, são utilizados links com 16 bits de largura em cada direção (16 pares), operando a 800 MHz (nas placas soquete 754) ou 1.0 GHz (nas placas soquete 939, 940 e AM2). Como são realizadas duas transferências por ciclo, podemos dizer também que a frequência "efetiva" é de (respectivamente) 1.6 GHz e 2.0 GHz

As placas AM3, cujo lançamento está previsto para 2008, trarão um barramento atualizado, capaz de operar a 2.0 GHz (4.0 GHz efetivos). Esta expansão não traz ganhos diretos, já que apenas aumenta a banda disponível para comunicação entre o processador e o chipset, sem influenciar na velocidade dos demais componentes. Ela serve apenas para evitar potenciais gargalos no futuro, conforme novos dispositivos e barramentos forem sendo incorporados.



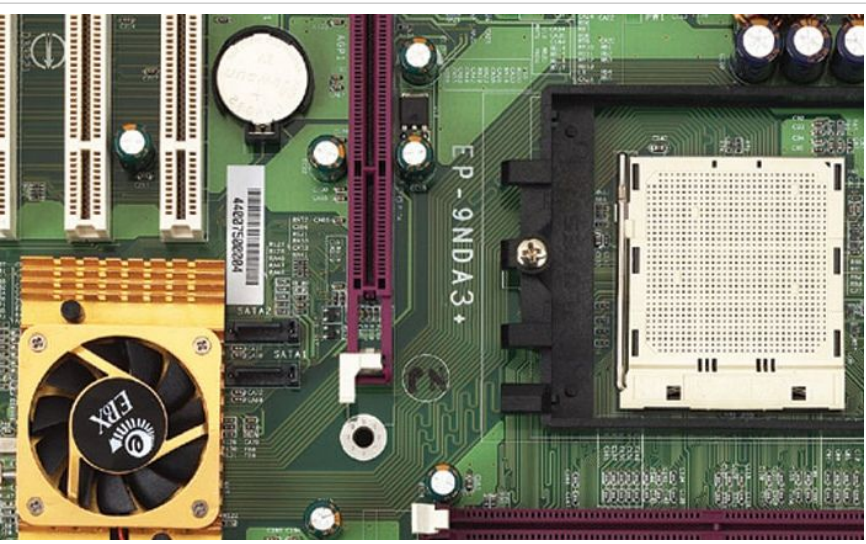
O **HyperTransport** é um barramento ponto a ponto, muito similar ao PCI Express sob diversos pontos de vista. A principal idéia é criar um barramento bastante rápido e de baixa latência utilizando poucas trilhas de dados.

Um link HyperTransport é sempre bidirecional, composto por dois links com de 2 a 32 bits de largura cada. Os dados são transmitidos duas vezes por ciclo (como nas memórias DDR) e a frequência de operação pode ser de 200 MHz a 2.6 GHz. Só para efeito de comparação, o barramento PCI opera a apenas 33 MHz, com apenas um



Com 16 bits de largura e operando a 800 MHz, com duas transferências por ciclo, temos um total de 6.4 GB/s de transferência de dados (3.2 GB/s em cada direção), o que é uma marca espantosa. Ao aumentar a frequência para 1.0 GHz, a taxa de transmissão sobe para 8 GB/s e, ao atingir 2.0 GHz, a taxa sobe para nada menos do que 16 GB/s.

Apesar da brutal taxa de transferência, o link HyperTransport é composto por um número relativamente pequeno de trilhas na placa mãe. Em muitas placas, é possível ver claramente as trilhas do HyperTransport ligando o processador ao chipset:



Trilhas referentes ao barramento HyperTransport interligando o processador e o chipset

Continuando, temos a questão do gerenciamento de energia. Com exceção do Sempron 2800+, todos os processadores da linha Athlon 64 suportam o Cool'n'Quiet, um sistema de gerenciamento de energia bastante eficiente que, assim como o SpeedStep da Intel e o PowerNow usado pela AMD nos processadores mobile, é capaz de ajustar a frequência de operação do processador de forma dinâmica de acordo com o uso. A principal diferença entre o Cool'n'Quiet e o PowerNow é que ele é otimizado para uso em desktops, de forma que utiliza um sistema menos agressivo de gerenciamento, que tem como prioridade não prejudicar o desempenho.

Um Athlon 64 3500+, por exemplo, pode trabalhar a 2.2 GHz (a frequência normal), 2.0, 1.8 ou 1.0 GHz, de acordo com os aplicativos em uso. Operando na frequência mínima, a tensão do processador cai de 1.4 para 1.1v, gerando uma economia adicional. No final, um Athlon 64, baseado no core Venice, ou mais atual, chega a consumir menos de 10 watts quando ocioso, operando na frequência e tensão mínimas.

Como a frequência do processador muda muito rapidamente, de acordo com a demanda (segundo a AMD, até 30 vezes por segundo), você mal percebe a redução de clock.

Demora exatamente o mesmo tempo para ripar um DVD, renderizar uma imagem 3D, aplicar um filtro no Photoshop, pois logo que o processador percebe o aumento na demanda de processamento, passa a operar na frequência máxima automaticamente.

As áreas onde você acaba percebendo alguma perda são justamente nas operações mais simples e rápidas, como fechar entre as janelas e abrir menus dentro dos programas. Estas operações são muito rápidas para disparar o aumento na frequência, de forma que, se prestar atenção, você realmente percebe alguma diferença, embora muito pouco.

De uma forma geral, o Cool'n'Quiet é um recurso muito interessante, pois aumenta a vida útil do equipamento (mesmo componentes como o HD e a placa mãe duram mais ao operar numa temperatura mais baixa, sem falar do cooler, que acumula menos poeira e assim precisa de menos manutenção) e gera economia na conta de luz, em troca de uma redução de desempenho muito pequena.

Com o passar dos meses, a economia pode somar um valor significativo. Em um PC com um Athlon 64 3500+ que ficasse ligado continuamente, teríamos uma economia de aproximadamente 30 kW por mês. Em dois anos, isso representa uma economia de mais de R\$ 350 na conta de luz, suficiente para pagar parte do próximo upgrade. :)

Para utilizar o Cool'n'Quiet, basta manter o recurso ativado no setup e instalar os drivers da placa. No caso do Linux, o suporte é oferecido através do módulo "powernow-k8" do Kernel e o serviço "powernowd" (instalado através do pacote de mesmo nome). Com os dois componentes disponíveis, a configuração é bem simples, basta carregar o módulo e ativar o serviço, como em:

```
# modprobe powernow-k8
# /etc/init.d/powernowd start
```

Você pode acompanhar a oscilação na frequência do processador utilizando um monitor de sistema qualquer. No Linux, você pode utilizar também o comando "cat /proc/cpuinfo".

Com o Athlon 64, a AMD voltou a utilizar o heat spreader sobre o núcleo do processador, assim como na época do K6-2. Devido a isso, não é mais possível diferenciar rapidamente as diferentes famílias do Athlon 64 apenas olhando para o tamanho e a posição do núcleo do processador, como na época do Athlon XP:



Athlon 64 e Athlon XP

Assim como o clock do processador, técnica de produção e quantidade de cache, o controlador de memória é também usado como um diferencial entre as diferentes famílias de processadores.

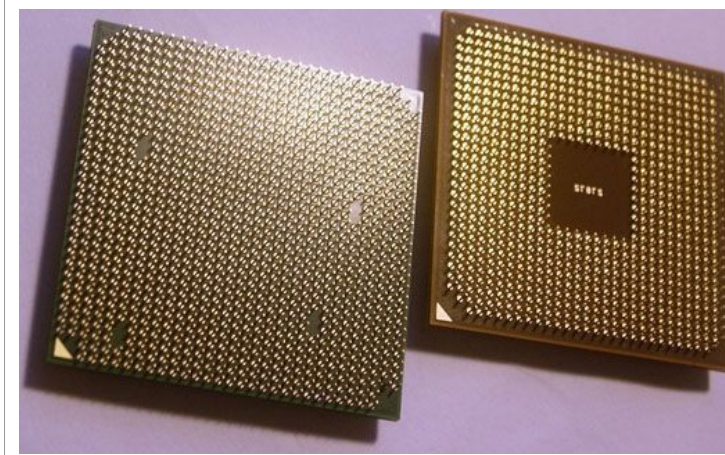
Tudo começou com o **Opteron**, o primeiro processador baseado na arquitetura Hammer, destinado a servidores. Ele utilizava um controlador de memória dual-channel e oferecia suporte apenas a memórias registered, módulos especiais que incluem um pequeno buffer que estabiliza o sinal, permitindo que sejam utilizados mais módulos. Os módulos registered são mais caros e mais lentos que os módulos DDR tradicionais que utilizamos em micros domésticos, mas eles são uma necessidade nos servidores, já que eles precisam utilizar um número muito maior de módulos. Devido ao brutal número de contatos utilizados pelo controlador de memória, o soquete utilizado pelo Opteron tinha nada menos do que 940 pinos.

A partir do momento que passou a ser capaz de produzir um volume maior de processadores, a AMD lançou os modelos destinados ao mercado doméstico, que incluem as versões iniciais do Athlon 64 e Athlon 64 FX.

A **Athlon 64 FX** era uma variação do Opteron, que vinha com 1 MB de cache L2 e mantinha o controlador de memória dual-channel. A grande mudança em relação ao Opteron é que ele passou a oferecer suporte a memórias

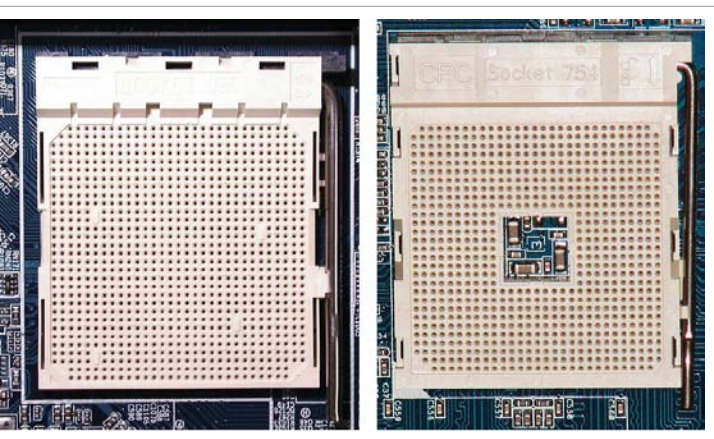
DDR comuns, o que barateou muito o custo dos micros. Para diferenciar as placas para o Opteron e para o Athlon 64 FX, a AMD criou o soquete 939, que era virtualmente idêntico ao soquete 940 do Opteron. A posição de alguns dos pinos foi alterada apenas para indicar a mudança no tipo de memória suportado e impedir que o Opteron fosse usado nas placas para o Athlon 64 FX e vice-versa. Mais tarde, o soquete 939 foi usado também pelas versões iniciais do Athlon X2.

Em seguida temos as versões "domésticas" do Athlon 64, que vinham com apenas 512 KB de cache L2 e utilizavam um controlador de memória simplificado, sem suporte a dual-channel. Devido ao menor número de contatos utilizados pelo controlador de memória, ele passaram a utilizar o soquete 754.



Athlon 64 FX (soquete 939) e Athlon 64 soquete 754 (à direita)





*Soquete 939 (à esquerda) e soquete 754*

Considerando dois processadores do mesmo clock, a diferença de desempenho entre um Athlon 64 "normal" e um Athlon 64 FX não é tão grande quanto se poderia imaginar. O controlador de memória dual-channel e o maior cache ajudam em alguns aplicativos, mas a diferença não era tão grande a ponto de justificar pagar US\$ 1000 pelo processador, como chegaram a custar algumas versões do FX. Vale lembrar que as primeiras versões do Athlon 64 FX competiam com o Pentium 4 Extreme Edition; ambos eram processadores "de vitrine", destinados ao público entusiasta e vendidos a preços exorbitantes. Mais tarde, foram lançadas versões do Athlon 64 soquete 754 com 1 MB de cache L2, que praticamente anularam as vantagens do FX.

Finalmente, temos os processadores recentes, que utilizam o soquete AM2.

Eles trazem um controlador de memória remodelado, com suporte a memórias DDR2. Apesar de também possuir 940 contatos, o soquete AM2 é incompatível com o soquete 940 utilizado pelo Opteron, além de ser incompatível com todos os processadores anteriores.

Os módulos de memória DDR2 utilizam 240 contatos, contra apenas 184 contatos dos módulos DDR. Com o controlador de memória integrado, os contatos vão diretamente para o processador, através de pinos no soquete. A AMD conseguiu fazer a migração das memórias DDR para as DDR2 sem aumentar o número de pinos do soquete remanejando uma série de contatos sem uso no processador.

Apesar do ganho de desempenho não ser tão grande quanto poderia parecer à primeira vista, as memórias DDR2 oferecem a vantagem de trabalharem utilizando uma tensão mais baixa, apenas 1.8v, contra 2.5v das memórias DDR. Isto faz com que tanto os módulos de memória, quanto o próprio controlador de memória integrado ao processador consumam menos energia (e dissipem menos calor), o que ajuda na questão da refrigeração. Atualmente as memórias DDR2 são produzidas em maior quantidade e são consequentemente mais baratas que as DDR, de forma que a migração para o AM2 foi um passo natural. As placas soquete 754 ficaram inicialmente relegadas à posição de plataforma de baixo custo e eventualmente foram completamente substituídas.

A partir da segunda metade de 2006, até mesmo os Semprons passaram a ser fabricados em versão M2.

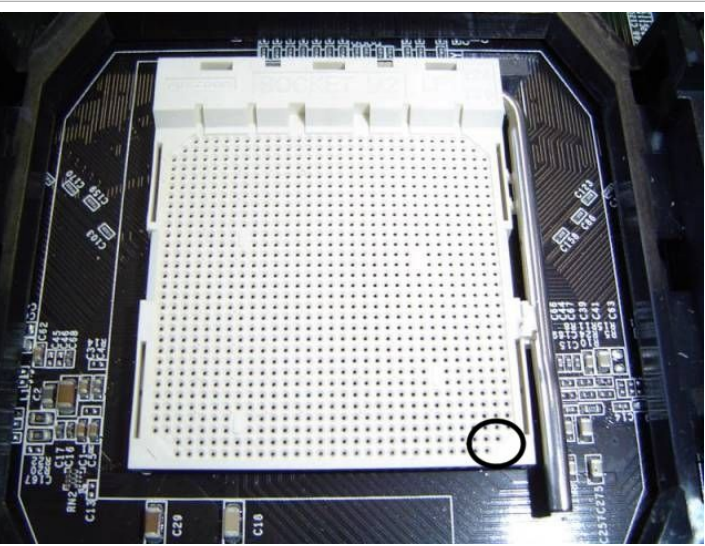
O AM2 trouxe também um novo suporte de retenção do cooler (a "gaiola" em volta do processador). Ele é preso à placa mãe por quatro parafusos (ao invés de apenas dois), o que tornou o mecanismo de retenção mais firme. O problema é que ele passou a usar um formato octagonal (ao invés do formato quadrado usado no soquete 754 e 939), o que quebrou a compatibilidade com a maior parte dos coolers antigos.



*Cooler NPH para processadores AM2*

Visualmente, o soquete AM2 é praticamente idêntico ao soquete 939, mudando apenas a posição de alguns dos pinos. O contato adicional pode ser visto no canto inferior direito do soquete:





Soquete AM2, com o pino adicional em destaque

A principal desvantagem de utilizar o controlador de memória integrado é que passou a ser necessário lançar uma versão atualizada do processador e criar um novo soquete cada vez que é necessário fazer alterações no controlador de memória, ou oferecer suporte a uma nova tecnologia, como no caso das memórias DDR2.

A partir do final de 2007, serão lançadas placas e processadores baseados no soquete AM2+, uma pequena atualização do AM2, que oferecerá um barramento HyperTransport atualizado, operando a 2.0 GHz, o dobro da frequência do HyperTransport utilizado no AM2 e oferecerão suporte a um sistema avançado de gerenciamento de energia, que permitirá controlar as frequências de operação do processador e do controlador de memória de forma independente.

Estas são inovações técnicas, que não possuem um impacto direto sobre o desempenho. Os processadores AM2+ poderão ser usados nas placas AM2 atuais e vice-versa, mas neste caso os novos recursos permanecem desativados.

A partir de 2008 a AMD passará a utilizar o soquete AM3 que, como o nome sugere, trará compatibilidade com as memórias DDR3. O soquete AM3 manterá a mesma pinagem do AM2 e será eletricamente com ele. Processadores AM3 terão um controlador de memória compatível tanto com memórias DDR3, quanto com memórias DDR2. Os processadores AM3 poderão ser usados em placas AM2 compatíveis (neste caso limitados ao uso de memórias DDR2, por limitações da placa), mas não será possível utilizar os processadores AM2 atuais em conjunto com as placas AM3, já que eles não oferecem suporte às memórias DDR3.

---

*Ou seja, com o lançamento do AM2+ e do AM3, teremos (desconsiderando o soquete F, utilizado pelo Opteron e Quad FX) um total de 6 padrões diferentes para os processadores AMD de 64 bits, exigindo atenção na hora de comprar ou fazer upgrade.*

---

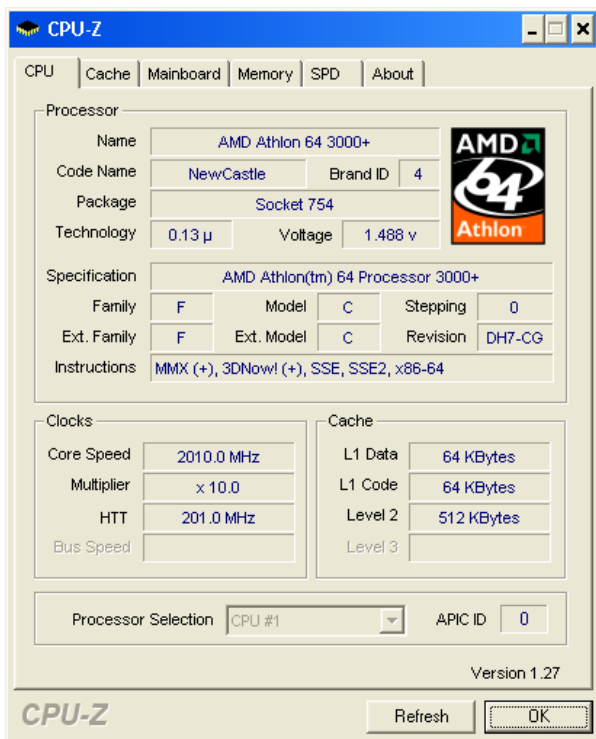
Por outro lado, a vida dos fabricantes de chipsets passou a ser muito mais simples, já que não precisam mais se preocupar com o controlador de memória. O uso dos diferentes soquetes não demanda mudanças no chipset, já que mudam apenas o processador e o encaixe (e trilhas) na placa mãe.

## Os modelos

Assim como na questão dos soquetes, a lista de modelos do Athlon 64 apresenta uma variedade espantosa de arquiteturas, variações na quantidade de cache, soquete usado, frequência do HyperTransport e assim por diante. O índice de desempenho dos processadores é calculado com base no desempenho médio, de forma que existem casos de processadores com clocks, caches e arquiteturas diferentes vendidos sob o mesmo índice e muitas vezes pelo mesmo preço.

Todo processador inclui uma pequena área de memória com detalhes sobre a arquitetura usada, quantidade de cache, frequência, tensão, conjuntos de instruções suportados e assim por diante. Estes códigos de identificação são acessíveis através de uma série de programas de diagnóstico, que podem ser usados para identificar o processador. Um dos mais práticos (para Windows) é o CPU-Z, disponível no:

<http://www.cpuid.com/cpuz.php>.



No Linux você pode usar o comando "*cat /proc/cpuinfo*", deixando que o próprio Kernel do sistema leia os códigos de identificação e mostre as informações na tela

## Athlon 64 e Athlon 64 FX

Tudo começou com os cores ClawHammer e SledgeHammer, que são as versões originais do K8, produzidas em uma técnica de 0.13 micron. Ambos tinham originalmente 1 MB de cache L2, a principal diferença entre eles é que o SledgeHammer vinha com ambos os controladores de memória ativados, enquanto

o ClawHammer vinha com apenas um, suportando apenas o modo single-channel.

Muitos modelos do ClawHammer vinham com metade do cache L2 desativado (o que permitia aproveitar processadores com defeitos no cache) e existiram também versões com suporte a dual-channel (virtualmente idênticas ao core SledgeHammer), vendidas em versão soquete 939. Na época a AMD tinha problemas para produzir os processadores em quantidade, de forma que foi lançando novas versões do jeito que conseguia, sem muita organização ou lógica.

O core **ClawHammer** foi utilizado nas primeiras versões do Athlon 64, sendo vendido nas versões **2800+** (1.8 GHz, 512 KB, soquete 754), **3000+** (2.0 GHz, 512 KB, soquete 754), **3200+** (2.0 GHz, 1 MB, soquete 754), **3400+** (2.2 GHz, 1 MB, soquete 754), **3500+** (2.2 GHz, 512 KB, soquete 939), **3700+** (2.4 GHz, 1 MB, soquete 754) e **4000+** (2.4 GHz, 1 MB, soquete 939).

Ele foi utilizado ainda no Athlon 64 **FX-53** (2.4 GHz, 1 MB, soquete 939) e no **FX-55** (2.6 GHz, 1 MB, soquete 939). Todas as versões do Athlon FX vinham com o multiplicador destravado, de forma a facilitar os overlocks. Isto era utilizado pela AMD como um diferencial para justificar a diferença de preço entre os FX e os modelos "normais" do Athlon 64. Note que o FX não é vendido sob o índice de desempenho, os números

seguem apenas uma sequência crescente que indica a "posição hierárquica" do processador, mas sem uma relação direta com seu desempenho.

Além de ser utilizado no Opteron, a versão do Athlon 64 destinada a servidores, o core **SledgeHammer** foi utilizado nos Athlon 64 **FX-51** (2.2 GHz, 1 MB) e **FX-53** (2.4 GHz, 1 MB), ambos vendidos apenas em versão soquete 940, o que adicionava o custo de utilizar memórias registered. Estas duas versões (lançadas em 2003, quase um ano antes do FX-53 e FX-55 baseados no ClawHammer) foram processadores "pra inglês ver", vendidos em quantidade muito limitada. O principal motivo da existência deles foi manter a guerra de benchmarks com a Intel.

O seguinte foi o core Newcastle, que ainda era fabricado usando a técnica de 0.13 micron, mas utilizava apenas 512 KB de cache L2, o que o tornava mais barato de produzir em relação ao **ClawHammer**. Ele foi vendido nas versões **2800+** (1.8 GHz, 512 KB, soquete 754), **3000+** (2.0 GHz, 512 KB, soquete 754), **3000+** (1.8 GHz, 512 KB, soquete 939), **3200+** (2.2 GHz, 512 KB, soquete 754), **3200+** (2.0 GHz, 512 KB, soquete 939), **3400+** (2.4 GHz, 512 KB, soquete 754), **3500+** (2.2 GHz, 512 KB, soquete 939) e **3800+** (2.4 GHz, 512 KB, soquete 939).

Note que, assim como no caso do ClawHammer, as versões soquete 939 são vendidas sob um índice de desempenho

mais alto, muito embora o uso do dual-channel tenha pouca influência sobre o desempenho das versões de baixo clock do Athlon 64. Isso faz com que um 3000+ soquete 754 acabe sendo mais rápido que um soquete 939 (que opera a apenas 1.8 GHz) na maioria das aplicações. Se você fosse montar um PC de baixo custo, na época, sem dúvida valeria mais à pena comprar um 3000+ soquete 754 e usar um único pente de memória.

O **Winchester** foi a primeira versão a ser fabricada em uma técnica de 0.09 micron. Ele é virtualmente idêntico ao Newcastle, incluindo os 512 KB de cache, mas oferece uma dissipação térmica muito mais baixa. Enquanto o TDP dos modelos baseados no Newcastle é de 89 watts, nos baseados no Winchester é apenas 67 watts. O Winchester foi usado apenas nos modelos **3000+** (1.8 GHz, 512 KB, soquete 939), **3200+** (2.0 GHz, 512 KB, soquete 939) e **3500+** (2.2 GHz, 512 KB, soquete 939).

O **Venice** mantém a técnica de 0.09 micron e os 512 KB de cache, mas adiciona suporte ao conjunto de instruções SSE3, o que resultou em um pequeno ganho de desempenho, sobretudo em games já otimizados. As primeiras versões chegaram ao mercado em abril de 2005, substituindo rapidamente os processadores baseados nos cores anteriores.

Ele foi vendido nas versões **3000+** (2.0 GHz, 512 KB, soquete 754), **3000+** (1.8 GHz, 512 KB, soquete 939), **3200+** (2.2 GHz, 512 KB, soquete 754), **3200+** (2.0 GHz, 512 KB, soquete 939), **3400+** (2.4 GHz, 512 KB, soquete 754), **3400+** (2.2 GHz, 512 KB, soquete 939), **3500+** (2.2 GHz, 512 KB, soquete 939) e **3800+** (2.4 GHz, 512 KB, soquete 939).

Existiu também o **Manchester**, uma revisão do Venice com um consumo elétrico pouca coisa inferior. Ele existiu em versões **3200+** (2.0 GHz, 512 KB, soquete 939) e **3500+** (2.2 GHz, 512 KB, soquete 939).

O próximo é o core **San Diego**, uma versão do core Manchester com 1 MB de cache L2. Ele foi utilizado tanto em modelos do Athlon 64 (um deles com metade do cache desabilitado, assim como na época do ClawHammer) e também em dois modelos do Athlon FX.

Dentro da linha Athlon 64, o San Diego foi utilizado nos modelos **3500+** (2.2 GHz, 512 KB, soquete 939), **3700+** (2.2 GHz, 1 MB, soquete 939) e **4000+** (2.4 GHz, 1 MB, soquete 939)

Dentro da linha FX, ele foi utilizado nos modelos **FX-55** (2.6 GHz, 1 MB, soquete 939) e **FX-57** (2.8 GHz, 1 MB, soquete 939). Estes dois modelos eram processadores escolhidos a dedo dentro da produção do San Diego, incluindo apenas os processadores mais perfeitos, que eram capazes de operar estavelmente

a frequências mais altas. Além do clock (e preço) mais alto, a única vantagem deles sobre os da linha Athlon 64 era o fato de virem com o multiplicador destravado.

O **Orleans** é mais um core fabricado utilizando a técnica de 0.09 micron. Apesar disso, ele traz três vantagens sobre os anteriores. A primeira é o suporte ao soquete **AM2** (e o consequente suporte a memórias DDR2), utilizado por todos os modelos. A segunda é o suporte ao **AMD-V** (Pacifica), um sistema de virtualização muito similar ao Intel-VT. Assim como o Intel-VT, o AMD-V pode ser utilizado para melhorar o desempenho de virtualizadores como o VMware e o Xen. A terceira é uma série de otimizações e o uso de uma tensão mais baixa, que reduzem o TDP de 89 watts (no Venice 3800+) para apenas 62 watts em todos os modelos.

Ele foi utilizado nos modelos **3000+** (1.8 GHz, 512 KB, soquete AM2), **3200+** (2.0 GHz, 512 KB, soquete AM2), **3500+** (2.2 GHz, 512 KB, soquete AM2), **3800+** (2.4 GHz, 512 KB, soquete AM2) e **4000+** (2.6 GHz, 512 KB, soquete AM2).

Concluindo, temos o core Lima, o primeiro fabricado usando a nova técnica de 0.065 micron. Um dos destaques é o baixo consumo: as duas versões possuem um TDP de apenas 45 watts.

Enquanto escrevo, ele é usado em apenas duas versões do Athlon 64: **3500+** (2.2 GHz, 512 KB, soquete AM2) e **3800+** (2.4 GHz, 512 KB, soquete AM2).



## Athlon 64 X2

O Athlon 64 X2 é a versão dual-core do Athlon 64. Quando ele foi lançado, em 2005, a Intel ainda see debatia com a plataforma NetBurst, produzindo primariamente o Pentium D, que era um processador dual-chip (onde temos dois processadores separados, dentro do mesmo encapsulamento) e não um processador dual-core "de verdade". No Pentium D, os dois cores comunicam-se entre si e com o chipset utilizando um barramento externo, relativamente lento.

No caso do Athlon X2, temos os dois núcleos construídos sobre o mesmo waffer. A comunicação entre os dois núcleos é feita internamente, utilizando um barramento interno, batizado de SRI (System Request Interface), o que garante tempos de latência mais baixos e pequenos ganhos de desempenho ao rodar muitos aplicativos simultaneamente (que é justamente a área em que os processadores dual-core oferecem mais vantagens), além de permitir que o processador seja capaz de operar a frequências um pouco mais altas.

Na camada mais externa, temos o "Crossbar Switch", responsável por dividir as requisições entre os dois cores, controlar o acesso de ambos ao controlador de memória (e ao barramento HyperTransport), além de monitorar o nível de carregamento de cada um, informação utilizada, entre outras coisas, pelo sistema de gerenciamento de energia.

Este diagrama da AMD mostra o conceito:

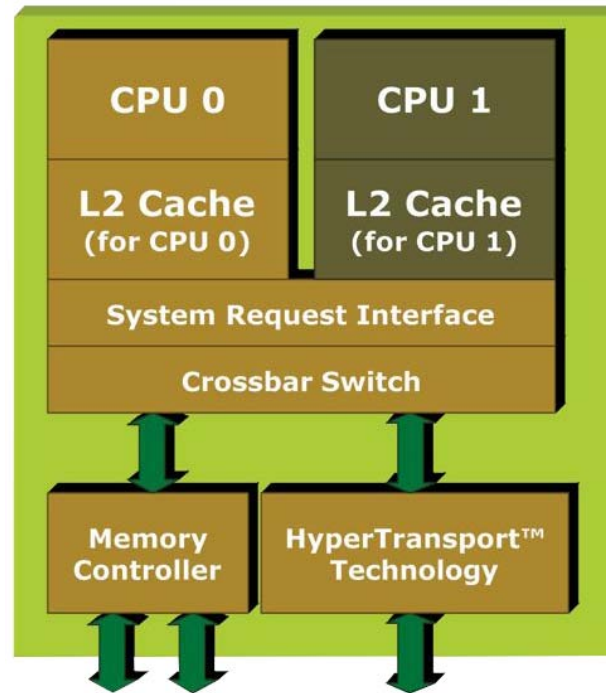
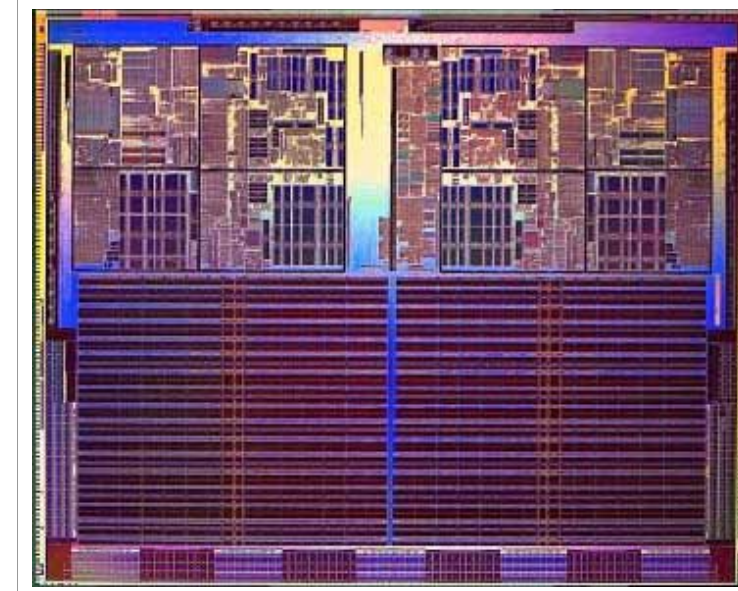


Diagrama de blocos do X2

Cada processador possui seu próprio cache L2, mas ambos compartilham o mesmo controlador de memória e o mesmo barramento HyperTransport, através do qual é feita toda a comunicação com o chipset e os demais componentes do PC. Olhando uma foto ampliada do processador, você pode notar que os dois núcleos ficam no centro, envolvidos pelos circuitos referentes ao controlador de memória e HyperTransport. Além dos circuitos compartilhados, os dois núcleos também não são exatamente cópias idênticas um do outro:



Athlon 64 X2

O uso de dois núcleos torna o processador mais dependente do barramento com a memória. Devido a isso, o Athlon 64 X2 foi lançado apenas em versões soquete 939 e AM2 (futuramente serão lançadas também versões AM2+ e AM3), já que uma versão soquete 754 seria penalizada pelo barramento single-channel. O lançamento do Athlon X2 acelerou a migração para as placas soquete 939, transformando a plataforma 754 em uma opção de baixo custo, destinada ao Sempron e às versões mais lentas do Athlon 64.

Ao contrário do que teríamos ao utilizar dois processadores separados em SMP, um processador dual-core atual, como o Athlon X2, Pentium D ou Core 2 Duo não consomem o dobro de energia que

as versões single-core. Isto é possível graças a sistemas de gerenciamento de energia incluídos no processador, que reduzem a frequência ou mesmo desativam completamente o segundo núcleo quando o processador está ocioso. No caso do Athlon X2, muitos componentes são compartilhados entre os dois processadores, o que aumenta a economia.

Comparado com um Athlon 64 4000+ com core San Diego (que também opera a 2.4 GHz), um Athlon X2 4800+ Toledo (2.4 GHz, 1 MB) consome cerca de 12 watts a mais enquanto o sistema está ocioso e 24 watts a mais ao rodar um benchmark. Considerando que o TDP do San Diego é de 89 watts, o aumento não é muito significativo.

Um aplicativo que conseguisse utilizar simultaneamente todo o processamento de ambos os cores poderia, em teoria, fazer com que o consumo chegasse a ser, momentaneamente, próximo do dobro, mas em situações reais isto não acontece com frequência.

A carreira do X2 começa com o core **Manchester**, lançado em Maio de 2005. Ele era fabricado usando uma técnica de 0.09 micron, com 512 KB de cache L2 por core (1 MB no total) e suporte às instruções SS3. A versão mais lenta tinha metade do cache L2 desabilitado, de forma a aproveitar os cores com defeitos no cache.

O Manchester foi usado nos Athlon 64 **X2 3600+** (2.0 GHz, 2x 256 KB, soquete 939),

**3800+** (2.0 GHz, 2x 512 KB, soquete 939), **4200+** (2.2 GHz, 2x 512 KB, soquete 939) e **4600+** (2.4 GHz, 2x 512 KB, soquete 939)

O seguinte foi o core **Toledo**, ainda fabricado em uma técnica de 0.09 micron, mas agora com 1 MB de cache por core, totalizando 2 MB. Ele foi lançado simultaneamente com o Manchester e os modelos do X2 baseados em ambos conviveram durante mais de um ano. Apenas três dos modelos produzidos utilizando o core Toledo vieram com todo o cache ativo. O demais vinham com metade do cache desativado, o que os tornava praticamente indistinguíveis dos baseados no Manchester.

O Toledo foi utilizado nos modelos **3800+** (2.0 GHz, 2x 512 KB, soquete 939), **4200+** (2.2 GHz, 2x 512 KB, soquete 939), **4400+** (2.2 GHz, 2x 1 MB, soquete 939), **4600+** (2.4 GHz, 2x 512 KB, soquete 939) e **4800+** (2.4 GHz, 2x 1 MB, soquete 939).

Quase um ano depois, em maio de 2006, foi lançado o **Windsor**, que passou a utilizar o soquete **AM2** e adicionou suporte ao **AMD-V**, mantendo a mesma técnica de produção de 0.09 micron e o uso de 2x 1 MB de cache. Assim como no Toledo, uma grande parte dos modelos vinham com metade do cache L2 desabilitado e eram vendidos sob índices de desempenho mais baixos que os "completos".

O Windsor foi utilizado nos modelos **3800+** (2.0 GHz, 2x 512 KB, soquete AM2), **4000+** (2.0 GHz, 2x 1 MB, soquete AM2), **4200+** (2.2 GHz, 2x 512 KB, soquete AM2), **4400+** (2.2 GHz, 2x 1 MB, soquete AM2), **4600+** (2.4 GHz, 2x 512 KB, soquete AM2), **5000+** (2.6 GHz, 2x 512 KB, soquete AM2), **5200+** (2.6 GHz, 2x 1 MB, soquete AM2), **5400+** (2.8 GHz, 2x 512 KB, soquete AM2), **5600+** (2.8 GHz, 2x 1 MB, soquete AM2) e **6000+** (3.0 GHz, 2x 1 MB, soquete AM2).

Todos os modelos baseados no core Windsor possuem um TDP de 89 watts, assim como a maior parte das versões baseadas nos cores Manchester e Toledo. A exceção fica por conta do 6000+, que apenas da diferença de apenas 200 MHz em relação ao 5600+, possui um TDP de 125 watts. Esta diferença tão grande é causada pelo gate leakage, o mesmo problema que a Intel enfrentou com o Pentium 4 ao cruzar a barreira dos 3.0 GHz e que às vezes enfrentamos ao fazer overlocks agressivos.

A partir de uma certa frequência (que varia de acordo com a arquitetura), os transistores passam a desperdiçar cada vez mais energia, de forma que é necessário utilizar uma tensão cada vez mais alta para estabilizar o processador (aumento que aumenta ainda mais o desperdício de energia), criando um ciclo vicioso que se intensifica a cada novo aumento de frequência. O 6000+ é basicamente uma versão overclocada do 5600+, onde a tensão foi aumentada

de 1.30 para 1.35v para estabilizar o processador. Ele é um modelo que deve ser evitado na hora da compra.

Como uma opção para quem prefere um processador mais econômico e frio, a AMD lançou a série "**Energy Efficient**", onde são usadas técnicas derivadas dos processadores mobile, como o uso de tensões mais baixas e de transistores de chaveamento mais lentos, que resultam em processadores com um consumo mais baixo, mas que em compensação não são capazes de atingir frequências de operação tão altas quanto os das séries principais.

A série Energy Efficient do Windsor, por exemplo, abrange modelos do 3600+ (2.0 GHz, 2x 1MB, soquete AM2) ao 5200+ (2.6 GHz, 2x 1MB, soquete AM2), todos com tensão de 1.2v e TDP de 65 watts. Eles são mais caros que os da série regular, com o preço variando de acordo com o clock (nas frequências mais altas, a diferença pode ser de mais de 200 reais, enquanto nos modelos populares pode cair para menos de 30).

De uma forma geral, os modelos Energy Efficiency valem à pena em casos onde a diferença de preço é pequena, pois a economia de energia acaba se pagando com o passar do tempo. Em um PC que fica ligado continuamente, uma economia de 24 watts acaba resultando em uma redução de mais de 200 reais na conta de luz ao longo de dois anos.

Por último, temos o **Brisbane**, a primeira série do X2 a ser produzida usando uma técnica de 0.065 micron. Uma informação importante é que, apesar da redução do espaço ocupado por cada processador, devido à nova técnica de fabricação, todos os modelos baseados no Brisbane possuem apenas 512 KB de cache por core. Esta redução foi feita para aumentar o índice de aproveitamento, já que, como em toda nova técnica de produção, o índice de defeitos é inicialmente muito grande, de forma que produzir processadores menores permite aumentar substancialmente o número de processadores "bons" por waffer.

Uma má notícia é que a AMD aumentou os tempos de latência do cache L2 do Brisbane de 12 para 14 tempos, visando facilitar a produção de futuros modelos com mais cache L2, além de permitir o uso de frequências de clock ligeiramente maiores. Como o Brisbane não inclui melhorias em relação aos modelos anteriores, ele acaba sendo mais **lento** que um Manchester do mesmo clock (e com a mesma quantidade de cache). Embora a diferença seja pequena, inferior a 2% na maioria dos aplicativos, ela existe.

O Brisbane foi usado em 6 modelos da série Energy Efficient, todos eles com TDP de 65 watts: **3600+** (1.9 GHz, 2x 512 KB, soquete AM2), **4000+** (2.1 GHz, 2x 512 KB, soquete AM2), **4400+** (2.3 GHz, 2x 512 KB, soquete AM2), **4800+** (2.5 GHz, 2x 512 KB, soquete AM2), **5000+** (2.6 GHz,

2x 512 KB, soquete AM2) e **5200+** (2.7 GHz, 2x 512, soquete AM2).

Como de praxe, o uso de um processador dual-core tem seus prós e contras. O principal benefício de usar um processador dual-core é o melhor desempenho ao rodar muitos aplicativos pesados simultaneamente. Se você é do tipo que abre 50 abas do navegador, ouve música, comprime um DVD, retoca imagens no Photoshop (ou Gimp ;) para o cartaz que está diagramando no Corel e ainda por cima quer abrir 3 máquinas virtuais do VMware, tudo ao mesmo tempo, um processador dual-core, acompanhado por 2 ou 4 GB de memória DDR2 (ou DDR3) é uma necessidade.

Por outro lado, para usuários que rodam um ou dois aplicativos por vez, que usam o PC predominantemente para games (sem executar outras tarefas simultaneamente, como por exemplo deixar o PC comprimindo um DVD em segundo plano) ou que rodam apenas aplicativos leves, um processador single-core mais barato, ou com uma frequência ligeiramente maior ofereceria uma relação custo-benefício melhor.

Comparando um X2 4800+ (2.4 GHz, 2x 1 MB, core Toledo) com um Athlon 64 single-core 4000+ (2.4 GHz, 1 MB, core San Diego) temos o X2 4800+ ganhando por uma margem de 17% no Winstone 2004 (Content Creation), 41% no SYS-Mark 2004 (3D Content Creation), 1% no Photoshop 7 e 2% no Premiere 6.5 (testes



do World Bench 2005), 1% no Doom3 (FPS a 1024x768), 2% no Half Life 2 e 3% no Unreal 2004.

Você pode notar que, com exceção do SYSMark 2004, todos os demais benchmarks e games mostram um ganho de desempenho muito inferior ao sugerido pelo aumento de 800 pontos no índice de desempenho da AMD. Isso acontece por que estes testes levam em conta o desempenho ao executar apenas uma tarefa de cada vez. Como disse, um processador dual-core traz benefícios consideráveis apenas ao rodar vários aplicativos simultaneamente.

Rodando um aplicativo por vez, ou rodando apenas benchmarks, existem até casos onde o 4800+ perde do 4000+ por uma pequena margem (1 ou 2%). É o caso de aplicativos que não conseguem obter nenhum benefício do segundo core e acabam tendo o desempenho penalizado pelo overhead de dividir a carga entre os dois.

## Sempron

Apesar de serem processadores completamente diferentes que os Semprons soquete A baseados no Barton, Thorton e Thoroughbred-B, a AMD continuou utilizando a marca "Sempron" ao lançar a linha de processadores de baixo custo, baseada na arquitetura K8.

Além da questão do cache menor, as versões iniciais do Sempron vinham sem suporte a AMD64, ou seja, sem suporte às instruções de 64 bits, incluindo os registradores extra e as outras melhorias trazidas pela arquitetura. Embora fossem baseados na arquitetura K8, eles eram processadores de 64 bits, sem praticamente nenhuma das inovações trazidas por ela.

Outra diferença é que o índice de desempenho do Sempron é calculado com relação ao desempenho do Celeron D e não ao Pentium 4. Isso faz com que exista uma diferença significativa entre o desempenho de um Sempron "3000+" e de um Athlon 64 também "3000+".

Por exemplo, o Athlon 64 3000+ baseado no core Newcastle (2.0 GHz, soquete 754, com 512 KB de cache) tem um desempenho parecido com o de um Sempron de 2.2 GHz com 256 KB de cache, baseado no core Palermo. O problema é que o Palermo de 2.2 GHz recebe o índice "3400+", 400 pontos acima do índice do Athlon 64 equivalente.

A primeira encarnação do Sempron K8 foi o core **Paris**, uma versão simplificada do ClawHammer, produzido em uma técnica de 0.13 micron, que possuía nativamente apenas 256 KB de cache e vinha sem suporte às instruções de 64 bits, ao Cool'n'Quiet e também às instruções SSE3.

Ele foi lançado em julho de 2004 e foi vendido por pouco mais de um ano.

Existiram duas versões: **3000+** (1.8 GHz, 128 KB, soquete 754) e **3100+** (1.8 GHz, 256 KB, soquete 754).

Como pode ver, o 3000+ vinha com metade do cache desabilitado. Como o Sempron já possui pouco cache nativamente, esta nova redução acabava tendo um efeito muito maior sobre o desempenho do que a diferença de 100 pontos no índice sugere. O Sempron com core Paris é um modelo a se evitar, mas entre os dois o 3000+ é o pior.

A partir de 2005, o core Paris foi rapidamente substituído pelo **Palermo**, produzido usando uma técnica de 0.09 micron. Nativamente, o core Palermo possui 256 KB de cache e oferece suporte às extensões de 64 bits, Cool'n'Quiet e também às instruções SSE3. O problema é que, a fim de aproveitar o maior número possível de processadores saídos da linha de produção, a AMD passou a desativar não apenas metade do cache, como de praxe, mas também o suporte às instruções de 64 bits e so SSE3 nos cores que eram produzidos com defeitos nos componentes relacionados a estes recursos, o que resultou em uma grande salada de modelos com índice de desempenho similar, mas com conjuntos muito diferentes de recursos.

A chave para diferenciar os Palermos é o código de identificação dos processadores. Os processadores cujo código termina com "BA" são os piores, pois não suportam nem instruções de 64 bits, nem SSE3.

Os que terminam com "BO" suportam SSE3, mas ainda não suportam instruções de 64 bits, enquanto os terminados com "BX" são as versões completas, com ambos os recursos ativados.

Com exceção do 2500+, todos os modelos foram fabricados nas três versões e a presença do suporte a 64 bits ou SSE3 não influenciava no índice de desempenho ou preço, de forma que era importante checar antes de comprar. A isto soma-se a questão do Cool'n'Quiet, que não era suportado pelo core Paris e está disponível apenas nos Palermos 3000+ em diante.

O Palermo foi usado nas versões **2500+** (1.4 GHz, 256 KB, soquete 754), **2600+** (1.6 GHz, 128 KB, soquete 754), **2800+** (1.6 GHz, 256 KB, soquete 754), **3000+** (1.8 GHz, 128 KB, soquete 754), **3100+** (1.8 GHz, 256 KB, soquete 754), **3300+** (2.0 GHz, 128 KB, soquete 754) e **3400+** (2.0 GHz, 256 KB, soquete 754).

Pouco depois, no final de 2005, foi lançada uma versão atualizada do core Palermo, com suporte ao soquete 939 (e consequentemente a dual-channel). Surgiram então os modelos **3000+** (1.8 GHz, 128 KB, soquete 939), **3200+** (1.8 GHz, 256 KB, soquete 939), **3400+** (2.0 GHz, 128 KB, soquete 939) e **3500+** (2.0 GHz, 256 KB, soquete 939). Estes 4 modelos suportavam SSE3, mas apenas as séries com final "BW" ofereciam suporte às instruções de 64 bits. A séries com final "BP" vinham com o suporte a 64 bits desabilitado e devem ser evitadas.

Finalmente, temos o Sempron com core **Manila**, ainda fabricado usando a técnica de 0.09 micron, mas agora com suporte ao soquete AM2. Ao contrário dos Palermos, todos os Manilas incluem suporte às instruções de 64 bits e SSE3, mas o Cool'n'Quiet é suportado apenas versões 3200+ em diante.

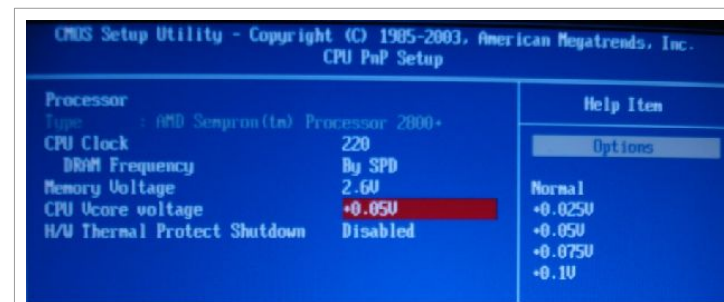
O core Manila foi usado nas versões **2800+** (1.6 GHz, 128 KB, soquete AM2), **3000+** (1.6 GHz, 256 KB, soquete AM2), **3200+** (1.8 GHz, 128 KB, soquete AM2), **3400+** (1.8 GHz, 256 KB, soquete AM2), **3500+** (2.0 GHz, 128 KB, soquete AM2), **3600+** (2.0 GHz, 256 KB, soquete AM2) e **3800+** (2.2 GHz, 256 KB, soquete AM2). Como de praxe, as versões de 256 KB oferecem um desempenho sensivelmente superior na maior parte dos aplicativos, por isso devem ser preferidos na hora da compra.

Esta confusão com relação aos modelos do Sempron e do Athlon 64 fez com que a AMD perdesse alguns potenciais compradores e permitiu que a Intel recuperasse parte do espaço anteriormente perdido, passando a oferecer versões do Celeron D a preços bastante competitivos.

No início de 2007, os Celeron D baseados no core Cedar Mill-512 (os modelos 347, 352, 356, 360 e 365) eram especialmente atraentes, pois possuíam 512 KB de cache, eram fabricados usando a técnica de 0.065 micron e suportavam overlocks generosos.

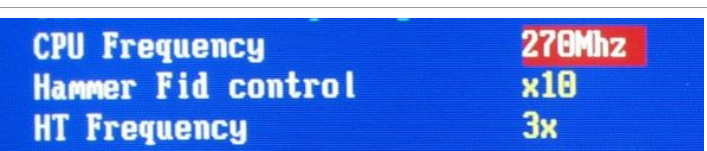
Overclock: Assim como os demais processadores dentro da família do Athlon 64, o Sempron não é particularmente generoso com relação aos overlocks. Os melhores dentro da família são os com core Manila, do 2800+ ao 3200+ (1.6 a 2.0 GHz). Eles utilizam uma tensão dual de 1.25v - 1.4v e, na maioria dos casos, podem ser overclocados para até 2.4 GHz com estabilidade, desde que a placa mãe suporte a frequência necessária e você aumente a tensão do processador em 0.1v.

A maioria das placas baratas, baseadas em chipsets SiS e VIA podem trabalhar com o FSB a até 220 MHz com estabilidade, o que permitiria fazer overclock das versões de 1.6, 1.8 e 2.0 GHz para, respectivamente, 1.76, 1.98 e 2.2 GHz. Neste caso o overclock é simples e na maioria dos casos você não vai precisar sequer aumentar a tensão do processador (embora um aumento de 0.05v possa ajudar a manter a estabilidade):



A partir daí, as possibilidades dependem das opções disponíveis no Setup da placa mãe. O primeiro problema é a frequência do barramento HyperTransport (HT Frequency), que é particularmente sensível. Nas placas soquete 754 o HyperTransport trabalha a 800 MHz (200 MHz x 4) e nas soquete 939 ou AM2 ele trabalha a 1.0 GHz (200 MHz x 5). Em ambos os casos, ele não suporta trabalhar com estabilidade a mais do que, respectivamente, 900 ou 1100 MHz, o que nos restringe ao humilde overclock de 10% do passo anterior.

Para ir além, é necessário que a placa mãe ofereça a opção de ajustar o multiplicador do barramento HyperTransport. Desta forma, você pode reduzir o multiplicador de 4x ou 5x para, respectivamente, 3x e 4x, o que permitirá aumentar o FSB para até 250 MHz com relativa segurança. Usando um multiplicador de 4x e 250 MHz em uma placa AM2, o HyperTransport trabalharia a 1.0 GHz, sua frequência padrão. Em uma placa soquete 754, você poderia aumentar o FSB para até 270 MHz e ainda assim (com multiplicador 3x) o HyperTransport estaria operando a 810 MHz.

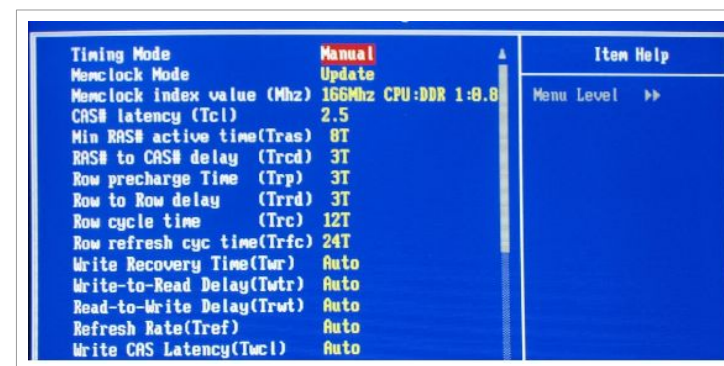


Ajuste do FSB e multiplicador do barramento HyperTransport

Reduzindo o multiplicador do HyperTransport e aumentando a tensão do processador em 0.1v, você poderia aumentar a frequência do FSB para 240 MHz e assim obter 1.92, 2.16 e 2.4 GHz. No caso das versões de 1.6 e 1.8 GHz, você poderia ir um pouco mais longe e ajustar o FSB para 250 MHz, obtendo 2.0 e 2.25 GHz, respectivamente.

A até 2.4 GHz, o processador deve funcionar sem problemas utilizando um cooler padrão, desde que o gabinete seja bem ventilado. A partir daí, os resultados variam muito de acordo com a série do processador, placa mãe, cooler usado e até mesmo com a qualidade da fonte de alimentação, já que quanto maior a frequência, maior o consumo elétrico e consequentemente maior o stress sobre ela.

Praticamente todas as placas soquete 754, 939 e AM2 oferecem a opção de manter a memória operando de forma assíncrona. Isso permite que você aumente a frequência do FSB (e consequentemente do processador) sem mexer na frequência da memória. Depois de testar o overclock do processador, você pode obter um pequeno ganho adicional aumentando a frequência, ou baixando os tempos de espera da memória. Neste caso as possibilidades dependem das opções oferecidas pela placa e da qualidade dos módulos de memória instalados.



Ajuste da frequência e tempos de espera da memória

Para overlocks mais extremos, outra opção útil é o "PCI/AGP Lock", opção suportada por placas com chipset nVidia e algumas com chipset VIA, onde é possível justar a frequência do FSB de forma independente da frequência dos demais barramentos da placa. assim como no caso da memória.

## Reconhecendo o processador

A gigantesca variedade de variações dentro da família Athlon 64 torna realmente complicado identificar os processadores visualmente. Um "3000+" pode ser desde um ClawHammer de 2.0 GHz, soquete 754, até um Orleans de 1.8 GHz soquete AM2, passando por praticamente todas as outras famílias. Não existe sequer uma distinção clara entre os processadores single-core e dual-core.



É relativamente simples distinguir os processadores baseado no soquete usado, já que os processadores soquete 754 possuem a área central sem pinos e os AM2 possuem um pino extra em relação aos soquete 939. Com base no soquete usado, data aproximada de fabricação, índice de desempenho e clock do processador, você pode ter então uma idéia de qual família ele pertence, mas para ter certeza mesmo você precisaria montar o micro e verificar usando o CPU-Z ou outro software.

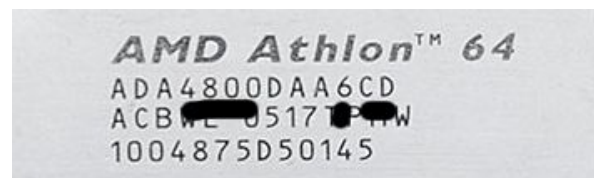
Mas, apesar de complicado, é possível diferenciar os Athlon 64 e X2 com base no código de identificação com 13 dígitos decalcado sobre o spreader do processador:



Código de identificação de um Athlon X2

Dentro do código, o campo mais óbvio são os 4 dígitos do número de identificação (4800 no exemplo). Como vimos, o mesmo código pode ser usado em processadores com diferentes combinações de clock, cache e soquete.

Os três dígitos finais ajudam a distinguir processadores com o mesmo índice, pois indicam a quantidade de cache e a arquitetura em que são baseados:



O antepenúltimo dígito (6 no exemplo) se refere ao cache. As possibilidades são as seguintes:

- 2 - 128 KB (Sempron)
- 3 - 256 KB (Sempron)
- 4 - 512 KB (1x 512 para os single-core ou 2x 256 KB para os dual-core)
- 5 - 1 MB (2x 512 KB no caso dos dual-core)
- 6 - 2 MB (2x 1 MB no caso dos dual-core)

Os dois últimos são um código de identificação, que indica o stepping, tecnologia de fabricação, soquete e também se o processador é single-core ou dual-core. A tabela para os processadores Athlon 64 e X2 em versão OEM é a seguinte:

**AP** - Stepping C0 (ClawHammer), single-core, 0.13 micron, soquete 754

**AR** - Stepping CG (ClawHammer), single-core, 0.13 micron, soquete 754

**AS** - Stepping CG (ClawHammer), single-core, 0.13 micron, soquete 939

**AW** - Stepping CG (Newcastle), single-core, 0.13 micron, soquete 939

**AX** - Stepping CG (Newcastle), single-core, 0.13 micron, soquete 754

**AZ** - Stepping CG (Newcastle), single-core, 0.13 micron, soquete 939

**BI** - Stepping D0 (Winchester), single-core, 0.09 micron, soquete 939

**BN** - Stepping E4 (San Diego), single-core, 0.09 micron, soquete 939

**BO** - Stepping E3 (Venice), single-core, 0.09 micron, soquete 754

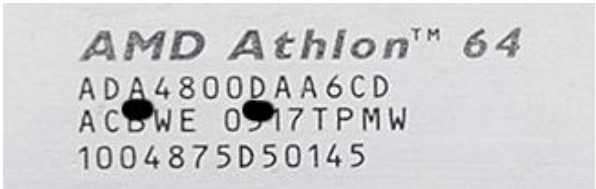
**BP** - Stepping E3 (Venice), single-core, 0.09 micron, soquete 939

**BV** - Stepping E4 (Manchester), dual-core, 0.09 micron, soquete 939

**BW** - Stepping E6 (Venice ou San Diego), single-core, 0.09 micron, soquete 939

<b>BX</b> - Stepping E6 (Venice), single-core, 0.09 micron, soquete 754	<b>DE</b> - Stepping G1 (Lima), single-core, 0.065 micron, soquete AM2	<b>CV</b> - Stepping D0 (Palermo), 0.09 micron, soquete 754
<b>BY</b> - Stepping E6 (Venice ou San Diego), single-core, 0.09 micron, soquete 939	<b>DH</b> - Stepping F3 (Lima), single-core, 0.065 micron, soquete AM2	<b>CW</b> - Stepping F2 (Manila), 0.09 micron, soquete AM2
<b>BZ</b> - Stepping E3 (Venice), single-core, 0.09 micron, soquete 939	<b>DL</b> - Stepping G1 (Brisbane), dual-core, 0.065 micron, soquete AM2	O terceiro dígito (A no exemplo) se refere ao TDP do processador. Ele não é uma indicação precisa, mas ajuda a identificar os processadores de baixo consumo e os modelos gastadores (como o 6000+ baseado no core Windsor). <i>As possibilidades são as seguintes:</i>
<b>CD</b> - Stepping E6 (Toledo), dual-core, 0.09 micron, soquete 939	<b>Para os Semprons, baseados na arquitetura K8, temos:</b>	
<b>CF</b> - Stepping E6 (Venice ou San Diego), single-core, 0.09 micron, soquete 939	<b>AX</b> - Stepping CG (Paris), 0.13 micron, soquete 754 (sem suporte a 64 bits e SSE3)	<b>A</b> - Standard (consumo normal dentro da arquitetura usada)
<b>CG</b> - Stepping E4 (Manchester ou San Diego), single-core, 0.09 micron, soquete 939	<b>BA</b> - Stepping D0 (Palermo), 0.09 micron, soquete 754 (sem suporte a 64 bits e a SSE3)	<b>D</b> - TDP de 35 watts
<b>CN</b> - Stepping F2 (Orleans), single-core, 0.09 micron, soquete AM2	<b>BI</b> - Stepping D0 (Palermo), 0.09 micron, soquete 939 (sem suporte a 64 bits)	<b>H</b> - TDP de 45 watts
<b>CS</b> - Stepping F2 (Windsor), dual-core, 0.09 micron, soquete AM2	<b>BO</b> - Stepping E3 (Palermo), 0.09 micron, soquete 754 (sem suporte a 64 bits)	<b>O</b> - TDP de 65 watts
<b>CU</b> - Stepping F2 (Windsor), dual-core, 0.09 micron, soquete AM2	<b>BP</b> - Stepping E3 (Palermo), 0.09 micron, soquete 939 (sem suporte a 64 bits)	<b>V</b> - Low power (os modelos Energy Efficient)
<b>CW</b> - Stepping F2 (Orleans), single-core, 0.09 micron, soquete AM2	<b>BW</b> - Stepping E6 (Palermo), 0.09 micron, soquete 939	<b>X</b> - TDP de 125 watts (como o Windsor 6000+)
<b>CZ</b> - Stepping F3 (Windsor), dual-core, 0.09 micron, soquete AM2	<b>BX</b> - Stepping E6 (Palermo), 0.09 micron, soquete 754	O oitavo dígito (logo após o índice de desempenho) também indica o soquete usado. Ele é útil em casos onde o processador já estiver instalado na placa-mãe ou não for possível examinar os contatos:
<b>DD</b> - Stepping G1 (Brisbane), dual-core, 0.065 micron, soquete AM2	<b>CN</b> - Stepping F2 (Manila), 0.09 micron, soquete AM2	<b>A</b> - Soquete 754
		<b>D</b> - Soquete 939

I - Soquete 940



Os dois primeiros dígitos (AD no exemplo) permitem confirmar a família da qual o processador faz parte:

- AD** - Athlon 64, Athlon 64 X2 ou Athlon 64 FX
- OS** - Opteron (servidores)
- AM** - Mobile Athlon 64 ou Mobile Athlon 64 X2
- SD** - Sempron
- SM** - Mobile Sempron
- TM** - Turion ou Turion X2

No caso dos processadores em versão boxed, o código de identificação termina com a sigla "BOX" e os dois dígitos depois do índice de desempenho indicam o código de identificação (dentro da tabela que vimos a pouco). No caso deles, o código não contém informações sobre o cache ou soquete utilizado.

Pelo código, podemos dizer que o Sempron da foto abaixo é um Sempron (SD),

da série de consumo standard (A), vendido sob o índice 3000+, soquete 939 (D), com 128 KB de cache (2) e que faz parte da série baseada no core Palermo, sem suporte a 64 bits (BP). O código não fala nada sobre a frequência real de operação, mas com base nas informações descobertas podemos deduzir que ele opera a 1.8 GHz, já que esta é a única frequência possível para um Sempron soquete 939 com 128 KB de cache. Como o código não termina com a sigla "BOX", sabemos também que ele é um processador OEM.



*Sempron com código SDA3000DIO2BP*

No caso do Athlon 64 a seguir, podemos dizer que trata-se de um Athlon 64 3800+, também soquete 939 (D), com 512 KB, e que é um processador single-core, baseado no core Newcastle (AW).

Com base nestas informações podemos deduzir que ele é o modelo de 2.4 GHz, já que não existem outras versões do Newcastle com 512 KB para o soquete 939 que receberam o índice 3800+.



*Athlon 64 com código ADA3800DEP4AW*

Carlos E. Morimoto.

É editor do site [www.guiadohardware.net](http://www.guiadohardware.net), autor de mais de 12 livros sobre Linux, Hardware e Redes, entre eles os títulos: "Redes e Servidores Linux", "Linux Entendendo o Sistema", "Linux Ferramentas Técnicas", "Entendendo e Dominando o Linux", "Kurumin, desvendando seus segredos", "Hardware, Manual Completo" e "Dicionário de termos técnicos de informática". Desde 2003 desenvolve o Kurumin Linux, uma das distribuições Linux mais usadas no país.