

TRAZENDO O MELHOR DE CADA EDIÇÃO

A evolução dos  
**COMPUTADORES**

Especial  
**KURUMIN 7**

Do **486** ao **Athlon**

Especial  
**Memória Ram**

Especial  
**HDS**

Especial  
**Redes**  
Por Carlos E. Morimoto

Montagem  
de Micros

Especial  
**NOTEBOOKS**

Configurando  
a rede no **Windows**



Colaboradores:

## 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.

## Júlio César Bessa Monqueiro

É especialista em Linux, participante de vários fóruns virtuais, atual responsável pelos scripts dos ícones mágicos do Kurumin, editor de notícias e autor de diversos artigos e tutoriais publicados no Guia do Hardware.

## Marcos Elias Picão

É produtor do Explorando e Aprendendo (<http://www.explorando.cjb.net>), um blog de informática que traz toda semana dicas de Windows, programas, sites, configurações e otimizações, para todos os níveis.

Iniciou sua vida digital em 2001, e aos poucos foi evoluindo, para frente e para trás, avançando nas novidades do mercado e, ao mesmo tempo, voltando ao passado para conhecer as "janelas" antigas, de vidro a vidro.

Mexe livremente com programação em Delphi, e mantém sites com dicas e tutoriais, além dos seus programas para Windows.

## Luciano Lourenço

Designer do Kurumin linux, trabalha com a equipe do Guia do Hardware.net executando a parte gráfica e de webdesing, editor da Oka do Kurumin onde desenvolve dicas para aplicações gráficas em SL, participa de projetos voltado a softwares livres como o "O Gimp", Inkscape Brasil e Mozilla Brasil.

## Contato Comercial:

Para anunciar no Guia do Hardware em revista escreva para:

[revista@guiadohardware.net](mailto:revista@guiadohardware.net)

## Participe do Fórum:

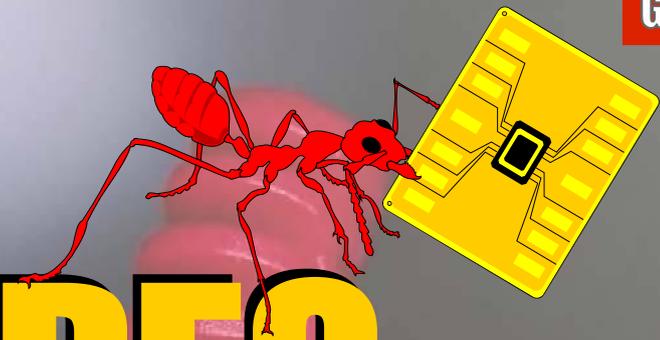
<http://guiadohardware.net/comunidade/>

# SUMÁRIO



# 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 usadas 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 começaram 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

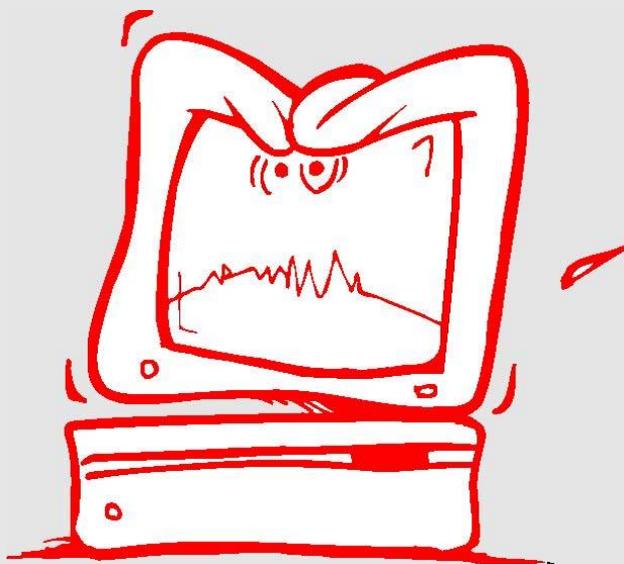
um grande número de capacitores, resistores e outros componentes. No total, ele pesava 30 toneladas e era tão volumoso que ocupava um grande galpão. Outro grave problema era o consumo elétrico: um PC típico atual, com um monitor LCD, consome cerca de 100 watts de energia, enquanto o ENIAC consumia incríveis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



**Relê**

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.

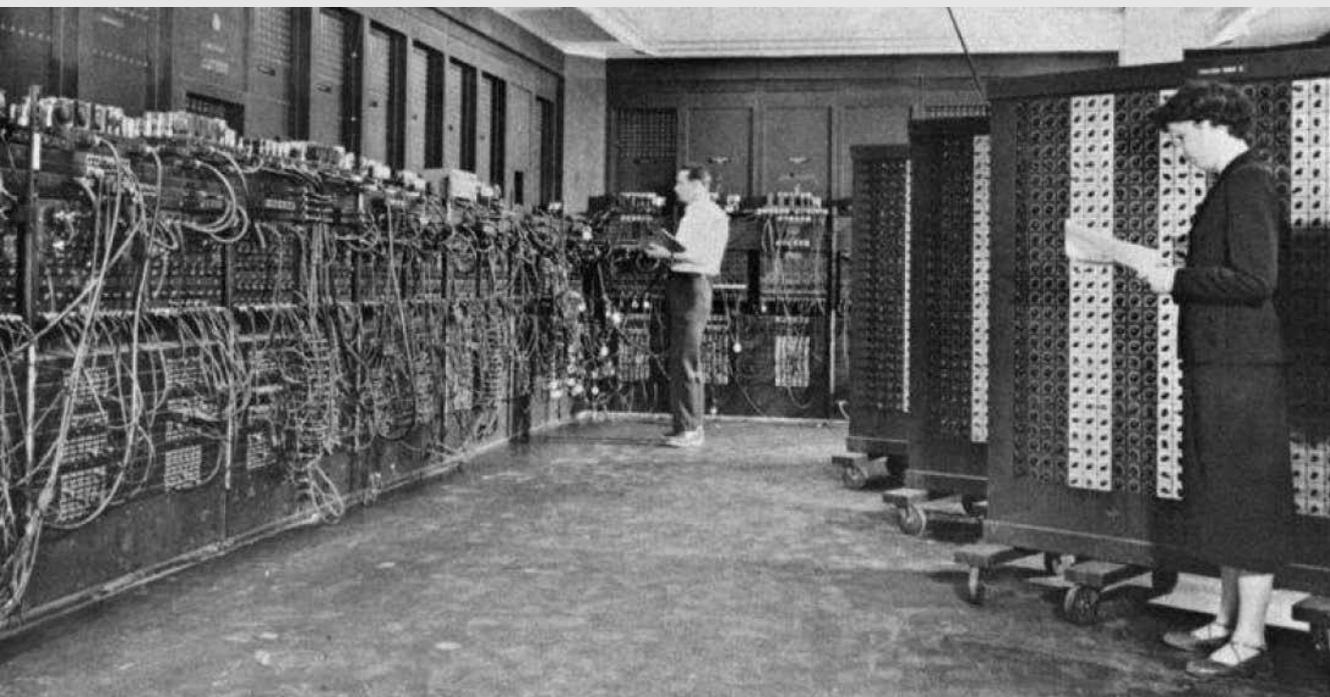


Foto do acervo do Exército dos EUA

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

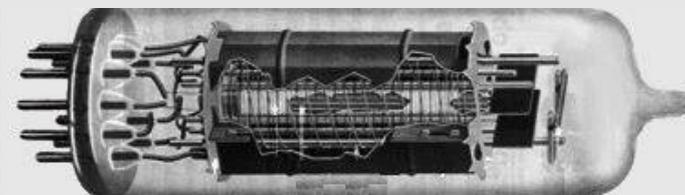
era feita através de cartões de cartolina perfurados, que armazenavam algumas poucas operações cada um.

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 freqüentes.

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 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 com um isolante, bloqueando a corrente. Porém, quando uma certa tensão elétrica era aplicada usando o filamento de controle, uma 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 até 100 MHz. Claro que esta era a frequência que podia ser

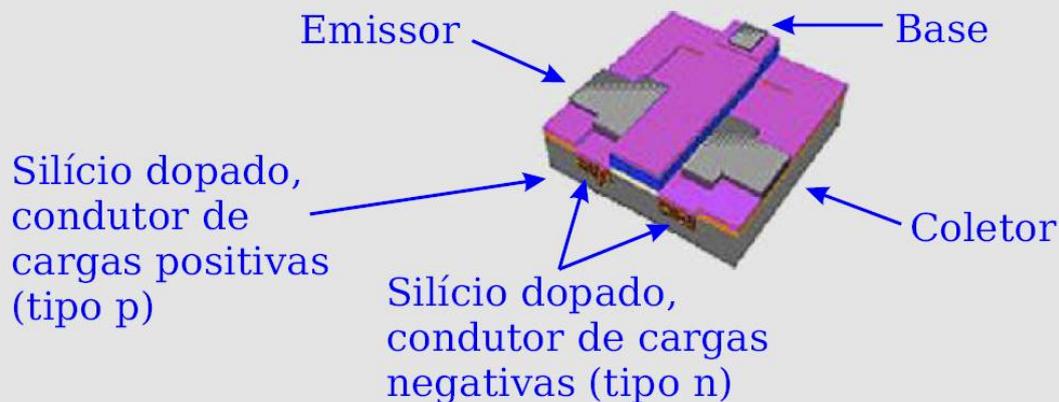
alcançada por um transistor sozinho, nos computadores da época, a frequência de operação era muito menor, já que em cada ciclo de processamento o sinal precisa passar por vários transistores.

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 transístor 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 transístor, que como vimos, pode estar ligado ou desligado. Veja como estes três componentes são agrupados num transístor moderno:



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

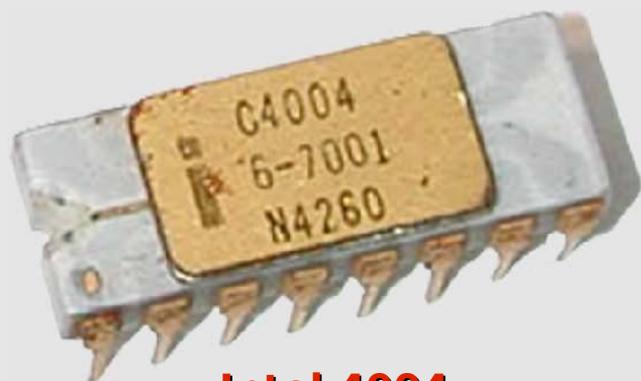
Cada transístor funciona como uma espécie de interruptor, que pode estar ligado ou desligado, como uma torneira que pode estar aberta ou fechada, ou mesmo como uma válvula. A diferença é que o transístor não tem partes móveis como uma torneira e é muito menor, mais barato, mais durável e muito mais rápido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

A mudança de estado de um transístor é feito através de uma corrente elétrica. Esta mudança de estado por sua vez pode comandar a mudança de estado de vários outros transístores ligados ao primeiro, permitindo 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 mais eles se aquecem e mais e mais energia consomem.

Um 386, por exemplo, consumia pouco mais de 1watt 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 o 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, enquanto processadores atuais já superam a barreira dos 3.0 Ghz.

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

O primeiro microship 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 de pouco mais de 2000 transistores 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 microships eram viáveis, criando uma verdadeira corrida evolucionária, em busca de processadores mais rápidos e avançados.

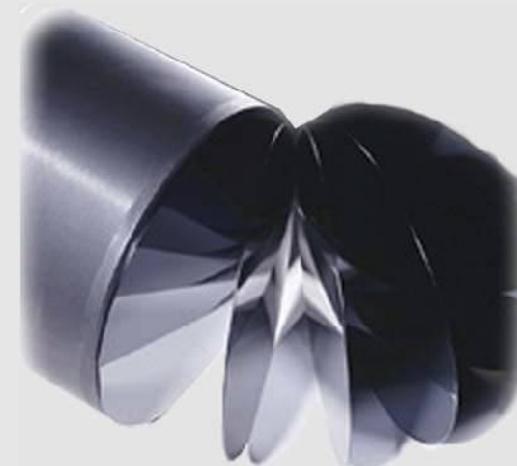
**Intel 4004**

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:



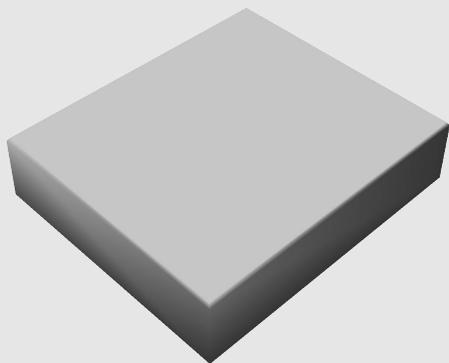
Estas “fatias” por sua vez são polidas e tratadas, obtendo os wafers de silício. A qualidade do wafer determinará o tipo de chip que poderá ser construído com base nele.

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 pre-ciso utilizar wafers de altíssima qualidade, que são extremamente caros.

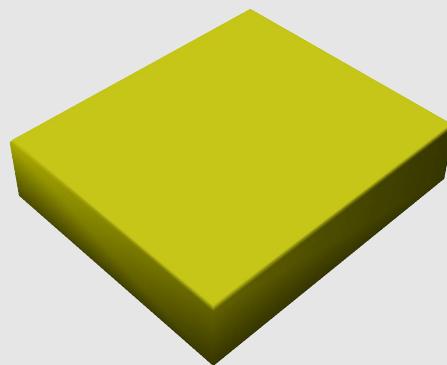
Embora o silício seja um material extremamente barato e abundante, toda a tecnologia necessária para produzir os wafers faz com que eles estejam entre os produtos mais caros produzidos pelo homem. Cada wafer de 30 centímetros custa de mais de 20 mil dólares para um fabricante como a Intel, mesmo quando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

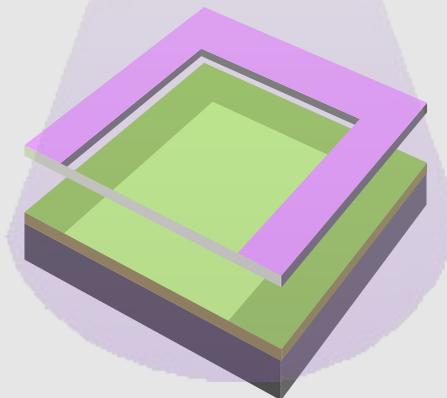


A primeira etapa do processo é oxidar a parte superior do wafer, transformando-a em dióxido de silício. Isto é obtido expondo o wafer a gases corrosivos e altas temperaturas. A fina camada de dióxido de silício que se forma é que será usada como base para a construção do transistor.



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

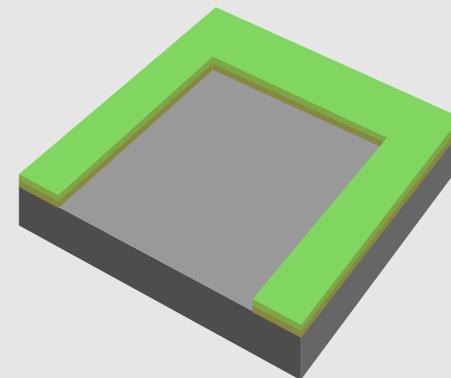
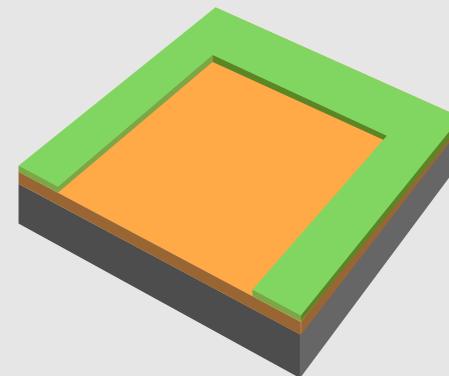
Usando uma máscara especial, é jogada luz ultravioleta apenas em algumas áreas da superfície. Esta máscara tem um padrão diferente para cada área do processador, de acordo com o desenho que se pretende obter. A técnica usada aqui é chamada de litografia óptica. Existem várias variações da tecnologia, 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.

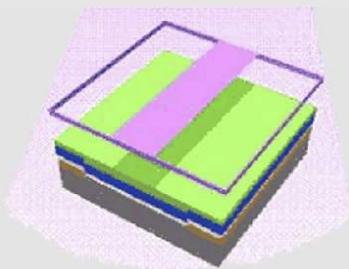
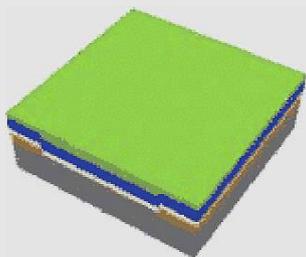
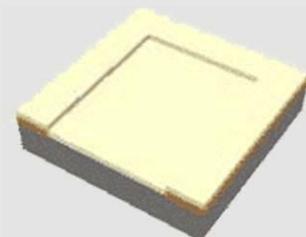
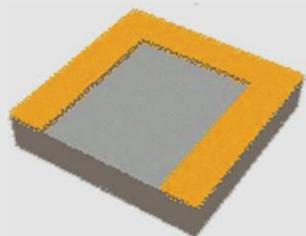


Finalmente, é removida a parte que restou da camada foto-sensível. Note que como temos substâncias diferentes é possível remover uma camada de cada vez, ora o dióxido de silício, ora a própria camada foto-sensível. Com isto é possível “desenhar” as estruturas necessárias para formar os transístores. Temos aqui pronta a primeira camada. Cada transístor é formado para várias camadas, dependendo do projeto do processador. Neste exemplo, temos um transístor simples, de apenas quatro camadas, mas os processadores atuais utilizam um numero muito maior de camadas, mais de vinte em alguns casos, dependendo da densidade que o fabricante pretende alcançar.

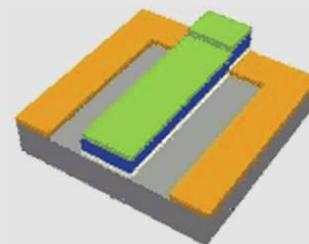
Começa então a construção da segunda camada do transístor. Inicialmente o wafer passa novamente pelo processo de oxidação inicial, sendo coberto por uma nova camada (desta vez bem mais fina) de dióxido de silício. Note que apesar da nova camada de dióxido, o desenho conseguido anteriormente é mantido.

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

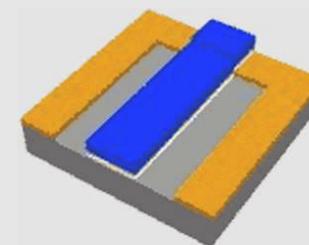
Novamente, o wafer passa pelo processo de litografia, desta vez utilizando uma máscara diferente.



Novamente, a parte da camada foto-sensível 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.

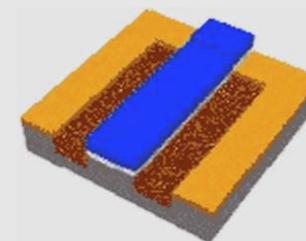


Como na etapa anterior, o que restou da camada foto-sensível é removida. Terminamos a construção da segunda camada do transístor.

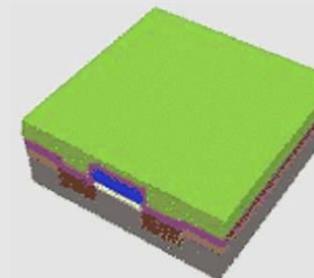


Chegamos a uma das principais etapas do processo de fabricação, que é a aplicação das impurezas, que transformarão partes do wafer de silício num material condutor.

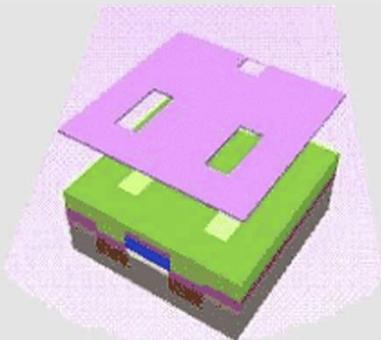
Estas impurezas também são chamadas de íons. Note que os íons aderem apenas à camada de silício que foi exposta no processo anterior e não nas camadas de dióxido de silício ou na camada de cristal de silício.



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

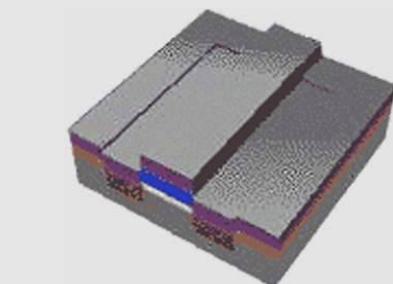
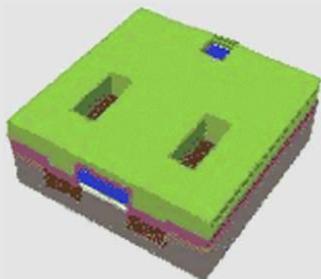


O wafer passa novamente pelo processo de litografia, usando mais uma vez uma máscara diferente.



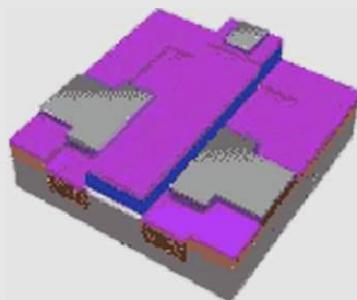
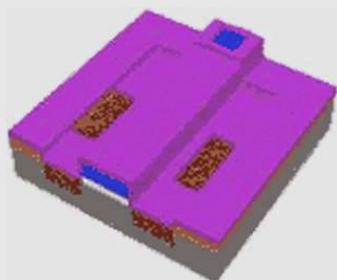
Uma finíssima camada de metal é aplicada sobre a estrutura anterior. Nos processadores atuais, que são produzidos através de uma técnica de produção de 0.065 micron, esta camada metálica tem o equivalente a apenas 3 átomos de espessura.

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



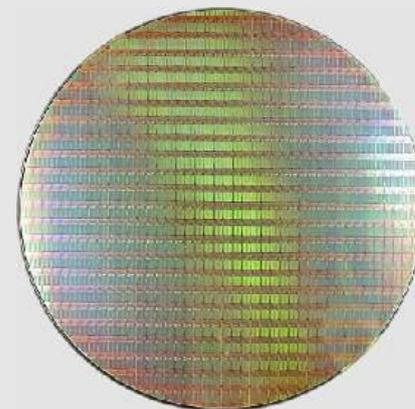
O processo de aplicação da camada fotosensível, de litografia e de remoção das camadas é aplicado mais uma vez, com o objetivo de remover as partes indesejadas da camada de metal. Finalmente temos o transistor pronto.

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

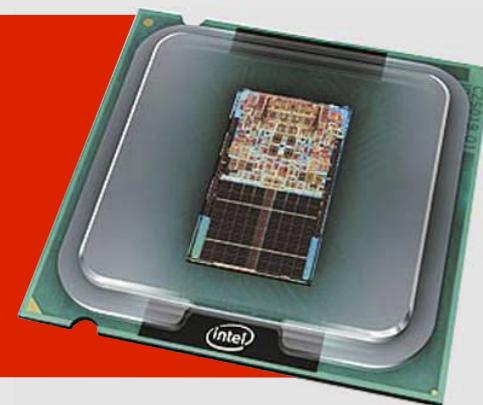


Cada processador é constituído por vários milhões de transistores, divididos em diversos grupos de componentes, entre eles as unidades de execução (onde as instruções são realmente processadas) e os caches. Como todo processador atual processa várias instruções por ciclo, são incluídos diversos circuitos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

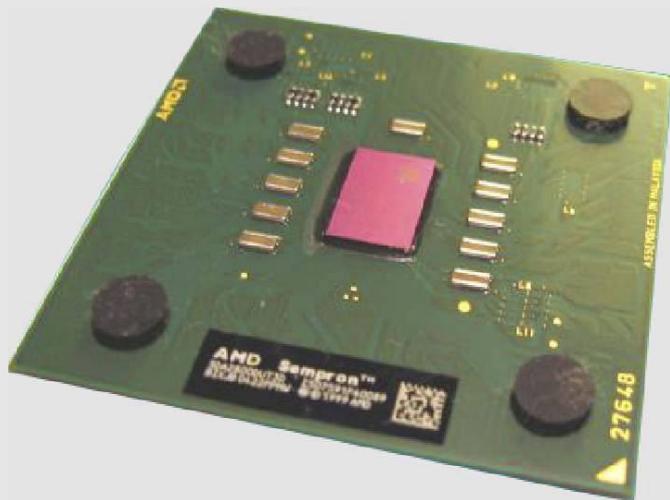
No final do processo, toda a área do wafer é coberta por processadores. Destes, muitos acabam sendo descartados, pois qualquer imperfeição na superfície do wafer, partícula de poeira, ou anomalia durante o processo de litografia acaba resultando numa área defeituosa. Temos também os processadores "incompletos", 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 transístor 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 transístor ocupa uma área 100 vezes menor. Enquanto o 4004 tinha apenas 2.000 transístores, o 486 tinha um milhão deles.

Como a velocidade de operação do transístor 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 transístores 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 transístores 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 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 comerciais 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 atingida 20 anos antes :).

No final da década de 70 sugeriram os supercomputares 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 370/195. 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 supercomputador 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 relacio-

nados 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 utilizava os ciclos de processamento 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 conexão 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 Electronics.

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.

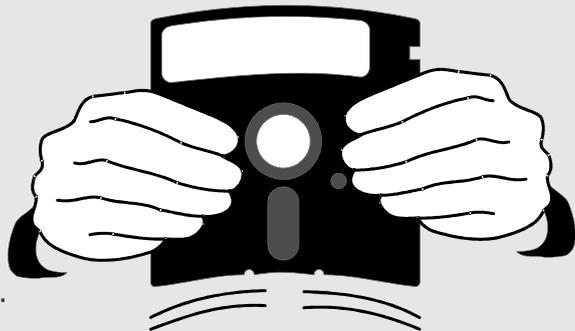
Em sua versão básica, o Altair não tinha muita utilidade prática, a não ser a de servir como fonte de aprendizado de eletrônica e programação.

Entretanto, pouco tempo depois, começaram a surgir vários acessórios para o Altair: um teclado que substituía 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 Electronics, 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.



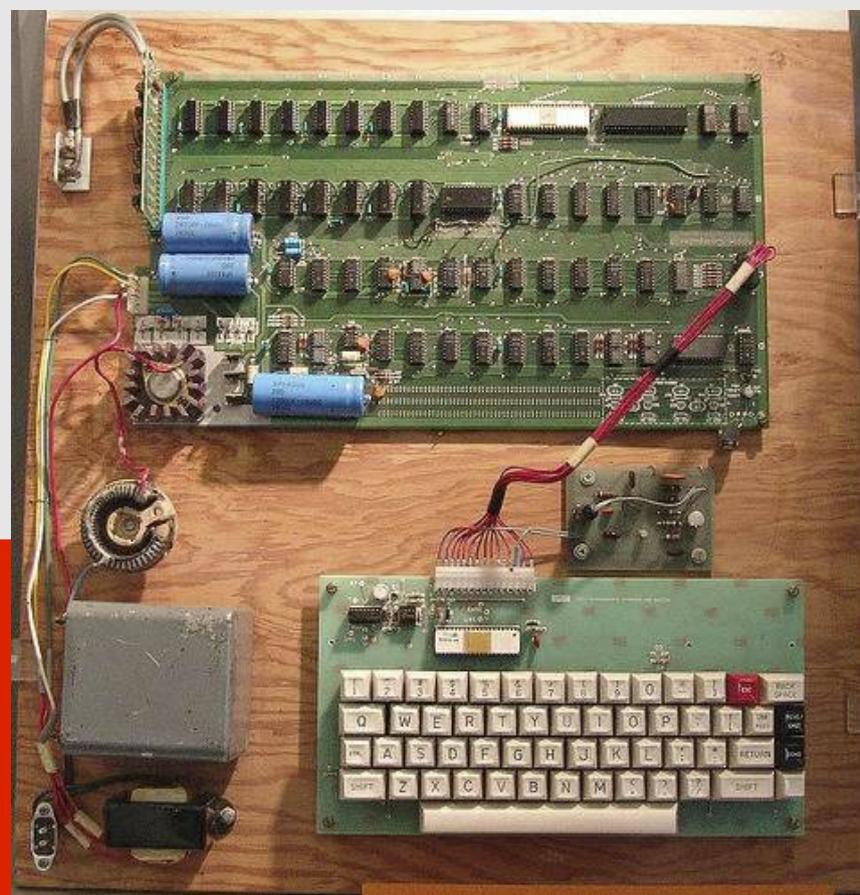
Pouco depois, em 1976, foi fundada a Apple, tendo como sócios Steve Jobs (que continua ativo até os dias de hoje) e Steve Wozniak. Na verdade, a Apple só foi fundada por que o projeto do Apple I (desenvolvido pelos dois nas horas vagas) foi recusado pela Atari e pela HP. Uma frase de Steve Jobs descreve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

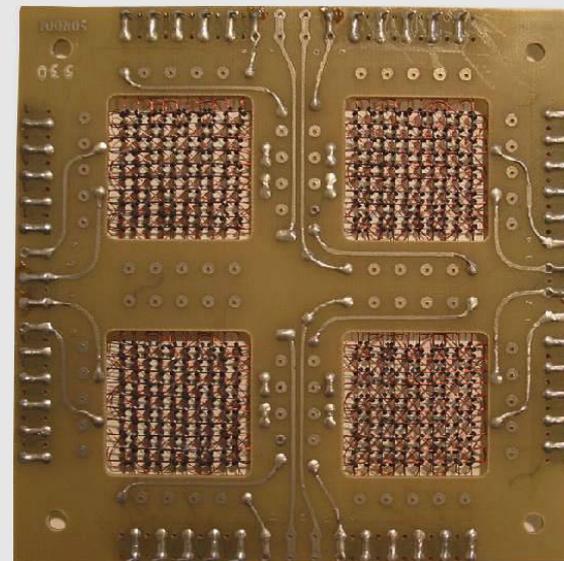
Ao longo das primeiras décadas, a memória RAM passou por duas grandes evoluções. No ENIAC, não existia uma unidade de memória dedicada. Parte das válvulas eram reservadas para armazenar as informações que estavam sendo processadas. Não existia unidade de armazenamento, além dos

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 miniatuizar 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 precisarem 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 transistores 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 regrava os dados a cada 64 milésimos de segundo (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 overclocado 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 videogame 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 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 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.

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.

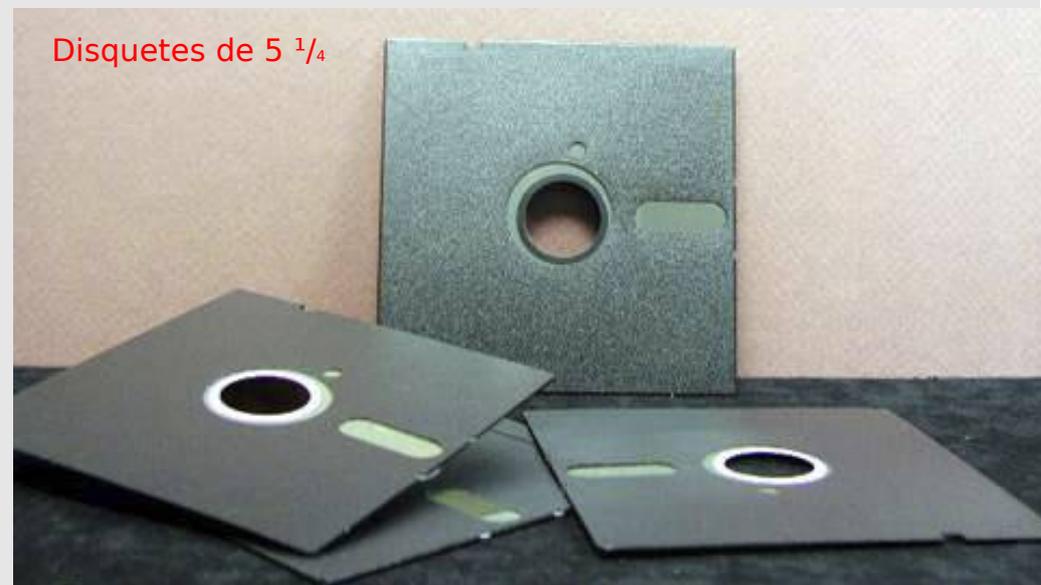
Na configuração básica, o PC custava "apenas" 1564 dólares da época, mas incluindo mais 48 KB de memória, dois drives de disquete e um monitor mono de 12", o preço chegava facilmente a 2500 dólares, que equivalem a mais de 7000 dólares em valores atuais.

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 freqüentemente.

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.

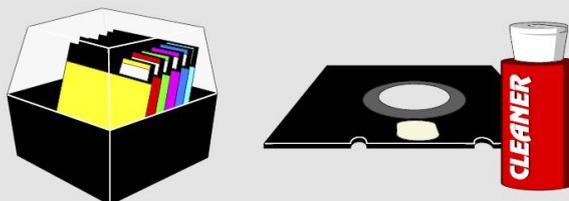
O sistema operacional usado no PC original era o MS-DOS 1.0 (na época ainda chamado de PC-DOS), que foi desenvolvido às pressas pela Microsoft com base num sistema operacional mais simples, chamado QDOS, comprado da Seattle Computers, uma pequena empresa desenvolvedora de sistemas. Na verdade, a Microsoft 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 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.

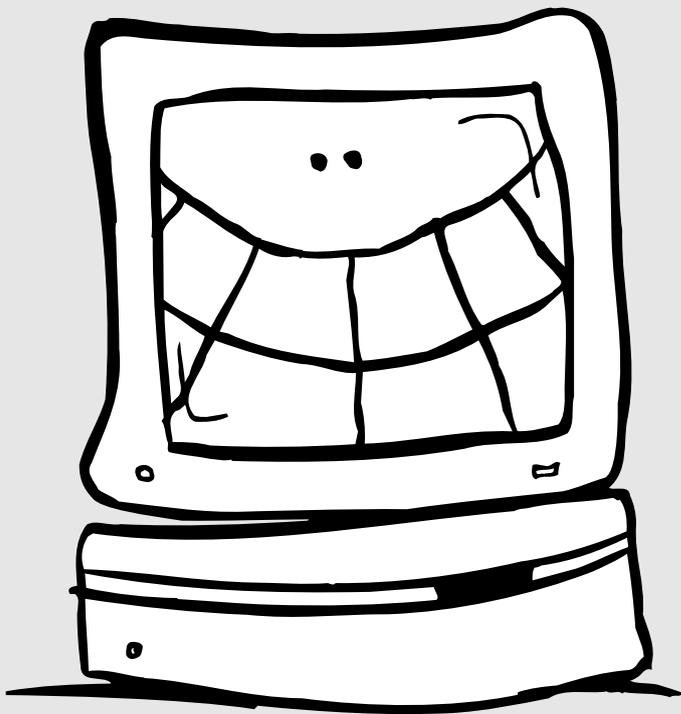
O processador 286 trouxe vários avanços sobre o 8088. Ele utilizava palavras binárias de 16 bits tanto interna quanto externamente, o que permitia o uso de periféricos de 16 bits, muito mais avançados do que os usados no PC original e no XT. O custo dos periféricos desta vez não chegou a ser um grande 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 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 DPMS (“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, executava 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 DPMS 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 DPMS. 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.

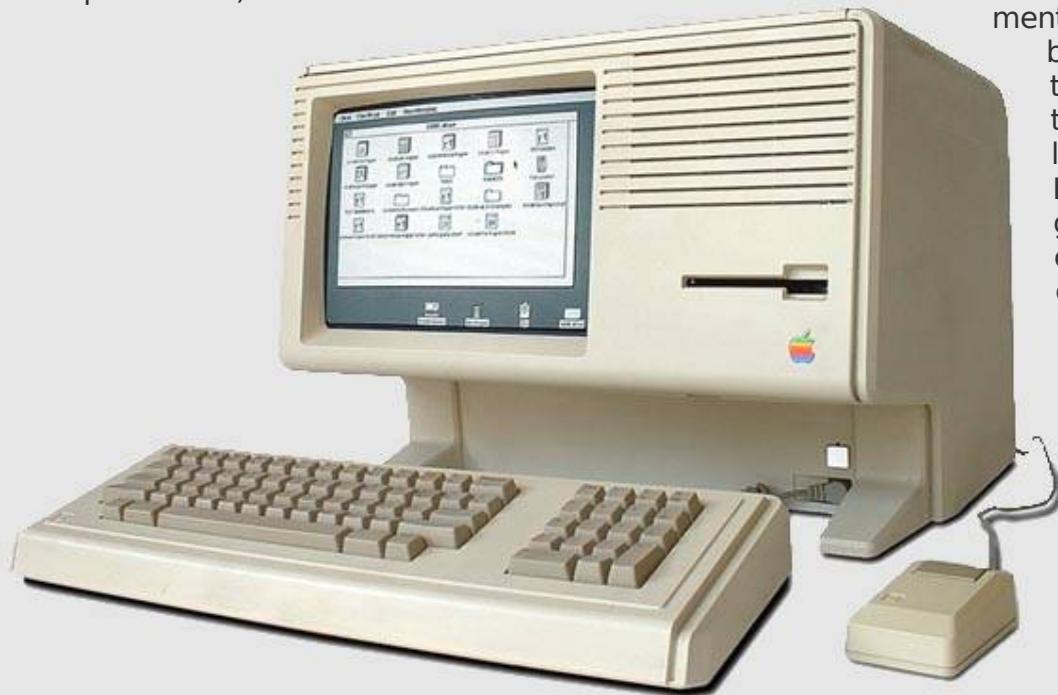
Voltando ao lançamento do 386, embora o processador tenha sido lançado em 1985, a IBM só foi capaz de lançar um PC baseado nele em 1987, dando

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).



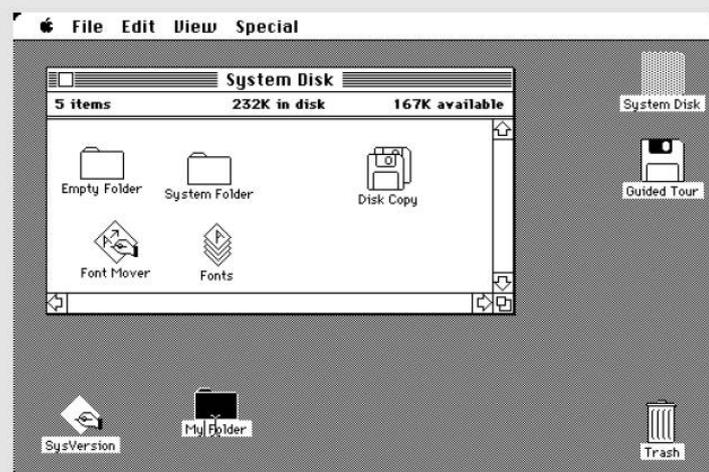
Embora não houvesse nada melhor no mercado, o Lisa acabou não atingindo o sucesso esperado. No total, foram produzidas cerca de 100.000 unidades em dois anos, porém a maior parte delas vendidas com grandes descontos, muitas vezes abaixo do preço de custo (como um lote de 5.000 unidades vendido para a 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 idéia 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.

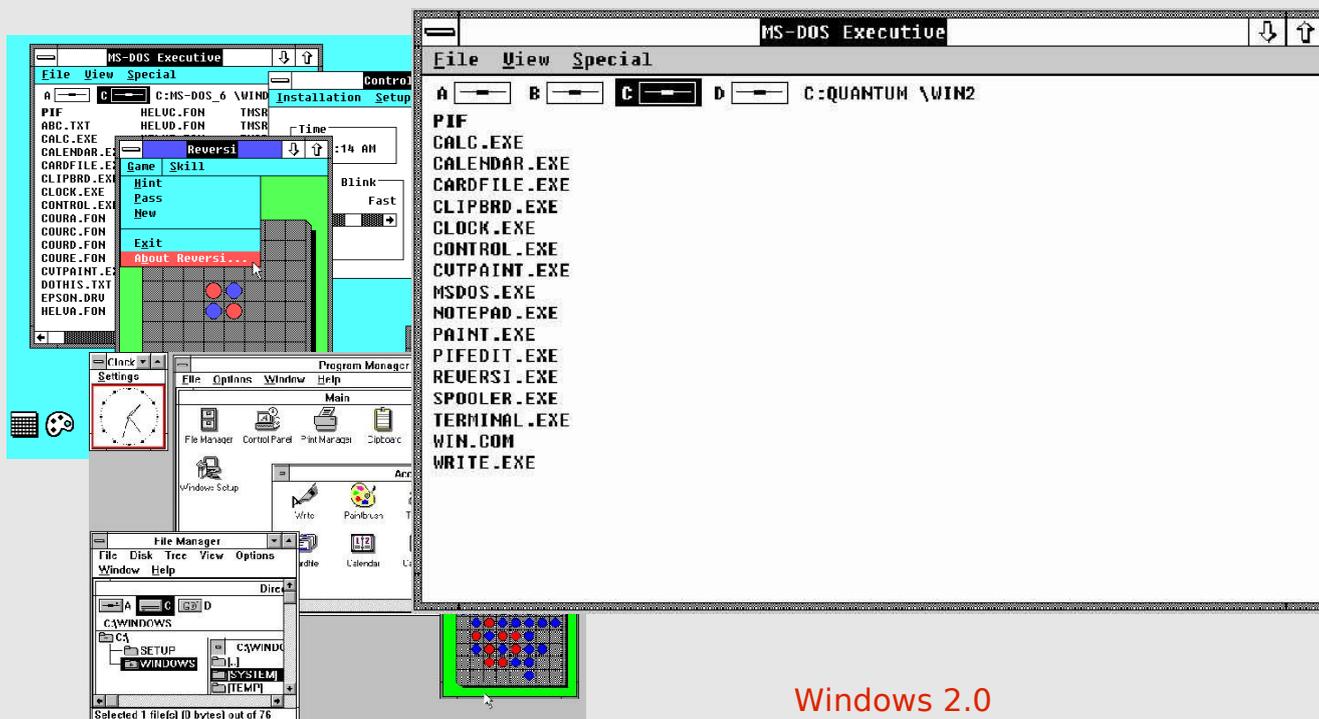
Neste meio tempo, a Microsoft teve tempo de desenvolver a primeira versão do Windows, anunciada em novembro de 1983. Ao contrário do MacOS, o Windows 1.0 era uma interface bastante primitiva, que fez pouco sucesso.

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.



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 co-meceram 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... ;)



1973: Xerox Alto

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.

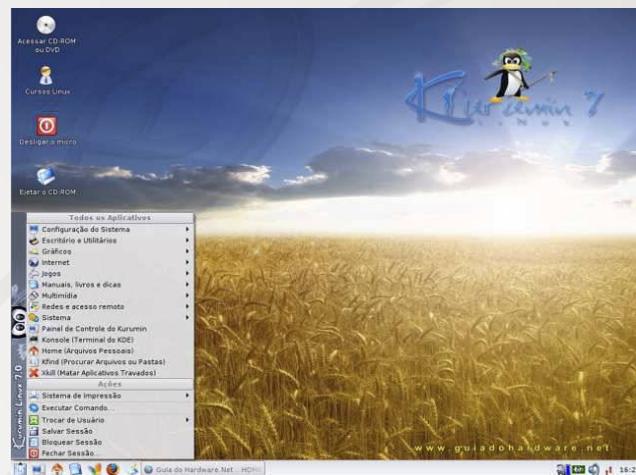
# KURUMIN 7

por Carlos E. Morimoto

O Kurumin é uma distribuição Linux destinada a desktops. Quando falo em desktops estou falando em um sistema destinado a uso geral, que você pode usar para acessar a Internet, trabalhar, assistir filmes, jogar e fazer todo tipo de tarefas. Existem muitas distribuições Linux destinadas a servidores, que é um porto seguro. Um servidor é uma máquina que fica o tempo todo ligada, sempre fazendo a mesma coisa. Existem vários tipos de servidores, como servidores web, servidores de arquivos, servidores de impressão, etc. Quase 70% dos servidores Web do mundo usam o Apache, a maioria deles rodando Linux. O Samba é mais rápido e estável que o Windows como servidor de arquivos e impressoras e por isso continua crescendo rapidamente. Quando se fala em compartilhar a conexão com a Web, novamente o Linux é o sistema mais usado e quando pesquisamos sobre um sistema robusto para rodar um banco de dados, como o Oracle, MySQL ou Postgre SQL, novamente o Linux é o sistema mais usado, o mais comentado e recomendado.



Mas, apesar de ser tão robusto, rápido e estável, o Linux ainda é pouco usado no ambiente doméstico: provavelmente você pode contar nos dedos os amigos (pelo menos os amigos fora da área técnica) que usam Linux no micro de casa. Isso ocorre porque as qualidades necessárias para construir um bom sistema para um servidor e um bom sistema para uso doméstico são muito diferentes, é como comparar um tanque de guerra com um carro popular. Um servidor precisa ser estável e seguro, o resto é secundário. Um sistema para uso doméstico, por outro lado, precisa ser fácil de usar, ser compatível com todo tipo de impressora, modem, scanner e outros periféricos, rodar todo tipo de programas e jogos. Lembre-se de que o Windows ganhou os desktops na época do Windows 3.11 e 95, quando não era nem estável, nem seguro. Existem muitas distribuições Linux recomendadas para uso em servidores, como o próprio Debian, Fedora e CentOS, para citar só algumas. Entretanto, existem poucas distribuições com ênfase nos usuários domésticos. É aqui que chegamos ao Kurumin.



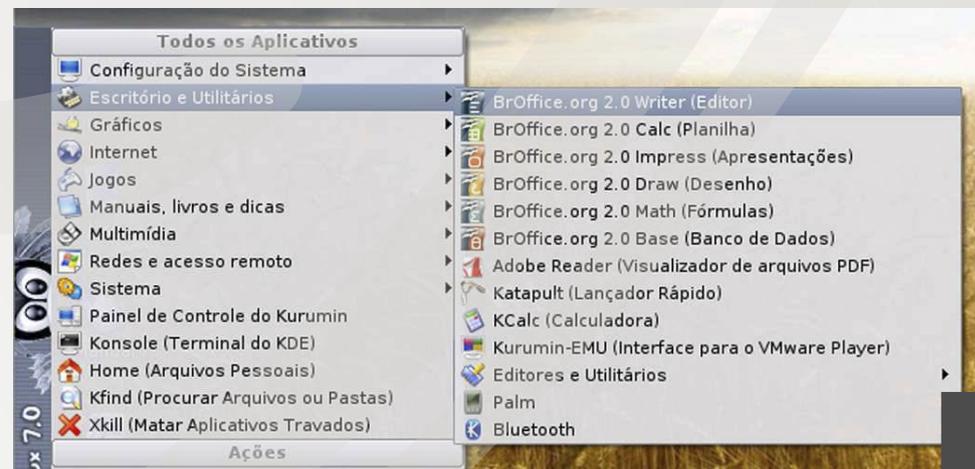
O Kurumin difere das outras distribuições por ser desenvolvido com foco na facilidade de uso. Ele roda diretamente a partir do CD, detectando o hardware da máquina e pode ser instalado rapidamente. Todos os scripts, ferramentas de configuração, menus, etc. são escritos diretamente em português do Brasil, ao invés de serem escritos em inglês e depois traduzidos. Isso faz com que tudo seja muito mais familiar. Muitas pessoas têm apontado o Kurumin como sendo não apenas mais fácil de usar que outras distribuições Linux, mas também mais fácil que o próprio Windows. O fato do sistema rodar a partir do CD o torna também uma boa opção na hora de mostrar o

sistema aos amigos ou testar uma nova versão, pois você pode usá-lo em qualquer micro, sem precisar fazer backup de tudo, particionar o HD e passar por um processo tedioso de instalação. Você simplesmente dá boot pelo CD e ele roda sem

alterar nada que está gravado no HD.

A primeira versão do Kurumin foi lançada em 14 de Janeiro de 2003. De lá pra cá foi um longo caminho :). Nas primeiras versões, o Kurumin era muito mais compacto, desenvolvido com o objetivo de caber em um mini CD. Como estas mídias armazenam apenas 193 MB, era preciso fazer di-

versas concessões com relação aos programas instalados, deixando de fora os aplicativos maiores. Conforme o sistema foi ganhando novos usuários, a idéia do mini CD acabou sendo abandonada, já que a maioria prefere um sistema completo, que já venha com todos os principais aplicativos pré-instalados. Atualmente, o Kurumin vem com um conjunto bastante completo de aplicativos. Na parte de escritório, temos o BrOffice, que é a versão Brasileira do Open-Office.org. A principal vantagem é que ele é personalizado para o público brasileiro e traz um corretor ortográfico bastante afinado. Como autor eu devo dizer que é muito mais confortável trabalhar com o Writer do que conviver com as esquisitices do Word. Temos ainda o Acrobat Reader, que é a ferramenta obrigatória para



visualizar arquivos PDF, além do suporte a Palms e Bluetooth. Se você tem um PDA com Bluetooth, você pode até mesmo compartilhar a conexão e navegar através dele :). O Kurumin-EMU é uma interface que permite instalar e configurar o VMware Player. Através dele, você pode rodar o Windows ou outros sistemas operacionais dentro de uma janela, muito útil em casos em que você precisa de determinados programas que só rodam no Windows.

Um quesito cada vez mais importante é a navegação web, já que, afinal, passamos a maior parte do tempo usando o navegador. Além do Firefox, que dispensa apresentações, temos o Konqueror, que é o navegador padrão do KDE. O Konqueror oferece muito mais recursos do que pode parecer à primeira vista. Além de ser um bom navegador e um poderoso gerenciador de arquivos, ele oferece uma série de plugins, que permitem transferir arquivos via SSH, acessar compartilhamentos de rede, criar galerias de imagens, entre várias outras coisas.

Por ser um componente do KDE, o Konqueror acaba sendo bem mais leve que o Firefox, por isso também é uma boa opção para quem usa micros de configuração mais modesta. Existe também uma grande preocupação com relação aos plugins nos navegadores, de forma que o suporte a streaming de vídeo, visualização de arquivos PDF, Flash e Java vêm pré-instalados de fábrica. Embora os webmails sejam cada vez mais populares, o Kurumin inclui também o Thunderbird, o cliente de e-mails primo do Firefox. O principal recurso do Thunderbird é um filtro anti-spam adaptável, que uma vez ativado, aprende a classificar as mensagens de acordo com os seus critérios. Nos primeiros dias, você vai marcando manualmente as mensagens que são spam e as que não são e logo ele aprende a eliminar os spams sozinho. Para usar MSN ou ICQ, você pode escolher entre o Kopete e o Gaim. O Kopete é o mensageiro do KDE, ele é o mais leve e o que possui mais funções. O Gaim, por sua vez, possui uma interface mais simples e, justamente por isso tam-

bém tem seus usuários fiéis. Ambos também suportam o Google Talk, Jabber e outros protocolos menos usados. Como não podia deixar de ser, é incluído também o Skype. Embora não venham instalados por padrão no sistema, você pode instalar o Gizmo e outros programas de VoIP que possuem versões para Linux rapidamente, basta baixar os pacotes para o Debian. Na categoria "Outros" você encontra também o KTorrent, um cliente gráfico para downloads via bittorrent, o Firestarter, que é um firewall gráfico, o GFTP, um cliente de FTP (que também suporta SFTP), entre outros.

O menu "Conectar na Internet ou Configurar a Rede" é na verdade uma cópia das

opções de configuração da rede disponíveis no Painel de Controle. Esta está disponível no menu Internet para facilitar as coisas para quem está usando o sistema pela primeira vez, tornando as opções mais fáceis de encontrar.

Na parte gráfica, temos o Gimp, o Inkscape e o Kolourpaint. Embora os três sejam editores de imagem, eles se enquadram em categorias completamente diferentes e por isso se complementam. O Gimp é um editor de imagens, no estilo do Photoshop, que permite tratar fotos e aplicar efeitos. Ele é muito avançado, com uma quantidade impressionante de opções e efeitos e por isso também é um pouco difícil de usar.



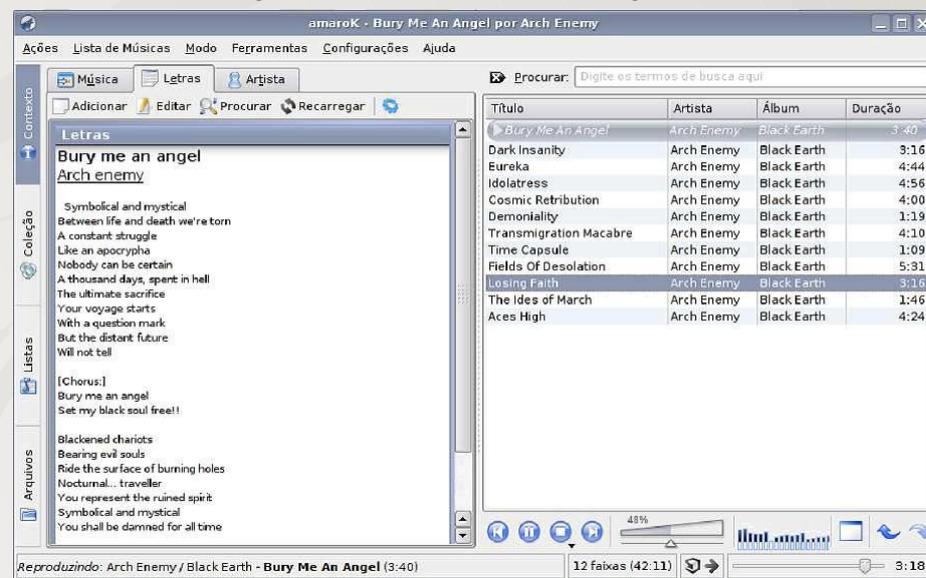
O Inkscape é um programa de desenho vetorial, similar ao Corel, usado para criar ilustrações, cartazes, banners, anúncios, etc., enquanto o Kolourpaint é um editor de imagem simples, no estilo do Paint, que permite fazer desenhos e retoques rápidos. Completando o time, temos também o Ksnapshot, um programa bastante flexível para tirar screenshots da tela, que permite especificar tempos de espera (de forma que você consiga chegar aos menus mais escondidos), capturar apenas a janela do aplicativo em foco e salvar os screenshots diretamente em vários formatos.



Se você tem uma câmera digital, experimente o Digikam. Ele é um dos aplicativos mais completos dentro da área, permitindo transferir fotos da câmera para o PC, organizá-las, tratar e aplicar efeitos, gerar álbuns em vários formatos, entre muitos outros recursos.

Na área de multimídia temos o trio Amarok, Kaffeine e K3B. O Amarok é um player de áudio "de nova geração", que oferece uma proposta bem mais completa que programas tradicionais, como o WinAMP e o XMMS. Ele permite organizar e classificar suas músicas de diversas formas, oferece opções para buscar letras de músicas e capas de CDs, suporte a iPods e outros players de áudio, entre muitos outros recursos. Ele pode parecer um pouco complexo demais primeira vista, mas depois de usá-lo por alguns dias você começa a perceber porque ele tem recebido tantas indicações favoráveis. Por exemplo, ao ouvir qualquer música (mesmo de uma

banda relativamente desconhecida), clique na aba "Letras" e ele baixa automaticamente a letra da música via web. Na aba "Música" ele mostra a capa do CD do qual a música faz parte, além de mostrar outras músicas da mesma banda que você tem na sua coleção, entre outras informações:



O Kaffeine é um media player mais tradicional, que oferece suporte a todo tipo de formato de vídeo e áudio, incluindo filmes em Divx, WMV, Quick Time, DVDs, VCDs, CDs de música e praticamente qualquer outro formato que você puder encontrar. Para que ele possa ler DVDs encriptados e abrir vídeos nos formatos mais complicados, é necessário que dois pacotes estejam instalados, o "libdvdcss2" (contém a biblioteca necessária para entender o sistema de encriptação usados nos DVDs) e o "w32codecs" (que contém as bibliotecas que adicionam suporte a formatos de vídeo adicionais). Os dois pacotes não podem ser incluídos diretamente no sistema, pois não podem ser distribuídos nos EUA e em alguns outros países, mas você é livre para usá-los se mora no Brasil. Para instala-los, use a opção "Instalar suporte a mais formatos de vídeo", ela dispara um script que faz a instalação via apt-get.



Finalmente, temos o K3B, um programa de gravação de CDs e DVDs, que supera o Nero em recursos e facilidade de uso. O único motivo para alguns ainda acharem que o Nero é superior é o fato de não conhecerem ou não saberem usar todos os recursos do K3B ;).



Uma prova disto é que o Nero possui uma versão Linux desde 2005, porém ela vem sendo virtualmente ignorada. Além dos próprios desenvolvedores, é possível contar os usuários nos dedos. O desinteresse é tanto que, segundo a busca do Google, na internet inteira só existem 62 referências ao link de download (<http://www.nero.com/ptb/linu-x3.html>), sendo que um deles é de uma dica que eu mesmo escrevi. Por ser o programa de gravação padrão do sistema, o K3B pode também ser acessado de outras formas.

Clicando com o botão direito sobre uma pasta com arquivos, você tem a opção "Criar CD de dados com o K3B". Clicando sobre um arquivo .iso, o K3B abre automaticamente, oferecendo a opção de gravá-lo em um CD ou DVD.

No menu "Outros", você encontra alguns programas adicionais. O Audacity é uma ferramenta de gravação e edição de áudio "semiprofissional" que inclui diversos efeitos, opções para juntar diversas faixas e muitos outros recursos. Apesar dos recursos, ele é bem simples de usar.

O XMMS é um player de áudio "da velha guarda", bem simples e fácil de usar, similar ao WinAMP do Windows. O gMplayer é outro player de mídia, que concorre diretamente com o Kaffeine; ele utiliza uma engine diferente, por isso sempre existem casos de vídeos em formatos mais exóticos que são exibidos corretamente em apenas um dos dois, por isso é interessante ter ambos instalados. Finalmente, temos o Kaudiocdcreator, que permite ouvir e ripar CDs de música, permitindo que você transfira as músicas para escutar no micro ou no seu MP3-Player. No mesmo menu você encontra os scripts que permitem configurar placas de TV. O menu "Redes e Acesso remoto" concentra as ferramentas para acesso remoto a outras máquinas, compartilhamento de arquivos e acesso a compartimentos de rede. Entre as opções, estão o SSH, VNC e NFS, o Smb4K (que permite visualizar e acessar os compartimentos em máquinas Windows da rede), o TSClient (que permite acessar máquinas Windows com o utilitário de assistência re-

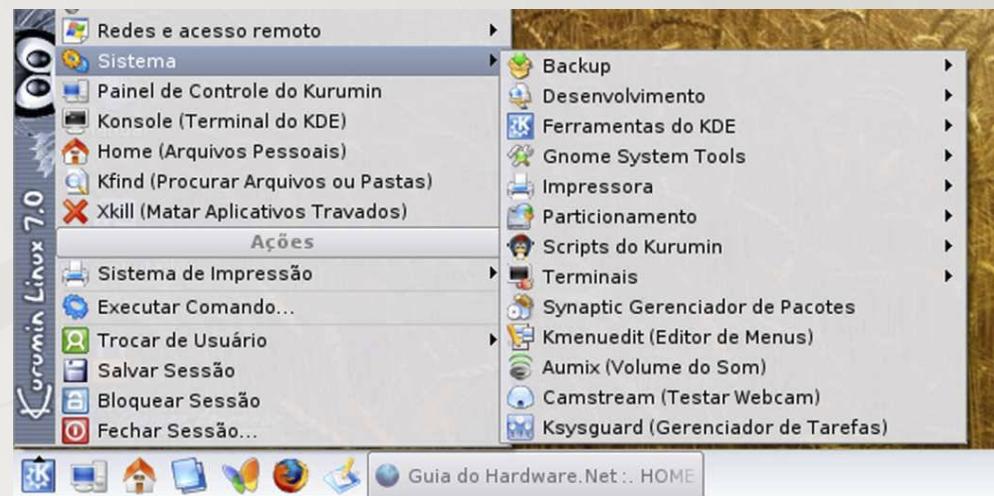
mota ativo) e o Synergy (que permite controlar vários micros usando apenas um teclado e mouse, como uma espécie de KVM via software). O Krfb, (disponível no submenu "Acesso Remoto") permite compartilhar o desktop, de forma similar à assistência remota do Windows. As máquinas da rede com o Krfb ativo podem ser acessadas remotamente usando o Krdc (disponível no mesmo menu), ou usando o cliente do VNC.



O menu "Sistema" é uma espécie de caixa preta, onde estão agrupados uma grande quantidade de utilitários para a configuração e manutenção do sistema. Só para citar alguns exemplos, temos o Partition Image (um concorrente do Ghost, que permite criar e restaurar imagens de partições), o Gparted (um particionador gráfico bastante poderoso, que é usado também durante a instalação do sistema) e o Testdisk, uma ferramenta de manutenção, que permite recuperar partições deletadas devido a acidentes, defeitos diversos ou à ação de vírus.



No menu "Scripts do Kurumin" você encontra o script para instalar o Clamav, um antivírus que permite desinfetar partições do Windows (muito útil para quem usa o sistema em dual boot).



O Clamav é o melhor exemplo dentro de um fenômeno interessante: ao contrário da crença popular, existem sim vários programas antivírus que rodam no Linux. Entretanto (apesar de rodarem no Linux), todos eles são destinados justamente a remover vírus em partições e arquivos do Windows, atendendo a quem usa os dois sistemas em dual boot ou precisa desinfetar arquivos antes de transferí-los a outras máquinas Windows da rede, por exemplo. Como pode ver, o Kurumin inclui uma variedade muito grande de programas. Muita gente até critica o sistema nesta questão, argumentando que seria melhor desenvolver um sistema mais " enxuto", com apenas os aplicativos mais bá-

icos. Esta é uma questão em que prefiro pecar pelo excesso do que pela omissão, já que o Kurumin é usado por toda classe de usuários, desde pessoas que estão comprando o primeiro micro, até usuários técnicos, que passam o dia no terminal. Para atender quem prefere um sistema mais compacto, foi criado o Kurumin Light (que pode ser encontrado na página de download do Kurumin). Ele é uma versão reduzida, com apenas 193 MB, que mantém a tradição do Kurumin gravado em mini CDs. Devido à restrição no tamanho, ele não inclui diversos recursos, como o suporte a Java, suporte

a impressão e aceleração 3D. Tudo isso pode ser instalado posteriormente, via apt-get, mas se você precisa destes recursos, a melhor opção é usar o Kurumin completo, que já vem com tudo pronto.

A partir do Kurumin 7, o projeto passou por outra mudança importante, passando a ser baseado nas versões estáveis do Debian, o que garante atualizações suaves e a possibilidade de usar a mesma instalação por muito tempo, sem se preocupar com atualizações ou correções de segurança. Basta atualizar o sistema via apt-get e atualizar os ícones mágicos, para manter sua instalação atualizada, sem dor de cabeça. Se você ainda usa uma das versões antigas, atualizar para o Kurumin 7 é fortemente recomendável :). Além dos aplicativos, temos o principal diferencial do Kurumin, que é a grande coleção de ferramentas e scripts de configuração. Alguns deles estão distribuídos pelo iniciar (como os scripts de configuração do Blue-tooth), mas a maioria está agrupada dentro do Painel de Controle:

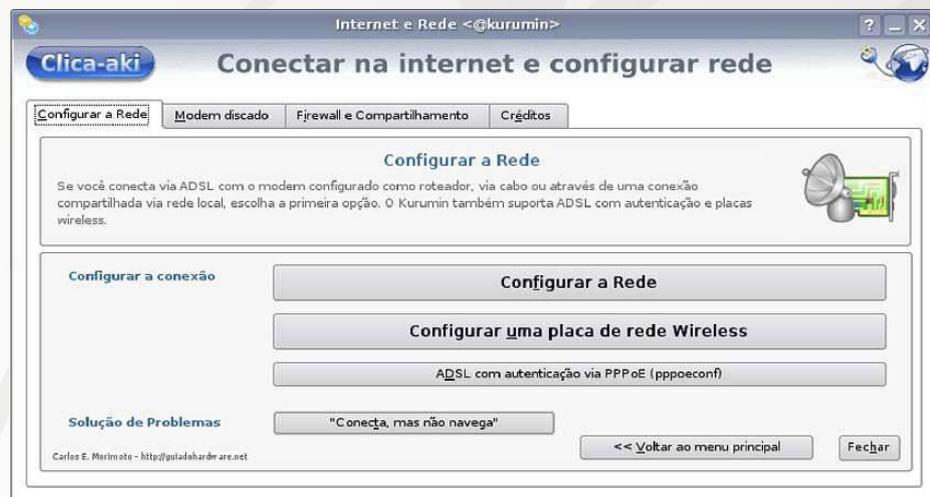


A grande quantidade de opções pode assustar à primeira vista, mas são justamente elas que tornam o Kurumin poderoso, permitindo que você extraia muito mais do sistema. Por exemplo, imagine que você tem dois micros antigos que estão sem uso. Você poderia aproveitá-los usando-os como terminais leves do seu micro, que tem memória e processador de sobra. O problema é que para isso você precisaria configurar um servidor LTSP, o que (se feito manualmente) é uma tarefa espinhosa.

Entretanto, se você usa o Kurumin, pode transformar seu micro em um servidor LTSP para os outros dois usando o script do Kurumin Terminal Server (Instalar e Configurar Servidores > Acesso Remoto). Por serem tão práticos, estes scripts permitem usar recursos "avançados", como o LTSP, em ambientes onde isso não seria viável, devido à dificuldade de instalação:



Entretanto, as coisas podem ficar um pouco mais complicadas se você acessa usando uma placa wireless ou usa acesso discado, daí as demais opções :). O suporte a redes wireless já foi um grande problema no Linux, mas hoje em dia existem drivers para a grande maioria das placas. Mesmo as placas que não possuem drivers podem ser configuradas usando o ndiswrapper, que permite ativar a placa usando o driver do Windows. Também estão disponíveis drivers para os softmodems que possuem drivers Linux. Infelizmente, no caso dos modems, a lista está longe de ser completa, por isso é importante testar antes de comprar. Pelo painel você pode também ativar o firewall. Embora o Linux seja bem mais seguro que o outro sistema, é sempre bom ter uma camada extra de segurança. Como costume dizer, as brechas de segurança são como minas terrestres: por mais improvável que possa parecer, você nunca sabe quando pode estar prestes a pisar em uma :).



Muitas das opções no painel de suporte a hardware são na verdade opções destinadas a solucionar problemas. Por exemplo, ao plugar um pendrive ou cartão de memória, um ícone é criado no desktop, permitindo que você acesse os arquivos. Nos raros casos onde isto não acontece automaticamente, você pode usar a opção "Detectar Pendrives e cartões de memória", que força uma nova busca.

O mesmo pode ser dito do Alsaconf, que obriga o sistema a redetectar sua placa de som. Algumas opções importantes são o script para instalar o driver 3D para placas nVidia, para compartilhar a impressora, para alterar a configuração do vídeo e para instalar o Kernel com suporte a SMP. O driver da nVidia não é obrigatório, mas sem ele sua cara GeForce FX vai ficar limitada a aplicativos 2D, o que é um grande desperdício. O mesmo se aplica ao Kernel SMP, que ativa o segundo processador em micros com processadores dual-core.



O Painel dos ícones mágicos é o recurso mais famoso do Kurumin. De uma forma geral, instalar programas no Linux é bem simples, pois (desde que o programa desejado esteja disponível nos repositórios oficiais) é fácil instalá-lo usando o Synaptic ou usando diretamente o apt-get, via terminal. Por exemplo, digamos que você ouviu falar de um programa de editoração muito bom, chamado Scribus. Você poderia muito bem pesquisar sobre ele no Google e procurar o pacote de instalação dentro da página do projeto, mas seria muito simples abrir o terminal e executar os três comandos abaixo:

```
$ su - <senha>
# apt-get update
# apt-get install scribus
```

O "su -" faz com que você se logue como root, de forma a poder instalar novos programas. O apt-get update faz com que o apt-get verifique a lista dos programas disponíveis e o "apt-get install scribus" baixa e instala o Scribus. Como pode ver, para instalar um novo programa, basta chamá-lo pelo nome. O mesmo se aplica se você estiver usando o Synaptic, que funciona como uma interface gráfica para o apt-get. As coisas podem se complicar um pouco no caso de programas que não estão disponíveis via apt-get.. Entretanto, um dos grandes problemas de quem começa a usar Linux é justamente o fato de não conhecer os programas disponíveis. Os ícones mágicos ajudam neste aspecto, oferecendo uma lista dos programas mais comuns, incluindo a descrição e um screenshot de cada um. Eles também automatizam a instalação de programas complicados, onde é necessário executar um conjunto de passos para fazer o trabalho.



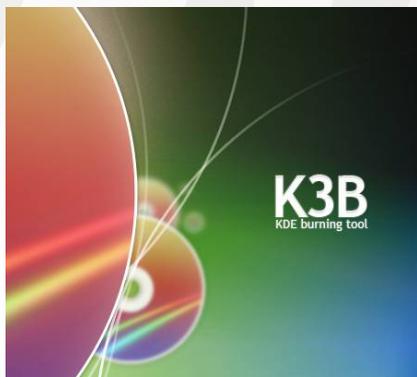
Finalmente, temos o painel de configuração de servidores, que é o mais complexo do grupo. Como costumam dizer, estes scripts não vão transformá-lo em um administrador de redes, mas vão permitir que faça muitas coisas que eles fazem :).

As opções mais usadas são as relacionadas com o compartilhamento da conexão. É muito fácil transformar um micro com duas placas de rede em um servidor Linux, compartilhando a conexão com os micros da rede local. Você começaria configurando a rede e ativando o firewall, no painel de configuração da rede, e em seguida usaria a opção "Compartilhar a conexão com a Internet" para compartilhar a conexão do servidor com os micros da rede local. O script configura também o servidor DHCP, de forma que os outros micros possam obter a configuração da rede automaticamente a partir do servidor. Para otimizar a conexão, você pode usar o Squid. Ele é um servidor proxy, que mantém um cache com as páginas e os arquivos já acessados, acelerando o acesso às páginas e reduzindo o uso do link. Use a opção de ativar o proxy transparente dentro do script; assim o proxy se integra ao compartilhamento da conexão, sem que você precise configurar cada micro manualmente para usá-lo.

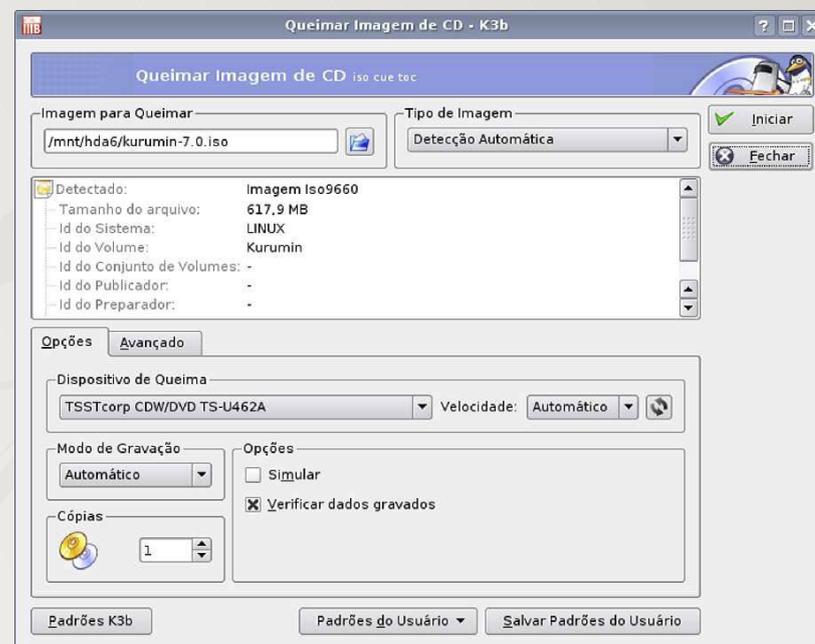


## Como baixar, gravar e dar boot

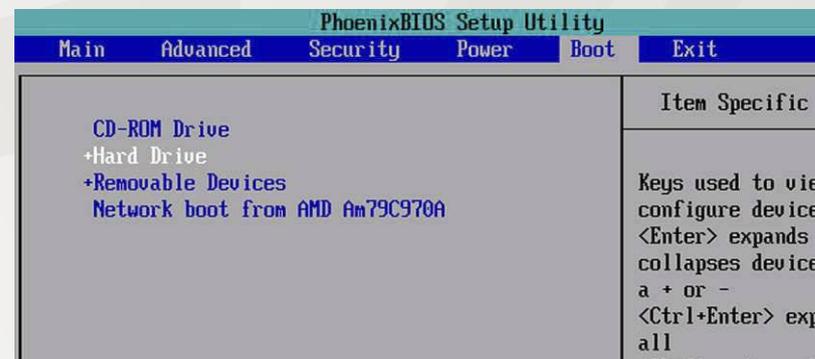
A forma mais popular de disponibilizar novas versões das distribuições é através de arquivos ISO, cópias binárias do conteúdo dos CDs ou DVD de instalação, que você pode gravar usando o Nero, K3B ou outro programa de gravação, obtendo um CD idêntico ao original. Gravar um arquivo ISO é diferente de gravar um arquivo qualquer no CD. Um arquivo ISO é uma imagem binária que deve ser copiada bit a bit no CD-ROM, e não simplesmente adicionado dentro de uma nova seção. Todos os bons programas de gravação suportam a gravação de arquivos ISO, veja como gravar o arquivo usando alguns programas populares. Ao usar o Easy CD Creator, abra o Easy CD Creator e clique em "File > Menu > Record CD from CD Image". Aponte o arquivo que será gravado. Marque a opção "ISO Image Files (\*.iso)" na janela de navegação e clique em "Start Recording" para gravar o CD. No Nero Burning Rom, clique em File > Burn Image, aponte o arquivo que será gravado e clique em "Write" para gravar o CD.



Ao usar o K3B, clique em Ferramentas > Gravar Imagem ISO (ou Queimar imagem de CD), aponte o arquivo, escolha a velocidade de gravação e clique em "Gravar". Ao usar o K3B, marque a opção "Verificar dados gravados"; ela verifica a mídia depois de concluída a gravação, avisando sobre erros de gravação.



Depois de gravado o CD, o próximo passo é configurar o setup da placa-mãe para dar boot através do CD-ROM. A maioria dos micros vêm configurados para dar boot preferencialmente através do CD-ROM. Neste caso basta deixar o CD na bandeja e você já cai na tela de boas-vindas do sistema. Se não for o seu caso, pressione a tecla DEL durante o teste de memória para entrar no Setup. Procure pela seção "Boot" e coloque o CD-ROM como dispositivo primário. Tudo pronto, agora é só salvar a configuração acessando o menu exit, escolhendo a opção "Save & Exit setup".



Ao reiniciar o micro sem o CD no drive, ele volta a carregar o Windows ou outro sistema que estiver instalado no HD. Esta alteração apenas faz com que ele passe a procurar primeiro no CD-ROM. Um hábito saudável é verificar a integridade do arquivo ISO antes de gravar o CD. Sempre é possível que o arquivo esteja incompleto ou venha corrompido, seja por problemas com a conexão ou no gerenciador de download usado. Você pode detectar este tipo de problema (e evitar gastar mídias à toa), verificando o MD5SUM do arquivo, um teste que soma todos os bits do arquivo e devolve uma "assinatura", um código de 32 dígitos que permite detectar qualquer mudança no arquivo. Os códigos de assinatura dos arquivos estão quase sempre disponíveis na página de download, como em: 11479ced7eea505506b5a3314d33ee70 kurumin-7.0.iso Você precisa apenas rodar o MD5SUM no arquivo baixado e ver se o resultado é igual ao número da página. No Linux (qualquer distribuição), acesse a pasta onde o arquivo foi baixado e digite: \$ md5sum kurumin-7.0.iso Se o número retornado for igual, você pode gravar a imagem sem medo, o arquivo está ok. Caso o número seja diferente, então o arquivo chegou corrompido ou incompleto. No Windows baixe o programa disponível no <http://www.md5summer.org/download.html>. Ele é gráfico, até mais fácil de usar que a versão Linux. Uma última dica é com relação aos famosos "Cloop Errors", erros de leitura que aparecem ao tentar rodar qualquer live-CD (que use o módulo cloop para compactar os dados) a partir de uma mídia ruim. O cloop possui um sistema próprio para verificar as integridade dos dados no CD, detectando e avisando sobre os erros de leitura. Nestes casos você verá vários "CLOOP READ erros" durante o boot, que indicam justamente que o sistema não está conseguindo ler os dados corretamente. Veja um exemplo:

```
CLOOP READ ERROR:AT POS 5233960 IN FILE/CDROM/KNOPPIX/KNOPPIX
CLOOP:ERROR-3 UNCOMPRESSING BLOCK 46065536/0/23207/05233960-
I/O ERROR DEV OB:00, SECTOR 17212 LINUXRC CANNOT
CREAT/VAN/RUN/VTMP. DIRECTORY NOEXISTENT
```

Os cloop erros podem ser causados por três fatores: a) O arquivo baixado está incompleto ou corrompido. b) O CD ou CD-RW está riscado/danificado e o sistema não está conseguindo ler os dados corretamente (o mais comum). Os CD-RW dão esse tipo de problema mais frequentemente, pois o índice de refração luminosa da mídia é mais baixo e a superfície de gravação é mais frágil, facilitando o aparecimento de problemas de leitura. c) O próprio leitor de CD/DVD ou o cabo IDE podem estar com problemas e por isso os dados não estão sendo lidos corretamente, embora a mídia esteja em bom estado.

Os CLOOP ERRORS são sempre causados direta ou indiretamente por um destes três fatores; problemas com o CD ou com o hardware do seu micro e não por problemas do software. O sistema pode apresentar outros tipos de problemas na sua máquina, como travar durante a detecção de algum componente, mas este erro específico é justamente um aviso de problemas com a leitura dos dados.

Em muitos casos, o sistema roda normalmente a partir do CD, mas apresenta problemas estranhos ao ser instalado (o modo gráfico não abre, alguns aplicativos não funcionam, mensagens de erro diversas reclamando da falta de aplicativos ou bibliotecas, etc.), causados por problemas de leitura durante a instalação ou mesmo devido a badblocks no HD. Prefira usar mídias CDR normais e depois doar suas cópias antigas para amigos que ainda não conheçam o sistema, assim você estará evitando dor de cabeça e ainda fazendo uma boa ação :).

## Requisitos mínimos

Falar em requisitos mínimos é sempre complicado, pois é sempre algo muito relativo. Por exemplo, os requisitos mínimos para rodar o Windows XP, publicados pela Microsoft, falam em um Pentium 233 com 64 MB de RAM (<http://www.microsoft.com/windowsxp/home/evaluation/sysreqs.mspx>) muito embora

o sistema fique extremamente lento e instável nesta configuração. Segundo eles, você só precisaria de um 386 com 4 MB de RAM para rodar o Windows 95! ;) Se fôssemos tomar como parâmetro estas fórmulas da Microsoft, poderia dizer que os requisitos mínimos para rodar o Kurumin seria um Pentium com 32 MB de RAM. Sim, é possível (com um pouco de malabarismo) instalar o sistema nesta configuração e rodar os aplicativos mais leves, porém com um desempenho muito ruim. Entretanto, se por "requisitos mínimos" você entende que estamos falando de uma configuração em que você possa instalar o sistema sem dificuldades e usar quase todos os seus recursos, então diria que o requisito mínimo é um Pentium II 266 (ou K6-2) com 196 MB de RAM. Você pode também usar micros com 128 MB sem muitos problemas, desde que tenha o cuidado de criar uma partição swap no HD. Ela pode ser criada usando o gparted, disponível no Iniciar > Sistema > Particionamento. Quando disponível, a partição swap é usada mesmo ao rodar o sistema a partir do CD.

Mesmo que você não tenha nenhuma distribuição Linux instalada no HD, é recomendável deixar uma partição swap reservada, mesmo que pequena, pois não apenas o Kurumin, mas também outras distribuições que rodam a partir do CD são capazes de detectar e ativar a partição swap no boot e usá-la automaticamente. Com a partição swap, o desempenho do sistema fica melhor, pois ele pode mover arquivos e bibliotecas que não estão em uso, mantendo mais memória livre para os programas. Você pode também criar um arquivo de swap "de emergência" usando as opções disponíveis no "Iniciar > Configuração do Sistema > Memória Swap". A opção para criar o arquivo em uma partição FAT pode ser usada quando você quiser usar uma partição do Windows, disponível no HD (é necessário que a partição esteja formatada em FAT32). Existe também uma opção para usar partições de outras distribuições Linux. Ambas as opções simplesmente criam um arquivo dentro da partição, sem alterar os demais arquivos:



Caso o micro tenha pouca RAM e ainda por cima não possua uma partição swap disponível, o sistema se oferecerá para usar o arquivo de swap do Windows (se o HD estiver formatado em FAT 32). Esta é uma medida desesperada para permitir que o sistema pelo menos consiga dar boot, mas como o arquivo de swap do Windows oferece um desempenho muito inferior ao de uma partição Linux swap, o desempenho do sistema ficará muito abaixo do normal. A memória swap (ou memória virtual) é um recurso usado por todos os sistemas operacionais atuais quando não existe memória RAM suficiente. Ele passa a armazenar os dados que não "cabem" na memória em um arquivo ou partição swap criada no HD. É o uso da memória swap que permite que o sistema continue funcionando, mesmo com pouca memória disponível.

Ao rodar a partir do CD, o Kurumin consome mais memória RAM do que ao ser instalado, pois o sistema precisa reservar parte da memória RAM para criar o ramdisk usado para armazenar os arquivos que precisam ser alterados durante o uso. Outro fator importante é que rodando do CD o desempenho do sistema fica em grande parte limitado ao desempenho do drive de CD-ROM, que é sempre muito mais lento que um HD. Ao dar boot em um micro com 128 MB, o sistema vai exibir uma mensagem de aviso no boot, dizendo que o micro não atende aos requisitos mínimos e avisando que tentará criar um arquivo de swap, usando uma partição disponível no HD. Apesar disso, o boot prossegue, parando apenas em caso de erro ou falta de memória. Ao ser instalado, o sistema consome

em média 60 MB de memória durante o boot. Isso significa que mesmo em um micro com 128 MB você ainda terá mais de 60 MB de memória livre. Com 196 MB de RAM a situação já fica bem mais tranqüila e você consegue rodar até mesmo os aplicativos mais pesados, como o OpenOffice. Evite apenas abrir muitos aplicativos simultaneamente, pois o sistema logo começará a usar memória swap, o que torna tudo muito lento. A configuração ideal, para alguém que usa vários programas ao mesmo tempo e pretende usar o VMware, seria um Athlon de 1.0 GHz com 512 MB de RAM. Hoje em dia, memória RAM é um item relativamente barato, por isso vale a pena usar 512 MB mesmo em micros relativamente antigos. Tenha em mente que na grande maioria dos aplicativos, uma quantidade generosa de memória RAM é mais importante do que um processador mais rápido.

Tendo 512 MB, você pode tranqüilamente abrir 10 abas no firefox, 3 ou 4 documentos grandes no OpenOffice, ler os e-mails no Thunderbird, assistir um DVD e rodar mais um punhado de aplicativos menores, sem que o sistema use uma quantidade significativa de swap. A partir daí, a configuração ideal varia de acordo com o tipo de aplicativos que quiser usar. Se você gosta de rodar vários sistemas simultaneamente usando o VMware, ou trabalha com modelagem 3D ou edição de vídeo, por exemplo, usar um micro mais parrudo, com um processador dual-core e 1 GB de RAM, traria grandes benefícios. Veja uma tabela rápida com relação à quantidade de memória RAM instalada:

<b>1 GB</b>	Ideal para rodar várias instâncias do VMware ou abrir muitos programas pesados simultaneamente. Recomendado para uso profissional, ou para quem faz questão do melhor desempenho possível.
<b>512 MB</b>	Ideal para navegar, rodar o OpenOffice, jogar e usar o VMware, desde que apenas uma máquina virtual por vez.
<b>256 MB</b>	Configuração mínima para rodar o sistema confortavelmente a partir do CD. Não é recomendado tentar usar o VMware.
<b>196 MB</b>	Configuração mínima recomendada. Não tente instalar programas com o sistema rodando do CD e instale no HD assim que possível, de forma a reduzir o uso de memória.
<b>128 MB</b>	Pesadas limitações ao rodar a partir do CD, mas o sistema pode ser usado de forma quase normal depois de instalado. Grande uso de swap ao abrir programas pesados como o Firefox ou OpenOffice.
<b>96 MB</b>	É possível instalar usando as dicas do capítulo 4 e usar o micro para rodar aplicativos leves.
<b>64 MB</b>	É possível instalar, mas o sistema roda com um desempenho muito ruim. Só aplicativos leves e, mesmo assim, com grande uso de memória swap. Seria recomendável instalar mais memória, ou usar o micro como um terminal LTSP.
<b>32 MB</b>	Embora até seja possível instalar o sistema (e usá-lo com pesadas limitações), seria mais recomendável transformar o micro em um terminal LTSP.

## Como o Kurumin funciona

Tradicionalmente, qualquer sistema operacional precisa ser instalado no HD antes de ser usado. Você dá boot usando o CD ou DVD de instalação e é aberto um sistema compacto, que roda o instalador e se encarrega de instalar e configurar o sistema principal. Depois de algum tempo respondendo perguntas e vendo a barra de progresso da cópia dos arquivos, você reinicia o micro e pode finalmente começar a usar o sistema. Isso é válido tanto para o Windows quanto para a maior parte das distribuições Linux. Os live-CDs são distribuições Linux que rodam diretamente a partir do CD-ROM, sem necessidade de instalar. Um dos pioneiros nesta área é o Knoppix (derivado do Debian), que até hoje é um dos live-CDs de maior sucesso. O Kurumin é um descendente direto dele, desenvolvido com base no Knoppix 3.1 (a versão mais recente no início de 2003) e desenvolvido de forma mais ou menos autônoma a partir daí, utilizando como base os pacotes do Debian, combinados com atualizações provenientes do Knoppix, Kanotix e de várias outras distribuições, além de um conjunto próprio de scripts e ferramentas de configuração, centralizadas na forma do Clica-aki:



Para quem já se acostumou com a idéia, pode parecer natural rodar o sistema a partir do CD e até mesmo instalar novos programas sem precisar modificar as informações salvas no HD, mas na época o aparecimento do Knoppix foi considerado um verdadeiro marco dentro do mundo Linux. A base de tudo é um módulo de Kernel chamado cloop, um hack que permite que o sistema rode a partir de um sistema de arquivos compactado, gravado no CD-ROM. Os dados são descompactados "on-the-fly", conforme são necessários. É algo que funciona de forma similar às partições compactadas pelo smartdrive no Windows 95/98 (ainda lembrado pelos saudosistas :), mas com um desempenho melhor e otimizado para CD-ROMs.

Graças à compressão, uma distribuição como o Kurumin pode incluir cerca de 1.6 GB de programas em uma imagem ISO com menos de 600 MB. Além de reduzir o tamanho do sistema, comprimir os arquivos melhora também a taxa de transferência do CD-ROM, diminuindo a perda de desempenho causada pela baixa taxa de leitura.

A idéia é que um CD-ROM de 40X é capaz de ler a em média 4 MB/s, pois como o CD gira sempre na mesma velocidade, as informações gravadas nas trilhas da parte externa do CD (mais longas) são lidas a mais ou menos o dobro da velocidade das do centro (que são mais curtas). Um CD-ROM de 40x lê a 6 MB/s nas trilhas externas mas a apenas 3 MB/s nas internas. Como o CD-ROM é gravado a partir do centro, na maior parte do tempo ele lê os dados a 3, 4 ou 5 MB/s.

No entanto, ao ler 4 MB/s de dados compactados a uma razão de 3x, ele estará lendo, na prática, a quase 12 MB/s, quase a mesma taxa de transferência de um HD de uma década atrás. Naturalmente ainda existem outros problemas, como o tempo de busca, que é muito mais alto em um CD-ROM, mas o problema principal é bastante amenizado. Se não fosse o sistema de compressão, os live-CDs seriam três vezes maiores e três vezes



mais lentos ao rodar do CD, o que os tornariam sistemas muito menos atrativos. Em contrapartida, por causa da compressão o trabalho do processador passa a ser maior, pois, além de processar os dados referentes aos programas, ele tem que, ao mesmo tempo, descompactar os dados lidos pelo CD-ROM. Por isso, mais do que em outras distribuições, o desempenho (ao rodar do CD) aumenta de acordo com o poder de processamento da máquina. A primeira etapa do boot é a tela de boas-vindas e uma linha onde você pode fornecer parâmetros para o boot. Logo depois é carregado o Kernel, que por sua vez inicializa o hardware, cria um ramdisk usando uma parte (pequena) da memória RAM onde são armazenados arquivos de configuração e outros dados que precisam ser alterados durante o uso. Depois disso entra em ação o hwsetup, o programa de detecção que, junto com um conjunto

de outros scripts, se encarrega de detectar a placa de vídeo, som, rede, modem e outros periféricos suportados. Durante o boot, o sistema exibe vários detalhes sobre os componentes da máquina e mostra como eles estão sendo detectados pelo sistema. Estas mensagens são úteis para identificar a configuração da máquina e já saber de antemão detalhes como o processador, quantidade de memória RAM e placa de vídeo instalada. Imagine o caso de um técnico que instala o sistema em vários micros diferentes, por exemplo:

```


Looking for CDROM in: /dev/scd16
Looking for CDROM in: /dev/hda
Looking for CDROM in: /dev/hdb
Looking for CDROM in: /dev/hdc
Acessando CD do Kurumin em: /dev/hdc...

Setting paths...
Total memory found: 254836 kB

Creating /ramdisk (dynamic size=193876k) on shared memory...

Creating unionfs and symlinks on ramdisk...
>> UnionFS: Uniao do CD/DVD (ro) com o ramdisk (rw) real...
Done.

Starting init process.
INIT: version 2.78-knoppix booting
Usando o Kernel 2.6.11-kanotix-7.
Processador 0 modelo: AMD Sempron(tm) 2800+ 2004MHz, Z
Ativando o suporte a ACPI (economia de energia): ac batt
n processor thermal video
Placa PCMCIA encontrada, iniciando o cardmgr.
controlador USB detectado, ativando o hotplug: (Re-)sc...
me.
Autoconfiguring devices... ████ |

```

Este trabalho de detecção é justamente o grande trunfo. Em poucos segundos o sistema é capaz de detectar, configurar e ativar todos os periféricos suportados na máquina, baseado nos códigos de identificação dos dispositivos, sem nenhuma intervenção do usuário. Claro, sempre existem casos de problemas. Algumas placas-mãe travam durante a detecção do ACPI, alguns notebooks travam durante a etapa inicial, quando o sistema procura por placas SCSI e assim por diante. Mas, na grande maioria dos casos, estes problemas podem ser resolvidos desativando as partes da detecção que causam problemas, usando opções de boot. Durante o boot, o Kurumin tenta sempre configurar automaticamente a rede, obtendo o IP e outros dados a partir de um servidor DHCP disponível na rede. Se a máquina acessar a internet através de uma conexão compartilhada dentro da rede local ou através de um modem ADSL configurado como roteador, ele já será capaz de acessar a web logo após o boot, caso contrário você pode configu-

rar a rede manualmente, usando o painel de controle. Existem inclusive opções para ativar softmodems e placas wireless que não possuem suporte nativo no Linux, neste caso usando o ndiswrapper, que permite ativar a placa usando o driver do Windows XP. Uma questão importante ao usar qualquer live-CD é a questão da memória RAM. Como o sistema por padrão não utiliza as partições do HD, tudo é feito usando a memória RAM, incluindo a instalação de novos programas.

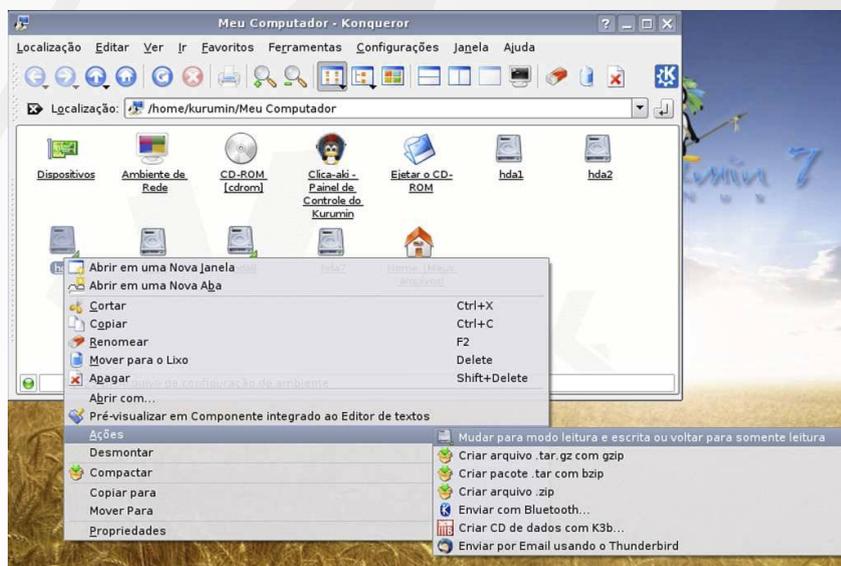
O ramdisk que é criado durante o boot vai crescendo conforme são feitas modificações. Em micros com pouca RAM você verá uma mensagem de "disco cheio" (quando na verdade o que acabou foi o espaço no ramdisk) ou até mesmo efeitos diversos por falta de memória RAM disponível. A instalação de novos programas é possibilitada pelo UnionFS, mais um hack que monitora tentativas de alterações nos arquivos do CD (originalmente impossíveis, já que o CD é somente leitura) e engana os programas, fazendo-as no ramdisk, per-

mitindo que você altere arquivos e instale novos programas, mesmo ao rodar o sistema a partir do CD. Este recurso está disponível a partir do Kurumin 5.0. Você pode rodar o Kurumin tranquilamente em micros com 196 MB de RAM, mas para usar o apt-get e instalar novos programas enquanto o sistema está rodando do CD, você precisa ter 512 MB. Em máquinas mais modestas, é recomendável instalar o sistema assim que possível, já que depois de instalado o sistema fica mais rápido e consome menos memória, por não precisar mais do ramdisk. Outro recurso importante é o acesso às partições do HD, incluindo, naturalmente, o acesso às partições do Windows. Muitas vezes existe a falsa impressão de que os aplicativos Linux utilizam formatos de arquivos próprios e você teria que começar tudo de novo. Isto não poderia estar mais longe da verdade :). Todos os formatos de arquivos padronizados, como imagens em .jpg, .png, .gif, .bmp, arquivos de áudio e vídeo em .mp3, .avi., etc., documentos em vários formatos e

assim por diante são abertos sem problemas dentro do Linux. O sistema já vem configurado para abrir as extensões mais comuns nos programas apropriados quando você clica sobre os arquivos, de forma que acabam não existindo tantas diferenças assim em relação ao Windows. Mesmo os arquivos do Word e Excell são abertos sem grandes problemas no OpenOffice/BrOffice. O maior problema são os formatos proprietários, usados por alguns programas. Arquivos .cdr, por exemplo, podem ser abertos apenas no próprio Corel Draw, mas você tem a opção de exportar seus trabalhos para um formato neutro (como o .svg), de forma que possa abri-los no Inkscape. É por questões como esta que é sempre recomendável manter o Windows em dual-boot nos primeiros meses (ou rodá-lo dentro do VMware), de forma que você possa converter seus documentos quando necessário. Voltando à questão do acesso aos arquivos, o Kurumin detecta as partições disponíveis no HD durante o boot e adiciona ícones para elas dentro do "Meu Computador", que pode ser acessado através do ícone no Desktop:

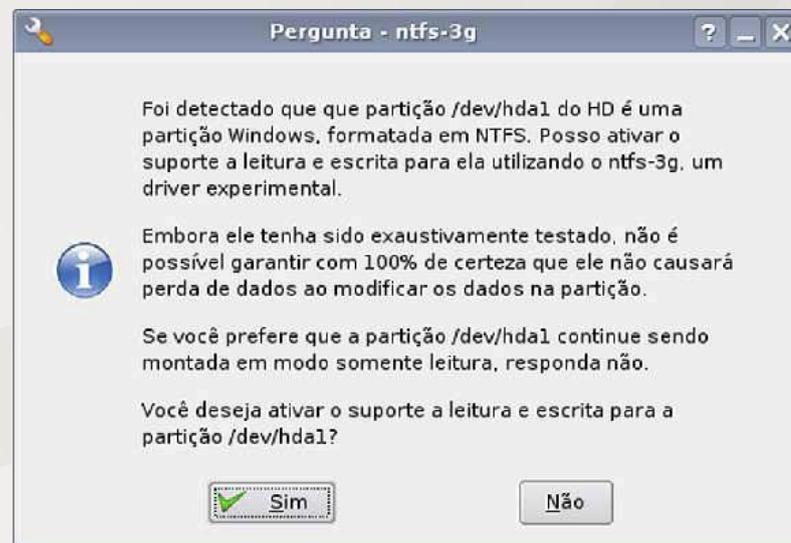


Por padrão, ao clicar sobre o ícone de uma partição, ela é montada em modo somente leitura, onde você consegue acessar os arquivos, mas não consegue fazer alterações. Para mudar o modo de acesso, clique com o botão direito sobre o ícone e acesse a opção "Ações > Mudar para o modo leitura e escrita ou voltar para somente leitura":



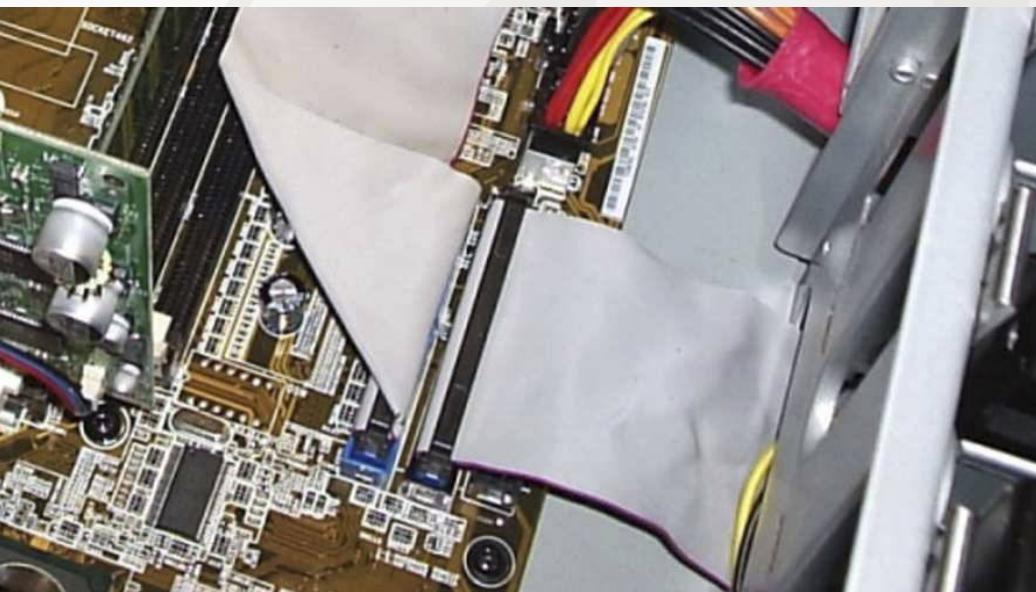
Para que todas as partições sejam acessadas em modo leitura e escrita por padrão, clique no ícone "Montar as partições em leitura e escrita" dentro do "Meu Computador". Ele muda automaticamente o modo de acesso de todos os ícones. Você precisa clicar nele apenas uma vez para que a alteração se torne definitiva. Um problema clássico do Linux era a falta de suporte à escrita em partições do Windows formatadas em NTFS.

Isso foi resolvido no Kurumin 7, com a inclusão do NTFS-3G, um novo driver que derruba esta limitação, oferecendo acesso completo aos arquivos. Ao clicar sobre o "Montar as partições em leitura e escrita", o script detecta que você possui partições NTFS no HD e exibe uma mensagem perguntando se você deseja ativar o NTFS-3G. Como de praxe, é exibido um alerta avisando que existe uma pequena possibilidade de perda de arquivos (afinal, você pode perder arquivos até mesmo utilizando o próprio Windows...) e confirmando se você quer continuar:



Respondendo "Sim", a alteração é feita e você passa a ter acesso completo às partições NTFS do Windows. Muita gente utiliza este recurso para dar manutenção em máquinas Windows quando o sistema (Windows) deixa de dar boot. Afinal, você pode dar boot com o Kurumin, acessar a partição do Windows, modificar ou substituir arquivos danificados, fazer backup dos dados ou até mesmo usar o Clamav para remover os vírus. Aproveitando que estamos aqui, vou falar um pouco mais sobre como o sistema identifica as partições do HD.

Cada dispositivo ou partição é acessado pelo sistema através de um device, um arquivo especial criado dentro do diretório `"/dev"`. Para entender a ordem usada para nomear estes dispositivos é preciso usar algumas noções de hardware. Na placa-mãe você encontra duas portas IDE (primária e secundária), que são usadas para instalar o HD e CD-ROM. Cada uma das duas permite conectar dois dispositivos, de forma que podemos instalar um total de 4 HDs ou CD-ROMs na mesma placa. Os drives IDE "tradicionais", que usam os cabos de 40 ou 80 vias, são chamados de "PATA", contração de "parallel ATA".



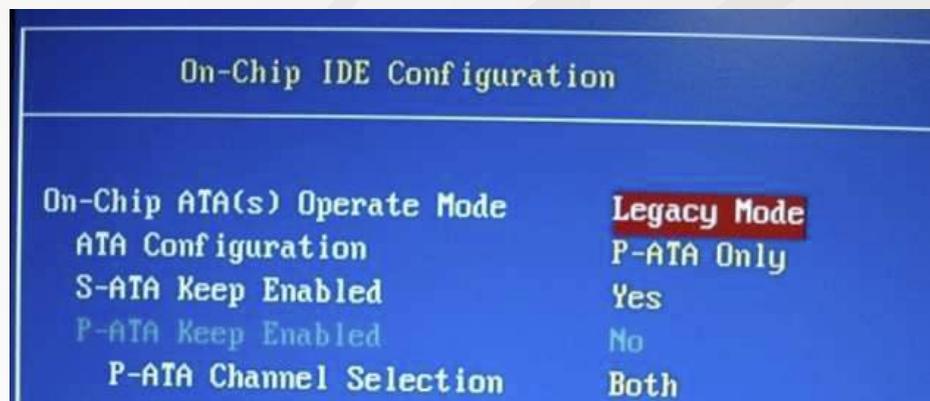
Cada par de drives é instalado na mesma porta. Para diferenciar os dois é usado um jumper, que permite configurar cada drive como master (mestre) ou slave. O mais comum é usarmos apenas um HD e mais um CD-ROM ou DVD, cada um instalado em sua própria porta e ambos configurados como master. Ao adicionar um segundo HD, você poderia escolher entre instalar na primeira ou segunda porta IDE, mas de qualquer forma precisaria configurá-lo como slave, mudando a posição do jumper. Independentemente de ser um HD, CD-ROM ou qualquer outro tipo de dispositivo, os drives são detectados pelo sistema da seguinte forma:

```
IDE primária, master: /dev/hda
IDE primária, slave: /dev/hdb
IDE secundária, master: /dev/hdc
IDE secundária, slave: /dev/hdd
```

Os HDs Serial ATA (SATA) são vistos pelo sistema da mesma forma que HDs SCSI. Isso também se aplica a pendrives e outros dispositivos USB. Aqui entra uma história interessante: como o código é aberto, é muito comum que novos módulos sejam baseados ou utilizem código de outros módulos já existentes. O suporte a drives SCSI no Kernel é tão bom que ele passou a ser usado (com pequenas adaptações) para dar suporte a outros tipos de dispositivos. Na época do Kernel 2.4, até os gravadores de CD eram vistos pelo sistema como drives SCSI.

O primeiro dispositivo SCSI é detectado como `"/dev/sda"`, o segundo como `"/dev/sdb"` e assim por diante. Se você tiver um HD SATA ou pendrive, o drive é visto como `"/dev/sda"` e não como `"/dev/hda"`, como seria se fosse um drive IDE. Se você tiver um HD SATA e um pendrive instalados na mesma máquina, então o HD será visto como `"/dev/sda"` (pois é inicializado primeiro, logo no início do boot) e o pendrive como `"/dev/sdb"`. Se você plugar um segundo pendrive, ele será visto como `"/dev/sdc"` e assim por diante.

Ao contrário dos dispositivos IDE, os devices são definidos seqüencialmente, conforme o sistema vai detectando os dispositivos. Quem chega primeiro leva. Se você tiver um HD IDE e um pendrive, então o HD será visto como `"/dev/hda"` e o pendrive como `"/dev/sda"`. Uma observação é que em muitas placas você encontrará uma opção dentro do Setup que permite colocar as portas SATA em modo de compatibilidade (Legacy Mode ou Compatibility Mode, dependendo da placa). Ao ativar esta opção, seu HD SATA passará a ser visto pelo sistema como `"/dev/hda"`, como se fosse um HD IDE normal. Esta opção é útil ao instalar distribuições antigas, que ainda não oferecem um bom suporte a HDs SATA.

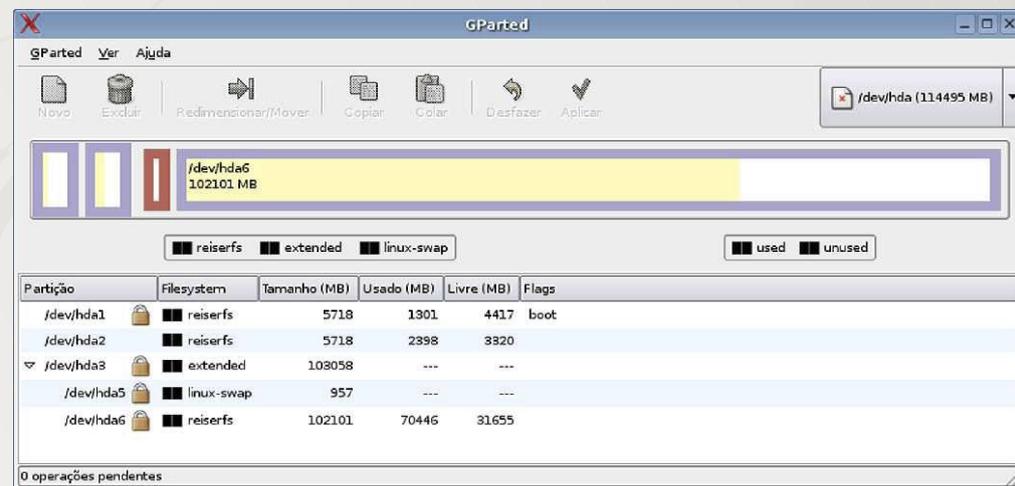


Em seguida vem a questão das partições. Ao invés de ser um espaço único e indivisível, um HD é como uma grande sala comercial, que pode ser dividida em vários escritórios e ambientes diferentes. Ao instalar o sistema operacional, você tem a chance de particionar o HD, onde é feita esta divisão. É sempre recomendável usar pelo menos duas partições separadas, uma para o sistema e outra para seus arquivos. Isto permite reinstalar o sistema sempre que necessário, sem perder seus arquivos e configurações.

No Linux existe ainda a necessidade de criar uma partição separada para a memória swap. Esta partição utiliza uma organização própria, otimizada para a tarefa.

Embora um pouco mais complicada, esta abordagem faz com que o acesso seja mais rápido que no Windows, onde o swap é feito dentro de um arquivo, criado na partição de instalação de sistema.

Existem diversos programas de particionamento, os mais usados no Linux são o cfdisk, o gparted e o qtparted. Muitas distribuições incluem particionadores próprios. O Mandriva, por exemplo, inclui o diskdrake.



Acima temos um screenshot do Gparted. Como pode ver, cada partição recebe um número e é vista pelo sistema como um dispositivo diferente. A primeira partição do `"/dev/hda"` é vista como `"/dev/hda1"` e assim por diante. O mesmo acontece com os pendrives, que do ponto de vista do sistema operacional são uma espécie de HD em miniatura. O sistema nunca acessa os dados dentro da partição diretamente. Ao invés disso, ele permite que você "monte" a partição em uma determinada pasta e acesse os arquivos dentro da partição através dela, o que é feito usando o comando "mount". Por baixo dos panos, é justamente isso que acontece quando você clica sobre os ícones das partições dentro do "Meu Computador".

A sintaxe básica inclui o dispositivo e a pasta onde ele será acessado, como em:

```
# mount /dev/hda2 /mnt/hda2
```

O mais comum é que as partições "extras" sejam montadas dentro da pasta "/mnt", que é própria para a tarefa, mas isso não é uma regra; você pode montar as partições em qualquer pasta vazia. Não se esqueça de criar a pasta desejada (se necessário), usando o comando "mkdir". No caso do CD-ROM, citamos apenas o dispositivo, sem incluir a partição (já que um CD-ROM não pode ser particionado, como um HD). Você pode tanto usar o dispositivo correto, como "/dev/hdc" ou "/dev/hdd", quanto usar o "/dev/cdrom", um link que é criado pelo sistema apontando para a localização correta:

```
# mount /dev/cdrom /mnt/cdrom
```

Se quiser trocar o CD que está na bandeja, você deve primeiro "desmontar" o CD-ROM, com o comando "umount /mnt/cdrom". O mesmo se aplica a pendrives e HDs externos: é sempre necessário desmontar antes de remover o dispositivo.

No caso dos pendrives e HDs, desmontar é fundamental, pois as alterações não são necessariamente salvas imediatamente por causa do cache de disco. Removendo sem desmontar, existe uma probabilidade muito grande das últimas alterações serem perdidas. É muito comum as pessoas gravarem arquivos no pendrive, desplugarem logo depois (sem desmontar) e, ao tentar acessá-los depois, perceberem que os arquivos simplesmente não foram gravados. Um problema comum em relação ao CD-ROM é que o sistema (por segurança) não permite que você ejete o CD enquanto qualquer programa o estiver acessando. Isto é frustrante às vezes, pois muitas vezes você está com muitas janelas abertas e não se lembra qual programa usou para acessá-lo.

O ícone "Ejetar o CD-ROM" dentro do "Meu Computador" é justamente uma forma emergencial de ejetar o CD, "custe o que custar". Ele procura e encerra todos os programas que estiverem acessando o CD-ROM e em seguida desmonta e ejeta o CD.

Os comandos executados por ele, caso queira fazer manualmente, são:

```
# fuser -k /mnt/cdrom
# umount /mnt/cdrom
# eject /mnt/cdrom
```

Se por acaso você tiver um drive de disquetes (em que século você vive? :), o comando para montá-lo manualmente é "mount /dev/fd0 /mnt/floppy" e, paradesmontá-lo, "umount /mnt/floppy". Assim como no caso dos pendrives, é importante desmontar antes de remover o disquete do drive. Os pontos de montagem, ou seja, as pastas onde as partições serão montadas, podem ser configurados através do arquivo "/etc/fstab". Quase sempre este arquivo é configurado durante a instalação, incluindo referências a todas as partições e CD-ROMs disponíveis, de forma que você pode montar as partições digitando apenas "mount /mnt/hda6" (por exemplo), sem precisar usar o comando completo.

Uma observação importante sobre as partições é que, a partir do Kernel 2.6.20, teremos uma grande mudança no conjunto de drivers para por-

tas IDE. Os drivers para HDs IDE serão integrados dentro do mesmo conjunto de drivers que são usados para os HDs SATA. Isso fará com que todos os HDs passem a ser nomeados como "/dev/sda" e "/dev/sdb", como se fossem HDs SATA.

## O UnionFS

Até o Kurumin 4.2, as opções para instalar novos programas e instalar servidores só funcionavam com o Kurumin instalado no HD, por causa da limitação óbvia de não ser possível instalar novos programas com o sistema rodando a partir do CD-ROM, já que ele é somente leitura. No entanto, a partir do Kurumin 5.0, esta última limitação foi derrubada, com a inclusão do UnionFS. Graças a ele, passou a ser possível usar o apt-get e os ícones mágicos para instalar novos programas e mexer em todos os arquivos de configuração do sistema, mesmo

com o sistema rodando do CD. Isto permite testar os recursos do sistema com muito mais liberdade, sem precisar instalar. Você pode dar boot, instalar um servidor Apache e Squid, os drivers da nVidia, o VMware e outros programas que quiser testar e no final reiniciar o micro, deixando tudo como se nada tivesse acontecido. O UnionFS funciona de uma forma bastante engenhosa, uma daquelas idéias aparentemente simples, que resolvem problemas complexos. Com o Kurumin rodando a partir do CD, os arquivos armazenados no diretório home e alguns arquivos de configuração, que precisam ser alterados durante o boot, são armazenados em um ramdisk (um disco virtual, criado usando uma parte da memória RAM); mas, fora isso, tudo é acessado dentro do arquivo /cdrom/KNOPPIX, que, além de fazer parte do CD, está compactado em um formato que não permite alterações, completamente selado. Para permitir esta "mágica", o UnionFS permite juntar dois (ou mais) diretórios em um, estabelecendo uma hierarquia entre eles.

O "Union" vem justamente de "união". Temos então o arquivo compactado do CD em um nível hierárquico mais baixo, montado como somente leitura e um ram-disk, que originalmente está quase vazio, mas que vai armazenando todas as alterações. Os dois são montados em uma única pasta, a /UNIONFS, que contém o conteúdo do arquivo compactado e do ramdisk. Os links que tradicionalmente apontariam para a pasta "/KNOPPIX", onde fica montado o arquivo compactado, são todos recriados apontando para ela. Na hora de ler um arquivo, o sistema verifica se existe uma versão mais recente armazenada no ramdisk, caso contrário lê no arquivo principal. Na hora de gravar, as alterações são sempre armazenadas no ramdisk, de forma automática e transparente. No final, você acaba podendo instalar programas e fazer qualquer tipo de alteração no sistema, da mesma forma que se ele estivesse instalado. As limitações neste caso são que todas as modificações são salvas no ramdisk. Para conseguir instalar programas grandes com o sistema rodando a partir do

CD, você precisa ter 512 MB de RAM. Caso contrário, você pode instalar alguns programas pequenos de cada vez e ir reiniciando o micro para testar outros, conforme a memória for sendo ocupada. O UnionFS é ativado por padrão durante o boot, quando é exibida a mensagem "UnionFS: União do CD/DVD (ro) com o ramdisk (rw) realizada com sucesso". Não é preciso usar nenhuma opção de boot, basta atualizar a lista do apt-get, executando o comando "sudo apt-get update" e começar a instalar programas.

```
Creating unionfs and symlinks on ramdisk...
>> UnionFS: Uniao do CD/DVD (ro) com o ramdisk (rw) realizada com sucesso.
Done.
```

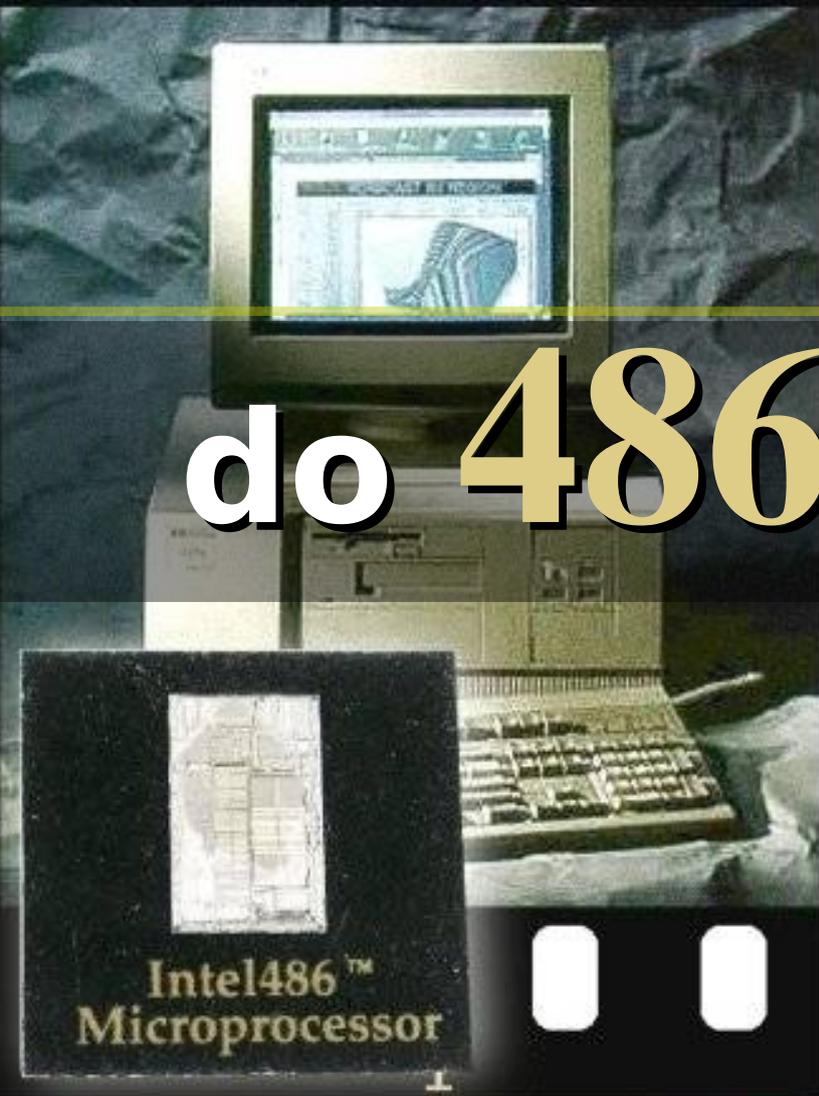


Carlos E. Morimoto.

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.

# do 486 ao Athlon

por Carlos E. Morimoto



*A poucos meses publiquei um guia sobre a história da informática, que começava com o Eniac e terminava com o 386. Esta é uma continuação, que conta a história da plataforma PC, do 486 ao Athlon.*

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.

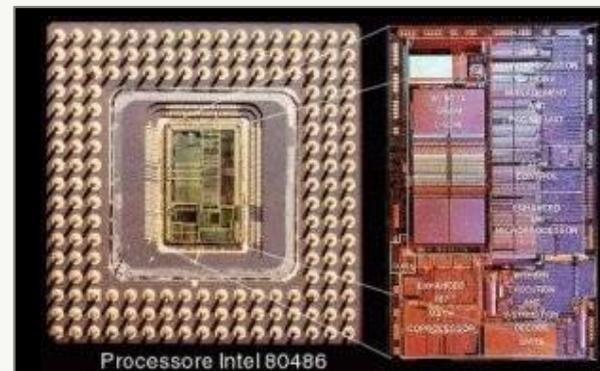
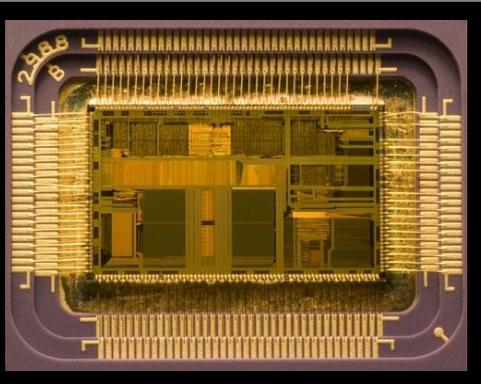
Este brutal aumento de complexidade pode ser justificado por três inovações introduzidas pelo 486. Em primeiro lugar, ele usa um co-processador aritmético integrado, ao invés de um chip separado, como no 386. Em segundo, ele incorpora 8 KB de cache ultrarrápido diretamente no processador, complementando 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.

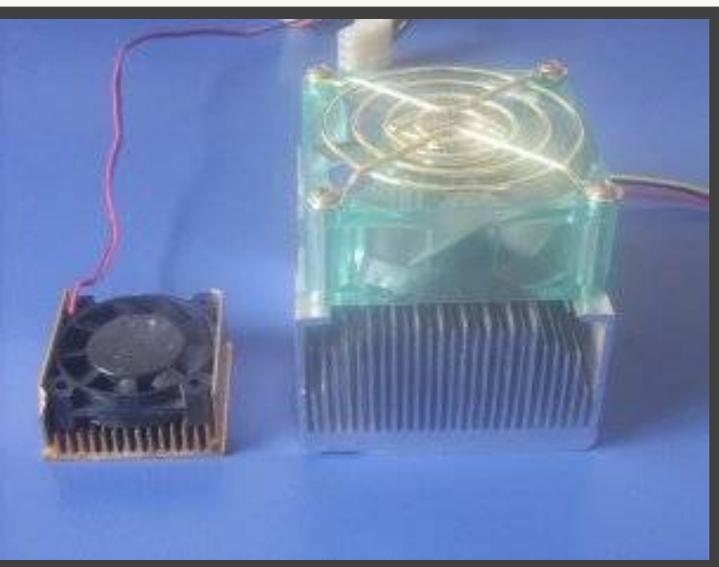
Voltando ao 486, foi introduzido também o processamento de instruções em etapas, recurso que é utilizado até os dias de hoje. A unidade de execução do 486 é composta por um pipeline de 5 estágios, que possuem funções distintas. Cada instrução passa sucessivamente por cada um dos 5 estágios, sendo que 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 fluido, 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 parudo), 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 sobrevivência 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.

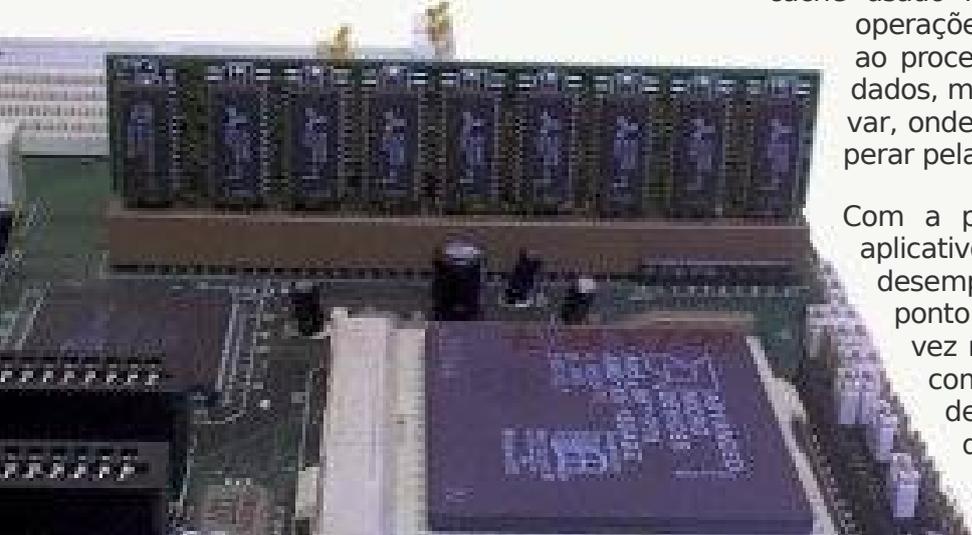
Apesar do clock de 133 MHz, o 5x86 concorria com o Pentium 75 em termos de desempenho. Era comum overclocar estes processadores para 160 MHz (aumentando a frequência da placa mãe de 33 para 40 MHz), resultando num desempenho próximo ao de um Pentium 90.

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 usado.

Embora 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.

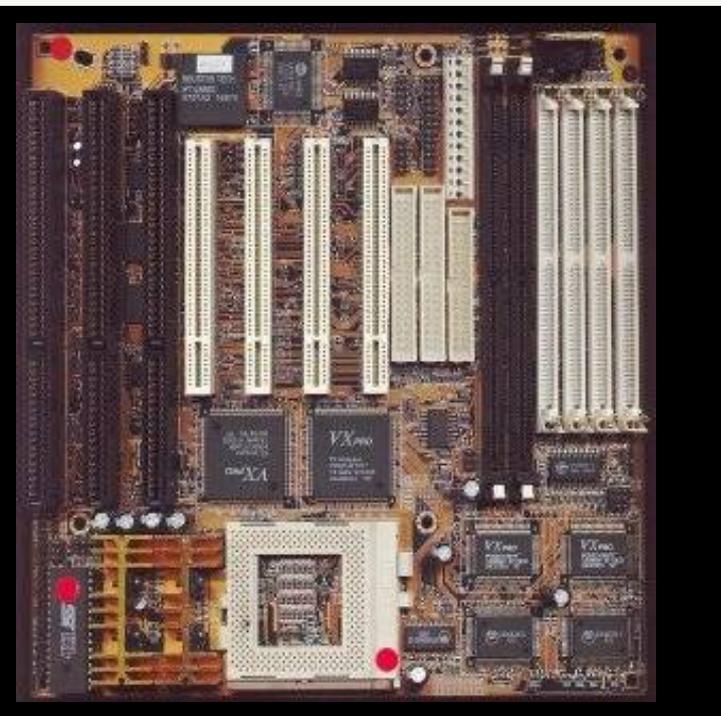
Para não depender apenas das novas instruções, a Intel aumentou o cache L1 do processador, de 16 para 32 KB. Com isto, o MMX passou a ser um pouco mais rápido do que um Pentium 1 da mesma frequência, mesmo nos aplicativos sem otimização. Lembre-se de que naquela época o cache L2 do processador ainda 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:



Um das prioridades da AMD foi manter a compatibilidade com as placas soquete 7 antigas. Por isso, optaram por vender o K6-3 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

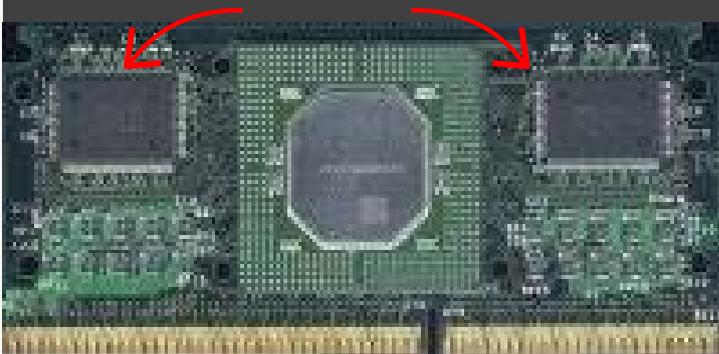
encapsulamento de cerâmica, temos uma placa de circuito, que traz o processador e o cache L2 integrado. Protegendo esta placa, temos uma capa plástica, formando um cartucho muito parecido com um cartucho de video-game, onde o cooler era instalado na parte de trás. O encaixe para ele foi batizado de slot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

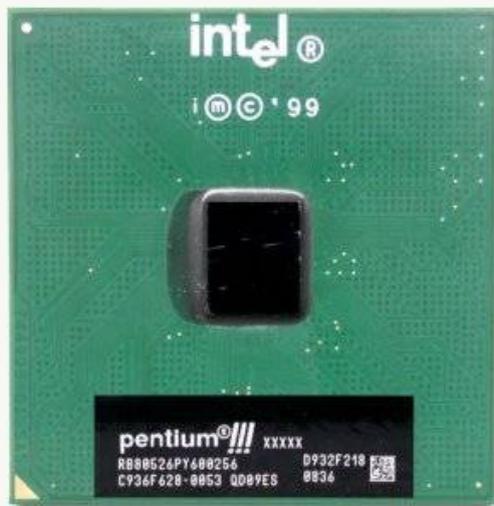
Não poderia ser mais simples: bastava investir num cooler de boa qualidade e instalar o Celeron Mendocino numa placa mãe configurada para operar a 100 MHz.

O Celerom 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.

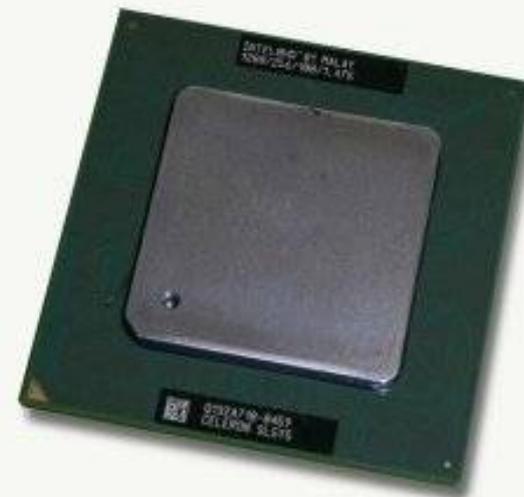
Quando ocorre um defeito no cache, em geral apenas alguns poucos bits são afetados, geralmente bits fisicamente próximos. Antes de saírem de fábrica, todos os processadores são rigorosamente testados, e os que apresentam defeitos no cache são separados. O Pentium III foi projetado de tal maneira que 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:**

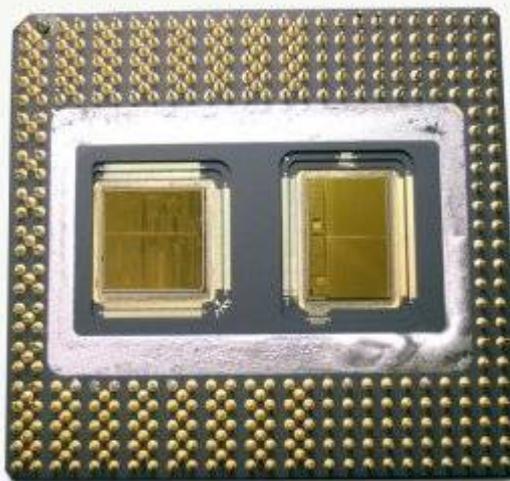


Uma questão interessante é que os processadores Pentium II e Pentium III são derivados do **Pentium Pro**, um processador destinado a workstations e servidores, lançado pela Intel em 1995, em versões de 150 a 200 MHz. Embora seja contemporâneo do Pentium I, o Pentium Pro era baseado numa arquitetura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

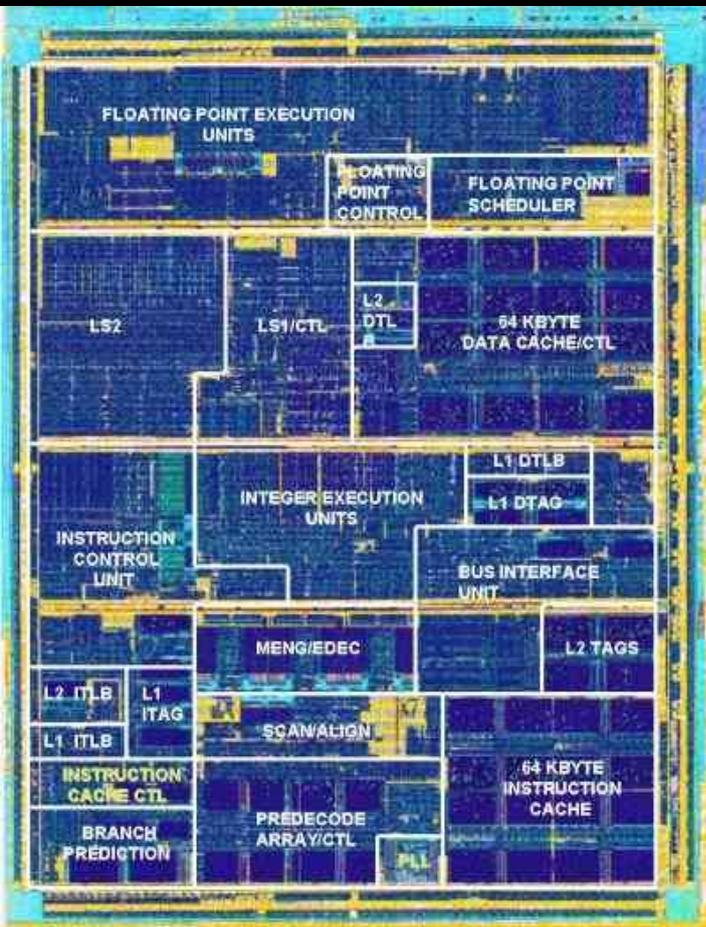
As versões iniciais do Athlon Thunderbird utilizavam barramento de 100 ou 133 MHz, assim como o Pentium III. Entretanto, as versões seguintes (com core Thoroughbred e Barton) passaram a utilizar bus de 166 ou 200 MHz, memórias DDR e tensões mais baixas, quebrando a compatibilidade com as placas antigas.

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.

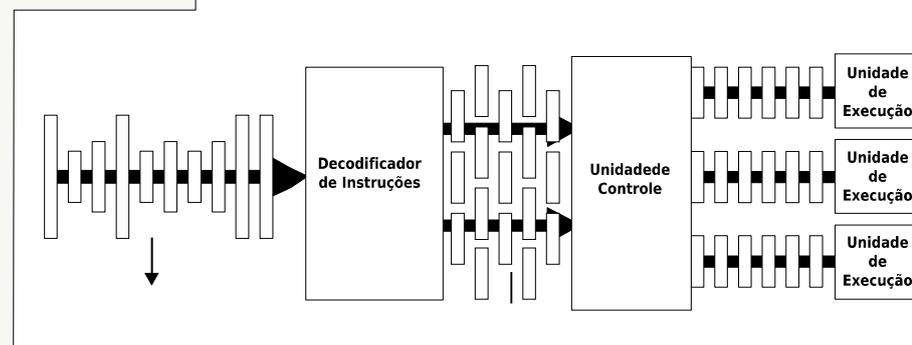
Uma operação de divisão é conseguida executando-se uma seqüência de operações de subtração e todos os demais cálculos, mesmo os cálculos mais complexos, executados pelo co-processador aritmético, são resolvidos usando apenas as quatro operações, obtidas a partir das simples instruções de soma e atribuição.

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.*



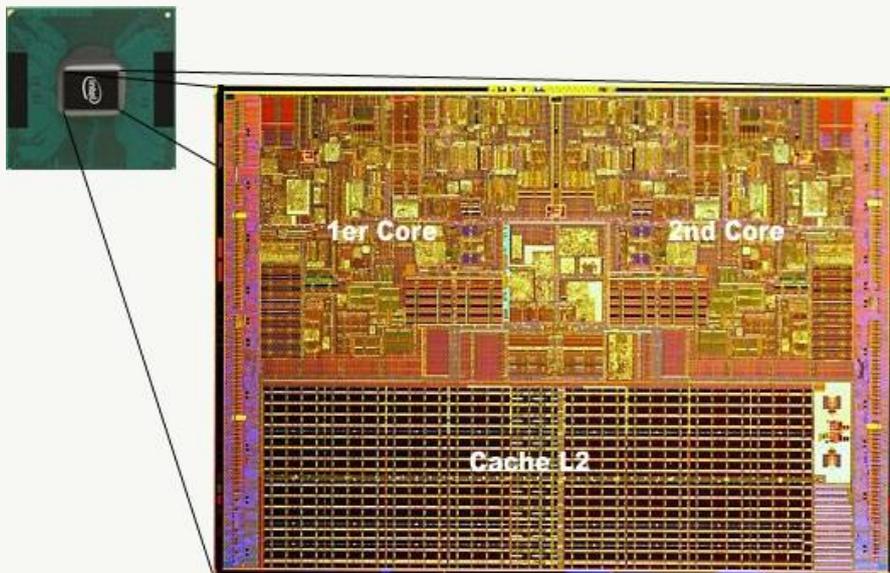
A lógica é que quanto mais unidades de execução tivermos trabalhando ao mesmo tempo, mais instruções todas juntas serão capazes de processar e quanto mais circuitos de decodificação e controle tivermos, mais eficiente será a decodificação das instruções, resultando em um processador mais rápido.

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 deci-

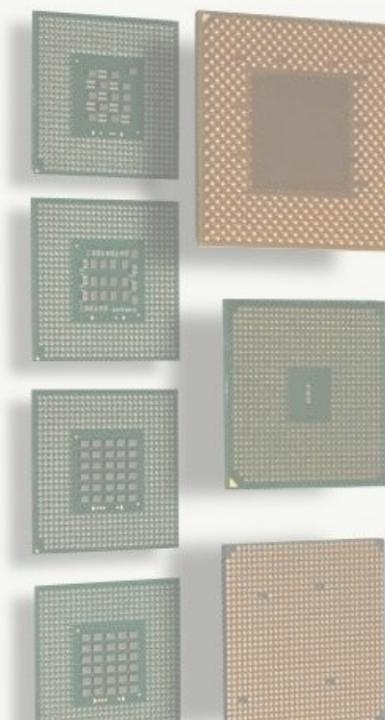
dir o que vai fazer depois. Graças aos circuitos de branch-prediction, os processadores são capazes de ir "adiantando o serviço", processando outras instruções mais adiante, enquanto a operação de tomada de decisão é solucionada. De qualquer forma, existe um limite para quanto trabalho o processador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

# Memória RAM

por **Carlos E. Morimoto**



A memória RAM é um componente essencial não apenas nos PCs, mas em qualquer tipo de computador. É necessária sempre uma certa quantidade de memória, usada para armazenar programas e dados que estão sendo processados.

O mais comum é que o processador carregue o programa a partir do HD ou outra unidade de armazenamento (um CD-ROM ou um pendrive, por exemplo), copie-o para a memória RAM, juntamente com outros dados necessários, e depois vá salvando os arquivos de volta no HD, conforme eles forem sendo modificados.

A memória RAM é bastante rápida e oferece tempos de acesso brutalmente mais baixos que o HD, mas possui a desvantagem de perder os dados armazenados quando o micro é desligado, daí a necessidade de salvar os arquivos periodicamente.

É também por causa disso que o processo de boot é refeito cada vez que você liga o micro. Durante o boot, o sistema operacional, drivers, bibliotecas e aplicativos são novamente copiados para a memória, junto com suas configurações e preferências.

A única forma de evitar repetir o demorado processo de boot é manter a memória RAM ativa, ou salvar seu conteúdo no HD, recuperando-o no próximo boot. Estas são as estratégias usadas pelas opções de suspender e hibernar, disponíveis tanto no Windows, quanto em várias distribuições Linux.

Ao suspender, a maioria dos componentes do sistema são desligados, incluindo o HD, placa de vídeo e a maior parte dos componentes da placa mãe. Mesmo o processador fica num estado latente, onde opera numa frequência muito baixa e mantém apenas os caches e alguns componentes essenciais ativos. Praticamente, os únicos componentes que continuam realmente

ativo durante o estado de hibernação são os pentes de memória; graças a isso o PC acaba consumindo menos de 20 watts de energia e pode voltar ao estágio original muito rapidamente.

Ao hibernar, o conteúdo da memória RAM é copiado para uma área reservada do HD e o micro é realmente desligado. Ao ligar novamente, o conteúdo da memória é restaurado e novamente temos o sistema de volta sem precisar passar pelo processo normal de boot. O problema da hibernação é que a restauração demora muito mais tempo, já que é necessário ler 512 MB, 1 GB ou mesmo 4 GB de dados (equivalentes à quantidade de memória RAM instalada) a partir do HD, o que muitas vezes demora mais do que um boot completo :).

Além dos diferentes tipos de memória RAM, existem também outras tecnologias de memórias de acesso aleatório, como as SRAM e mais recentemente as MRAM. Temos ainda as onipresentes memórias flash, que concorrem com os HDs como mídia de armazenamento, como veremos em detalhes mais adiante.

O tipo mais comum de memória RAM, aquela que compramos no forma de pentes e instalamos na placa mãe, é chamada de DRAM, ou "dynamic RAM".

A memória DRAM passou a ser usada apenas a partir do final da década de 70, substituindo os chips de memória SRAM, que eram muito mais caros. Com o passar do tempo, as memórias DRAM viraram o padrão, de forma que geralmente dizemos apenas "memória RAM" e não "memória DRAM" ;).

### *Num chip de memória DRAM, cada bit é formado pelo conjunto de um transistor e um capacitor.*

O transistor controla a passagem da corrente elétrica, enquanto o capacitor a armazena por um curto período. Quando o capacitor contém um impulso elétrico, temos um bit 1 e quando ele está descarregado, temos um bit 0.

Quando falo em "capacitor", tenha em mente que não estamos falando em nada similar aos capacitores eletrolíticos da placa mãe. Os "capacitores" usados nos chips de memória são extremamente pequenos e simples, basicamente dois pequenos blocos de metal ligados ao transistor, que conservam o impulso elétrico por apenas uma fração de segundo.

Para evitar a perda dos dados, a placa mãe inclui um circuito de refresh, que é responsável por regravar o conteúdo da memória várias vezes por segundo (a cada 64 milissegundos ou menos), algo similar ao que temos num monitor CRT, onde o canhão de elétrons do monitor precisa atualizar a imagem várias vezes por segundo para evitar que as células de fósforo percam seu brilho.

O processo de refresh atrapalha duplamente, pois consome energia (que acaba sendo transformada em calor, contribuindo para o aquecimento do micro) e torna o acesso à memória mais lento. Pesar disso, não existe muito o que fazer, pois a única solução seria passar a usar memória SRAM, que é absurdamente mais cara :).

A principal diferença é que na memória SRAM cada célula é formada por 4 ou 6 transistores, ao invés de apenas um. Dois deles controlam a leitura e gravação de dados, enquanto os demais formam a célula que armazena o impulso elétrico (a célula continua armazenando um único bit). As memórias SRAM são muito mais rápidas e não precisam de refresh, o que faz com que também consumam pouca energia. Além de ser usada como memória cache, a memória SRAM é muito usada em palmtops e celulares, onde o consumo elétrico é uma questão crítica.

Seria perfeitamente possível construir um PC que usasse memória SRAM como memória principal, mas o custo seria proibitivo. Foi por causa do custo que as memórias DRAM passaram a ser utilizadas em primeiro lugar.

Mesmo utilizando um único transistor por bit, os pentes de memória RAM são formados por um número assustador deles, muito mais que os processadores e outros componentes. Um pente de memória de 1 GB, por exemplo, é formado geralmente por 8 chips de 1 gigabit cada um (8 gigabits = 1 gigabyte). Cada chip possui nada menos do que 1 bilhão de transistores e capacitores e o módulo inteiro acumula um total de 8 bilhões de conjuntos.

Apesar dessa brutal quantidade de transistores, os chips de memória são relativamente simples de se produzir, já que basta repetir a mesma estrutura indefinidamente. É muito diferente de um processador, que além de ser muito mais complexo, precisa ser capaz de operar a frequências muito mais altas.

Com a evolução nas técnicas de fabricação, os pentes de memória foram ficando cada vez mais baratos com o passar das décadas. Na época dos micros 486, chegava-se a pagar 40 dólares por megabyte de memória, valor que hoje em dia compra um pente de 512 MB (ou até mais).

O problema é que os requisitos dos sistemas operacionais e aplicativos também aumentaram, quase que na mesma proporção. Enquanto o MS-DOS rodava bem com 2 ou 4 MB de memória, o Windows 95 já precisava de pelo menos 16 MB. O Windows XP (assim como a maioria das distribuições Linux atuais) não roda bem com menos de 256 MB, enquanto no Vista o ideal é usar 1 GB ou mais.

Na maioria das situações, ter uma quantidade suficiente de memória RAM instalada é mais importante que o desempenho do processador, pois sem memória RAM suficiente o sistema passa a utilizar memória swap, que é absurdamente mais lenta.

Enquanto uma seqüência de 4 leituras num módulo de memória DDR2-800 demora cerca de 35 bilionésimos de segundo, enquanto que um acesso a um setor qualquer do HD demora pelo menos 10 milésimos. A taxa de transferência nominal do mesmo módulo de memória é de 6.4 GB/s, enquanto mesmo um HD rápido, de 7200 RPM tem dificuldades para superar a marca de 60 MB/s, mesmo lendo setores seqüenciais. Ou seja, a memória RAM neste caso possui um tempo de acesso quase 300.000 vezes menor e uma taxa de transferência contínua mais de 100 vezes maior que o HD.

Se lembramos que a memória RAM já é muito mais lenta que o processador (justamente por isso temos os caches L1 e L2), fica fácil perceber o quanto o uso de memória swap por falta de memória RAM física pode prejudicar o desempenho do sistema.

É fácil monitorar o uso de swap. No Windows XP ou Vista basta pressionar Ctrl+Alt+Del e acessar o gerenciador de tarefas, enquanto no Linux você pode usar o comando "free" ou um aplicativo de gerenciamento, como o ksysguard.

No caso do Windows Vista é possível usar um pendrive como memória adicional, através do ReadyBoost. Este recurso pode ajudar em micros com pouca memória RAM, pois o pendrive oferece tempos de acesso mais baixos que o HD (embora a taxa de transferência continue baixa) e também reduzir o tempo de carregamento dos programas. É uma opção para casos em que você já tem o pendrive e procura um uso para ele, mas não espere milagres. Em se tratando de memória, não existe o que inventar: ou você procura um sistema operacional e programas mais leves, ou compra mais memória. Não dá para ficar em cima do muro ;).

Como disse a pouco, embora seja brutalmente mais rápida que o HD e outros periféricos, a memória RAM continua sendo muito mais lenta que o processador. O uso de caches diminui a perda de desempenho, reduzindo o número de acessos à memória; mas, quando o processador não encontra a informação de que precisa nos caches, precisa recorrer a um doloroso acesso à memória principal, que pode demorar o equivalente a mais de 100 ciclos do processador.

Para reduzir a diferença (ou pelo menos tentar impedir que ela aumente ainda mais), os fabricantes de memória passaram a desenvolver um conjunto de novas tecnologias a fim de otimizar o acesso aos dados. Acompanhando estas mudanças, tivemos mudanças físicas no formato dos módulos, de forma que podemos classificar os módulos de memória de duas formas:

- **Quanto ao formato usado** (SIMM, DIMM, etc.)

- **Quanto à tecnologia usada** (EDO, SDRAM, DDR, DDR2, etc.)

## Os Formatos

Nos micros XT, 286 e nos primeiros 386, ainda não eram utilizados chips de memória. Ao invés disso, os chips de memória eram instalados diretamente na placa mãe, encaixados individualmente em colunas de soquetes (ou soldados), onde cada coluna formava um banco de memória.

Este era um sistema antiquado, que trazia várias desvantagens, por dificultar upgrades de memória ou a substituição de módulos com defeito. Imagine você, fazendo um upgrade de memória numa placa como esta:

Não é só você que não achou muito atraente a idéia de ficar catando chips de memória um a um. Foi questão de tempo até que alguém aparecesse com uma alternativa mais prática, capaz de tornar a instalação fácil até mesmo para usuários inexperientes.

Os módulos de memória são pequenas placas de circuito onde os chips DIP são soldados, facilitando o manuseio e instalação.

Os primeiros módulos de memória criados são chamados de módulos SIMM, sigla que significa "Single In Line Memory Module", justamente por que existe uma única via de contatos, com 30 vias.

*Formato de memória usada em micros mais antigos*



Apesar de existirem contatos também na parte de trás do módulo, eles servem apenas como uma extensão dos contatos frontais, de forma a aumentar a área de contato com o soquete. Examinando o módulo, você verá um pequeno orifício em cada contato, que serve justamente para unificar os dois lados.



*Sim de 30 vias*

Estes módulos de 30 vias possuíam sempre 8 ou 9 chips de memória. Cada chip fornecia um único bit de dados em cada transferência, de forma que 8 deles formavam um módulo capaz de transferir 8 bits por ciclo. No caso dos módulos com 9 chips, o último era destinado a armazenar os bits de paridade, que melhoravam a confiabilidade, permitindo identificar erros.

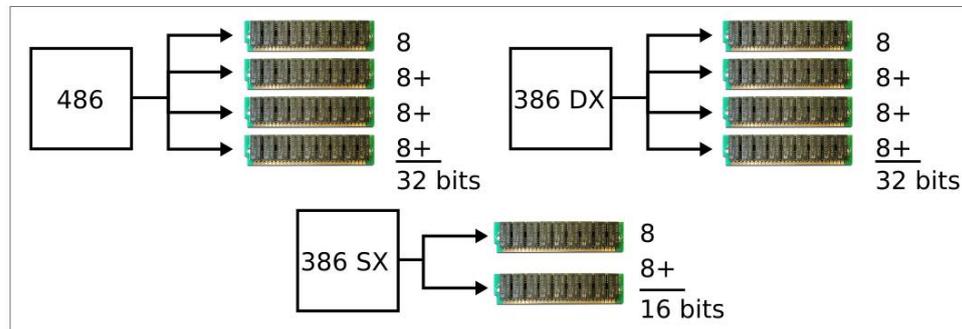
Hoje em dia os módulos de memória são mais confiáveis, de forma que a paridade não é mais usada. No lugar dela, temos o ECC, um sistema mais avançado, usado em pentes de memória destinados a servidores.

Os módulos de 30 vias foram utilizados em micros 386 e 486 e foram fabricados em varias capacidades. Os mais comuns foram os módulos de 1 MB, mas era possível encontrar também módulos de 512 KB, 2 MB e 4 MB. Existiram também módulos de 8 e 16 MB, mas eles eram muito raros devido ao custo.

Os processadores 386 e 486 utilizavam um barramento de 32 bits para o acesso à memória, era necessário combinar 4 pentes de 30 vias para formar um banco de memória.

Os 4 pentes eram então acessados pelo processador como se fossem um só. Era preciso usar os módulos em quartetos: 4 módulos ou 8 módulos, mas nunca um número quebrado.

A exceção ficava por conta dos micros equipados com processadores 386SX, onde são necessários apenas 2 módulos, já que o 386SX acessa a memória usando palavras de 16 bits.



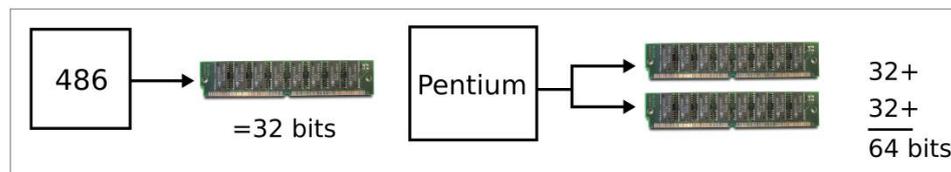
Apesar de serem muito mais práticos do que os chips DIP, os módulos SIMM de 30 vias ainda eram bastante inconvenientes, já que era preciso usar 4 módulos idênticos para formar cada banco de memória. Eles foram desenvolvidos pensando mais na questão da simplicidade e economia de custos do que na praticidade.

Para solucionar o problema, os fabricantes criaram um novo tipo de módulo de memória SIMM, de 32 bits, que possui 72 vias. Os módulos de 72 vias substituíram rapidamente os antigos nas placas para 486 e se tornaram o padrão nos micros Pentium, sendo em seguida substituídos pelos módulos de 168 vias.



*Sim de 72 vias*

Ao invés de quatro módulos, é preciso apenas um módulo SIMM de 72 vias para formar cada banco de memória nos micros 486. Como o Pentium acessa a memória usando palavras de 64 bits, são necessários 2 módulos em cada banco. É por isso que nos micros Pentium precisamos sempre usar os pentes de memória em pares:



O acesso de 64 bits à memória foi introduzido para permitir que o processador conseguisse acessar grandes quantidades de dados mais rapidamente. O processador é tão mais rápido que a memória RAM, que depois de esperar vários ciclos para poder acessá-la, o melhor a fazer é pegar a maior quantidade de dados possível e guardar tudo no cache. Naturalmente os dados serão processados em blocos de 32 bits, mas a poupança ajuda bastante.

Dentro de um banco, todos os módulos são acessados ao mesmo tempo, como se fossem um só, por isso era sempre recomendável usar dois pentes iguais. Ao usar quatro pentes, o importante era que cada par fosse composto por dois pentes iguais. Não existia problema em usar dois pares de pentes diferentes, como ao usar dois pentes de 16 MB e mais dois de 8 MB para totalizar 48 MB, por exemplo.

Uma curiosidade é que algumas placas mãe para Pentium, podem trabalhar com apenas um módulo de 72 vias. Neste caso, a placa engana o processador, fazendo dois acessos de 32 bits consecutivos, e entregando os dados de uma só vez para o processador. Apesar de funcionar, este esquema reduz bastante a velocidade do micro, pois a velocidade de acesso à memória fica reduzida à metade.

Finalmente, temos os módulos DIMM, usados atualmente. Ao contrário dos módulos SIMM de 30 e 72 vias, os módulos DIMM possuem contatos em ambos os lados do módulo, o que justifica seu nome, “Double In Line Memory Module” ou “módulo de memória com duas linhas de contato”.

Todos os módulos DIMM são módulos de 64 bits, o que eliminou a necessidade de usar 2 ou 4 módulos para formar um banco de memória. Muitas placas mãe oferecem a opção de usar dois módulos (acessados simultaneamente) para melhorar a velocidade de acesso. Este recurso é chamado de dual-channel e melhora consideravelmente o desempenho, sobretudo nas placas mãe com vídeo on-board, onde a placa de vídeo disputa o acesso à memória RAM com o processador principal. De qualquer forma, mesmo nas placas dual-channel, usar os módulos em pares é opcional; você pode perfeitamente usar um ou três módulos se preferir.

Existem três formatos de memória DIMM. Os mais antigos são os módulos de memória SDR, com de 168 vias, que eram utilizados a até poucos anos atrás. Em seguida, temos os pentes de memória DDR, que possuem 184 contatos e os módulos DDR2, que possuem 240.

Apesar do maior número de contatos, os módulos DDR e DDR2 são exatamente do mesmo tamanho que os módulos SDR de 168 vias, por isso foram introduzidas mudanças na posição dos chanfros de encaixe, de forma que você não consiga encaixar os módulos em placas incompatíveis.

Os módulos SDR possuem dois chanfros, enquanto os DDR (abaixo) possuem apenas um chanfro, que ainda por cima é colocado numa posição diferente:



*Diferença no chanfro da memória SDR e DDR*

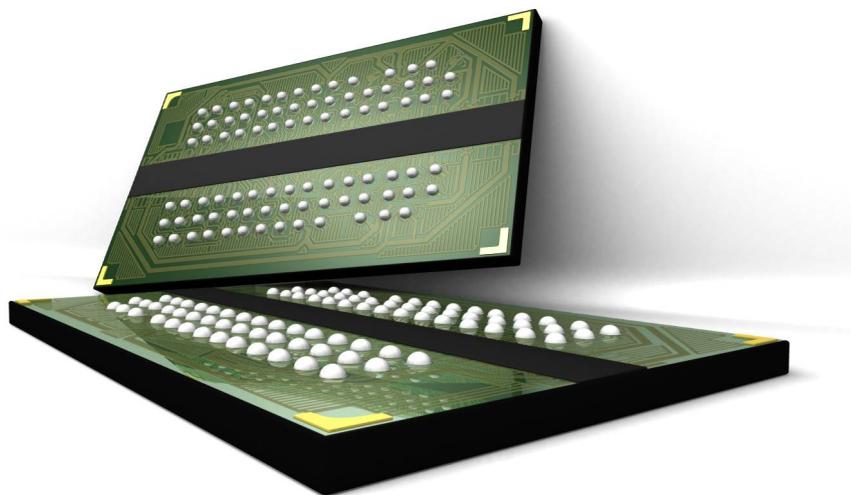
Os módulos DDR 2 também utilizam um único chanfro, mas ele está posicionado mais à esquerda que o usado nos módulos DDR, de forma que é novamente impossível encaixar um módulo DDR2 numa placa antiga:



*Memória DDR2*

Isto é necessário, pois além das mudanças na forma de acesso, os pentes DDR 2 utilizam tensão de 1.8V, enquanto os módulos DDR usam 2.5. Se fosse possível instalar um módulo DDR2 numa placa antiga, a maior tensão queimaria o módulo rapidamente.

Outra diferença é que os chips DDR2 utilizam o encapsulamento BGA (Ball Grid Array), ao invés do encapsulamento TSOP (Thin Small-Outline Package), usado nos chips SDR e DDR. A grande diferença é que no BGA os pontos de solda são posicionados diretamente na parte inferior dos chips, ao invés de serem usadas as "perninhas" laterais. Isso reduz a distância que o sinal elétrico precisa percorrer, além de reduzir o nível de interferências, permitindo que os módulos sejam capazes de operar a frequências mais altas. Esta imagem ilustrativa da Micron mostra bem como os chips se parecem:



Mais recentemente estão surgindo no mercado alguns pentes de memória DDR que também utilizam chips BGA, mas eles são menos comuns.



Outra característica que torna os módulos DDR2 diferentes é a presença de um terminador resistivo dentro de cada chip de memória. O terminador é necessário para "fechar o circuito", evitando que os sinais elétricos retornem na forma de interferência ao chegarem ao final do barramento. Nos módulos DDR os terminadores ao instalados na placa mãe, o que torna a terminação menos eficiente. Como os módulos DDR2 operam a frequências muito mais altas, a presença do terminador dentro dos próprios chips se tornou uma necessidade, já que torna o sinal mais estável e livre de ruídos.

Existem também os módulos SODIMM (Small Outline DIMM), destinados a notebooks. Eles são basicamente versões miniaturizadas dos pentes destinados a desktops, que utilizam os mesmos tipos de chips de memória.

Os módulos SODIMM SDR possuem 144 pinos, enquanto os módulos DDR e DDR2 possuem 200 pinos. Nos pentes SDR o chanfro fica próximo ao centro do módulo, enquanto nos DDR e DDR2 ele fica à esquerda. Assim como nos pentes para desktops, existe uma pequena diferença no posicionamento do chanfro entre os pentes DDR e DDR2, que impede o encaixe incorreto, já que ambos são incompatíveis. Abaixo temos um pente SODIMM DDR2:



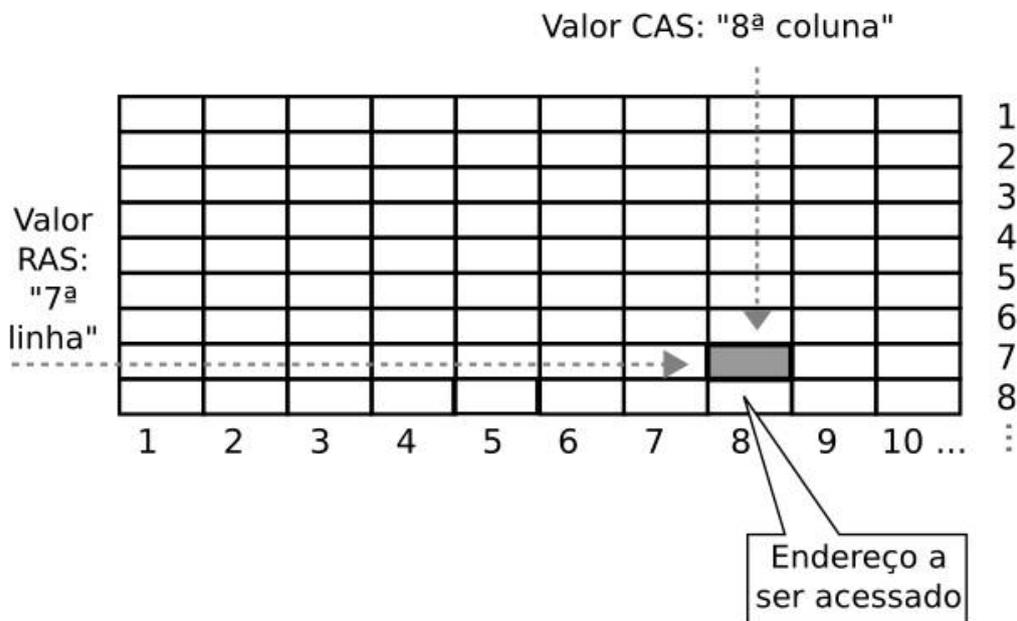
## As Tecnologias

Um chip de memória é um exército de clones, formado por um brutal número de células idênticas, organizadas na forma de linhas e colunas, de uma forma similar a uma planilha eletrônica.

O chip de memória em si serve apenas para armazenar dados, não realiza nenhum tipo de processamento. Por isso, é utilizado um componente adicional, o controlador de memória, que pode ser incluído tanto no chipset da placa mãe, quanto dentro do próprio processador, como no caso do Athlon 64.

Para acessar um determinado endereço de memória, o controlador primeiro gera o valor RAS (Row Address Strobe), ou o número da linha da qual o endereço faz parte, gerando em seguida o valor CAS (Column Address Strobe), que corresponde à coluna.

Quando o RAS é enviado, toda a linha é ativada simultaneamente; depois de um pequeno tempo de espera, o CAS é enviado, fechando o circuito e fazendo com que os dados do endereço selecionado sejam lidos ou gravados:



Não existe um caminho de volta, ligando cada endereço de volta ao controlador de memória. Ao invés disso, é usado um barramento comum, compartilhado por todos os endereços do módulo. O controlador de memória sabe que os dados que está recebendo são os armazenados no endereço X, pois ele se "lembra" que acabou de acessá-lo.

Antigamente (na época dos módulos de 30 vias), cada chip de memória se comportava exatamente desta forma, lendo um bit de cada vez. Apesar disso, o processador lia 32 bits de dados a cada ciclo, de forma que eram usados 4 pentes, com 8 chips cada um.

Do ponto de vista do processador, não existia divisão, os chips eram acessados como se fossem um só. O processador não via 32 endereços separados, em 32 chips diferentes, mas sim um único endereço, contendo 32 bits.

Nos módulos DIMM atuais são geralmente usados 8 chips de 8 bits cada um, formando os 64 bits fornecidos ao processador. Existem ainda módulos com 16 chips de 4 bits cada, ou ainda, módulos com 4 chips de 16 bits (comuns em notebooks). Do ponto de vista do processador, não faz diferença, desde que somados os chips totalizem 64 bits.

Imagine que o controlador de memória envia seqüências com 4, 8 ou 16 pares de endereços RAS e CAS e recebe de volta o mesmo número de pacotes de 64 bits. Mesmo em casos onde o processador precisa de apenas alguns poucos bytes, contendo uma instrução ou bloco de dados, ele precisa ler todo o bloco de 64 bits adjacente, mesmo que seja para descartar os demais.

No caso das placas dual-channel, continuamos tendo acessos de 64 bits, a única diferença é que agora a placa mãe é capaz de acessar dois endereços diferentes (cada um em um pente de memória) a cada ciclo de clock, ao invés de apenas um. Isso permite transferir o dobro de dados por ciclo, fazendo com que o processador precise esperar menos tempo ao transferir grandes quantidades de dados.

Na verdade, nos PCs atuais, praticamente qualquer dispositivo pode acessar a memória diretamente, através dos canais de DMA (Direct Memory Access) disponíveis através do barramento PCI, AGP, PCI Express e até mesmo a partir das portas SATA, IDE e USB. Naturalmente, todos os acessos são coordenados pelo processador, mas como a memória é uma só, temos situações onde o processador precisa esperar para acessar a memória, por que ela está sendo acessada por outro dispositivo.

Existem várias formas de melhorar o desempenho da memória RAM.

O primeiro é aumentar o número de bits lidos por ciclo, tornando o barramento mais largo, como o aumento de 32 para 64 bits introduzida pelo Pentium 1, que continua até os dias de hoje. O problema em usar um barramento mais largo é que o maior número de trilhas necessárias, tanto na placa mãe, quanto nos próprios pentes de memória, aumentam o custo de produção.

A segunda é acessar dois ou mais módulos de memória simultaneamente, como nas placas dual-channel. O problema é que neste caso precisamos de dois pentes, além de circuitos e trilhas adicionais na placa mãe.

A terceira é criar pentes de memória mais rápidos, como no caso das memórias DDR e DDR2. Esta questão da velocidade pode ser dividida em dois quesitos complementares: o número de ciclos por segundo e a latência, que é o tempo que a primeira operação numa série de operações de leitura ou escrita demora para ser concluída. O tempo de latência poderia ser comparado ao tempo de acesso de um HD.

É aqui que entram as diferentes tecnologias de memórias que foram introduzidas ao longo das últimas décadas, começando pelas memórias regulares, usadas nos XTs e 286, que evoluíram para as memórias FPM, usadas em PCs 386 e 486, em seguida para as memórias EDO, usadas nos últimos micros 486s e nos Pentium. Estas três primeiras tecnologias foram então substituídas pelas memórias SDRAM, seguidas pelas memórias DDR e DDR2 usadas atualmente.

## Memórias Regulares

As memórias regulares são o tipo mais primitivo de memória RAM. Nelas, o acesso é feito da forma tradicional, enviando o endereço RAS, depois o CAS e aguardando a leitura dos dados para cada ciclo de leitura.

Isto funcionava bem nos micros XT e 286, onde o clock do processador era muito baixo, de forma que a memória RAM era capaz de funcionar de forma sincronizada com ele. Em um 286 de 8 MHz, eram usados chips com tempo de acesso de 125 ns (nanossegundos) e em um de 12 MHz eram usados chips de 83 ns.

O problema era que a partir daí as memórias da época atingiram seu limite e passou a ser necessário

fazer com que a memória trabalhasse de forma assíncrona, onde o processador trabalha a uma frequência mais alta que a memória RAM.

A partir do 386, a diferença passou a ser muito grande, de forma que as placas mãe passaram a trazer chips de memória cache, dando início à corrida que conhecemos.

## Memórias FPM

A primeira melhora significativa na arquitetura das memórias veio com o FPM (Fast-Page Mode, ou "modo de paginação rápida").

A idéia é que, ao ler um bloco de instruções ou arquivo gravado na memória, os dados estão quase sempre gravados seqüencialmente. Não seria preciso então enviar o endereço RAS e CAS para cada bit a ser lido, mas simplesmente enviar o endereço RAS (linha) uma vez e em seguida enviar uma seqüência de até 4 endereços CAS (coluna), realizando uma série rápida de 4 leituras.

O primeiro ciclo de leitura continua tomando o mesmo tempo, mas as 3 leituras seguintes passam a ser bem mais rápidas. Graças a esta pequena otimização, as memórias FPM conseguem ser até 30% mais rápidas que as memórias regulares, sem que fosse necessário fazer grandes alterações nos chips de memória.

O burst de 4 leituras pode ser prolongado para 8, ou até mesmo 16 leituras consecutivas, desde que lendo dados gravados em endereços adjacentes, da mesma linha.

As memórias FPM foram utilizadas em micros 386, 486 e nos primeiros micros Pentium, na forma de módulos SIMM de 30 ou 72 vias, com tempos de acesso de 80, 70 ou 60 ns, sendo as de 70 ns as mais comuns.

Instaladas em uma placa mãe soquete 7, que trabalhe com bus de 66 MHz, os intervalos de espera de memórias FPM podem ser de até 6-3-3-3, o que significa que o processador terá de esperar cinco ciclos da placa mãe para a memória efetuar a primeira leitura de dados e somente mais 3 ciclos para cada leitura subsequente. Os tempos de espera das memórias podiam ser configurados no setup, através da opção "Memory Timing" ou similar, onde ficavam disponíveis opções como "slow", "normal" e "fast", que substituem os valores numéricos.

No caso das placas para 486, que operavam a clocks mais baixos (30, 33 ou 40 MHz), os tempos de espera podiam ser configurados com valores mais baixos, como 4-3-3-3 ou 3-2-2-2, já que com menos ciclos por segundo, é natural que os tempos de resposta dos módulos correspondam a um número menor de ciclos da placa mãe.

## Memórias EDO

As memórias EDO (Extended Data Output) foram introduzidas a partir de 1994 e trouxeram mais uma melhoria significativa no modo de acesso a dados. Nas memórias FPM, uma leitura não pode ser iniciada antes que a anterior termine, mesmo dentro do burst de 4 leituras dentro da mesma linha. O controlador precisa esperar que os dados referentes à leitura anterior cheguem, antes de poder ativar endereço CAS seguinte.

Nas memórias EDO, o controlador faz a leitura enviando o endereço RAS, como de costume, e depois enviando os 4 endereços CAS numa frequência pré-definida, sem precisar esperar que o acesso anterior termine. Os sinais chegam às células de memória na seqüência em que foram enviados e, depois de um pequeno espaço de tempo, o controlador recebe de volta as 4 leituras.

O resultado acaba sendo exatamente o mesmo, mas passa a ser feito de forma mais rápida. Usadas em uma placa soquete 7, operando a 66 MHz, as memórias EDO são capazes de trabalhar com tempos de acesso de apenas 6-2-2-2, ou mesmo 5-2-2-2 (nos módulos de 60 ns).

Nos bursts de 8 ou mais leituras, o ganho acaba sendo ainda maior, com o módulo FPM realizando a leitura dos 8 endereços em 27 ciclos (6-3-3-3-3-3-3-3) e o EDO em 20 (6-2-2-2-2-2-2-2). Veja que o ganho é maior em leituras de vários endereços consecutivos, por isso alguns aplicativos se beneficiam mais do que outros.

Os chips de memória EDO foram produzidas em versões com tempos de acesso 70, 60 e 50 ns, com predominância dos módulos de 60 ns. Elas foram usadas predominantemente na forma de módulos de 72 vias, usados nos micros 486 e Pentium fabricados a partir do ano de 1995.

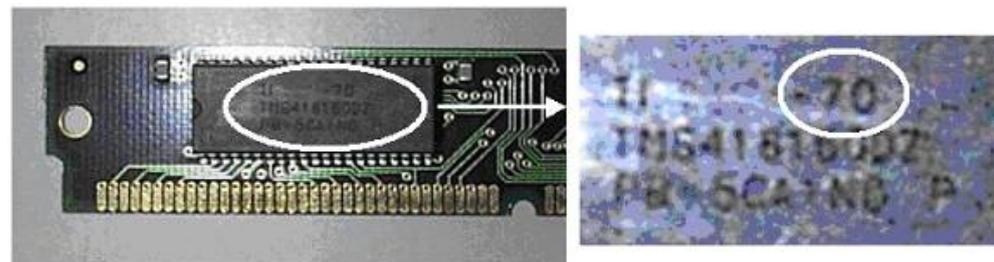
Existiram ainda alguns módulos DIMM de 168 com memória EDO. Eles foram bastante raros, pois foram logo substituídos pelos pentes de memória SDRAM.

As melhorias na arquitetura das memórias EDO tornaram-nas incompatíveis com placas mãe equipadas com chipsets mais antigos. Basicamente, apenas as placas para processadores Pentium e algumas placas mãe para 486 com slots PCI (as mais recentes) aceitam trabalhar com memórias EDO. Existem também placas para 486 "tolerantes" que funcionam com memórias EDO, apesar de não serem capazes de

tirar proveito do modo de acesso mais rápido, e finalmente, as placas incompatíveis, que nem chegam a inicializar caso sejam instaladas memórias EDO.

Todos os módulos de 30 vias são de memórias FPM, enquanto (com exceção de alguns módulos antigos) todos os de 168 vias são de memórias SDRAM. A confusão existe apenas nos módulos de 72 vias, que podem ser tanto de memórias EDO quanto de memórias FPM. Para saber quem é quem, basta verificar o tempo de acesso. Todo módulo de memória traz seus dados estampados nos chips, na forma de alguns códigos; o tempo de acesso é indicado no final da primeira linha. Se ela terminar com -7, -70, ou apenas 7, ou 70, o módulo possui tempo de acesso de 70 ns. Se por outro lado a primeira linha terminar com -6, -60, 6 ou 60 o módulo é de 60 ns.

Como quase todos os módulos de 70 ns são de memórias FPM, e quase todos os módulos de memórias EDO são de 60 ns, você pode usar este método para determinar com 95% de certeza o tipo de memória usada.



## Memórias SDRAM

Tanto as memórias FPM quanto as memórias EDO são assíncronas, isto significa que elas trabalham em seu próprio ritmo, independentemente dos ciclos da placa mãe. Isso explica por que memórias FPM que foram projetadas para funcionar em placas para processadores 386 ou 486,

funcionam sem problemas em placas soquete 7, que trabalham a 66 MHz. Na verdade, a memória continua trabalhando na mesma velocidade, o que muda são os tempos de espera que passam a ser mais altos. Assim, ao invés de responder a cada 2 ciclos da placa mãe, elas podem passar a responder a cada 3 ou 4 ciclos, por exemplo.

As memórias SDRAM (Synchronous Dynamic RAM) por sua vez, são capazes de trabalhar sincronizadas com os ciclos da placa mãe, sem tempos de espera. Isto significa, que a temporização das memórias SDRAM é sempre de uma leitura por ciclo. Independentemente da frequência de barramento utilizada, os tempos de acesso serão sempre de 6-1-1-1, ou mesmo 5-1-1-1.

Veja que o primeiro acesso continua tomando vários ciclos, pois nele é necessário realizar o acesso padrão, ativando a linha (RAS) e depois a coluna (CAS). Apenas a partir do segundo acesso é que as otimizações entram em ação e a memória consegue realizar uma leitura por ciclo, até o final da leitura.

O burst de leitura pode ser de 2, 4 ou 8 endereços e existe também o modo "full page" (disponível apenas nos módulos SDRAM), onde o controlador pode especificar um número qualquer de endereços a serem lidos seqüencialmente, até um máximo de 512. Ou seja, em situações ideais, pode ser possível realizar a leitura de 256 setores em 260 ciclos! :). Só para efeito de comparação, se fossem usadas memórias regulares, com tempos de acesso similares, a mesma tarefa tomaria pelo menos 1280 ciclos.

Outra característica que ajuda as memórias SDRAM a serem mais rápidas que as EDO e FPM é a divisão dos módulos de memória em vários bancos. Um módulo DIMM pode ser formado por 2, 4, ou mesmo 8 bancos de memória, cada um englobando parte dos endereços disponíveis. Apenas um dos bancos pode ser acessado de cada vez, mas o controlador de memória pode aproveitar o tempo de ociosidade para fazer algumas operações nos demais, como executar os ciclos de refresh e também a pré-carga dos bancos que serão acessados em seguida. Nos módulos EDO e FPM, todas estas operações precisam ser feitas entre os ciclos de leitura, o que toma tempo e reduz a frequência das operações de leitura.

---

*A partir da memória SDRAM, tornou-se desnecessário falar em tempos de acesso, já que a memória trabalha de forma sincronizada em relação aos ciclos da placa mãe.*

---

.As memórias passaram então a ser rotuladas de acordo com a frequência em que são capazes de operar.

No caso das memórias SDRAM temos as memórias PC-66, PC-100 e PC-133, no caso das DDR temos as PC-200, PC-266, PC-333, PC-400 (e assim por diante), enquanto nas DDR2 temos as PC-533, PC-666, PC-800, PC-933, PC-1066 e PC1200.

Um pente de memória PC-133 deve ser capaz de operar a 133 MHz, fornecendo 133 milhões de leituras por segundo. Entretanto, esta velocidade é atingida apenas quando o pente realiza um burst de várias leituras. O primeiro acesso, continua levando 5, 6 ou mesmo 7 ciclos da placa mãe, como nas memórias antigas.

Ou seja, o fato de ser um pente PC-100, não indica que o módulo possui um tempo de acesso de 10 ns ou menos (nem mesmo os módulos DDR2 atuais atingem esta marca). Pelo contrário, a maioria dos módulos PC-100 trabalhavam com tempos de acesso de 40 ns. Mas, graças a todas as otimizações que vimos, as leituras podiam ser paralelizadas, de forma que no final o módulo suporta bursts de leitura onde, depois de um lento ciclo inicial, o módulo consegue realmente entregar 64 bits de dados a cada 10 ns.

Independentemente da frequência de operação, temos também os módulos CL2 e CL3, onde o "CL" é abreviação de "CAS latency", ou seja, o tempo

de latência relacionado ao envio do valor CAS, durante o primeiro acesso de cada burst.

Em módulos CL2, o envio do valor CAS toma 2 ciclos, enquanto nos CL3 toma 3 ciclos.

A eles, somamos um ciclo inicial e mais dois ciclos relacionados ao envio do valor RAS, totalizando 5 (nos módulos CL2) ou 6 (nos CL3) ciclos para o acesso inicial.

A diferença acaba sendo pequena, pois os acessos seguintes demoram sempre apenas um ciclo.

Um módulo CL2 realizaria um burst de 8 leituras em 12 ciclos (5-1-1-1-1-1-1-1), enquanto o CL3 demoraria 13 ciclos (6-1-1-1-1-1-1-1-1). Ou seja, embora os módulos CL2 sejam celebrados e sejam alvo de um grande esforço de marketing por parte dos fabricantes, a diferença de performance é realmente muito pequena para justificar pagar mais caro num módulo CL2.

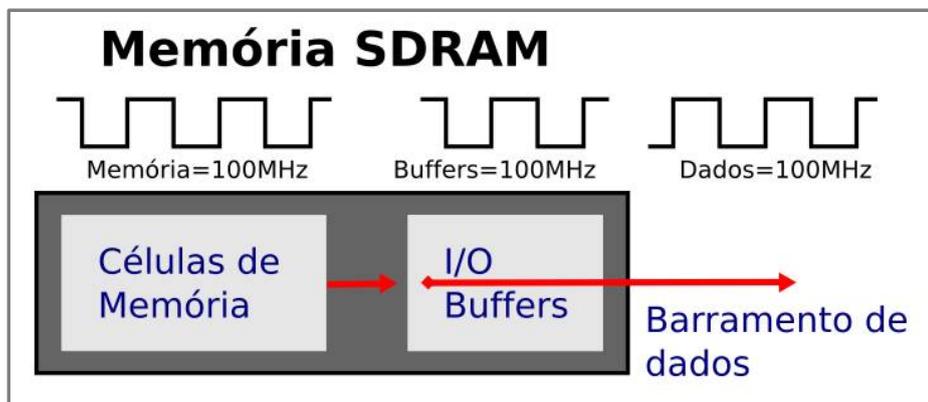
Apesar disso, os módulos CL2 trabalham com tempos de acesso um pouco mais baixo e por isso suportam melhor o uso de frequências mais altas que o especificado, dando mais margem para overclock.

*Vêja que das memórias regulares, até as SDRAM, foi possível multiplicar a velocidade das memórias sem fazer alterações fundamentais nas células, que continuam seguindo o mesmo projeto básico, com um transistor e um capacitor para cada bit armazenado. Desde a década de 80, as reduções nos tempos de acesso foram apenas incrementais, acompanhando as melhorias nas técnicas de fabricação. O que realmente evoluiu com o passar do tempo foram os circuitos em torno dos módulos, que otimizaram o processo de leitura, extraindo mais e mais performance.*

Chegamos então às memórias DDR e DDR2 usadas atualmente, que levam este processo crescente de otimização a um novo nível:

## Memórias DDR

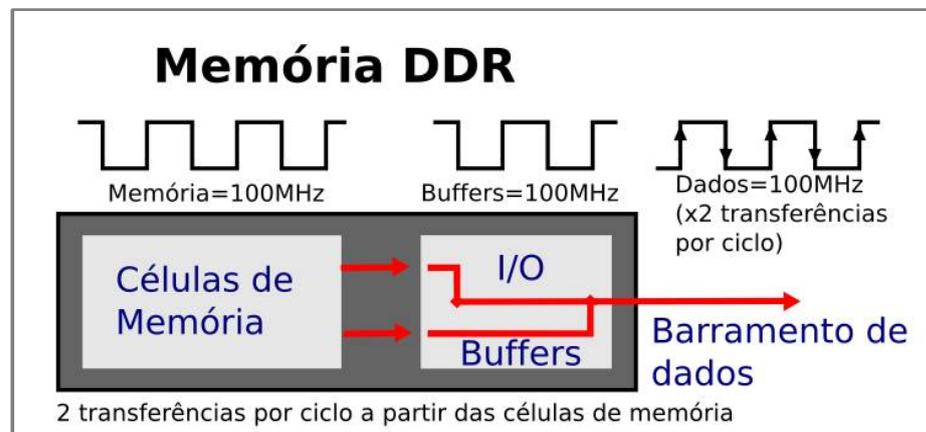
Apesar das otimizações, os módulos de memória SDRAM continuam realizando apenas uma transferência por ciclo, da forma mais simples possível. Depois de decorrido o longo ciclo inicial, as células de memória entregam uma leitura de dados por ciclo, que passa pelos buffers de saída e é despachada através do barramento de dados. Todos os componentes trabalham na mesma frequência:



As memórias DDR implementam um novo truque, que as torna capazes de realizarem duas transferências por ciclo e serem quase duas vezes mais rápidas que as memórias SDRAM, mesmo mantendo a mesma frequência de operação e a mesma tecnologia básica. Vem daí o termo "DDR", que significa "Double Data Rate", ou duplo fluxo de dados. Com o lançamento das memórias DDR, as SDRAM passaram a ser chamadas de "SDR", ou "Single Data Rate".

Os chips de memória DDR incluem circuitos adicionais, que permitem gerar comandos de acesso e receber os dados referentes às leituras duas vezes por ciclo de clock, executando uma operação no início do ciclo e outra no final. Como são utilizadas as mesmas trilhas para realizar ambas as transferências, não foi necessário fazer grandes modificações nem nos módulos, nem nas placas mãe.

Apesar disso, as células de memória propriamente ditas continuam operando na mesma frequência. Num módulo DDR-266, por exemplo, elas operam a apenas 133 MHz, da mesma forma que num pente PC-133. O pulo do gato é fazer com que cada um dos dois comandos de leitura (ou gravação) sejam enviados para um endereço diferente, na mesma linha. Estas duas leituras são enviadas através do barramento de dados na forma de duas transferências separadas, uma realizada no início e a outra no final do ciclo de clock:



O maior problema é que o ciclo inicial continua demorando o mesmo tempo que nas memórias SDRAM, de forma que o ganho aparece apenas em leituras de vários setores consecutivos e a taxa de transferência nunca chega realmente a dobrar, variando bastante de acordo com o tipo de aplicativo usado.

A temporização para um burst de 8 leituras, usando memórias DDR, seria  $5-1/2-1/2-1/2-1/2-1/2-1/2-1/2$  (8.5 ciclos) ao invés de  $5-1-1-1-1-1-1-1$  (12 ciclos) como num módulo SDR. A diferença é menor em bursts menores, de apenas 2 ou 4 leituras.

Apesar disso, as memórias DDR acabaram sendo um excelente negócio, pois tornaram possível obter ganhos perceptíveis de performance sem um aumento considerável no custo. Justamente por isso elas se popularizaram rapidamente, substituindo as memórias SDRAM num espaço de menos de um ano.

Os módulos DDR podem ser vendidos tanto segundo sua frequência de operação, quanto segundo sua taxa de transferência.

DDR-200 (100 MHz) =	PC1600
DDR-266 (133 MHz) =	PC2100
DDR-333 (166 MHz) =	PC2700
DDR-400 (200 MHz) =	PC3200
DDR-466 (233 MHz) =	PC3700
DDR-500 (250 MHz) =	PC4000

Assim como no caso dos módulos SDRAM, existem módulos de memória DDR CL2 e CL3, sendo que nos CL2 o tempo do acesso inicial é reduzido em um ciclo, resultando em um pequeno ganho de desempenho. Como as DDR realizam duas operações por ciclo, surgiram também os módulos CL2.5, que ficam no meio do caminho.

As especificações dos módulos indicam a frequência máxima para a qual seu funcionamento foi comprovado. Nada impede que você use o módulo a uma frequência mais baixa que o especificado; você pode

usar um módulo DDR-400 numa placa mãe configurada para trabalhar a 133 MHz, por exemplo, mas neste caso não existe ganho de desempenho com relação a um módulo DDR-266, com exceção de pequenas diferenças relacionadas ao valor CAS ou à temporização dos dois módulos.

Quase sempre, é possível também usar o módulo a frequências um pouco mais altas que o especificado, fazendo over-clock. O módulo DDR-400 poderia funcionar então a 215 MHz, por exemplo. Fazer over-clock sem aumentar a tensão da memória, não traz perigo para os módulos (mesmo a longo prazo), porém você também não tem garantia de estabilidade. Normalmente os módulos CL2 ou CL2.5 suportam melhor os overclocks, já que o controlador tem mais margem para aumentar a temporização dos módulos para compensar o aumento na frequência.

Ao misturar dois módulos de especificações diferentes, é necessário nivelar por baixo, usando a frequência suportada pelo módulo mais lento. Justamente por isso, nem sempre é conveniente aproveitar os módulos antigos ao fazer upgrade, pois você acaba subutilizando o novo módulo.

Em casos onde você tem, por exemplo, um módulo de 128 MB de memória DDR-266, vale mais a pena vender o módulo antigo e comprar um módulo DDR-400 ou DDR-466 de 512 MB ou mais, do que usar o antigo em conjunto com um módulo novo. Verifique apenas se a placa mãe permite ajustar a frequência de forma assíncrona, sem aumentar junto a frequência do FSB.

Continuando, quase todos os pentes de memória SDRAM ou DDR possuem um chip de identificação chamado de "SPD" (Serial Presence Detect), que armazena

os códigos de identificação do módulo, detalhes sobre a frequência, tempos de acesso, CAS latency e outras especificações. Estas informações são exibidas por programas de identificação, como o CPU-Z e o Sandra. No Linux, você pode ler as informações gravadas no chip usando o script "decodedimms.pl" (você pode encontrá-lo usando o comando "locate"), que faz parte do pacote lmsensors. Ele retorna uma longa lista de informações sobre cada um dos módulos instalados na máquina, como neste exemplo:

```
Memory Serial Presence Detect Decoder
By Philip Edelbrock, Christian Zuckschwerdt, Burkart Lingner,
Jean Delvare and others
Version 2.10.1
Decoding EEPROM: /sys/bus/i2c/drivers/eeprom/0-0050
Guessing DIMM is in                bank 1

---=== SPD EEPROM Information ===---

EEPROM Checksum of bytes 0-62      OK (0x8C)
# of bytes written to SDRAM EEPROM 128
Total number of bytes in EEPROM    256
Fundamental Memory type            DDR SDRAM
SPD Revision                        0.0

---=== Memory Characteristics ===---

Maximum module speed              400MHz (PC3200)
Size                               512 MB
tCL-tRCD-tRP-tRAS                  3-3-3-8
Supported CAS Latencies           3, 2.5, 2
Supported CS Latencies               0
Supported WE Latencies               1
Minimum Cycle Time (CAS 3)        5 ns
```

```

Maximum Access Time (CAS 3)      0.65 ns
Minimum Cycle Time (CAS 2.5)     6 ns
Maximum Access Time (CAS 2.5)    0.7 ns
Minimum Cycle Time (CAS 2)     7.5 ns
Maximum Access Time (CAS 2)      0.75 ns

```

---=== Manufacturing Information ===---

```

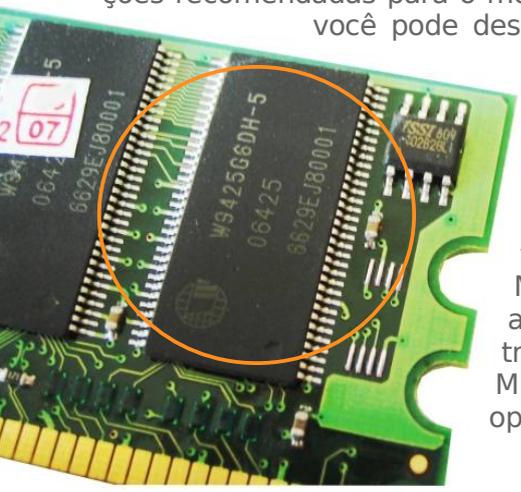
Manufacturer                Kingston
Manufacturing Location Code       0x04
Part Number                       K
Manufacturing Date                0x001E
Assembly Serial Number            0x6B376D48

```

Pelas informações, podemos ver que se trata de um módulo DDR-400 (PC3200) de 512 MB da Kingston. Veja que o módulo suporta o uso de CAS 3, 2.5 ou 2, mas em seguida é especificado que o tempo mínimo de acesso usando CAS 3 são 5 ns e usando CAS 2 são 7.5 ns. Ou seja, o módulo só é capaz de usar CAS 2 em frequências mais baixas. Ao operar a 200 MHz, sua frequência nominal, ele passa automaticamente a usar CAS 3. Ou seja, apesar das especificações serem um pouco confusas, elas indicam que na verdade tenho em mãos um módulo CL3.

O SPD é um pequeno chip de memória EEPROM, com apenas 128 ou 256 bytes, que pode ser localizado facilmente no módulo:

Graças a ele, a placa mãe pode utilizar automaticamente as configurações recomendadas para o módulo, facilitando a configuração. Mas, você pode desativar a configuração automática (By SPD) e especificar sua própria configuração. A maioria das placas atuais permitem que a memória opere de forma assíncrona com o clock da placa mãe, permitindo que a placa mãe opere a 166 MHz, enquanto a memória opera a 200 ou 233 MHz, por exemplo. Ao usar um módulo antigo, também é possível fazer o contrário, mantendo a placa mãe a 200 MHz, mas configurando a memória para operar a 133 MHz, por exemplo.



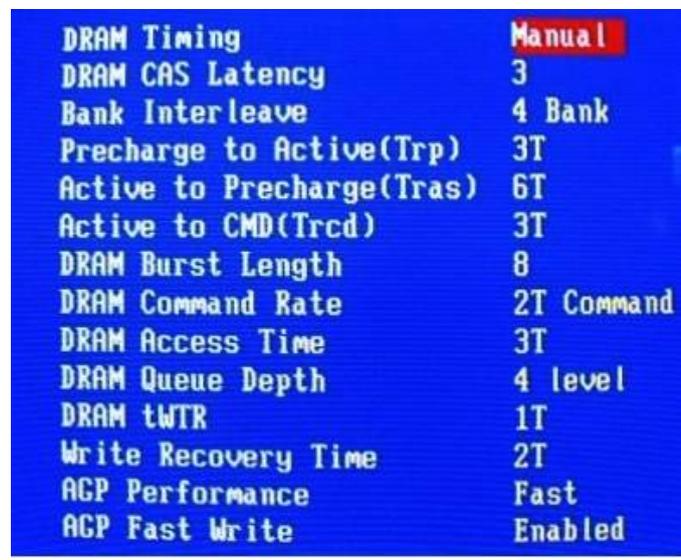
Muitas placas vão além, permitindo que você ajuste manualmente o valor CAS da memória. Isso pode ser útil ao fazer overclock, pois um módulo DDR-400, pode não conseguir trabalhar estavelmente a 233 MHz (por exemplo), mantendo o CAS em 2 tempos, mas pode funcionar perfeitamente se o tempo for aumentado para 3 tempos.

O inverso também é possível. Um módulo DDR-400 CAS3 poderia vir a trabalhar estavelmente com CAS 2 se a frequência fosse reduzida para 166 MHz, por exemplo, oferecendo uma boa flexibilidade para quando você tem tempo disponível e quer chegar ao melhor desempenho possível.

Algumas placas vão ainda mais longe, oferecendo conjuntos completos de ajustes:

Brincar com a frequência e tempos de acesso da memória não oferece riscos para o equipamento. No máximo você pode precisar limpar o setup, para que o micro volte a inicializar depois de tentar usar uma configuração não suportada pelos módulos.

O maior risco está em aumentar a tensão usada pelos módulos (Memory Voltage). É comprovado que pequenos aumentos



na tensão aumentam a possibilidade dos módulos trabalharem estavelmente a frequências mais altas, sobretudo nos módulos DDR2, que dissipam mais calor. O problema é que isso também pode abreviar a vida útil dos módulos, por isso nem sempre é uma boa idéia.

Aumentos de até 5 a 6% estão dentro do limite de tolerância dos circuitos e não oferecem grandes riscos. Você pode usar 2.65v num módulo DDR ou 1.9v num módulo DDR2, mas a partir daí existe prejuízo para a vida útil. Aumentos a partir de 20% podem realmente queimar os módulos em poucas horas, por isso as opções normalmente não ficam disponíveis.

## Memórias DDR2

Seguindo a tendência inaugurada pelas memórias DDR, as DDR2 novamente duplicam a taxa de transferência, realizando agora 4 operações por ciclo. Novamente, as células de memória continuam trabalhando na mesma frequência anterior e o acesso inicial continua demandando aproximadamente o mesmo tempo. Entretanto, as demais operações dentro do burst passam a ser realizada em apenas um quarto de ciclo de clock. Usando memórias DDR2, um burst de 8 leituras demoraria apenas 6.75 ciclos de clock ( $5\frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4}$ ), contra 8.5 ciclos nas DDR e 12 nas SDR.

Como você pode ver, a diferença é maior em aplicativos que precisam manipular grandes blocos de dados e menor em aplicativos que lêem pequenos blocos de dados espalhados. Em nenhuma situação prática a transferência chega realmente a dobrar. Dizer que as "DDR2 são duas vezes mais rápidas" é apenas uma figura de linguagem ;).

---

*Em 2005, quando os primeiros módulos DDR2-533 chegaram ao mercado, eles rapidamente ganharam a fama de "lentos", pois eram comparados a módulos DDR-400 ou DDR-466, que já estavam entrincheirados.*

---

Embora um módulo DDR2 ganhe de um DDR da mesma frequência em todos os quesitos (um DDR2-800 contra um DDR-400, por exemplo), o mesmo não acontece se compararmos módulos de frequências diferentes. Um DDR2-533 opera a apenas 133 MHz, por isso acaba realmente perdendo para um DDR-400 (200 MHz) na maioria das aplicações, pois a ganho de realizar 4 operações por ciclo

acaba não sendo suficiente para compensar a diferença na frequência de operação das células de memória. Vale lembrar que um módulo DDR2-533 trabalha com tempos de latência similares a um módulo DDR-266.

Realizar bursts de leituras rápidas pode não ser a forma mais perfeita de criar memórias mais rápidas (por causa do lento ciclo inicial), mas é sem dúvida a mais simples e barata. A frequência de operação das memórias aumenta de forma gradual, conforme são melhoradas as técnicas de produção.

Assim como no caso dos processadores, não é possível criar um processador capaz de operar ao dobro do clock de uma hora para a outra, mas é possível criar um processador dual-core, por exemplo. No caso das memórias é mais simples, pois você pode ler vários endereços simultaneamente (ou quase), fazendo apenas mudanças nos circuitos controladores.

Dependendo da fonte, você pode ler tanto que as memórias DDR2 operam ao dobro da frequência que as DDR, quanto que elas realizam quatro transferências por ciclo ao invés de duas. Nenhuma das duas explicações estão erradas, mas ambas são incompletas.

Como disse, as células de memória continuam trabalhando na mesma frequência das memórias SDR e DDR, mas os buffers de entrada e saída, responsáveis por ler os dados, passaram a operar ao dobro da frequência. É justamente esta frequência que é "vista" pelo restante do sistema, de forma que a maioria dos programas de diagnóstico mostra a frequência dobrada usada pelos circuitos de entrada e não a frequência real da memória.

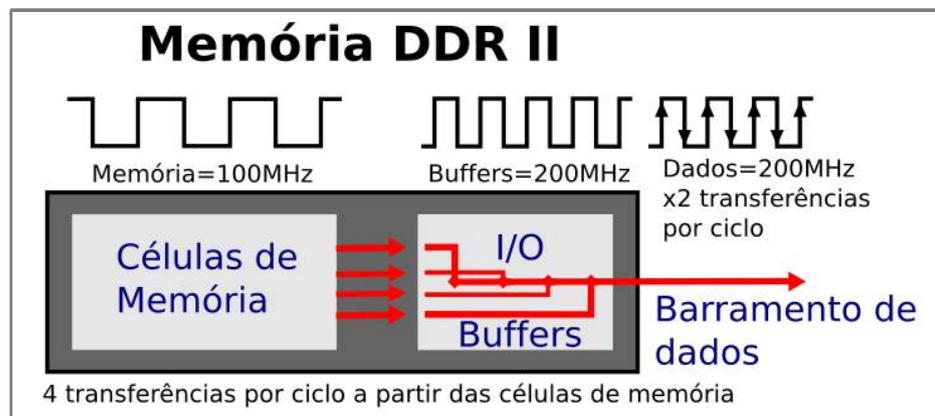
Devido a esta ambigüidade, não é errado dizer que os módulos DDR2 operam ao dobro da frequência dos DDR (os buffers e outros circuitos de apoio realmente operam), nem que são realizadas 4 leituras por ciclo (já que as células de memória continuam operando à mesma frequência).

Ao realizar uma leitura, o controlador de memória gera quatro sinais distintos, que ativam a leitura de quatro endereços adjacentes (4-bit prefetch). As quatro leituras são feitas simultaneamente e os dados entregues ao buffer, que se encarrega de despachá-los através do barramento principal.

Presumindo que o módulo DDR2 de exemplo operasse a 100 MHz, teríamos as células de memória ainda operando na mesma frequência, mas agora entregando 4 leituras de setores sequenciais por ciclo.

Os buffers e o barramento de dados operam agora a 200 MHz, de forma que as 4 leituras podem ser enviadas em 2 ciclos, com duas transferências por ciclo.

Os dois ciclos do barramento são realizados no mesmo espaço de tempo que apenas um ciclo das células de memória:



Como vimos, as células de memória podem ser grosseiramente comparadas a uma planilha eletrônica, com inúmeras linhas e colunas. Não existe uma grande dificuldade em ler vários endereços diferentes simultaneamente, desde que o fabricante consiga desenvolver os circuitos de controle necessários. Graças a isso, o desenvolvimento das memórias tem sendo focado em realizar mais leituras pro ciclo, combinada com aumentos graduais nas frequências de operação.

Quando as memórias DIMM surgiram, ainda na época do Pentium II, os pentes mais rápidos operavam a 100 MHz (os famosos módulos PC-100).

Atualmente temos chips de memória de até 300 MHz que, combinados com as 4 leituras por ciclo, resultam em módulos com transferência teórica de até 9.6 GB/s:

DDR2-533 (133 MHz)	=	PC2-4200
DDR2-667 (166 MHz)	=	PC2-5300
DDR2-800 (200 MHz)	=	PC2-6400
DDR2-933 (233 MHz)	=	PC2-7500
DDR2-1066 (266 MHz)	=	PC2-8500
DDR2-1200 (300 MHz)	=	PC2-9600

O CAS latency dos módulos DDR é medido em termos de ciclos do circuito controlador, por isso são normalmente o dobro do que nos módulos DDR. É como duas "duas unidades" ou "quatro metades", no final dá no mesmo ;). Um módulo DDR2-800 com CAS latency 4, possui o mesmo tempo de acesso que um DDR-400 com CAS latency 2.

Normalmente, as especificações das memórias DDR2 incluem não apenas o CAS latency (**tCL**), mas também o RAS to CAS delay (**tRCD**), Row Precharge Time (**tRP**) e RAS Activate to Charge (**tRAS**). Estes mesmos valores podem ser encontrados nas especificações de módulos DDR e SDR, mas com as memórias DDR2 os fabricantes passaram a divulgá-los de forma mais aberta, usando qualquer redução nos valores para diferenciar seus módulos dos concorrentes. Temos então módulos DDR2-800 "4-4-4-12" ou "5-5-5-15", por exemplo.

O primeiro número é o CAS latency, que já conhecemos. O seguinte é o RAS to CAS delay, que é o tempo que o controlador precisa esperar entre o envio do endereço RAS e o CAS. Para realizar uma leitura, o controlador envia o sinal RAS, espera o tempo referente ao RAS to CAS delay, envia o sinal CAS, aguarda o número de ciclos referente a ele e então finalmente tem a leitura. Num módulo DDR2 4-4-4-12, tanto o tCL quanto o tRCD demoram 4 ciclos, de forma que o acesso inicial demoraria um total de 8 ciclos. Num módulo 5-5-5-15, o tempo subiria para 10 ciclos.

É importante lembrar (mais uma vez ;) que aqui estamos falando de ciclos dos circuitos de

acesso, que trabalham ao dobro da frequência. Os 8 ciclos de um módulo DDR2 equivalem ao mesmo espaço de tempo consumido por 4 ciclos de um módulo DDR ou SDR.

Junto com o ciclo inicial, o controlador pode realizar um burst de mais 7 leituras (totalizando 8). Cada uma destas leituras adicionais consome o equivalente a meio ciclo do controlador (ou a um quarto de ciclo das células de memória). Caso ele precise de mais dados dentro da mesma linha, ele repete o envio do sinal CAS e realiza um novo burst de leituras.

Note que o controlador só precisa enviar o sinal RAS ao mudar a linha ativa, de forma que o tRCD só entra na conta no primeiro acesso. Para os seguintes, temos apenas o tempo referente ao tCL. Caso o controlador precise realizar 24 leituras (dentro da mesma linha), num módulo DDR2 4-4-4-12, teríamos 11.5 ciclos (8+3.5) para as 8 primeiras leituras e mais 15 ciclos (4+3.5+4+3.5) para as 16 leituras subsequentes. É por causa desta peculiaridade que os módulos DDR e DDR2 não possuem mais o "full-page mode" suportado pelas memórias SDRAM; ele deixou de ser necessário.

O Row Precharge Time (tRP) entra em ação quando o controlador precisa alternar entre diferentes linhas. Cada linha inclui 512 endereços de memória, o equivalente a 4 KB de dados. As linhas são divididas em 4 ou 8 páginas, de forma que um pente DDR2 de 1 GB teria 8 páginas de 32.768 linhas, ou 4 páginas de 65.536 linhas. Com tantas linhas e páginas diferentes, não é difícil imaginar que o chaveamento entre elas é muito freqüente :). Quando falo em linhas e páginas, tenha em mente que esta é apenas a forma como o controlador de memória "enxerga" o módulo.

*Fisicamente, mesmo os bits de uma mesma linha estão espalhados pelos vários chips do módulo.*

Antes de poder acessar uma determinada linha, o controlador de memória precisa carregá-la (precharge). Isso consiste em recarregar os capacitores dentro das células de memória, facilitando a leitura dos dados. O Row Precharge Time (tRP) é justamente o tempo necessário para fazer o carregamento, necessário antes de chavear para outra linha, seja no mesmo banco, seja num banco diferente.

Sendo assim, mesmo que seja necessário ler um único setor, a leitura demorará (num módulo 4-4-4-12), 4 ciclos para o tRP, 4 ciclos para o tRCD, 4 ciclos para o tCL, totalizando 12 ciclos. Estes 12 ciclos são justamente o tempo referente ao RAS Activate to Charge (tRAS), que é o tempo mínimo para realizar uma leitura completa. O tRAS é sempre proporcional aos três primeiros valores, pois é justamente a soma dos três. É por isso que ele sempre é mais alto em módulos com CAS latency mais alto.

É possível reduzir o tRAS utilizando um recurso chamado Additive Latency, onde o comando para iniciar o precharge do banco seguinte pode ser enviado antes que a leitura atual termine. Isso faz com que o tempo total da leitura seguinte seja reduzido em 1 ou até mesmo 2 ciclos. Este é o caso dos módulos 5-4-4-11 ou 4-4-4-11, por exemplo. Em outros casos é necessário um ciclo adicional para fechar o banco, que aumenta o tRAS ao invés de diminuir. De qualquer forma, o tRAS é dos quatro o parâmetro que menos influi no desempenho, pois só faz alguma diferença real quando o sistema precisa realizar séries de acessos rápidos, a linhas diferentes.

Bem, esta última parte exigiu uma explicação mais complexa que o habitual. Como você pode ver, os tempos de acesso dos módulos DDR2 é um pouco mais complexo do que pode parecer à primeira vista.

Entretanto, o mais importante dos 4 valores continua sendo o primeiro (o bom e velho CAS latency, ou tCL), que é o tRCD e o tRP são quase sempre iguais a ele e o tRAS é a soma dos três. Ou seja, se o CAS latency é mais baixo, automaticamente os demais valores também são.

O lançamento das memórias DDR teve um impacto diferente para a Intel e a AMD. Para a Intel, a migração para as memórias DDR2 foi mais simples, já que o controlador de memória é incluído no chipset, de forma que aderir a uma nova tecnologia demanda apenas modificações nas placas. Para a AMD, a mudança foi mais tortuosa, já que o Athlon e derivados utilizam um controlador de memória embutido diretamente no processador. Foi necessário atualizar toda a linha de processadores, além de introduzir um novo soquete, o AM2, que exigiu também mudanças nas placas.



# Memória Flash

Diferentemente da memória RAM e também das SRAM, a memória Flash permite armazenar dados por longos períodos, sem precisar de alimentação elétrica. Graças a isso, a memória Flash se tornou rapidamente a tecnologia dominante em cartões de memória, pendrives, HDs de estado sólido (SSDs), memória de armazenamento em câmeras, celulares e palmtops e assim por diante.

Se a memória Flash não existisse, todas estas áreas estariam muito atrasadas em relação ao que temos hoje. Os celulares e palmtops provavelmente ainda utilizariam memória SRAM para armazenar os dados e seriam por isso mais caros e perderiam os dados quando a bateria fosse removida. Os pendrives simplesmente não existiriam e os cartões de memória estariam estagnados nos cartões compact-flash, utilizando microdrives ou pequenas quantidades de memória SRAM alimentada por uma pequena bateria. Formatos mais compactos, como os cartões SD e mini SD simplesmente não existiriam.

Existem dois tipos de memória Flash. A primeira tecnologia de memória Flash a se popularizar foi o tipo NOR, que chegou ao mercado em 1988.



Os chips de memória Flash NOR possuem uma interface de endereços similar à da memória RAM.

Graças a isso, eles rapidamente passaram a ser usados para armazenar o BIOS da placa mãe e firmwares em dispositivos diversos, que antes eram armazenados em chips de memória ROM ou EEPROM. Nos primeiros PCs, por exemplo, o BIOS da placa mãe era gravado em um chip de memória ROM e por isso não era atualizável, a menos que o chip fosse fisicamente substituído.

O problema com as memórias NOR é que elas são muito caras e, embora as leituras sejam rápidas, o tempo de gravação das células é muito alto. Num chip de memória NOR típico, as operações de gravação demoram cerca de 750 ns, ou seja, teríamos pouco mais de 1000 operações de gravação por segundo!

No caso do BIOS da placa mãe, isso não é um grande problema, pois você só precisa atualizá-lo esporadicamente. Mas, imagine um palmtop que tentasse utilizar apenas memória NOR com memória de trabalho... O sistema rodaria tão lentamente que a idéia acabaria sendo abandonada mais cedo ou mais tarde. :)

Apesar disso, a memória Flash do tipo NOR é bastante usada até hoje em palmtops, celulares e diversos tipos de dispositivos, para armazenar o sistema operacional (neste caso chamado de firmware), que é carregado durante o boot, sem ser alterado. A vantagem neste caso é o XiP (execute in place), onde o sistema pode rodar diretamente a partir do chip de memória, sem precisar ser primeiro copiado para a memória RAM.

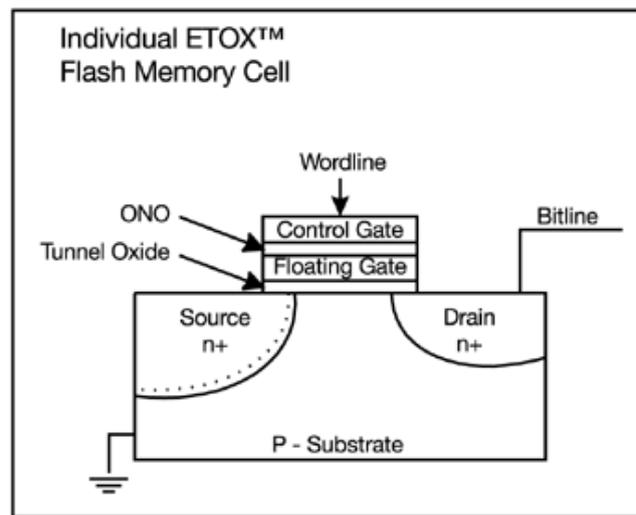
O chip de memória NOR é complementado por uma pequena quantidade de memória SRAM ou DRAM, que é usada como memória de trabalho. Em muitos casos, a memória é usada também para armazenar dados e configurações que, justamente por isso, podem ser perdidos quando a carga da bateria se esgota completamente.

As memórias Flash NOR chegaram a ser utilizadas nos primeiros cartões de memória PCMCIA e Compact Flash, mas elas desapareceram deste ramo quando foram introduzidas as memórias **NAND**, que são de longe o tipo mais usado atualmente.

Nelas, cada célula é composta por dois transistores, com uma fina camada de óxido de silício precisamente posicionada entre

os dois, que armazena cargas negativas. Isto cria uma espécie de armadilha de elétrons, que permite manter os dados por longos períodos de tempo, sem que seja necessário manter a alimentação elétrica (como nas memórias SRAM), ou muito menos fazer um refresh periódico (como na memória DRAM). Isto simplifica muito o design dos cartões, pendrives e outros dispositivos, pois eles precisam incluir apenas os chips de memória Flash NAND, um chip controlador e as trilhas necessárias. Nada de baterias, circuitos de refresh ou qualquer coisa do gênero.

*Aqui temos um diagrama da Intel que mostra uma célula de memória Flash NAND:*



Pelo diagrama você pode notar que embora mais complexa que uma célula de memória RAM (onde temos apenas um transistor e um capacitor), a célula de memória flash ocupa pouco espaço, pois o segundo transistor é posicionado sobre o primeiro. Graças ao tamanho reduzido das células, cada chip de memória Flash NAND armazena uma quantidade muito maior de dados, o que faz com que o preço por megabyte seja muito mais baixo.

Além de mais baratas que as NOR, as memórias NAND também são muito mais rápidas na hora de gravar dados. A principal limitação é que elas são endereçadas usando páginas de 2 KB e acessadas através de um barramento serial. Ou seja, do ponto de vista do sistema, um cartão de memória flash NAND está mais para um HD do que para uma unidade de memória. Você pode usá-lo para guardar dados, mas na hora que o sistema precisa rodar um programa, precisa primeiro copiá-lo para a memória RAM, da mesma forma que faria ao usar um HD.

De alguns anos para cá, os palm-tops e smartphones passaram a cada vez mais utilizar memória Flash NAND como área de armazenamento de dados e programas, substituindo a memória SRAM. Isso se tornou possível graças a um conjunto de truques feitos via software, onde o sistema utiliza uma quantidade menor de memória SRAM como área de trabalho e vai lendo e os arquivos na memória Flash conforme eles são necessários. Esse esquema é muito similar ao que temos num PC, onde os arquivos são salvos no HD, porém processados usando a memória RAM.

Um dos primeiros aparelhos a aderir a este sistema foi o Treo 650, lançado em 2004.



Atualmente ele é utilizado na grande maioria dos modelos, pois, além de cortar custos, melhora a confiabilidade do aparelho, já que os dados não são mais perdidos ao remover a bateria.

O grande boom da memória Flash aconteceu entre 2004 e 2005, quando uma combinação de dois fatores fez com que os preços por MB caíssem rapidamente:

O primeiro foi o brutal aumento na produção e a concorrência entre os fabricantes, que empurraram os preços para baixo. Além de gigantes como a Samsung e a Toshiba, até mesmo a Intel e AMD investiram pesadamente na fabricação de memória Flash.

O segundo foi a introdução da tecnologia MLC (mult-level cell), onde cada célula passa a armazenar dois ou mais bits ao invés de apenas um. Isso é possível graças ao uso de tensões intermediárias. Com 4 tensões diferentes, a célula pode armazenar 2 bits, com 8 pode armazenar 3 bits e assim por diante. O MLC foi implantado de forma mais ou menos simultânea pelos diversos fabricantes e permitiu reduzir drasticamente o custo por

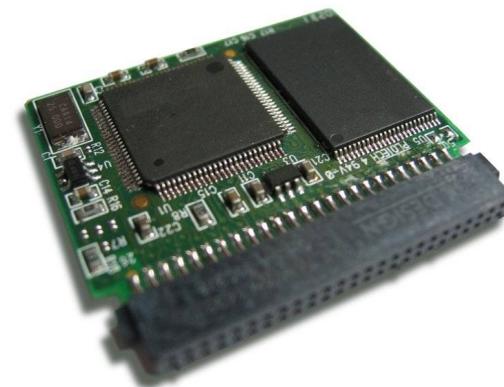
megabyte, quase que de uma hora para a outra. Outra tecnologia similar é o MBC (Multi-Bit Cell), desenvolvido pela Infineon.

Os chips "tradicionais", que armazenam um único bit por célula passaram a ser chamados de "SLC" (single-bit cell) e ainda são produzidos com o objetivo de atender o mercado de cartões de alto desempenho (sobretudo os cartões CF destinados ao mercado profissional). Embora muito mais caros, eles oferecem um melhor desempenho e são mais duráveis.

Outra tecnologia usada pelos fabricantes para cortar custos e ao mesmo tempo permitir a criação de chips de maior densidade, é o "Die-Stacking", onde dois ou mais chips são "empilhados", conectados entre si e selados dentro de um único encapsulamento, que possui o mesmo formato e contatos que um chip tradicional. Como uma boa parte do custo de um chip de memória flash corresponde justamente ao processo de encapsulamento, o uso do Die-Stacking permite mais uma redução substancial do custo.

Como de praxe, a popularização das memórias Flash deu início a uma guerra entre diversos formatos de cartões, alguns abertos e outros proprietários.

**Compact Flash:** Excluindo os jurássicos cartões de memória PCMCIA, o primeiro formato de cartão foi o Compact Flash (CF), onde é utilizada uma interface muito similar à interface IDE usada pelos HDs, com nada menos que 50 pinos. Aqui temos um cartão CF aberto:



De um dos lados temos o chip controlador e um dos chips de memória e no outro temos espaço para mais dois chips, totalizando até 3 chips de alta capacidade. Graças a este design, os cartões CF oferecem boas taxas de transferência, mas em compensação são caros e volumosos, o que explica a decadência do formato.

Os cartões Compact Flash ainda são produzidos e sobrevivem em alguns nichos. Eles são usados por algumas câmeras da Canon, voltadas para o segmento profissional (onde a boa taxa de transferência dos cartões CF

presta bons serviços) e em diversos tipos de sistemas embarcados. Devido à similaridade entre os dois barramentos, existem adaptadores que permitem instalar cartões CF numa porta IDE, substituindo o HD.

**Smart Media:** Em 1995 a Toshiba lançou o formato Smart Media (SM), um formato muito mais simples, onde o chip de memória é acessado diretamente, sem o uso de um chip controlador. O chip de memória é encapsulado dentro de um cartucho plástico, com apenas 0.76 mm de espessura e os contatos externos são ligados diretamente a ele. Nesta foto você pode ver um cartão Smart Media em comparação com um cartão MMC e um Memory Stick:



Apesar de finos, os cartões SM eram relativamente grandes, o que levou os fabricantes a abandonarem o formato. Surgiram então os formatos xD, MMC, SD e Memory Stick. Surpreendentemente, os leitores de cartões USB passaram oferecer suporte para todos os formatos simultaneamente. Isto foi possível graças ao desenvolvimento de chips controladores "tudo em um", capazes de converter cada um dos protocolos nos comandos suportados pelo padrão USB. Existem também os leitores incluídos nos notebooks, que lêem cartões SD e Memory Stick. Do ponto de vista do sistema operacional, eles são diferentes dos leitores USB, pois são ligados ao barramento PCI

(ou PCI Express) ao invés de usarem o barramento USB e a maioria das funções são executadas via software (como num softmodem), graças ao driver instalado.

**Cartões xD:** O próximo da lista é o xD, um formato proprietário, usado em câmeras da Olympus e da Fujifilm. Eles são relativamente rápidos se comparados com os Smart Media e com os cartões MMC, mas são bem mais lentos que os cartões SD usados atualmente. Existiram duas atualizações para o formato: o "xD M" (que permitiu o desenvolvimento de cartões com mais de 512 MB) e o "xD H" (que melhorou a velocidade de transferência). Apesar disso, ambos acabaram sendo pouco usados, devido à concorrência dos cartões SD.

Assim como nos cartões SM, os contatos são ligados diretamente no chip de memória, sem o uso de um chip controlador. Isso em teoria baratearia os cartões, mas devido à pequena demanda (e conseqüentemente aos baixos volumes de produção), os cartões xD são atualmente bem mais caros.



Cartões xD

Isso acaba prejudicando a competitividade das câmeras dos dois fabricantes, que perdem mercado por insistirem no padrão.

**Cartões MMC:** O MMC é um padrão "quase aberto", onde é necessário pagar uma taxa inicial para obter as especificações e mais um valor anual à MMC Association, além de seguir um conjunto de restrições. Os cartões MMC possuem exatamente as mesmas dimensões dos cartões SD atuais e são compatíveis com a maior parte das câmeras e outros dispositivos, além de utilizarem o mesmo encaixe que eles nos adaptadores. As únicas diferenças visíveis é que os cartões MMC são um pouco mais finos (1.4 mm, contra 2.1 mm dos SD) e possuem apenas 7 pinos, enquanto os SD possuem dois pinos extras, totalizando 9.

O maior problema é que os cartões MMC são lentos, pois utilizam um antiquado barramento serial para a transferência de dados, que transfere um bit por vez a uma frequência máxima de 20 MHz.

Em teoria, os cartões MMC poderiam transferir a até 2.5 MB/s, mas a maioria dos cartões ficam muito longe desta marca. Os cartões mais antigos utilizam um modo de transferência ainda mais lento, limitado a 400 KB/s.

Como não existe praticamente nenhuma diferença de custo entre produzir um cartão MMC ou SD, os fabricantes migraram rapidamente para o padrão mais rápido, fazendo com que o MMC entrasse em desuso. Mais recentemente foram lançados os padrões RS-MMC, MMC Plus e SecureMMC, versões atualizadas do padrão MMC, que visam reconquistar seu lugar no mercado.

Chegamos então aos dois padrões que sobreviveram à guerra: o SD, que é o padrão "parcialmente aberto", apoiado pela grande maioria dos fabricantes e o Memory Stick, o padrão proprietário da Sony.

**Memory Stick:** Embora tenha conseguido atingir uma sobrevivência surpreendente, o Memory Stick ficou restrito aos produtos da Sony e por isso seu futuro é incerto. Além do padrão original, existem também os formatos Memory Stick Duo, Pro, Pro Duo, Micro e Pro-HG.

Tanto o padrão original, quanto o Memory Stick Duo estão limitados a 128 MB, por isso ficaram rapidamente obsoletos e são usados apenas por dispositivos antigos, fabricados até o início de 2003. O principal diferença entre os dois formatos é o tamanho reduzido dos cartões Memory Stick Duo, que são um pouco menores que os cartões SD.

Em seguida temos os cartões Memory Stick Pro e Memory Stick Pro Duo (ambos lançados em 2003), que substituem diretamente os dois padrões anteriores. Além do melhor desempenho, eles trouxeram um padrão atualizado de endereçamento, que permite o desenvolvimento de cartões de até 32 GB. Aqui temos uma foto mostrando os 4 formatos:



O Memory Stick Micro (ou M2) é um formato miniaturizado, desenvolvido para uso em celulares (mais especificamente nos Sony Ericsson), que mede apenas 1.5 x 1.2 cm. Os cartões normalmente são vendidos em conjunto com um adaptador, que permite usá-los em qualquer dispositivo ou leitor que use cartões Memory Stick Pro.



Concluindo, temos o Memory Stick Pro-HG, que utiliza um novo barramento de dados, que transmite 8 bits por ciclos a uma frequência de 60 MHz, o que permite uma taxa de transferência de até 60 MB/s (contra 20 MB/s dos padrões anteriores). Embora na prática a taxa de transferência dependa mais dos chips de memória Flash usados,

o barramento mais rápido coloca os cartões Pro\_HG em vantagem em relação aos cartões SD, já que eles estão limitados a um máximo de 20 MB/s pelo barramento usado.

**Cartões SD:** Finalmente, temos os cartões SD (Secure Digital), que acabaram se tornando o formato dominante. Como o nome sugere, os cartões SD oferecem um sistema de proteção de conteúdo (o CPRM), que é implementado diretamente no chip controlador. Ele se destina a atender o lobby das gravadoras, oferecendo uma forma de "proteger" arquivos de áudio e outros tipos de conteúdo contra cópias não autorizadas. Os cartões Memory Stick implementam um sistema similar (o Magic Gate), mas felizmente ambos são pouco usados.

Existem três formatos de cartões SD. Além do formato padrão, temos os cartões miniSD e microSD, versões miniaturizadas, que são eletricamente compatíveis com o padrão original e podem ser encaixados num slot para cartões SD regulares usando um adaptador simples.

Os cartões SD suportam 3 modos de transferência. O 4 bits mode é o modo "padrão", onde o cartão transfere 4 bits por ciclo, a uma frequência de até 50 MHz, resultando em taxas de transferência de até 25 MB/s (desde que os chips de memória usados acompanhem, naturalmente). O segundo é o 1 bit mode, onde é transferido um único bit por ciclo, a uma frequência de no máximo 20 MHz. Este modo é usado para manter compatibilidade com os cartões MMC. É graças a ele que você pode usar cartões MMC em câmeras e leitores para cartões SD e vice-versa. Finalmente, existe o modo SPI (ainda mais lento), que é utilizado por algumas câmeras antigas e também em diversos tipos de dispositivos embarcados.

É por causa dos três modos de operação que um mesmo cartão SD pode ser acessado a velocidades bem diferentes de acordo

com o dispositivo onde ele é usado. Muitas câmeras antigas que permitem acessar o conteúdo do cartão quando ligadas a uma porta USB transferem a velocidades muito baixas, muitas vezes inferiores a 300 KB/s. O driver "sdhci" (no Linux), que dá suporte aos leitores de cartões incluídos em notebooks, por exemplo, é (pelo menos até o Kernel 2.6.21) limitado ao modo SPI, por isso é bastante lento em relação ao driver Windows, que é capaz de utilizar o modo 4 bits. Ou seja, o leitor do seu notebook funciona, mas a uma velocidade muito baixa e com uma grande utilização do processador.



O modo SPI é o preferido pelos desenvolvedores de sistemas embarcados e drivers open-source, pois ele é muito simples e por isso pode ser emulado via software, sem a necessidade de usar um controlador adicional.

No modo SPI 4 são usados 4 pinos do cartão: um para enviar o sinal de clock, outro para enviar comandos, o terceiro para selecionar qual chip dentro do cartão será acessado e o último para transferir dados, um bit de cada vez. Desde que você possa controlar o uso dos 4 pinos, é fácil escrever uma função ou driver para acessar o cartão.

O modo SPI é o mais lento, mas é suficiente para muitas aplicações. Imagine o caso de um sensor de temperatura que usa o cartão apenas para armazenar um log das variações, gravando alguns poucos bits por vez, por exemplo. :)

**Controladores:** Com exceção dos antigos cartões Smart Media e xD, que vimos a pouco, todos os cartões de memória Flash incluem um chip controlador, que é encarregado do gerenciamento dos endereços e todas as operações de leitura e gravação, além de executarem funções de manutenção diversas.

Os cartões atuais utilizam o sistema wear levelling para ampliar a vida útil das células. As células de memória Flash NAND suportam de 100.000 a 1.000.000 de operações de leitura ou gravação, de acordo com a qualidade

dos chips. Pode parecer bastante a princípio, mas a maioria dos sistemas de arquivos (especialmente FAT e EXT) realizam atualizações frequentes na tabela de endereçamento da partição. Se nada fosse feito a respeito, as gravações sucessivas iriam rapidamente inutilizar as células responsáveis pelo armazenamento da tabela, inutilizando o cartão. Graças ao wear levelling é feito uma espécie de "rodízio" dos endereços mais acessados entre as células do cartão, evitando a fadiga de alguns endereços isolados.

Outra função é remapear os endereços defeituosos, onde um setor de uma área reservada passa a ser usado em seu lugar. Isto é muito similar ao sistema utilizado nos HDs modernos, onde a controladora também é capaz de remapear os bad-blocks automaticamente.

Você pode então se perguntar como o controlador faz para descobrir os endereços defeituosos. A resposta é que, além dos dados e dos setores da área reservada, a memória armazena também alguns bytes adicionais (tipicamente 64 bytes adicionais para cada bloco de 2048 bytes), usados para guardar códigos ECC. Estes códigos permitem não apenas identificar, mas também corrigir erros simples nos dados gravados.

Como o controlador não tem como descobrir exatamente em qual célula ocorreu o erro, normalmente todo o bloco de 2048 bytes é remapeado.

Grande parte dos cartões de memória Flash já saem de fábrica com alguns setores defeituosos remapeados (assim como os HDs). Isso permite que os fabricantes aproveitem módulos que de outra forma precisariam ser descartados, reduzindo o custo de forma considerável.

Até certo ponto, o controlador também é responsável pelas taxas de transferência suportadas pelo cartão, já que é ele quem determina os modos de acesso e frequência de clock suportadas. Mesmo que os chips de memória sejam suficientemente rápidos, a taxa de transferência máxima pode ser limitada pelo controlador. Por exemplo, muitos cartões microSD utilizam controladores limitados a 20 MHz, que são capazes de transferir a no máximo 10 MB/s, enquanto muitos dos novos já utilizam controladores capazes de operar a 50 MHz, como nos cartões SD regulares.

A velocidade dos cartões é comparada pelos fabricantes à velocidade dos drives de CD-ROM. Um cartão "133x" é um cartão que usa um controlador capaz de transferir

a 20 MB/s, um "155x" é um capaz de transferir a 25 MB/s e assim por diante. As taxas reais são normalmente mais baixas (sobretudo nas operações de gravação), pois ficam limitadas também à velocidade dos chips, por isso não leve o índice muito a sério, ele é apenas uma ferramenta de marketing. De qualquer forma, é conveniente evitar cartões que não fazem menção à velocidade de transferência, pois eles normalmente são limitados a 33x ou menos. Note que os cartões SDHC adotam um índice diferente, como veremos a seguir.

**miniSD e microSD:** Embora pequenos em relação aos cartões Compact Flash e Smart Media, os cartões SD ainda são grandes demais para algumas aplicações, sobretudo uso nos celulares e câmeras mais compactas. Para solucionar o problema foram criados dois formatos miniaturizados, o miniSD e o microSD, que são menores e mais finos.

O miniSD mede 2.15 x 2.0 cm, com apenas 1.4 mm de espessura. Embora os cartões ainda sejam um pouco mais caros que os SD padrão, o formato está ganhando popularidade rapidamente, usado no Nokia N800 e no E62, por exemplo. De uma forma geral, todos os aparelhos onde o cartão é instalado internamente (embaixo da bateria, por exemplo), estão passando a utilizá-lo.

O microSD é um formato ainda menor, concorrente do M2, destinado a

celulares, MP3 players e outros dispositivos onde as dimensões reduzidas e o baixo consumo são importantes. Ele mede apenas 1.5 x 1.1 cm, com apenas 1 mm de espessura. Na maioria dos casos, o cartão acompanha um adaptador SD. Como os dois padrões são compatíveis eletricamente, o adaptador é apenas um dispositivo passivo, muito barato de se produzir:



*Cartões miniSD e microSD*

Você pode se perguntar como é possível que os cartões microSD sejam tão compactos, já que qualquer cartão SD precisa de pelo menos dois chips (o chip de memória e o controlador) e um cartão microSD mal temos espaço para um. A resposta está no die-stacking, tecnologia que comentei a pouco. Num cartão microSD temos um ou mais chips de memória e o próprio controlador "empilhados", formando um único encapsulamento.

Ela é instalada pelo fabricante numa placa de circuito que contém os contatos externos e em seguida selado dentro da cobertura externa. O mesmo se aplica aos cartões Memory Stick Micro, que possuem dimensões similares.

Não existe como desmontar um microSD e, mesmo que você quebre um no meio, não vai conseguir diferenciar os chips, pois eles são produzidos usar wafers muito finos (até 0.025 mm de espessura nos mais recentes) e juntados de forma muito precisa. Os primeiros microSDs de 4 GB foram produzidos usando nada menos do que 8 chips de 512 MB empilhados. É provável que no futuro seja possível utilizar um número ainda maior.

**SDHC:** Inicialmente, o padrão de cartões SD previa o desenvolvimento de cartões de até 2 GB, formatados por padrão em FAT16. Você pode reformatar o cartão em outros sistemas de arquivos, mas neste caso a maior parte das câmeras e outros dispositivos deixam de conseguir acessá-lo, embora você ainda consiga acessar o cartão normalmente se conectá-lo a um PC usando um adaptador USB.

Quando o limite de 2 GB foi atingido, os fabricantes passaram a criar extensões para permitir a criação de cartões de 4 GB, usando hacks para modificar o sistema de endereçamento e passando a usar FAT32 (ao invés de FAT16) na formatação.

Estes cartões de 4 GB "não padronizados" são compatíveis com a maioria dos dispositivos antigos, mas você pode enfrentar problemas diversos de compatibilidade, já que eles não seguem o padrão.

Para colocar ordem na casa, foi criado o padrão SDHC (Secure Digital High Capacity), onde a tabela de endereçamento foi expandida e passou a ser oficialmente usado o sistema de arquivos FAT32. Todos os cartões que seguem o novo padrão carregam o logotipo "SDHC" (que permite diferenciá-los dos cartões de 4 GB "não oficiais") e trazem um número de classe, que indica a taxa de transferência mínima em operações de escrita. Os cartões "Class 2" gravam a 2 MB/s, os "Class 4" a 4 MB/s, os "Class 6" a 6 MB/s e assim por diante. O mesmo se aplica também aos cartões miniSD e microSD.

Note que a numeração não diz nada sobre a velocidade de leitura, mas ela tende a ser proporcionalmente maior. Veja um exemplo de cartão com o logotipo:



## Outras tecnologias

*Concluindo, temos aqui mais algumas tecnologias de memória que merecem ser citadas:*

**Memórias BEDO:** As memórias BEDO (Burst EDO) foram desenvolvidas no final da era Pentium 1 para concorrer com as memórias EDO. Elas utilizam uma espécie de pipeline para permitir acessos mais rápidos que as EDO. Em um Bus de 66 MHz, as memórias BEDO são capazes de funcionar com temporização de 5-1-1-1, quase 30% mais rápido que as memórias EDO convencionais.

No papel as memórias BEDO eram interessantes, mas elas nunca foram usadas em grande escala. A tecnologia era propriedade da Micron, que ansiava por cobrar royalties dos demais fabricantes, caso a tecnologia fosse adotada em grande escala. Os fabricantes de memória trabalham com margens de lucro incrivelmente apertadas, de forma que a palavra "royalties" gera calafrios. Ao invés de caírem na armadilha da Micron, eles se apressaram em adotar as memórias SDRAM, que além de serem um padrão aberto, eram tecnicamente superiores.

**Memórias Rambus (RDRAM):** Assim como as memórias BEDO, as RDRAM são um tipo proprietário de memória, que acabou não ganhando popularidade.

Os módulos de memórias Rambus são chamados de "Rambus Inline Memory Modules" ou RIMMs. Como pode ser visto na ilustração acima, os módulos RIMM são bem semelhantes aos módulos DIMM, mas em geral eles vem com uma proteção de metal sobre os chips de memória, que também serve para facilitar a dissipação de calor, já que os módulos RIMM aquecem bastante devido à alta frequência de operação.

Uma particularidade era a necessidade de instalar módulos terminadores em todos os slots não populadados da placa mãe, para reduzir o nível de ruído eletromagnético:



Em 1996 a Intel fechou um acordo com a Rambus Inc., uma então pequena empresa que desenvolvia um tipo de memória otimizada para sistemas que precisam de um largo barramento de dados com a memória. As memórias Rambus foram utilizadas no Nintendo 64 e no Playstation 2, e o plano era que elas fossem adotadas em larga escala nos PCs, com a ajuda da Intel. A Rambus Inc. receberia royalties dos fabricantes e a Intel ficaria com parte do bolo, na forma de incentivos e descontos.

A Intel introduziu o suporte às memórias Rambus a partir do chips i820, ainda na época do Pentium III, e continuou tentando empurrar a tecnologia com o chipset i850, usado na primeira geração de placas para Pentium 4.

O problema era que o chipset i850 suportava somente memórias Rambus, sem opção de usar memórias SDRAM ou DDR (que eram novidade na época). Na época do lançamento do Pentium 4, um módulo RIMM de 64 MB custava US\$ 99, enquanto um módulo de memória PC-133 da mesma capacidade custava apenas US\$ 45. Isto significava gastar US\$ 216 (ao comprar 256 MB) a mais, só de memória, sem contar a diferença de preço do processador Pentium 4 e da placa-mãe,

que na época ainda eram consideravelmente mais caros.

As memórias Rambus utilizam um barramento de dados de apenas 16 bits de largura, em oposição aos 64 bits utilizados pelos módulos de memória SDRAM, suportando em compensação, frequências de barramento de até 400 MHz com duas transferências por ciclo (como o AGP 2x), o que na prática equivale a uma frequência de 800 MHz. Esta organização lembra um pouco o barramento PCI Express, onde o uso de menos trilhas de dados permitem a operação a frequências mais altas.

Trabalhando a 400 MHz com duas transferências por ciclo, sua velocidade máxima, as memórias Rambus permitem uma banda total de 1.6 Gigabytes por segundo. O i850 era capaz de acessar dois módulos simultaneamente, proporcionando um barramento total de 3.2 GB/s. Esta é uma marca respeitável, comparável à de um módulo DDR-400, porém atingida em 2001.

O grande problema era que apesar da boa taxa de transferência, os módulos trabalhavam com tempos de latência muito altos. Isso prejudicava muito o desempenho, fazendo com que um Pentium III espetado numa placa mãe i820 com um pente RIMM acabasse sendo mais lento na maioria

das aplicações que um PC simular equipado com memórias SDRAM PC-133. Mesmo em conjunto com o Pentium 4, que incluía uma série de otimizações (incluindo o uso de módulos RIMM em pares), as memórias Rambus falhavam em oferecer algum ganho tangível de performance em relação às memórias DDR.

Pouca gente comprou as versões iniciais do Pentium 4 e quem se arriscou, acabou com um abacaxi nas mãos. Isto obrigou a Intel a modificar a plataforma, passando a utilizar memórias DDR padrão. Esta demora gerou um vácuo, que permitiu que a AMD aumentasse consideravelmente sua participação no mercado, já que contava com o Athlon Thunderbird, um processador mais barato e mais eficiente.

No final, as memórias DDR (seguidas pelas DDR2) ganharam a briga, tornando-se o padrão de memória dominante. Mais uma vez a indústria rejeitou um padrão proprietário de memória, em favor de um padrão aberto.

**Registered DIMM:** Os módulos de memória que usamos nos micros domésticos são chamados de unbuffered. Eles usam um layout simples e eficiente, onde o controlador de memória tem acesso direto aos chips de memória, garantindo tempos de latência mais baixo.



A desvantagem é que o uso dos registradores retarda a transmissão dos sinais, aumentando a latência e consequentemente reduzindo o desempenho dos módulos. A maioria das placas com suporte a módulos registered não suporta módulos unbuffered, de forma que seu uso não é uma opção. Também não é possível usar módulos registered, muito menos misturá-los com módulos unbuffered nas placas para desktop que não os suportam.

Os suporte a módulos registered está disponível apenas em placas mãe destinadas a servidores e workstations, onde a possibilidade de usar mais memória supera as desvantagens. É possível encontrar tanto módulos de memória SDRAM, quando módulos DDR e DDR2 em versão registered. Por utilizarem componentes adicionais e serem produzidos em pequena quantidade, eles normalmente custam o dobro do preço dos módulos unbuffered, de forma que você só deve considerar seu uso quando realmente necessário.

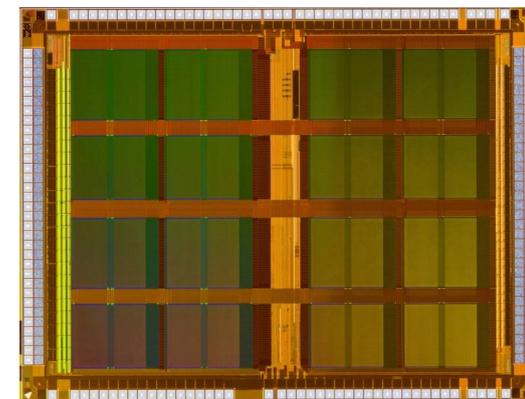
**MRAM:** As memórias MRAM (Magnetoresistive RAM) utilizam células magnéticas para armazenar dados, ao invés de células que armazenam eletricidade, como nas memórias DRAM, SRAM ou Flash. O layout básico lembra um pouco um módulo de memória DRAM, onde temos um transistor para cada bit de dados. A grande diferença é que, ao invés de um capacitor, é usada uma célula magnética, que pode ser gravada e lida usando eletricidade e conserva seus dados por longos períodos (assim como nos HDs) sem precisar de refresh ou alimentação elétrica.

As memórias MRAM são quase tão rápidas quanto as memórias SRAM, consomem menos energia e suportam um número quase ilimitado de ciclos de leitura, ao contrário das memórias flash. Elas são uma espécie de "Santo Graal" da informática, uma tecnologia que, se fosse suficientemente barata, poderia vir a substituir, com vantagens, a maioria dos demais tipos de memórias.

O problema é que as memórias MRAM são difíceis de fabricar e até o momento nenhum fabricante foi capaz de produzir chips com densidades similares à memória RAM ou Flash.

Embora se fale nas memórias MRAM desde a década de 90, os primeiros chips disponíveis comercialmente foram produzidos apenas em 2006 (pela Freescale).

O ponto positivo é que os chips trabalham com tempo de acesso de apenas 35 ms, tanto para leitura, quanto para gravação, o que bate de longe os chips de memória Flash e rivaliza com os chips de memória SRAM usados em palmtops e no cache de HDs, oferecendo a vantagem de não perderem os dados armazenados e não precisarem de alimentação elétrica. O problema é que armazenavam apenas 4 megabits (512 KB) e custavam US\$ 25 cada:



Estes chips foram produzidos em pequena escala, usando técnicas obsoletas de produção, por isso o preço e a densidade atingida tendem a melhorar conforme a tecnologia avança e os chips passam a ser produzidos em maior escala. Diversas empresas, entre elas a IBM e a Samsung tem investido no desenvolvimento de memórias MRAM, por isso devemos ter progressos nos próximos anos.

De início, o concorrente das memórias MRAM são justamente os chips de memória SRAM, que são o tipo mais rápido e caro de memória em uso atualmente. As aplicações são óbvias: HDs com caches que não perdem os dados quando o micro é desligado no botão, além de palmtops e celulares menores e com uma maior autonomia de energia.

A longo prazo, pode ser que as memórias MRAM passem a ser usadas em PCs, substituindo a memória RAM. Um PC que utilizasse memórias MRAM como memória principal poderia manter o estado anterior depois de desligado, sem precisar de um novo processo de boot. Não haveria mais problema de perda de dados por causa de desligamentos incorretos, pois ao ligar o PC novamente, tudo estaria como antes.

A partir daí, quem sabe, novas técnicas de produção permita que

passem a concorrer com as memórias flash, mas por enquanto, isso ainda é exercício de futurologia. A menos que alguma grande revolução aconteça, as memórias MRAM devem demorar pelo menos mais 4 ou 5 anos para se tornarem competitivas com as memórias SRAM e pelo menos uma década para começarem a substituir as memórias DRAM em alguns nichos.

**DDR3:** As memórias DDR3 estão para as DDR2 como elas estão para as DDR tradicionais. O buffer de endereços passou a ter 8 bits, permitindo que o controlador realize 8 leituras simultâneas. Assim como nas memórias DDR2, foi aumentada apenas a frequência dos circuitos externos, de forma a absorver a maior taxa de transferência, mas não a frequência de operação das células de memória.

Num módulo DDR3-1600 (que seria o concorrente direto dos módulos DDR2-800 e dos DDR-400), temos as células de memória operando a apenas 200 MHz e por isso o ciclo inicial continua demorando tanto quanto nos módulos antigos. O truque é ele é seguido por bursts de 8 leituras por ciclo, realizadas em endereços adjacentes. As 8 leituras são transferidas simultaneamente aos buffers de saída.

Como eles agora operam a 800 MHz e realizam duas transferências por ciclo, eles são capazes de transferir as 8 leituras (para o controlador de memória) num tempo equivalente a apenas um ciclo das células de memória; bem a tempo de receber o próximo bloco de 8 leituras.

Os primeiros módulos de memória DDR3 chegaram ao mercado no início de 2007, mas o Intel Bearlake, o primeiro chipset com suporte a elas será lançado apenas no terceiro trimestre. Ou seja, as expectativas mais otimistas são que a primeira leva de placas e módulos com preços competitivos chegue ao mercado apenas no final do ano. No caso da AMD a migração demorará mais um pouco, já que será necessária uma nova atualização no controlador de memória incluído nos processadores e possivelmente também alguma atualização no soquete e na arquitetura das placas. O roadmap da AMD fala em suporte às memórias DDR3 apenas a partir de 2008.

Com certeza os fabricantes vão se apressar em atualizar suas linhas de produtos, de forma a mantê-las atualizadas, mas o ganho de desempenho proporcionado pelas memórias DDR3 (em relação às DDR2) é apenas incremental, por isso não existe necessidade de apressar os upgrades.

A tensão de operação também foi reduzida, de 1.8v (nas DDR2) para apenas 1.5v, resultando num consumo elétrico consideravelmente menor. Embora num desktop o consumo elétrico dos pentes de memória não pareça um fator importante, ele é uma questão significativa nos notebooks e outros dispositivos compactos. O consumo mais baixo permitirá também que as DDR3 passem a competir em nichos antes dominados pelas memórias SRAM, usadas justamente devido ao baixo consumo.

**Paridade e ECC:** Por melhor que seja a qualidade, todos os tipos de memória são passíveis de erros, que podem ser causados por inúmeros fatores, desde variações na tensão da tomada que não são completamente absorvidos pela fonte de alimentação, estática, diversos tipos de interferências eletromagnéticas e, por incrível que possa parecer até mesmo raios cósmicos, que num PC doméstico causam um soft-error em média a cada poucos meses:

<http://www-1.ibm.com/servers/eserver/pseries/campaigns/chipkill.pdf>

Ao contrário dos "hard-errors", que são danos físicos nos módulos de memória, causados por eletricidade estática ou outros tipos de descargas, os soft-erros são erros momentâneos, onde um ou alguns poucos bits são alterados, sem que os chips de memória sejam danificados.

Eles podem causar os mais diversos efeitos colaterais, como travamentos de programas, pequenos danos em arquivos salvos, e assim por diante. Num desktop eles não costumam ser catastróficos, mas podem causar efeitos sérios em sistemas que manipulam informações sensíveis, como no caso dos bancos, por exemplo, onde um soft-error poderia mudar o saldo da sua conta bancária ;).

Para aumentar o grau de confiabilidade dos sistemas, foram criados métodos de diagnóstico e correção de erros. Tudo começou com os sistemas de paridade, usados em muitos pentes de 30 e 72 vias.

A paridade é um método mais antigo, que somente é capaz de identificar alterações nos dados depositados nas memórias, sem condições de fazer qualquer tipo de correção. A paridade consiste na adição de mais um bit para cada byte de memória, que passa a ter 9 bits, tendo o último a função de diagnosticar alterações nos dados.

A operação de checagem dos dados na paridade é bem simples: são contados o número de bits "1" de cada byte. Se o número for par, o bit de paridade assume um valor "1" e caso seja ímpar, o 9º bit assume um valor "0". Quando requisitados pelo processador, os dados são checados pelo circuito de paridade que verifica se o número de bits "1" corresponde ao depositado no 9º bit.

Caso seja constatada alteração nos dados, ele envia ao processador uma mensagem de erro. Claro que este método não é 100% eficaz, pois não é capaz de detectar a alteração de um número de bits que mantenha a paridade. Caso por exemplo, dois bits zero retornassem alterados para bits um, o circuito de paridade não notaria a alteração nos dados. Felizmente, a possibilidade da alteração de dois ou mais bits ao mesmo tempo é remota.

Exemplo de Byte de dados	Número de Bits "1" no Byte	Bit de paridade
00000000	0	1
10110011	5	0
00100100	2	1
11111111	8	1

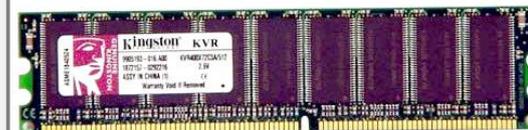
O uso da paridade não torna o computador mais lento, pois os circuitos responsáveis pela checagem dos dados são independentes do restante do sistema. Seu único efeito colateral, é o encarecimento dos módulos de memória, que ao invés de 8 ou 16 chips, passam a ter 9 ou 18, tornando-se pelo menos 12% mais caros.

Além do aumento no custo, o grande problema da paridade é que ela apenas permite identificar erros, mas sem corrigi-los. Isso acaba fazendo com que ela tenha pouca utilidade, pois ao receber um erro suas únicas opções são ignorá-lo, ou parar tudo e reiniciar o micro. Conforme os módulos de memória foram tornando-se mais confiáveis, os módulos com paridade entraram em desuso.

Em seguida temos o **ECC**, o sistema atual, que permite não apenas identificar, mas também corrigir erros simples. O ECC acaba sendo a solução perfeita, pois permite que um servidor continue funcionando, sem interrupções e de forma confiável, mesmo com um grande número de soft-errors, causados por fatores diversos.

O número de bits necessários para implementar o ECC decresce conforme aumenta a largura do barramento usado pelo módulo. Num módulo de 32 bits (como os antigos módulos de 72 vias), são necessários 7 bits adicionais para cada 32 bits de memória, mas nos módulos DIMM de 64 bits atuais, são necessários apenas 8 bits para cada 64 bits de memória, ou seja, o mesmo que seria necessário para usar paridade.

Os módulos DIMM com ECC são fáceis de identificar, pois eles possuem 5, 9 ou 18 chips, ao invés de 4, 8 ou 16. O uso de ECC é mais comum em módulos registered, que são específicos para servidores, mas também é possível encontrar alguns módulos unbuffered com ECC:



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.

# Tudo o que você sempre quis saber sobre os HDs (e não tinha coragem de perguntar :)

por *Carlos E. Morimoto*



Sem dúvida, o disco rígido foi um dos componentes que mais evoluiu na história da computação. O primeiro disco rígido (o IBM 350) foi construído em 1956, e era formado por um conjunto de nada menos que 50 discos de 24 polegadas de diâmetro, com uma capacidade total de 4.36 MB (5 milhões de caracteres, com 7 bits cada um), algo espantoso para a época. Comparado com os discos atuais, este pioneiro custava uma verdadeira fortuna: 35 mil dólares. Porém, apesar de inicialmente, extremamente caros, os discos rígidos foram tornando-se populares nos sistemas corporativos, pois forneciam um meio rápido de armazenamento de dados.

Foto original: Matt & Kim Rudge's photos  
<http://www.flickr.com/photos/mattandkim/>

Foram produzidas cerca de 1000 unidades do IBM 350 entre 1956 e 1961, quando a produção foi descontinuada em favor de versões mais modernas. Esta foto rara, cortesia do museu digital da IBM dá uma idéia das suas dimensões:



IBM 350

Como você pode ver, o IBM 350 não era exatamente um "disco rígido" dentro da concepção que temos hoje em dia. O gabinete tinha 1.70m de altura e quase o mesmo de comprimento e pesava quase uma tonelada. Na época ele era chamado de "unidade de disco" (termo ainda usado hoje em dia por alguns) e podia ser acoplado a diversos computadores produzidos pela IBM. O termo "disco rígido" só surgiu duas décadas depois, junto com os modelos mais compactos.

De lá pra cá, tivemos uma evolução notável. Hoje em dia os HDs já ultrapassaram a marca de 1 TB, utilizam gravação perpendicular e interfaces SATA 300. São brutalmente mais rápidos que os modelos antigos e também mais baratos. Mesmo com o barateamento da memória Flash, os HDs ainda continuam imbatíveis na hora de armazenar grandes quantidades de dados.

## Como um HD funciona

Dentro do disco rígido, os dados são gravados em discos magnéticos, chamados de **platters**. O nome "disco rígido" vem justamente do fato dos discos internos serem extremamente rígidos.

Os platters são compostos de duas camadas. A primeira é chamada de **substrato**, e nada mais é do que um disco metálico, feito de ligas de alumínio. Mais recentemente, alguns fabricantes passaram a utilizar também vidro, que oferece algumas vantagens, como a maior dureza, embora também seja mais difícil de se trabalhar. Os primeiros HDs com discos de vidro foram os IBM Deskstar 75GXP, lançados em 2001.

Independentemente do material usado, o disco precisa ser completamente plano. Como os discos giram a grandes velocidades e as cabeças de leitura trabalham extremamente próximas da superfície magnética, qualquer variação seria fatal.

Para atingir a perfeição necessária, o disco é polido em uma sala limpa, até que se torne perfeitamente plano. Finalmente, vêm a parte final, que é a colocação da superfície magnética nos dois lados do disco.

Como a camada magnética tem apenas alguns microns de espessura, ela é recoberta por uma fina camada protetora, que oferece alguma proteção contra pequenos impactos. Esta camada é importante, pois apesar dos discos serem encapsulados em salas limpas, eles internamente contêm ar, com pressão ambiente.

Os discos são montados em um eixo também feito de alumínio, que deve ser sólido o suficiente para evitar qualquer vibração dos discos, mesmo a altas rotações. Este é mais um componente que passa por um processo de polimento, já que os discos devem ficar perfeitamente presos e alinhados. No caso de HDs com vários discos, eles ao separados usando espaçadores, novamente feitos de ligas de alumínio.

Finalmente, temos o motor de rotação, responsável por manter uma rotação constante. O motor é um dos maiores responsáveis pela durabilidade do disco rígido, pois uma grande parte das falhas graves provém justamente do motor.

Os HDs mais antigos utilizavam motores de 3.600 rotações por minuto, enquanto que atualmente, são utilizados motores de 5.400, 7.200 ou 10.000 RPM.

Nos HDs de notebook ainda são comuns motores de 4.200 RPM, mas os de 5.400 RPM já são maioria. Embora não seja o único, a velocidade de rotação é sem dúvidas o fator que influencia mais diretamente o desempenho.

Para ler e gravar dados no disco, são usadas **cabeças de leitura** eletromagnéticas (heads) que são presas a um **braço móvel** (arm), o que permite seu acesso a todo o disco. O braço de leitura é uma peça triangular, também feita de ligas de alumínio, para que seja ao mesmo tempo leve e resistente. O mecanismo que movimenta o braço de leitura é chamado de **actuator**.

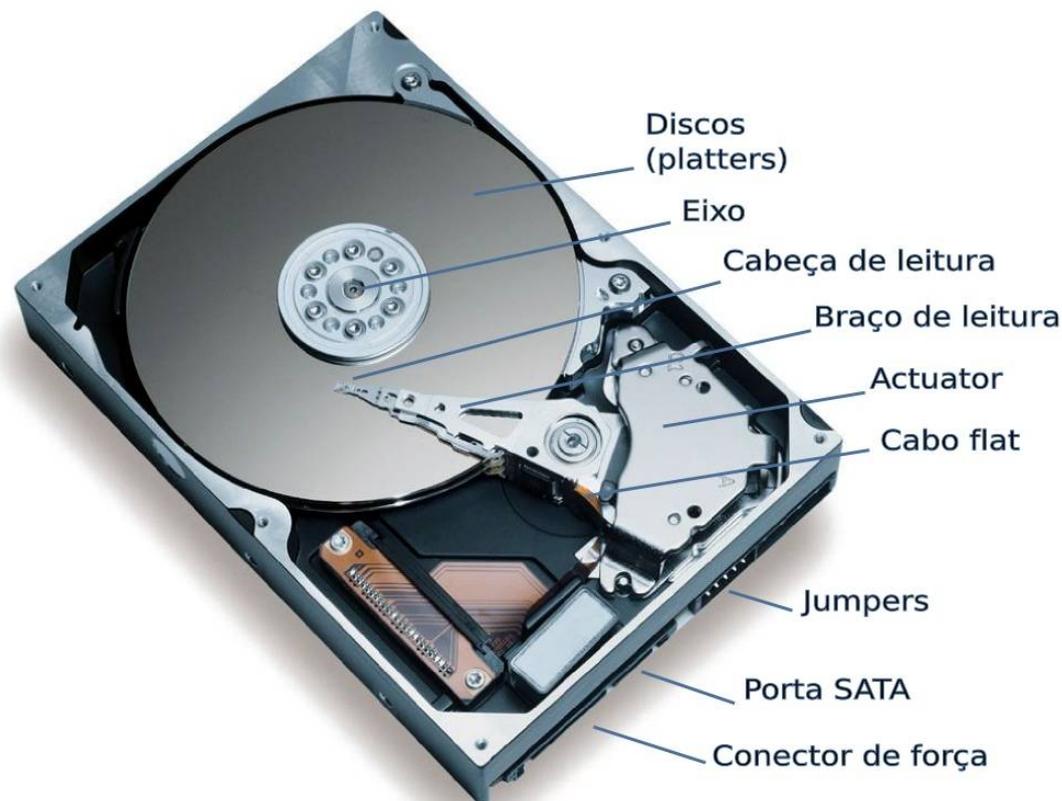
Nos primeiros discos rígidos, eram usados motores de passo para movimentar os braços e cabeças de leitura. Eles são o mesmo tipo de motor usado nos drives de disquete, onde ao receber um impulso elétrico o motor move o braço por uma curta distância, correspondente ao comprimento de uma trilha. O problema é que eles eram muito suscetíveis a problemas de desalinhamento e não permitiam densidades de gravação muito altas.

Os discos contemporâneos (qualquer coisa acima de 80 MB) utilizam um mecanismo bem mais sofisticado para esta tarefa, composto por um dispositivo que atua através de atração e repulsão eletromagnética, sistema chamado de voice coil.

Basicamente temos um eletroímã na base do braço móvel, que permite que a placa controladora o movimente variando rapidamente a potência e a polaridade do ímã. Apesar de parecer suspeito à primeira vista, esse sistema é muito mais rápido, preciso e confiável que os motores de passo. Para você ter uma idéia, os HDs do início da década de 80, com motores de passo, utilizavam apenas 300 ou 400 trilhas por polegada, enquanto um Seagate ST3750640AS (de 750 GB) atual utiliza nada menos do que 145.000.

Para que o HD possa posicionar a cabeça de leitura sobre a área exata referente à trilha que vai ser lida, existem sinais de feedback gravados na superfícies do disco, que orientam o posicionamento da cabeça de leitura. Eles são sinais magnéticos especiais, gravados durante a fabricação dos discos (a famosa formatação física), que são protegidos através de instruções de bloqueio incluídas no firmware do HD contra alteração posterior. Estes sinais eliminam os problemas de desalinhamento que existiam nos primeiros HDs.

*Aqui temos um diagrama mostrando os principais componentes do HD:*



Ao ler um arquivo, a controladora posiciona a cabeça de leitura sobre a trilha onde está o primeiro setor referente a ele e espera que o disco gire até o setor correto. Este tempo inicial, necessário para iniciar a leitura é chamado de tempo de acesso e mesmo em HDs atuais de 7.200 RPM fica em torno de 12 milésimos de segundo, o que é uma eternidade em se tratando de tempo computacional. O HD é relativamente rápido ao ler setores seqüenciais, mas ao ler vários pequenos arquivos espalhados pelo HD, o desempenho pode cair assustadoramente. É por isso que existem programas desfragmentadores, que procuram reorganizar a ordem dos arquivos, de forma que eles sejam gravados em setores contínuos.

Outro dado interessante é a maneira como as cabeças de leitura lêem os dados, sem tocar na camada magnética. Se você tiver a oportunidade de ver um disco rígido aberto, verá que, com os discos parados, as cabeças de leitura são pressionadas levemente em direção ao disco, tocando-o com uma certa pressão. Aqui temos o braço de leitura de um HD, depois de removido.

Apesar disso, quando os discos giram à alta rotação, forma-se uma espécie de colchão de ar, que repele a cabeça de leitura, fazendo com que ela fique sempre a alguns nanômetros de distância dos discos. É o mesmo princípio utilizado na asa de um avião; a principal diferença neste caso é que a cabeça de leitura é fixa, enquanto os discos é que se movem, mas, de qualquer forma, o efeito é o mesmo. Como veremos a seguir, os HDs não são fechados hermeticamente, muito menos a vácuo, pois é necessário ar para criar o efeito.

Esta foto mostra a cabeça de leitura "flutuando" sobre o disco em movimento. A distância é tão curta que mesmo ao vivo você tem a impressão de que a cabeça está raspando no disco, embora na realidade não esteja. Como a cabeça de leitura se movimentava rapidamente durante a operação do disco, é muito difícil tirar fotos. Para conseguir tirar esta, precisei "trapacear", desmontando o actuator e suavemente movendo a cabeça da área de descanso para o meio do disco :).



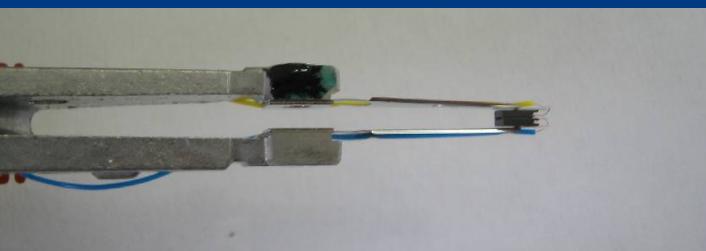
Os discos magnéticos são montados diretamente sobre o eixo do motor de rotação, sem o uso de correias ou qualquer coisa do gênero. É justamente este design simples que permite que os discos girem a uma velocidade tão grande.

Embora mais potente e muito mais durável, o motor de rotação usado nos HDs é similar aos usados nos coolers. Nos HDs antigos, eram usados motores sleeve bearing, o sistema mais simples e menos durável, que foi usado nos HDs de 3600 RPM. Em seguida, foram adotados motores ball-bearing, onde são usados rolamentos para aumentar a precisão e a durabilidade. Nos HDs modernos, é utilizado o sistema fluid-dynamic bearing, onde os rolamentos são substituídos por um fluido especial, que elimina o atrito, reduzindo o ruído e o nível de vibração.

*Aqui temos o mesmo HD da foto anterior completamente desmontado, mostrando o interior do motor de rotação:*



*Veja que mesmo sem o disco magnético entre elas, as duas cabeças de leitura pressionam-se mutuamente:*



Assim como a maioria dos modelos de baixa capacidade, este HD utiliza um único disco, mas a maioria dos modelos utiliza dois, três ou quatro, que são montados usando espaçadores. O HD possui duas cabeças de leitura para cada disco (uma para cada face), de forma que um HD com 4 discos utilizaria 8 cabeças de leitura, presas ao mesmo braço móvel.

Embora usar mais discos permita construir HDs de maior capacidade, não é comum que os fabricantes utilizem mais de 4, pois a partir daí torna-se muito difícil (e caro) produzir componentes com a precisão necessária para manter todos os discos alinhados. Antigamente, era comum que HDs de alta capacidade (e alto custo :), sobretudo os destinados a servidores, possuíssem 6, ou até mesmo 12 discos, mas eles saíram de moda a partir da década de 90, devido à baixa demanda. Desde então, os fabricantes padronizaram a produção em torno dos HDs com até 4 discos e quem precisa de mais capacidade compra vários e monta um sistema RAID.

Naturalmente, qualquer HD aberto fora de uma sala limpa acaba sendo impregnado por partículas de poeira e por isso condenado a começar a apresentar badblocks e outros defeitos depois de alguns minutos de operação.

Todo HD é montado e selado num ambiente livre de partículas, as famosas salas limpas. Apesar disso, eles não são hermeticamente fechados.

Em qualquer HD, você encontra um pequeno orifício para entrada de ar (geralmente escondido embaixo da placa lógica), que permite que pequenas quantidades de ar entrem e saiam, mantendo a pressão interna do HD sempre igual à do meio ambiente. Este orifício é sempre protegido por um filtro, que impede a entrada de partículas de poeira.

Devido a isso, a pressão do ar tem uma certa influência sobre a operação do HD. Os HDs são normalmente projetados para funcionar a altitudes de até 3.000 metros acima do nível do mar. Em altitudes muito elevadas, a pressão do ar é menor, comprometendo a criação do colchão de ar. Para casos extremos, existem HDs pressurizados, que podem trabalhar a qualquer altitude.

Internamente, o HD possui um segundo filtro, que continuamente filtra o ar movimentado pelos discos. Ele tem a função de capturar as partículas que se desprendam dos componentes internos durante o uso, devido a desgaste ou choques diversos.

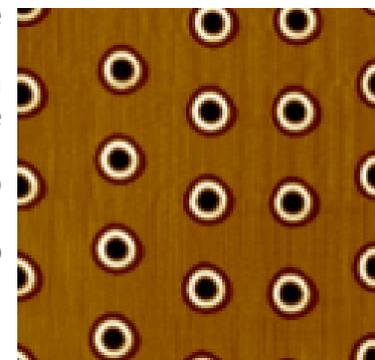
*Aqui temos uma foto de um, preso num dos cantos da parte interna do HD:*



Enquanto o HD está desligado, as cabeças de leitura ficam numa posição de descanso. Elas só saem dessa posição quando os discos já estão girando à velocidade máxima. Para prevenir acidentes, as cabeças de leitura voltam à posição de descanso sempre que não estão sendo lidos dados, apesar dos discos continuarem girando.

É justamente por isso que às vezes, ao sofrer um pico de tensão, ou o micro ser desligado enquanto o HD é acessado, surgem setores defeituosos. Ao ser cortada a energia, os discos param de girar e é desfeito o colchão de ar, fazendo com que as cabeças de leitura possam vir a tocar os discos magnéticos.

Para diminuir a ocorrência deste tipo de acidente, nos HDs modernos é utilizado um sistema que recolhe as cabeças de leitura automaticamente para a área de descanso quando a energia é cortada (tecnologia chamada de auto-parking). A área de descanso é também chamada de "landing zone" e engloba algumas das trilhas mais centrais do disco, uma área especialmente preparada para receber o impacto do "pouso" das cabeças de leitura. Uma das tecnologias mais populares é a LZT (Laser Zone Texture), desenvolvida pela IBM, onde um laser é usado para produzir pequenas cavidades ao longo da zona de pouso, que reduzem o atrito com a cabeça de leitura.



Outra técnica consiste em usar "rampas" feitas de material plástico, posicionadas na área externa dos discos, que suspendem as cabeças de leitura, evitando que elas toquem os discos mesmo quando eles param de girar. Esta tecnologia foi inicialmente usada em HDs de notebook, mas recentemente passou a ser usada também nos de 3.5" para desktops. Ela pode parecer simples, mas na verdade exige bastante tecnologia, devido à precisão necessária.

Aqui as rampas de material plástico que suspendem as cabeças de leitura:



Apesar de evitar danos físicos, o auto-parking nada pode fazer para evitar perda de dados ao desligar o micro incorretamente. Mesmo que todos os arquivos estejam salvos, ainda existem dados no cache de disco (criado pelo sistema operacional usando parte de memória RAM) e também no cache do HD, que utiliza memória SDRAM (também volátil). Para acelerar as operações de gravação, todos os arquivos (sobretudo os pequenos) são salvos inicialmente nos caches e depois transferidos para os discos magnéticos em momentos de ociosidade. Quando o micro é desligado abruptamente, os dados em ambos os caches

são perdidos, fazendo com que você sempre perca as últimas alterações, muitas vezes em arquivos que acreditava estarem salvos.

Por causa de tudo isso, é sempre importante usar um nobreak em micros de trabalho. A longo prazo, os dados perdidos e possíveis danos ao equipamento por causa de quedas de energia acabam custando muito mais do que um nobreak popular.

Mesmo assim, por melhores que sejam as condições de trabalho, o HD continua sendo um dispositivo baseado em componentes mecânicos, que tem uma vida útil muito mais curta que a de outros componentes do micro. De uma forma geral, os HDs para desktop funcionam de forma confiável por de dois a três anos (num PC usado continuamente). Depois disso, é melhor substituir o HD por um novo e mover o antigo para outro micro que não armazena informações importantes, pois a possibilidade de defeitos começa a crescer exponencialmente.

Fala-se muito sobre a vulnerabilidade dos HDs com relação a ímãs. Como os HDs armazenam os dados em discos magnéticos, colocar um ímã suficiente forte próximo a ele pode apagar rapidamente todos os dados. Existem inclusive "desmagnetizadores", que são eletroímãs ligados na tomada, que você passa sobre os HDs e outros discos magnéticos, justamente com a intenção de apagar os dados rapidamente.

Entretanto, se você abrir um HD condenado, vai encontrar dois ímãs surpreendentemente fortes instalados dentro do mecanismo que move a cabeça de leitura.

Naturalmente, estes ímãs não danificam os dados armazenados (senão não estariam ali ;). O principal motivo disto é que eles estão instalados numa posição perpendicular aos discos magnéticos. Se você remover os ímãs e colocá-los sobre outro HD, verá que no outro dia uma boa parte dos dados terão sido perdidos.

Se você (como todos nós) é do tipo que não consegue desmontar um micro sem deixar cair parafusos nos locais mais inacessíveis do gabinete, tem dificuldades em colocar os parafusos dos dois lados ao instalar o HD e ainda por cima nunca acha uma chave de fenda magnética para comprar, pode usar estes magnetos "roubados" do HD para transformar qualquer chave de fenda em uma chave magnética. Basta "encaixar" os ímãs nela quando quiser o efeito. Esses magnetos são feitos de uma liga contendo neodímio e, além de parafusos, permitem levantar objetos um pouco mais pesados, como martelos, por exemplo... ;)



Naturalmente, você deve tomar cuidado de não passá-los sobre discos magnéticos, a menos que queira intencionalmente apagá-los. Se você deixar a chave em contato com os ímãs por um longo período, ela continuará magnetizada (por algum tempo) mesmo depois de retirá-los.

Ao contrário da crença popular, chaves magnéticas não são perigosas para os HDs, pois os magnetos usados são muito fracos se comparados aos magnetos usados no mecanismo de leitura e no motor de rotação do HD. Os próprios discos magnéticos são relativamente resistentes a forças magnéticas externas, de forma que ímãs de baixa potência não oferecem grande perigo.

## A placa controladora

A placa lógica, ou placa controladora é a parte "pensante" do HD. Com exceção dela, o HD é um dispositivo relativamente simples, composto por uma série de dispositivos mecânicos. É a controladora que faz a interface com a placa mãe, controla a rotação do motor e o movimento das cabeças de leitura, de forma que elas leiam os setores corretos, faz a verificação das leituras, de forma a identificar erros e se possível corrigi-los usando os bits de ECC disponíveis em cada setor, atualizar e usar sempre que possível os dados armazenados no cache de disco (já que acessá-lo é muito mais rápido do que fazer uma leitura nas mídias magnéticas), e assim por diante.



Veja que a placa possui apenas três chips. O maior, no canto superior é um Samsung K4S641632H-UC60. Você pode notar que ele é muito semelhante a um chip de memória, e na verdade é :). Ele é um chip de memória SDRAM de 8 MB, que armazena o cache de disco. Até pouco tempo, os HD utilizavam chips de memória SRAM, mas os fabricantes passaram a utilizar cada vez mais chips de memória SDRAM convencional para reduzir o custo de produção. Na prática não muda muita coisa, pois apesar de ser mais lenta, a memória SDRAM oferece desempenho suficiente para a tarefa.

Assim como no caso dos processadores, o cache é um componente importante para o desempenho do HD. Ele armazena os dados acessados, diminuindo bastante o número de leituras. Dados armazenado no

cache podem ser transferidos quase que instantaneamente, usando toda a velocidade permitida pela interface SATA ou IDE, enquanto um acesso a dados gravados nos discos magnéticos demoraria muito mais tempo.

Continuando, temos o controlador principal, um chip Marvell 88i6525, que é quem executa todo o processamento. Este chip é na verdade um SOC (system on a chip), pois na verdade é um conjunto de vários chips menores, agrupados dentro do mesmo encapsulamento.

Por exemplo, este HD é um modelo SATA. A controladora da placa mãe se comunica com ele utilizando comandos padronizados, que são comuns a qualquer HD SATA. É por isso que você não precisa instalar um driver especial para cada modelo de HD, precisa apenas de um driver padrão, que sabe se comunicar com qualquer HD. Internamente, estes comandos SATA são processados e convertidos nos comandos que irão moder a cabeça de leitura, fazer girar os discos até o ponto correto e assim por diante. O sistema operacional não gerencia diretamente o cache de disco, quem faz isso é a própria controladora, que esforça para usá-lo da forma mais eficiente possível.

Naturalmente, tudo isso exige processamento, daí a complexidade interna do chip controlador.

Apesar de pequena, a placa controladora de um disco atual é muito mais sofisticada do que um micro antigo inteiro (um 286 por exemplo). Elas possuem mais poder de processamento e até mesmo mais memória, na forma do cache. Os HDs atuais usam de 8 a 32 MB de cache de disco, mais memória do que era usada em micros 386 e 486 e ainda por cima muito mais rápida! :)

Uma curiosidade é que muitos HDs antigos utilizavam um processador Intel 186 como controlador de discos. O 186 é, como você pode imaginar, o "elo perdido" entre o 8088 usados no PC XT e o 286. Ele é um chip que acabou não sendo usado nos micros PCs, mas fez um grande sucesso como microcontrolador para funções diversas.

Concluindo, temos um terceiro chip, escondido na parte inferior esquerda da foto. Ele é um Hitachi HA13645, um chip especializado, que controla o movimento das cabeças de leitura e também a rotação do motor. O chip principal envia comandos a ele, dizendo que quer acessar o setor X, ou que o motor deve entrar em modo de economia de energia, por exemplo, e ele os transforma nos impulsos elétricos apropriados. Estas funções mudam de um modelo de HD para o outro, por isso os fabricantes preferem usar um chip de uso geral como o Marvell 88i6525 como controlador principal, mudando apenas o controlador, que é um chip menor e mais barato.

A placa controladora é um componente "externo" do HD, que pode ser rapidamente substituído caso necessário. Grande parte (talvez até a maioria) dos casos onde o HD "queima" devido a problemas na rede elétrica, ou defeitos diversos, podem ser solucionados através da troca da placa controladora, permitindo recuperar os dados sem ter que recorrer aos caros serviços de uma empresa especializada.

O grande problema é justamente onde encontrar outra placa. Os fabricantes vendem placas avulsas em pequenas quantidades para empresas de recuperação, mas o fornecimento é muito restrito. Para técnicos autônomos e pequenas empresas, a única solução é usar placas doadas por outros HDs. Se o HD for um modelo recente, você pode simplesmente comprar outro, pegar a placa emprestada para fazer a recuperação dos dados e depois devolvê-la ao dono. Mas, no caso de HDs mais antigos, a única forma é procurar nos sites de leilão e fóruns em busca de uma placa usada. Existe um verdadeiro mercado paralelo de venda de placas avulsas, já que existem muitos casos de HDs inutilizados por problemas na mídia magnética, onde a placa ainda é utilizável.

É comum que os fabricantes utilizem a mesma placa lógica e os mesmos discos magnéticos em vários HDs da mesma família, variando apenas o número de discos usados. Assim, o modelo de 500 GB pode ter 4 discos, enquanto o modelo de 250 GB possui apenas dois, por exemplo. Nestes casos, é normal que a placa controladora de um funcione no outro.

Remover a placa é simples, basta usar uma chave torx para remover os parafusos e desencaixar a placa com cuidado. Na maioria dos HDs atuais, a placa é apenas encaixada sobre os contatos, mas em outros ela é ligada através de um cabo flat, que precisa ser desconectado com cuidado.

#### Remoção de uma placa lógica



Mais uma curiosidade é que os primeiros PCs utilizavam HDs com interfaces MFM ou RLL. Eles utilizavam controladoras externas, instaladas em um slot ISA e ligadas ao HD por dois cabos de dados. Este arranjo era muito ineficiente, pois a distância tornava a comunicação muito suscetível a interferências e corrupção de dados. Estes HDs possuíam várias peculiaridades com relação aos atuais, como a possibilidade de fazer uma "formatação física", onde as trilhas de dados eram realmente regravadas, o que permitia recuperar HDs com problemas de alinhamento.

Estes HDs jurássicos foram usados nos micros XT, 286 e sobreviveram até os primeiros micros 386, quando foram finalmente substituídos pelos HDs IDE, que por sua vez foram substituídos pelos HDs SATA que usamos atualmente, onde a controladora é parte integrante do HD.

Hoje em dia, a "formatação física" sobrevive apenas como um vício de linguagem. Muitos dizem que "fizeram uma formatação física" ao reparticionar o HD ou usar um programa que apaga os dados gravados (como o "zero-fill", ou o "dd" do Linux), embora uma coisa não tenha nada a ver com a outra.

## Os discos

A capacidade de um HD é determinada por basicamente dois fatores: a tecnologia utilizada, que determina sua densidade e o diâmetro dos discos, que determina a área útil de gravação.

A densidade de gravação dos HDs tem aumentado de forma surpreendente, com a introdução de sucessivas novas técnicas de fabricação. Para você ter uma idéia, no IBM 350 os discos eram simplesmente pintados usando uma tinta especial contendo limalha de ferro, um processo bastante primitivo.

Com o passar do tempo, passou a ser usado o processo de eletroplating, que é semelhante à eletrólise usada para banhar bijuterias à ouro.

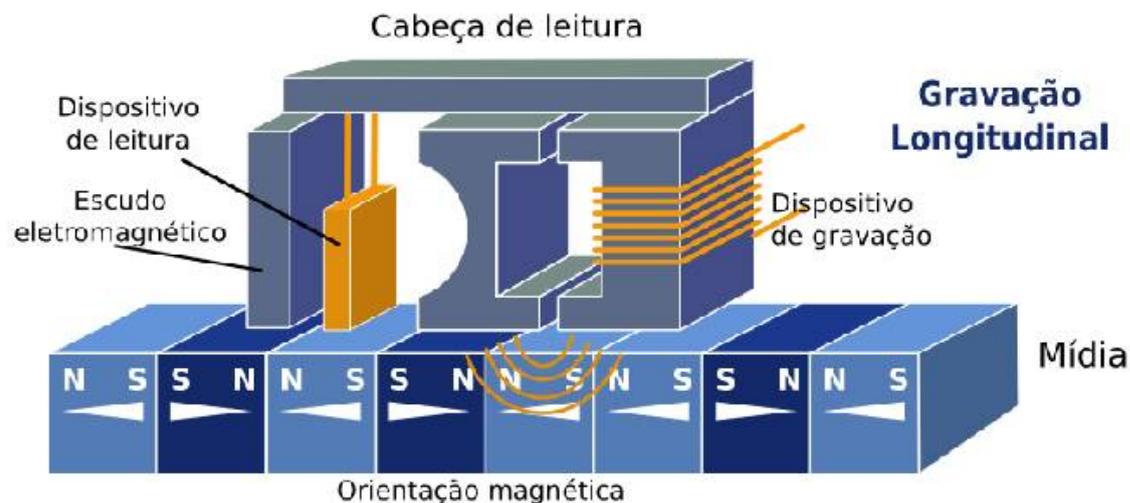
Esta técnica não permite uma superfície muito uniforme e justamente por isso só funciona em discos de baixa densidade. Ela foi usada até o final da década de 80.

A técnica usada atualmente (chamada de sputtering) é muito mais precisa. Nela a superfície magnética é construída depositando grãos microscópicos de forma incrivelmente uniforme. Quanto menores os grãos, mais fina e sensível é a superfície, permitindo densidades de gravação mais altas.

A densidade de gravação de um HD é medida em gigabits por polegada quadrada. Os HDs fabricados na segunda metade de 2006, por exemplo, utilizavam em sua maioria discos com densidade de 100 gigabits (ou 12.5 GB) por polegada quadrada. Neles, cada bit é armazenado numa área magnética com aproximadamente 200x50 nanômetros (uma área pouco maior que a de um transistor nos processadores fabricados numa técnica de 0.09 micron), e é composta por apenas algumas centenas de grãos magnéticos. Estes grãos medem apenas alguns nanômetros e são compostos por ligas de cobalto, cromo, platina, boro e outros materiais raros, muito longe da limalha de ferro utilizada pelos pioneiros.

Considerando que os discos giram a 7200 RPM e a cabeça de leitura lê os dados a mais de 50 MB/s (quando lendo setores seqüenciais), atingir densidades como as atuais é simplesmente impressionante.

*Este esquema mostra como funciona o processo de escrita e gravação em um HD:*



Como você pode ver, a cabeça é composta por dois dispositivos separados, um para gravação e outro para leitura. O dispositivo de gravação é similar a um eletroímã, onde é usada eletricidade para criar o capô magnético usado para realizar a gravação. Nos primeiros HDs, tínhamos um filamento de cobre enrolado sobre um corpo de ferro. Nos HDs atuais, os materiais usados são diferentes, mas o princípio de funcionamento continua o mesmo.

O dispositivo de leitura, por sua vez, faz o processo oposto. Quando ele passa sobre os bits gravados, capta o campo magnético emitido por eles, através de um processo de indução (nos HDs antigos) ou resistência (nos atuais), resultando em uma fraca corrente, que é posteriormente amplificada.

O dispositivo de gravação é protegido por um escudo eletromagnético, que faz com que ele capte apenas o campo magnético do bit que está sendo lido, e não dos seguintes. Você pode notar que não existe isolamento entre os dispositivos de leitura e gravação. Isso acontece por que apenas um deles é usado de cada vez.

Note que esta divisão existe apenas nos HDs modernos, que utilizam cabeças de leitura/gravação MR ou GMR. Nos antigos, que ainda utilizavam cabeças de leitura de ferrite, o mesmo dispositivo fazia a leitura e a gravação.

O grande problema é que, assim como em outras áreas da informática, a tecnologia avançou até o ponto em que se começou a atingir os limites físicos da matéria. Num HD, a área referente a cada bit armazenado funciona como um minúsculo ímã, que tem sua orientação magnética alterada pela cabeça de leitura. Quando ela é orientada em um sentido temos um bit 1 e no sentido oposto temos um bit 0. A área da superfície utilizada para a gravação de cada bit chamada de "magnetic element", ou elemento magnético.

*A partir de um certo ponto, a área de gravação torna-se tão pequena que a orientação magnética dos bits pode ser alterada de forma aleatória pela própria energia térmica presente no ambiente (fenômeno de superparamagnetismo) o que faz com que a mídia deixe de ser confiável.*

A tecnologia usada nos HDs fabricados até a primeira metade de 2007 é chamada de gravação longitudinal (longitudinal recording), onde a orientação magnética dos bits é gravada na horizontal, de forma paralela à mídia.

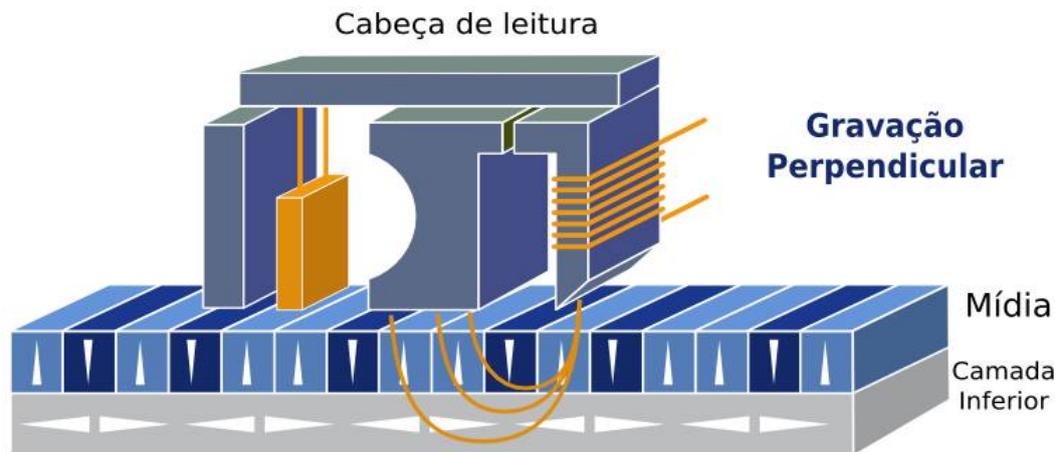
O problema é que a partir dos 100 gigabits por polegada quadrada, tornou-se muito difícil aumentar a densidade de gravação, o que acelerou a migração para o sistema de gravação perpendicular (perpendicular recording), onde a orientação magnética passa a ser feita na vertical, aumentando muito a densidade dos discos.

Estima-se que utilizando gravação longitudinal, seria possível atingir densidades de no máximo 200 gigabits por polegada, enquanto que utilizando gravação perpendicular seja possível atingir até 10 vezes mais. Isso significa que os fabricantes ainda terão margem para produzir HDs de até 10 terabytes antes de esgotar as possibilidades oferecidas pela nova tecnologia.

Na gravação perpendicular, a mídia de gravação é composta de duas camadas. Inicialmente temos uma camada de cromo, que serve como um indutor, permitindo que o sinal magnético gerado pelo dispositivo de gravação "atravesse" a superfície magnética, criando um impulso mais forte e, ao mesmo tempo, como uma espécie de isolante entre a superfície de gravação e as camadas inferiores do disco.

Ele poderia (até certo ponto) ser comparado à camada extra usada nos processadores fabricados com tecnologia SOI (silicon on insulator), onde uma camada isolante é criada entre os transistores e o wafer de silício, reduzindo a perda de elétrons e, conseqüentemente, o consumo elétrico do processador.

Sobre a camada de cromo, são depositados os grãos magnéticos. A diferença é que agora eles são depositados de forma que a orientação magnética seja vertical, e não horizontal. A cabeça de leitura e gravação também é modificada, de forma a serem capazes de lidar com a nova orientação:



Embora pareça uma modificação simples, o uso da gravação perpendicular em HDs é uma conquista técnica notável. Em termos comparativos, seria como se a NASA conseguisse enviar uma missão tripulada até Marte.

O processo de gravação perpendicular foi adotado rapidamente por todos os principais fabricantes. O primeiro foi a Fujitsu, que lançou um HD de 1.8" com gravação perpendicular ainda em 2005. O próximo foi a Seagate, que em Abril de 2006 anunciou o Barracuda 7200.10, um disco de 3.5" com 750 GB. Em Agosto de 2006 a Fujitsu anunciou um HD de 2.5" com 160 GB e em Janeiro de 2007 a Hitachi anunciou o Deskstar 7K1000, um HD de 3.5" com 1 TB que utiliza um design incomum, com 5 platters ao invés dos 4 comumente usados.

Em seguida, temos a questão do diâmetro dos discos. Como vimos, os primeiros HDs eram gigantescos, e utilizavam discos de até 24 polegadas de diâmetro. Com o passar das décadas, os discos foram encolhendo, até chegar ao que temos hoje.

Mas, como tudo na vida, existem receitas ideais para o tamanho dos discos magnéticos, de acordo com a área onde eles serão utilizados.

O problema em produzir discos muito compactos é que a superfície de gravação fica exponencialmente menor, permitindo gravar menos dados. Apesar disso, os demais componentes continuam custando quase o mesmo (ou até mais, dependendo da escala de miniaturização necessária). Isso faz com que o custo por megabyte cresça, conforme

o tamanho físico do HD diminui. Um exemplo prático disso é a diferença no custo dos HDs de 2.5" para notebooks e os modelos de 3.5" para desktops.

A partir de um certo ponto de miniaturização, o custo por megabyte se torna mais alto que o dos cartões de memória flash e os HDs deixam de ser viáveis. O melhor exemplo é o HD de 0.85" apresentado pela Toshiba em 2005, que tinha como objetivo atender o mercado de palmtops e smartphones. Ele era tão pequeno que podia ser produzido no formato de um cartão SD e possuía um consumo elétrico baixíssimo:



O problema é que ele seria lançado em versões de apenas 2 e 4 GB, com preços a partir de US\$ 150. Com a rápida queda no custo da memória flash, logo surgiram cartões de 2 e 4 GB que custavam menos, de forma que o mini-HD acabou não encontrando seu lugar no mercado e foi descontinuado silenciosamente.

O interessante é que o oposto também é verdadeiro. HDs com discos muito grandes também acabam sendo inviáveis, pois acabam sendo bem mais lentos e mais passíveis de problemas, o que se deve a vários fatores.

O primeiro é a questão da rotação, já que discos maiores são mais pesados e demandam um maior esforço do motor de rotação, consumindo mais energia e gerando mais calor e mais barulho. Discos maiores também acabam sendo menos rígidos, o que impede que sejam girados a velocidades muito altas e torna todo o equipamento mais sensível a impactos. Dobrar o diâmetro dos discos, faz com que a rigidez seja reduzida em até 75%.

O segundo é a dificuldade de produção. Com o avanço da tecnologia, a mídia de gravação precisa ser cada vez mais fina e uniforme. Quanto maior os discos, mais difícil é recobrir toda a superfície sem que haja um grande número de pontos defeituosos.

Como se não bastasse, temos o terceiro motivo, que é o maior tempo de acesso, já que com uma superfície maior, as cabeças de leitura demoram muito mais tempo para conseguir localizar os dados (justamente devido à maior distância a ser percorrida). Se combinarmos isso com a velocidade mais baixa de rotação, acabamos tendo uma redução muito grande no desempenho.

Isso explica por que os HDs com discos de 5.25" usados nos primeiros PCs foram rapidamente substituídos pelos de 3.5". O pico evolutivo dos HDs de 5.25" foram os Quantum Bigfoot, produzidos até 1999, em capacidades de até 18 GB. Embora eles armazenassem um maior volume de dados por platter, a velocidade de rotação era bem mais baixa (apenas 3600 RPM), os tempos de acesso eram maiores e, ainda por cima, a durabilidade era menor.

Os HDs de 3.5" e de 2.5" atuais parecem ser o melhor balanço entre os dois extremos. Os HDs de 3.5" oferecem um melhor desempenho, mais capacidade de armazenamento e um custo por megabyte mais baixo (combinação ideal para um desktop), enquanto os HDs de 2.5" são mais compactos, mais silenciosos, consomem menos energia e são mais resistentes a impactos, características fundamentais no caso dos notebooks.

Temos ainda os HDs de 1.8" (mais finos e do tamanho de um cartão PCMCIA), que são usados em notebooks ultraportáteis, além de mp3-players e alguns dispositivos de armazenamento portátil.

Uma quarta categoria são os microdrives, que utilizam discos de 1" (pequenos o suficiente para serem produzidos no formato de cartões compact flash) e podem ser utilizados em palmtops e mp3-players.

Eles foram utilizados no Palm Life Drive (4 GB) e também no iPod Nano (4 e 8 GB), mas acabaram perdendo seu espaço para os cartões de memória flash. A Hitachi chegou a anunciar o desenvolvimento de microdrives de 20 GB, utilizando tecnologia de gravação perpendicular, mas a produção em série acabou sendo cancelada, pois o preço de venda seria mais alto que o da mesma quantidade de memória flash.



Para organizar o processo de gravação e leitura dos dados, a superfície dos discos é dividida em trilhas e setores. As trilhas são círculos concêntricos, que começam no final do disco e vão se tornando menores conforme se aproximam do centro. É diferente de um CD-ROM ou DVD, onde temos uma espiral contínua.

Cada trilha recebe um número de endereçamento, que permite sua localização. A trilha mais externa recebe o número 0 e as seguintes recebem os números 1, 2, 3, e assim por diante. Para facilitar ainda mais o acesso aos dados, as trilhas se dividem em setores, que são pequenos trechos de 512 cada um, onde são armazenados os dados.

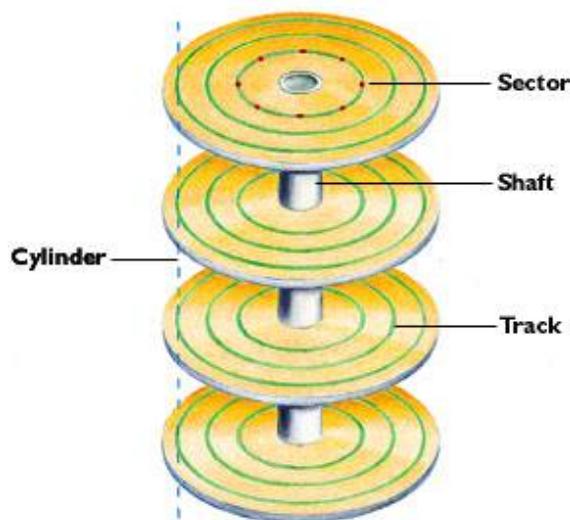
Além das trilhas e setores, temos também as faces de disco. Como vimos, os HDs atuais possuem de 1 a 4 discos. Como são utilizadas ambas as faces de cada disco, temos um total de 2 a 8 faces e o mesmo número de cabeças de leitura.

Como todas as cabeças de leitura estão presas no mesmo braço móvel, elas não possuem movimentos independentes. Para acessar um dado contido na trilha 199.982 da face do disco 3, por exemplo, a controladora do disco ativa a cabeça de leitura responsável pelo disco 3 e a seguir, ordena ao braço de leitura que se dirija à trilha correspondente. Não é possível que uma cabeça de leitura esteja na trilha 199.982 ao mesmo tempo que outra esteja na trilha 555.631 de outro disco, por exemplo.

Já que todas as cabeças de leitura sempre estarão na mesma trilha de seus respectivos discos, deixamos de chamá-las de trilhas e passamos a usar o termo "cilindro". Um cilindro nada mais é do que o conjunto de trilhas com o mesmo número nos vários discos. Por exemplo, o cilindro 1 é formado pela trilha 1 de cada face de disco, o cilindro 2 é formado pela trilha 2 de cada face, e assim por diante. A próxima ilustração mostra como funciona esta divisão.



### Tracks, Cylinders, and Sectors



*Esta antiga ilustração da Quantum mostra como funciona esta divisão*

A trilha mais externa de um disco rígido possui mais que o dobro de diâmetro da trilha mais interna e, conseqüentemente, possui capacidade para armazenar muito mais dados. Porém, nos primeiros discos rígidos, assim como nos disquetes, todas as trilhas do disco, independentemente de seu diâmetro, possuíam o mesmo número de setores, fazendo com que nas trilhas mais externas, os setores ocupassem um espaço muito maior do que os setores das trilhas mais internas.

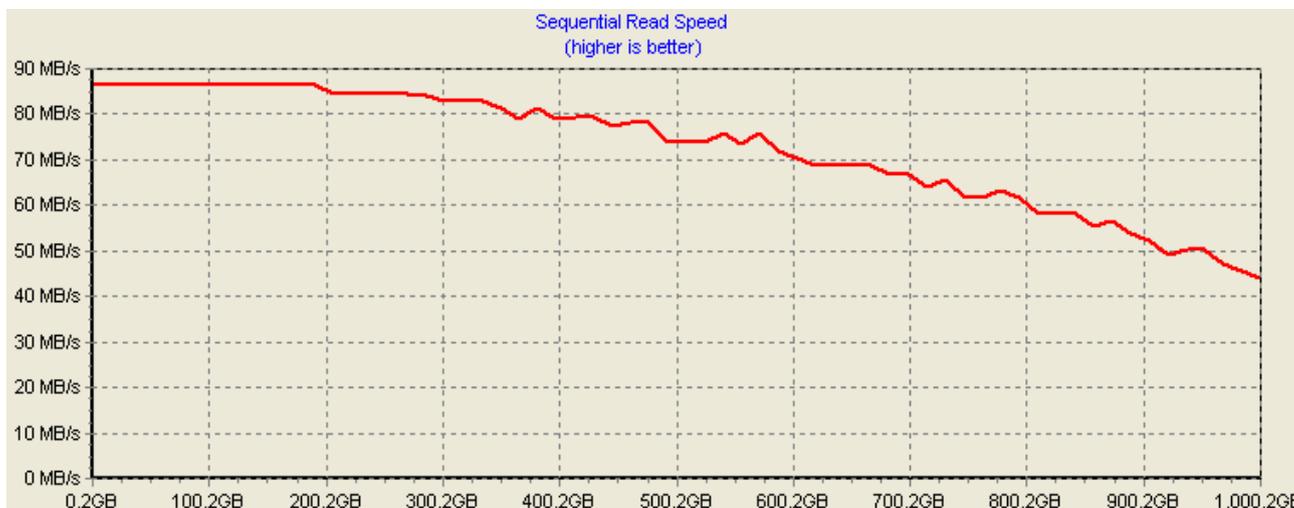
Tínhamos então um grande espaço desperdiçado, pois era preciso nivelar por baixo, fazendo com que todas as trilhas possuísem o mesmo número de setores permitido pelas trilhas mais internas, acabando por desperdiçar enormes quantidades de espaço nas primeiras trilhas do disco.

Atualmente, os HDs utilizam o Zoned bit Recording (ZBR), que permite variar a quantidade de setores por trilha, de acordo com o diâmetro da trilha a ser dividida, permitindo uma organização mais racional do espaço em disco e, conseqüentemente, uma maior densidade de gravação.

O HD pode ter então 1584 setores por trilha na área mais externa dos discos e apenas 740 na área mais interna, por exemplo. Como os discos giram sempre na mesma velocidade, isso causa um pequeno efeito colateral, que é uma considerável variação no desempenho de acordo com a área do disco que está sendo lida, proporcional ao número de setores por trilha.

Tocando em miúdos, o desempenho ao ler as trilhas mais externas acaba sendo mais que o dobro do obtido ao ler as mais internas. É por isso que em geral se recomenda colocar a partição com a instalação do sistema, ou com a partição swap no início do disco (que corresponde às trilhas mais externas) para obter o melhor desempenho.

Usando um programa de benchmark que permita realizar uma leitura seqüencial de toda a superfície do HD, como o HD Tach, você obterá sempre um gráfico similar ao da próxima página, onde a taxa de leitura começa num nível alto (trilhas externas) e vai decaindo até atingir o ponto mais baixo no final do teste (ao ler o conteúdo das trilhas mais internas).



*Queda na taxa de leitura variando de acordo com a localização da trilha (da mais externa para a mais interna)*

Um dos principais motivos do desempenho dos HDs não ter crescido na mesma proporção da capacidade ao longo das últimas décadas é que a densidade das trilhas aumentou numa escala muito maior que a dos setores dentro destas. Ou seja, as trilhas foram ficando mais "finas", mas o número de setores por trilha passou a aumentar em escala incremental. Aumentar o número de trilhas permite aumentar a área de armazenamento, mas é o número de setores por trilha, combinado com a velocidade de rotação do HD que determina a performance.

Um antigo Maxtor 7040A, de 40 MB, por exemplo, possuía uma taxa de leitura média em torno de 700 KB/s, o que permitia ler todo o conteúdo do disco em cerca de um minuto.

Um Seagate Barracuda 7200.10 atual, de 750 GB, é bem mais rápido, com uma taxa média de leitura de 64 MB/s, mas, apesar disso, como a capacidade é brutalmente maior, ler todos os dados do disco demoraria pelo menos 3:15 horas!

No futuro, esta tendência deve se manter, pois é muito mais simples para os fabricantes produzirem cabeças de leitura e sistemas de codificação capazes de lidar com trilhas mais finas, do que espremer mais dados dentro de cada trilha, já que elementos magnéticos mais curtos correspondem a um sinal magnético mais fraco e mais difícil de ser captado pela cabeça de leitura. Como um agravante, temos o problema do superparamagnetismo, que vimos a pouco.

## Correção de erros e bad blocks

Concluindo, temos a questão da detecção e correção de erros, que faz parte do processo de leitura e gravação.

Por melhor que seja sua qualidade, uma mídia magnética nunca é 100% confiável (como pode confirmar quem já teve o desprazer de trabalhar com disquetes ;). Pequenas falhas na superfície da mídia podem levar a erros de leitura, sobretudo quando ela possui uma densidade de gravação de mais de 100 gigabits por polegada quadrada e gira a 7.200 RPM ou mais, como nos HDs atuais.

Isso não significa que o seu HD vá pifar amanhã, mas que são comuns erros na leitura de um setor ou outro. Obviamente, como todos os nossos dados importantes são guardados no disco rígido, a possibilidade de erros na leitura de "um setor ou outro" não seria aceitável, principalmente no caso de máquinas destinadas a operações críticas. Imagine se neste "setor ou outro" do servidor de um grande banco, estivessem gravados os dados referentes à conta bancária de um cliente importante, por exemplo.

De modo a tornar os HDs uma forma de armazenamento confiável, os fabricantes utilizam sistemas de ECC para detectar e corrigir erros de leitura eventualmente encontrados. O ECC é o mesmo sistema utilizado em pentes de memória destinados a servidores e também em CD-ROMs, onde são usados alguns bits adicionais para cada bloco de dados.

Num HD, cada setor armazena, além dos 512 bytes de dados, mais algumas dezenas de bytes contendo os códigos ECC. A criação dos bytes de ECC, assim como sua utilização posterior é feita pela placa lógica, um processo automático que é feito de forma completamente transparente ao sistema operacional.

Quando um setor é lido pela cabeça de leitura, juntamente com os dados são lidos alguns dos códigos ECC, que visam apenas verificar se os dados que estão sendo lidos são os mesmos que foram gravados, uma técnica que lembra o sistema de paridade antigamente usado na memória RAM. Caso seja verificado um erro, são usados os demais códigos para tentar corrigir o problema. Na grande maioria dos casos, esta primeira tentativa é suficiente. Estes erros transitórios, que são corrigidos com a ajuda dos códigos ECC são chamados de "soft errors" e não causam nenhum efeito colateral além de um delay de alguns milissegundos na leitura.

Caso não seja possível corrigir o erro usando o ECC, a controladora faz uma nova tentativa de leitura do setor, pois é grande a possibilidade do erro ter sido causado por alguma interferência ou instabilidade momentânea. Caso o erro persista, ela fará várias tentativas sucessivas, reduzindo a velocidade de rotação dos discos e comparando o resultado de várias leituras, de forma a tentar recuperar os dados gravados no setor. Este processo gera aquele ruído característico de HD sendo "mastigado" e quase sempre indica o aparecimento de um bad block.

Por serem defeitos físicos na mídia magnética, não existe muito o que fazer com relação a eles. O jeito é marcar os bad blocks, de forma que eles não sejam mais usados.

Os HDs atuais são capazes de marcar automaticamente os setores defeituosos. A própria controladora faz isso, independentemente do sistema operacional. Existe uma área reservada no início do disco chamada "defect map" (mapa de defeitos) com alguns milhares de setores que ficam reservados para alocação posterior. Sempre que a controladora do HD encontra um erro ao ler ou gravar num determinado setor, ela remapeia o setor defeituoso, substituindo-o pelo endereço de um setor "bom", dentro do defect map. Como a alocação é feita pela própria controladora, o HD continua parecendo intacto para o sistema operacional.

De fato, é normal que os HDs já venham de fábrica com alguns setores remapeados, causados por pequenas imperfeições na superfície da mídia. Como eles não são visíveis para o sistema operacional, nem causam problemas no uso normal, acabam passando despercebidos.

Naturalmente, o defect map é uma área limitada, que corresponde normalmente a uma única trilha. Caso o HD possua algum problema crônico, eventualmente os endereços se esgotarão e os bad blocks realmente passarão a se tornar visíveis.

Deste ponto em diante, entram em cena utilitários como o scandisk (no Windows) e o badblocks (no Linux), que permitem realizar um exame de superfície, marcando os setores defeituosos encontrados. Estes setores não são marcados no defect map, mas sim em uma área reservada da partição.

Um grande número de setores defeituosos são indício de problemas graves, como envelhecimento da mídia, defeitos no mecanismo de leitura ou mesmo contaminação do HD por partículas provenientes do ambiente. O ideal nestes casos é fazer backup de todos os dados e substituir o HD o mais rápido possível.

Entretanto, mesmo para estes HDs condenados, às vezes existe uma solução. É comum a maioria dos setores aparecerem mais ou menos agrupados, englobando uma área relativamente pequena do disco. Se houverem muitos bad clusters em áreas próximas, você pode reparticionar o disco, isolando a área com problemas.

Se, por exemplo, você percebesse que a maioria dos defeitos se encontra nos últimos 20% do disco, bastaria abrir o particionador, deletar a partição atual e criar uma nova, englobando apenas 80% do disco. Neste caso, você perderia uma boa parte da área útil, mas pelo menos teria a possibilidade de continuar usando a parte "boa" do HD (em algum micro usado para tarefas secundárias, sem dados importantes), até que ele dê seus derradeiros suspiros.



## Entendendo as interfaces: IDE, SATA, SCSI e SAS

Assim como outros componentes, as interfaces usadas como meio de conexão para os HDs passaram por um longo caminho evolutivo.

As placas-mãe usadas nos primeiros PCs sequer possuíam interfaces de disco embutidas. Naquela época, as interfaces IDE ainda não existiam, de forma que novas interfaces eram vendidas junto com os HDs e instaladas em slots ISA disponíveis. A primeira interface foi criada pela Seagate, para uso em conjunto com o ST-506, um HD de 5 MB. Em seguida foi lançado o ST-412, de 10 MB. As duas interfaces são chamadas respectivamente de MFM e RLL devido ao método de codificação usado. Além da Seagate, estes HDs e interfaces foram produzidos também por outros fabricantes, como a Quantum e a Maxtor.

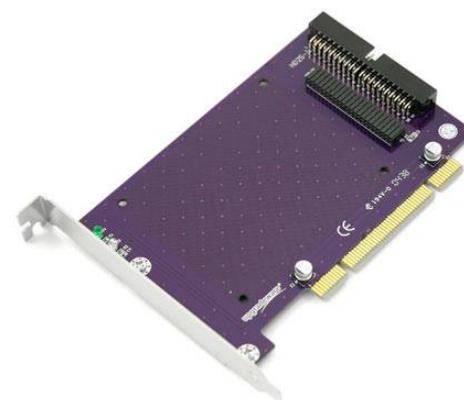
Em 1985 a Quantum lançou um produto bastante peculiar, o "Plus HardCard", que era um HD RLL de 20 MB onde tanto o HD, quanto a controladora eram integrados a uma única placa ISA e o "HD" era instalado diretamente no slot, sem ocupar uma das baias do gabinete.

Como você pode imaginar, estes HDs eram um tanto quanto problemáticos, pois a placa era presa por um único parafuso, o que causava problemas de vibração excessiva e barulho. Mesmo assim, estes HDs foram relativamente populares na época:



O padrão seguinte foi o ESDI (Enhanced Small Device Interface), criado por um consórcio de diversos fabricantes, incluindo a Maxtor. As interfaces ESDI ainda eram instaladas em slots ISA, mas trabalhavam a uma velocidade muito maior que as MFM e RLL, oferecendo um barramento teórico de 3 MB/s. É bem pouco para os padrões atuais, mas os HDs da época trabalhavam com taxas de transferências muito mais baixas, de forma que a velocidade acabava sendo mais do que satisfatória.

Quantum Plus HardCard





## IDE

Tanto as interfaces MFM e RLL, quanto a ESDI possuem algo em comum, que é o fato da controladora fazer parte da interface, e não ao próprio HD, como temos hoje em dia. Naturalmente, integrar a interface ao HD oferece diversas vantagens, pois elimina os problemas de sincronismo causados pelo uso de cabos longos e simplifica todo o design.

Não demorou para que os fabricantes percebessem isso. Surgiu então o padrão IDE "Integrated Drive Electronics" (que indica justamente o uso da controladora integrada), desenvolvido pela Quantum e Western Digital.

Os primeiros HDs e interfaces IDE chegaram ao mercado em 1986, mas inicialmente não existia um padrão bem definido, o que fez que os primeiros anos fossem marcados por problemas de compatibilidade entre os produtos dos diferentes fabricantes.

Em 1990 o padrão foi ratificado pelo ANSI, dando origem ao padrão ATA. Como o nome "IDE" já estava mais difundido, muita gente continuou usando o termo "IDE" e outros passaram a usar "IDE/ATA" ou simplesmente "ATA", fazendo com que os dois termos acabassem virando sinônimos.

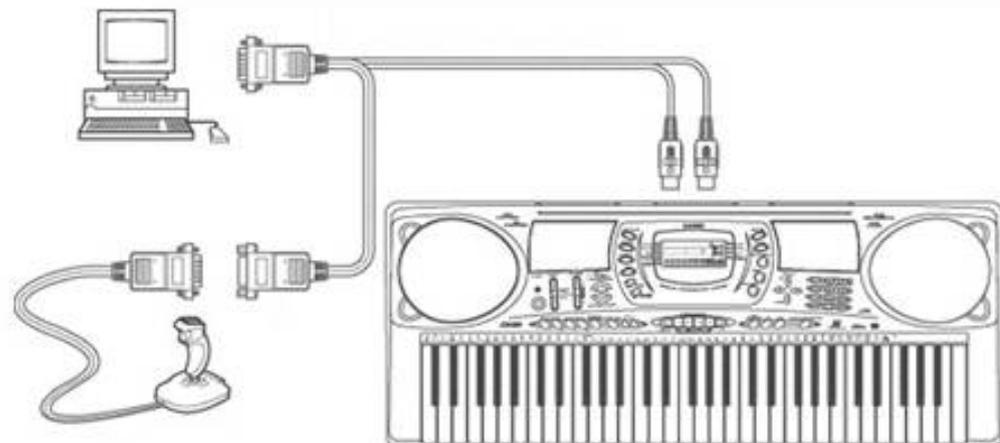
As primeiras placas IDE traziam apenas uma ou duas portas IDE e eram instaladas num slot ISA de 16 bits. Mas, logo os fabricantes passaram a integrar também outros conectores, dando origem às placas "super-ide", que eram usadas na grande maioria dos micros 386 e 486. As placas mais comuns incluíam uma porta IDE, uma porta FDD, duas portas seriais, uma paralela, além do e o conector do joystick.



*Controladora super IDE*

Como você pode ver, estas placas eram configuradas através de um conjunto de jumpers, já que na época ainda não existia plug-and-play :). Os jumpers permitiam configurar os endereços de IRQ, DMA e I/O usados, além de desativar os componentes individualmente. Se você precisasse de duas portas paralelas, por exemplo, utilizaria duas placas e configuraria uma delas para usar o IRQ 5 e endereço de I/O 378 e a outra para usar o IRQ 7 e o endereço de I/O 278.

A partir de um certo ponto, os fabricantes passaram a integrar os controladores diretamente no chipset da placa mãe, dando origem às placas com conectores integrados que conhecemos. A exceção ficou por conta do conector do joystick, que passou a ser integrado nas placas de som. Uma curiosidade é que o conector inclui também os pinos usados por dispositivos MIDI (como teclados musicais), que também são ligados no conector do joystick, através de um adaptador:

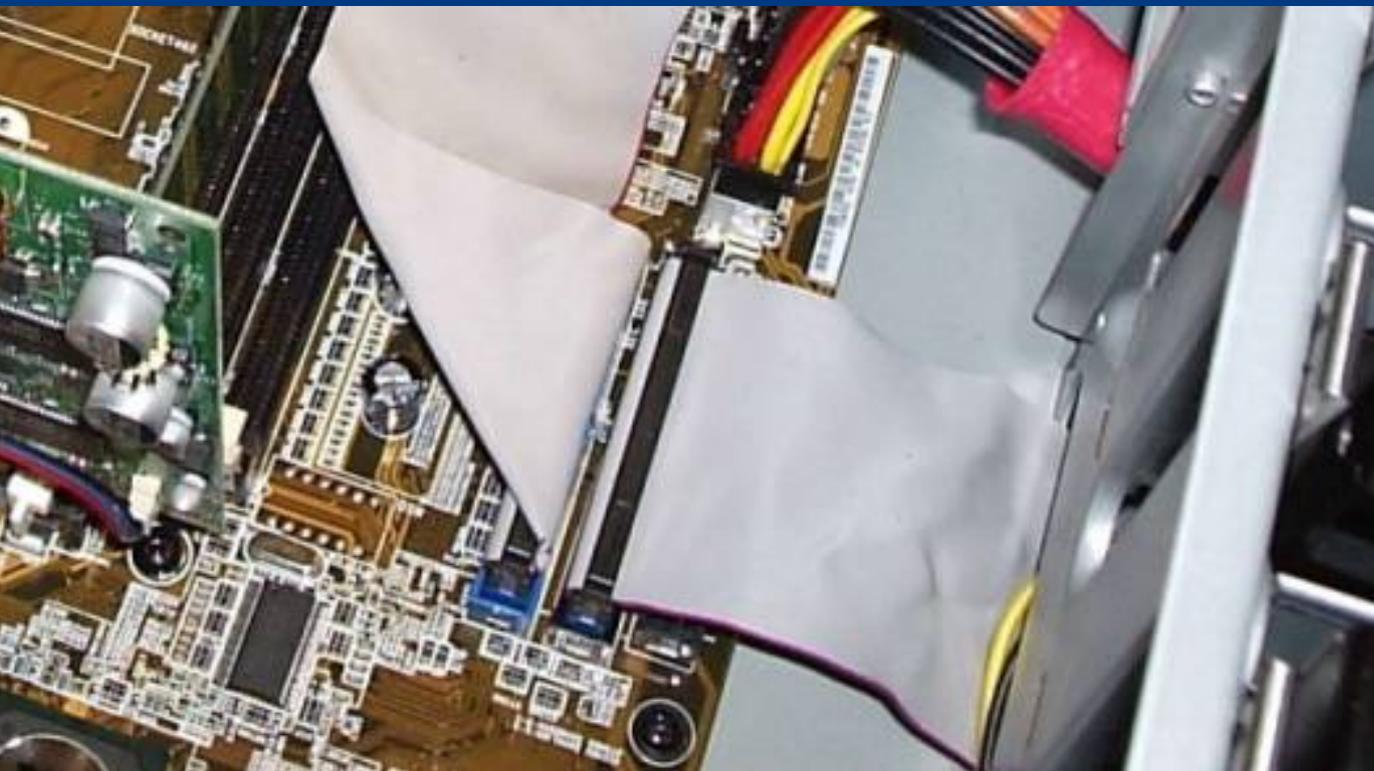


Inicialmente, as interfaces IDE suportavam apenas a conexão de HDs. Devido a isso, os primeiros drives de CD utilizavam interfaces proprietárias, incorporadas à placa de som, ou mesmo controladoras SCSI. Na época eram comuns os "kits multimídia", que incluíam o CD-ROM, placa de som, caixinhas e microfone.

Para solucionar o problema, foi desenvolvido o protocolo ATAPI (AT Attachment Packet Interface) que tornou-se rapidamente o padrão, riscando as interfaces proprietárias do mapa. É graças a ele que você pode comprar um drive de CD ou DVD e instalá-lo diretamente em uma das portas IDE, sem ter que comprar junto uma placa de som do mesmo fabricante :).

Na placa-mãe você encontra duas portas IDE (primária e secundária). Mesmo com a popularização das interfaces SATA, as portas IDE ainda continuam sendo incluídas nas placas recentes e devem demorar ainda mais alguns anos para desaparecerem completamente.

*Cada uma das portas permite instalar dois drives, de forma que podemos instalar um total de 4 HDs ou CD-ROMs na mesma placa*



Existem casos de placas mãe com 4 portas IDE (permitindo usar até 8 drives) e também controladoras IDE PCI, que incluem duas portas adicionais, que podem ser usadas em casos onde você precise usar mais do que 4 drives IDE no mesmo micro.

Para diferenciar os dois drives instalados na mesma porta, é usado um jumper, que permite configurar cada drive como master (mestre) ou slave.

Dois drives instalados na mesma porta compartilham o barramento oferecido por ela, o que acaba sempre causando uma pequena perda de desempenho. Por isso, quando são usados apenas dois drives (um HD e um CD-ROM, por exemplo), é preferível instalar cada um em uma das portas, deixando ambos jumpeados como master. Ao adicionar um terceiro, você poderia escolher entre instalar na primeira ou segunda porta IDE, mas, de qualquer forma, precisaria configurá-lo como slave, mudando a posição do jumper.

Usar cada drive em uma porta separada ajuda principalmente quando você precisa copiar grandes quantidades de dados de um HD para outro, ou gravar DVDs, já que cada drive possui seu canal exclusivo com o chipset.

No Windows, os drives são simplesmente identificados de forma seqüencial. O HD instalado como master da IDE primária apareceria no Windows Explorer como "C:" e o CD-ROM, instalado na IDE secundária como "D:", por exemplo.

Se você adicionasse um segundo HD, instalado como slave da primeira IDE, ele passaria a ser o "D:" e o CD-ROM o "E:".

*No Linux, os drives recebem endereços fixos, de acordo com a posição em que forem instalados:*

<b>IDE primária</b>	Master = /dev/hda
	Slave = /dev/hdb
<b>IDE secundária</b>	Master = /dev/hdc
	Slave = /dev/hdd

O cabo IDE possui três encaixes, um que é ligado na placa mãe e outro em cada dispositivo. Mesmo que você tenha apenas um dispositivo IDE, você deverá ligá-lo no conector da ponta, nunca no conector do meio. O motivo para isto, é que, ligando no conector do meio o cabo ficará sem terminação, fazendo com que os dados venham até o final do cabo e retornem na forma de interferência, prejudicando a transmissão.

*Como de praxe, as interfaces IDE/ATA passaram por um longo caminho evolutivo. As interfaces antigas, usadas em micros 386/486 e nos primeiros micros Pentium suportam (de acordo com seu nível de atualização), cinco modos de operação, que vão do PIO mode 0, ao PIO mode 4:*

<b>PIO mode 0</b>	3.3 MB/s
<b>PIO mode 1</b>	5.2 MB/s
<b>PIO mode 2</b>	8.3 MB/s
<b>PIO mode 3</b>	11.1 MB/s
<b>PIO mode 4</b>	16.6 MB/s

As mais recentes suportam também o Multiword DMA, que é um modo de acesso direto, onde o HD ou CD-ROM podem transferir dados diretamente para a memória, sem que o processador precise se envolver diretamente na transferência. O uso do DMA melhora bastante o desempenho e a responsividade do sistema, evitando que o micro "pare" enquanto um programa pesado está sendo carregado, ou durante a gravação de um CD, por exemplo.

Apesar disso, o Multiword DMA não chegou a ser muito usado, pois não era diretamente suportado pelo Windows 95, e os drivers desenvolvidos pelos fabricantes freqüentemente apresentavam problemas de estabilidade. Para piorar, muitos drives de CD e HDs antigos não funcionavam quando o DMA era ativado.

A solução veio com o padrão ATA-4, ratificado em 1998. Ele nada mais é do que o padrão Ultra ATA/33 (o nome mais popularmente usado) que é usado em placas para micros Pentium II e K6-2 fabricadas até 2000. Nele, a taxa de transferência máxima é de 33 MB/s e é suportado o modo UDMA 33, que permite transferências diretas para a memória também a 33 MB/s. É graças a ele que você pode assistir a filmes em alta resolução e DVDs no seu PC sem falhas

Você pode fazer uma experiência, desativando temporariamente o suporte a UDMA para o seu DVD-ROM para ver o que acontece.

No Linux, use o comando "hdparm -d0 /dev/dvd" (como root). No Windows, acesse o gerenciador de dispositivos, acesse as propriedades do drive e desmarque a opção referente ao DMA.

Tente agora assistir a um DVD. Você vai perceber que tanto o vídeo quanto o som ficam cheios de falhas, tornando a experiência bastante desagradável. Isso acontece por que, com o UDMA desativado, o processador precisa periodicamente parar o processamento do vídeo para ler mais dados no DVD. Quanto mais rápido o processador, mais curtas são as falhas, mas elas persistem mesmo num processador de 2 ou 3 GHz.

Para reverter, use o comando "hdparm -d1 /dev/dvd" ou marque novamente a opção do DMA, no caso do Windows.

Depois que o problema do DMA foi finalmente resolvido, os fabricantes se concentraram em aumentar a velocidade das portas. Surgiram então os padrões ATA-5 (Ultra ATA/66), ATA-6 (Ultra ATA/100) e ATA-7 (Ultra ATA/133), que é o usado atualmente.

Eles suportam (respectivamente), os modos UDMA 66, UDMA 100 e UDMA 133, além de manterem compatibilidade com os padrões anteriores.

Modo de Operação	Taxa de transferência:
<b>ATA-4 (Ultra ATA/33, UDMA 33)</b>	33 MB/s
<b>ATA-5 (Ultra ATA/66, UDMA 66)</b>	66 MB/s
<b>ATA-6 (Ultra ATA/100, UDMA 100)</b>	100 MB/s
<b>ATA-7 (Ultra ATA/133, UDMA 133)</b>	133 MB/s

As portas ATA/133 usadas nas placas atuais são uma necessidade por dois motivos. O primeiro é que os HDs atuais já superam a marca dos 70 ou 80 MB/s de taxa de transferência ao ler setores contínuos e a interface precisa ser substancialmente mais rápida que o HD, para absorver também as transferências feitas a partir do cache, que são bem mais rápidas. O segundo motivo é que só a partir das interfaces ATA/100 foi introduzido o suporte a HDs IDE com mais de 137 GB (decimais) de capacidade, como veremos em detalhes a seguir.

Para que os modos mais rápidos sejam utilizados, é necessário que exista também suporte por parte do HD e que o driver correto esteja instalado.

No caso do HD, não existe muito com o que se preocupar, pois os fabricantes são os primeiros a adotar novos modos de operação, de forma a manter seus produtos. Se você tem em mãos um HD antigo, que só suporta UDMA 33, por exemplo, pode ter certeza de que a taxa de transferência oferecida por ele é baixa, o que torna desnecessário o uso de uma interface mais rápida em primeiro lugar.

Ao contrário dos HDs, os drivers de CD e DVD ficaram estagnados no UDMA 33, pois como eles trabalham com taxas de transferência muito mais baixas, os padrões mais rápidos também não trazem vantagens. É possível que alguns fabricantes eventualmente passem a lançar drives "ATA/133", usando a interface mais rápida como ferramenta de marketing, mas isso não faria diferença alguma no desempenho.

Como de praxe, devo insistir na idéia de que a velocidade da interface determina apenas o fluxo de dados que ela pode transportar e não a velocidade real do dispositivo ligado a ela. Um CD-ROM de 52x lerá as mídias a no máximo 7.8 MB/s, independentemente da velocidade da interface. Funciona como numa auto-estrada: se houver apenas duas pistas para um grande fluxo de carros, haverá muitos congestionamentos, que acabarão com a duplicação da pista. Porém, a mesma melhora não será sentida caso sejam construídas mais faixas.

Continuando, junto com as interfaces Ultra ATA/66, veio a obrigatoriedade do uso de cabos IDE de 80 vias, substituindo os antigos cabos de 40 vias. Eles são fáceis de distinguir dos antigos, pois os fios usados no cabo são muito mais finos, já que agora temos o dobro deles no mesmo espaço:



A adição dos 40 fios adicionais é uma história interessante, pois eles não se destinam a transportar dados. Tanto os conectores, quanto os encaixes nos drives continuam tendo apenas 40 pinos, mantendo o mesmo formato dos cabos anteriores. Os 40 cabos adicionais são intercalados com os cabos de dados e servem como terras, reduzindo o nível de interferência entre eles. Este "upgrade" acabou sendo necessário, pois os cabos IDE de 40 vias foram introduzidos em 1986, projetados para transmitir dados a apenas 3.3 MB/s!

Os cabos de 80 vias são obrigatórios para o uso do UDMA 66 em diante. A placa mãe é capaz de identificar o uso do cabo de 80 vias graças ao pino 34, que é ligado de forma diferente.

Ao usar um cabo antigo, de 40 vias, a placa baixa a taxa de transmissão da interface, passando a utilizar o modo UDMA 33.

Veja que no caso dos CD-ROMs e DVDs, ainda é comum o uso dos cabos de 40 vias, simplesmente por que, como vimos, eles ainda utilizam o modo UDMA 33. Entretanto, se você precisar instalar um HD junto com o drive óptico, é interessante substituir o cabo por um de 80 vias, caso contrário o desempenho do HD ficará prejudicado.

Outra exigência trazida pelos novos padrões é o uso de cabos com no máximo 45 centímetros de comprimento, já que acima disso o nível de interferência e atenuação dos sinais passa a prejudicar a transmissão dos dados. O padrão ATA original (o de 1990) permitia o uso de cabos de até 90 centímetros (!) que não são mais utilizáveis hoje em dia, nem mesmo para a conexão do drive de CD/DVD.

Mais uma mudança introduzida pelos cabos de 80 vias é o uso de cores para diferenciar os três conectores do cabo. O conector azul deve ser ligado na placa mãe, o conector preto é ligado no drive configurado com master da interface, enquanto o conector do meio (cinza) é usado para a conexão do segundo drive, caso presente.

Os cabos de 80 vias também suportam o uso do sistema cable select (nos de 40 vias o suporte era opcional), onde a posição dos drives (master/slave) é determinada por em qual conector do cabo eles estão ligados, eliminando a possibilidade de conflitos, já que instalar dois drives configurados como master na mesma interface normalmente faz com que ambos deixem de ser identificados no setup.

*Para usar o cable select é preciso colocar os jumpers dos dois drives na posição "CS". Consulte o diagrama presente no topo ou na lateral do drive para ver a posição correta*



Os HDs IDE de 2.5", para notebooks utilizam um conector IDE miniaturizado, que possui 44 pinos. Os 4 pinos adicionais transportam energia elétrica, substituindo o conector da fonte usado nos HDs para desktop.

Existem ainda adaptadores que permitem instalar drives de 2.5" em desktops. Eles podem ser usados tanto em casos em que você precisa recuperar dados de um notebook com defeito, quanto quando quiser usar um HD de notebook no seu desktop para torná-lo mais silencioso.

Estes adaptadores são muito simples e baratos de fabricar, embora o preço no varejo varie muito, já que eles são um item relativamente raro:





## SATA

As interfaces IDE foram originalmente desenvolvidas para utilizar o barramento ISA, usado nos micros 286. Assim como no barramento ISA, são transmitidos 16 bits por vez e utilizados um grande número de pinos. Como é necessário manter a compatibilidade com os dispositivos antigos, não existe muita margem para mudanças dentro do padrão, de forma que, mesmo com a introdução do barramento PCI e do PCI Express, as interfaces IDE continuam funcionando fundamentalmente da mesma forma.

Mesmo quando foram introduzidas as interfaces UDMA, a única grande mudança foi a introdução dos cabos de 80 vias, desenvolvidos de forma a permitir taxas de transmissão maiores, sem contudo mudar o sistema de sinalização, nem mudar os conectores.

A partir de um certo ponto, ficou claro que o padrão IDE/ATA estava chegando a seu limite e que mudanças mais profundas só poderiam ser feitas com a introdução de um novo padrão. Surgiu então o SATA (Serial ATA).

Assim como o PCI Express, o SATA é um barramento serial, onde é transmitido um único bit por vez em cada sentido. Isso elimina os problemas de sincronização e interferência encontrados nas interfaces paralelas, permitindo que sejam usadas frequências mais altas.

Graças a isso, o cabo SATA é bastante fino, contendo apenas 7 pinos, onde 4 são usados para transmissão de dados (já que você precisa de 2 fios para fechar cada um dos dois circuitos) e 3 são terras, que ajudam a minimizar as interferências.



*Cabo e conector em um HD com interface SATA*

Os cabos SATA são bem mais práticos que os cabos IDE e não prejudicam o fluxo de ar dentro do gabinete. Os cabos podem ter até um metro de comprimento e cada porta SATA suporta um único dispositivo, ao contrário do padrão master/slave do IDE/ATA. Por causa disso, é comum que as placas mãe ofereçam 4 portas SATA (ou mais), com apenas as placas de mais baixo custo incluindo apenas duas.

No final, o ganho de desempenho permitido pela maior frequência de transmissão acaba superando a perda por transmitir um único bit por vez (ao invés de 16), fazendo com que, além de mais simples e barato, o padrão SATA seja mais rápido.

Existem três padrões de controladoras SATA, o SATA 150 (também chamado de SATA 1.5 Gbit/s ou SATA 1500), SATA 300 (SATA 3.0 Gbit/s ou SATA 3000) e também o padrão SATA 600 (ou SATA 6.0 Gbit/s), que ainda está em desenvolvimento. Como o SATA utiliza dois canais separados, um para enviar e outro para receber dados, temos 150 ou 300 MB/s em cada sentido, e não 133 MB/s compartilhados, como no caso das interfaces ATA/133.

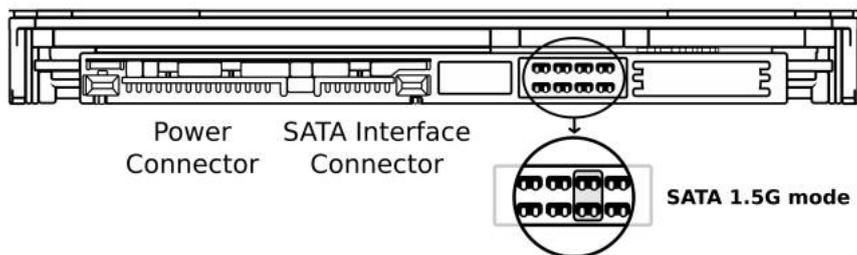
Os nomes SATA 300 e SATA 3000 indicam, respectivamente, a taxa de transferência, em MB/s e a taxa "bruta", em megabits. O SATA utiliza o sistema de codificação 8B/10B, o mesmo utilizado pelo barramento PCI Express, onde são adicionados 2 bits de sinalização para cada 8 bits de dados. Estes bits adicionais substituem os sinais de sincronismo utilizados nas interfaces IDE/ATA, simplificando bastante o design e melhorando a confiabilidade do barramento. Desta forma, a controladora transmite 3000 megabits, que, devido à codificação correspondem a apenas 300 megabytes. Ou seja, não é um arredondamento :).

As controladoras SATA 300 são popularmente chamadas de "SATA II", de forma que os dois termos acabaram virando sinônimos. Mas, originalmente, "SATA II" era o nome da associação de fabricantes que trabalhou no desenvolvimento dos padrões SATA (entre eles o SATA 300) e não o nome de um padrão específico.

Da mesma forma, o padrão de 600 MB/s chama-se SATA 600, e não "SATA III" ou "SATA IV". Mesmo os próprios fabricantes de HDs não costumam usar o termo "SATA II", já que ele é tecnicamente incorreto.

Outra curiosidade é que muitas placas mãe antigas, equipadas com controladoras SATA 150 (como as baseadas no chipset VIA VT8237 e também nas primeiras revisões dos chipsets SiS 760 e SiS 964), apresentam problemas compatibilidade com HDs SATA 300. Por causa disso, a maioria dos HDs atuais oferecem a opção de usar um "modo de compatibilidade" (ativado através de um jumper), onde o HD passa a se comportar como um dispositivo SATA 150, de forma a garantir a compatibilidade.

*Veja as instruções impressas na etiqueta de um HD da Samsung:*



If your S-ATA Controller or motherboard cannot support S-ATA 3.5 Gbps, change jumper pin setting with reference to the above picture

Inicialmente, os HDs e placas mãe com interfaces SATA eram mais caros, devido ao tradicional problema da escala de produção. Todo novo produto é inicialmente mais caro que a geração anterior simplesmente por que a produção é menor. A partir do momento em que passa a ser produzido em quantidade, os preços caem, até o ponto em que a geração anterior é descontinuada.

A partir do momento em que os HDs SATA se popularizaram, o preço caiu em relação aos IDE. Atualmente os HDs IDE são produzidos em escala cada vez menor e por isso se tornaram mais caros e mais difíceis de encontrar do que os HDs SATA.

No caso dos micros antigos, uma opção é instalar uma controladora SATA. As mais baratas, com duas portas e em versão PCI, já custam menos de 20 dólares no exterior e tendem a cair de preço também por aqui, tornando-se um item acessível, assim como as [controladoras USB](#). Note que o uso do barramento PCI limita a velocidade da controladora a 133 MB/s (um pouco menos na prática, já que o barramento PCI é compartilhado com outros dispositivos), mas isso não chega a ser um problema ao utilizar apenas um ou dois HDs.



Existem ainda conversores (chamados de bridges), que permitem ligar um HD IDE diretamente a uma porta SATA, mas eles são mais difíceis de encontrar e geralmente mais caros que uma controladora SATA PCI:



Com o lançamento do SATA, os HDs e controladoras IDE/ATA passaram a ser chamadas de "PATA", abreviação de "Parallel ATA", ressaltando a diferença.



## SCSI

As controladoras SCSI (pronuncia-se "iscâzi") são as tradicionais concorrentes das interfaces IDE. O primeiro padrão SCSI (SCSI 1) foi ratificado em 1986, na mesma época em que os primeiros HDs IDE chegaram ao mercado e consistiam em controladoras de 8 bits, que operavam a 5 MHz, oferecendo um barramento de dados de até 5 MB/s.

Em 1990, foi lançado o padrão Wide SCSI (SCSI 2). A frequência continuou a mesma, mas as controladoras passaram a utilizar um barramento de 16 bits, que dobrou a taxa de transmissão, que passou a ser de 10 MB/s.

Em seguida surgiram os padrões Fast SCSI (8 bits) e Fast Wide SCSI (16 bits), que operavam a 10 MHz e ofereciam taxas de transferência de, respectivamente 10 MB/s e 20 MB/s.

Em seguida surgiram os padrões Ultra SCSI (8 bits) e Ultra Wide SCSI (16 bits), que operavam a 20 MHz e ofereciam taxas de transferência de, respectivamente 20 MB/s e 40 MB/s.

A partir daí, surgiram os padrões Ultra SCSI (8 bits, 20 MHz = 20 MB/s), Wide Ultra SCSI (16 bits, 20 MHz = 40 MB/s), Ultra2 SCSI (8 bits, 40 MHz = 40 MB/s) e Wide Ultra2 SCSI (16 bits, 40 MHz = 80 MB/s). Veja que até a evolução foi bastante previsível, com um novo padrão simplesmente dobrando a frequência e, conseqüentemente, a taxa de transferência do anterior.

A partir daí, o uso de controladoras de 8 bits foi abandonado e surgiram os padrões Ultra160 SCSI, onde a controladora operava a 40 MHz, com duas transferências por ciclo, resultando num barramento de 160 MB/s e no Ultra 320 SCSI, que mantém as duas transferências por ciclo, mas aumenta a frequência para 80 MHz, atingindo 320 MB/s.

Além da diferença na velocidade, as antigas controladoras de 8 bits permitiam a conexão de apenas 7 dispositivos, enquanto as atuais, de 16 bits, permitem a conexão de até 15.

Diferentemente do que temos numa interface IDE, onde um dispositivo é jumpeado como master e outro como slave, no SCSI os dispositivos recebem números de identificação (IDs) que são números de 0 a 7 (nas controladoras de 8 bits) e de 0 a 15 nas de 16 bits. Um dos IDs disponíveis é destinado à própria controladora, deixando 7 ou 15 endereços disponíveis para os dispositivos.

O ID de cada dispositivo é configurado através de uma chave ou jumper, ou (nos mais atuais), via software. A regra básica é que dois dispositivos não podem utilizar o mesmo endereço, caso contrário você tem um conflito similar ao que acontece ao tentar instalar dois HDs jumpeados como master na mesma porta IDE :).

A maioria dos cabos SCSI possuem apenas 3 ou 4 conectores, mas existem realmente cabos com até 16 conectores, usados quando é realmente necessário instalar um grande número de dispositivos.



Modelo	Controladora de 8 Bits (Narrow SCSI)	Controladora de 16 Bits (Wide SCSI)
<b>SCSI 1</b>	5 MB/s	10 MB/s
<b>Fast SCSI (SCSI-2)</b>	10 MB/s	20 MB/s
<b>Ultra SCSI (SCSI-3)</b>	20 MB/s	40 MB/s
<b>Ultra2 SCSI (SCSI-4)</b>	40 MB/s	80 MB/s
<b>Ultra160 SCSI</b>	-	160 MB/s
<b>Ultra320 SCSI</b>	-	320 MB/s

No barramento SCSI temos também o uso de terminadores, que efetivamente "fecham" o barramento, evitando que os sinais cheguem à ponta do cabo e retornem na forma de interferência. Na maioria dos casos o terminador é encaixado no dispositivo, mas em alguns casos basta mudar a posição de uma chave. Também existem casos de cabos que trazem um terminador pré-instalado na ponta.



*Terminador SCSI*

Note que estou usando o termo "dispositivos" e não "HDs", pois (embora raro hoje em dia) o padrão SCSI permite a conexão de diversos tipos de dispositivos, incluindo CD-ROMs, impressoras, scanners e unidades de fita.

Os gravadores de CD SCSI foram populares nos anos 90, pois o barramento SCSI oferece transferências mais estáveis que as antigas portas ATA-2 e ATA-3, usadas até então. Naquela época ainda não existia burn-free, de forma que qualquer interrupção no fluxo de dados causava a perda da mídia. Com o surgimento das interfaces IDE com suporte a UDMA, a briga se equilibrou e os gravadores de CD IDE invadiram o mercado.

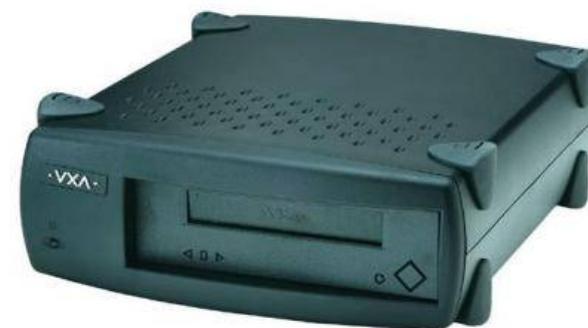
As impressoras e scanners SCSI também ganharam algumas batalhas, mas acabaram perdendo a guerra para os dispositivos USB.

As unidades de fita já foram o meio mais popular para fazer backup de grandes quantidades de dados, utilizando as famosas fitas DAT. Como a fita precisa ser gravada e lida seqüencialmente, o mais comum é gerar um arquivo compactado em .tar.gz, .tar.bz2, ou mesmo em .rar, contendo todos os arquivos do backup e gravá-lo na fita, de forma seqüencial. Um arquivo muito grande pode ser dividido em vários volumes e gravado em fitas separadas. O grande problema é que é preciso ler e descompactar todo o arquivo para ter acesso aos dados.

O problema com as unidades de fita é que, embora as fitas sejam relativamente baratas, as unidades de gravação são vendidas por preços salgados. Conforme os HDs foram crescendo em capacidade e caindo em custo, eles passaram a oferecer um custo por megabyte mais baixo, fazendo com que os sistemas RAID e servidores de backup se popularizassem roubando o mercado das unidades de fita.

Um drive VXA-320 da Exabyte, por exemplo, custa US\$ 1.250 e utiliza fitas de apenas 160 GB. É comum que os fabricantes dobrem a capacidade, dizendo que as fitas armazenam "320 GB comprimidos", mas a taxa de compressão varia de acordo com o tipo de dados.

A velocidade de gravação também é relativamente baixa, em torno de 12 MB/s (cerca de 43 GB reais por hora) e cada fita custa US\$ 80, o que dá um custo de US\$ 0.50 por GB. Como hoje em dia um HD de 300 GB custa (no Brasil) menos de R\$ 250, a unidade de fita simplesmente perde em todos os quesitos, incluindo confiabilidade e custo por megabyte. Ao invés de utilizar a unidade de fita, acaba sendo mais prático, rápido e barato fazer os backups usando HDs externos.



*Exabyte VXA-320*

Chegamos então na questão dos cabos. O SCSI permite tanto a conexão de dispositivos internos, quanto de dispositivos externos, com o uso de cabos e conectores diferentes para cada tipo. As controladoras de 8 bits utilizam cabos de 50 vias, enquanto as de 16 bits utilizam cabos de 68 vias. Este da foto é um **HD Ultra320 SCSI**, que utiliza o conector de 68 pinos.



As controladoras SCSI são superiores às interfaces IDE em quase todos os quesitos, mas perdem no mais importante, que é a questão do custo. Como a história da informática repetidamente nos mostra, nem sempre o padrão mais rápido ou mais avançado prevalece. Quase sempre, um padrão mais simples e barato, que consegue suprir as necessidades básicas da maior parte dos usuários, acaba prevalecendo sobre um padrão mais complexo e caro.

De uma forma geral, o padrão IDE tornou-se o padrão nos desktops e também nos servidores e estações de trabalho de baixo custo, enquanto o SCSI tornou-se o padrão dominante nos servidores e workstations de alto desempenho. Em volume de vendas, os HDs SCSI perdem para os IDE e SATA numa proporção de mais de 30 para 1, mas ainda assim eles sempre representaram uma fatia considerável do lucro líquido dos fabricantes, já que representam a linha "premium", composta pelos HDs mais caros e de mais alto desempenho.

É comum que novas tecnologias sejam inicialmente usadas em HDs SCSI sendo somente utilizadas nos discos IDE depois de tornarem-se mais baratas. Isto acontece justamente por causa do mercado de discos SCSI, que prioriza o desempenho muito mais do que o preço.

Além do custo dos HDs, existe também a questão da controladora. Algumas placas destinadas a servidores trazem controladoras SCSI integradas, mas na grande maioria dos casos é necessário comprar uma controladora separada.

As controladoras Ultra160 e Ultra320 seriam subutilizadas caso instaladas em slots PCI regulares (já que o PCI é limitado a 133 MB/s), de forma que elas tradicionalmente utilizam slots PCI-X, encontrados apenas em placas para servidores. Isto significa que mesmo que você quisesse, não poderia instalar uma controladora Ultra320 em seu desktop. Apenas mais recentemente passaram a ser fabricadas **controladoras PCI-Express**.



Como de praxe, vale lembrar que a velocidade da interface não corresponde diretamente à velocidade dos dispositivos a ela conectados. Os 320 MB/s do Ultra320 SCSI, por exemplo, são aproveitados apenas ao instalar um grande número de HDs em RAID.

Existem muitas lendas com relação ao SCSI, que fazem com que muitos desavisados comprem interfaces e HDs obsoletos, achando que estão fazendo o melhor negócio do mundo. Um HD não é mais rápido simplesmente por utilizar uma interface SCSI. É bem verdade que os HDs mais rápidos, de 15.000 RPM, são lançados apenas em versão SCSI, mas como os HDs ficam rapidamente obsoletos e tem uma vida útil limitada, faz muito mais sentido comprar um HD SATA convencional, de 7.200 ou 10.000 RPM, do que levar pra casa um HD SCSI obsoleto, com 2 ou 3 anos de uso.



## SAS

Com a introdução do Serial ATA, o barramento SCSI perdeu grande parte de seus atrativos, já que o SATA oferece uma grande parte das vantagens que antes eram atribuídas ao SCSI e, ao mesmo tempo, oferece um sistema de cabeamento mais simples.

Para preencher a lacuna, surgiu o SAS (Serial Attached SCSI), um barramento serial, muito similar ao SATA em diversos aspectos, que adiciona diversas possibilidades interessantes voltadas para uso em servidores. Ele preserva o mesmo conjunto de comandos e por isso é compatível a nível de software. Não estou falando aqui do Windows e programas como os que utilizamos em desktops, mas sim de aplicativos personalizados, complexos e caros, utilizados em grandes servidores.

Assim como o SCSI conviveu com o padrão IDE por mais de duas décadas, o SAS está destinado a concorrer com o SATA, com cada um entrancheirado em seu respectivo nicho: o SATA nos micros domésticos e servidores de baixo custo e o SAS em servidores maiores e estações de trabalho.

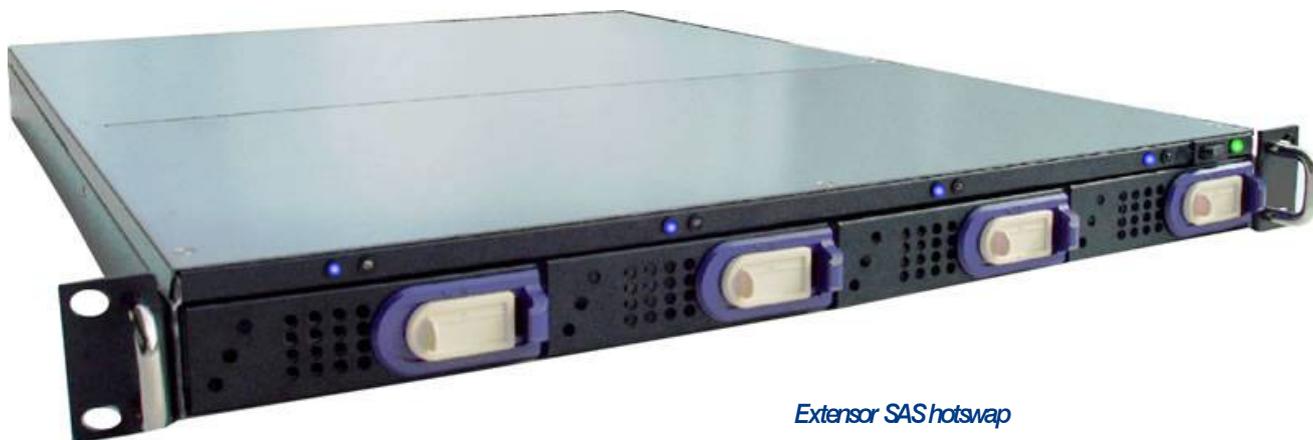
As versões iniciais do SAS suportavam taxas de transferência de 150 e 300 MB/s. Recentemente foi introduzido o padrão de 600 MB/s e passou a ser desenvolvido o padrão seguinte, de 1.2 GB/s.

A evolução é similar à do padrão SATA (note que as velocidades são as mesmas), porém o SAS tende a ficar sempre um degrau à frente.

A maior velocidade é necessária, pois o SAS permite o uso de extensores (expanders), dispositivos que permitem ligar diversos discos SAS a uma única porta. Existem dois tipos de extensores SAS, chamados de "Edge Expanders" e "Fanout Expanders". Os Edge Expanders permitem ligar até 128 discos na mesma porta, enquanto os Fanout Expanders permitem conectar até 128 Edge Expanders (cada um com seus 128 discos!), chegando a um limite teórico de até 16.384 discos por porta SAS.

Este recurso foi desenvolvido pensando sobretudo nos servidores de armazenamento. Com a popularização dos webmails e outros serviços, o armazenamento de grandes quantidades de dados tornou-se um problema. Não estamos falando aqui de alguns poucos gigabytes, mas sim de vários terabytes ou mesmo petabytes de dados. Imagine o caso do Gmail, por exemplo, onde temos vários milhões de usuários, cada um com mais de 2 GB de espaço disponível.

Os extensores SAS normalmente possuem a forma de um gabinete 1U ou 2U, destinados a serem instalados nos mesmos racks usados pelos próprios servidores. Em muitos, os discos são instalados em gavetas removíveis e podem ser trocados "a quente" (hotswap), com o servidor ligado. Isto permite substituir rapidamente HDs defeituosos, sem precisar desligar o servidor.



*Extensor SAS hotswap*

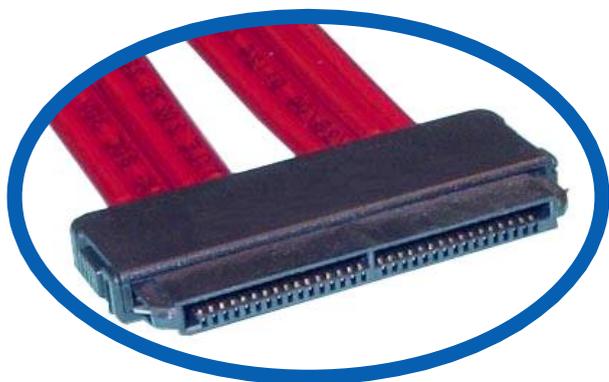
Nestes casos, seria utilizado um sistema RAID, onde parte do espaço e armazenamento é destinado a armazenar informações de redundância, que permitem restaurar o conteúdo de um HD defeituoso assim que ele é substituído, sem interrupção ou perda de dados. Ao contrário das controladoras RAID de baixo custo, encontradas nas placas mãe para desktop, que executam suas funções via software, as controladoras SAS tipicamente executam todas as funções via hardware, facilitando a configuração (já que deixa de ser necessário instalar drivers adicionais) e oferecendo um maior desempenho e flexibilidade.

Outra pequena vantagem é que o SAS permite o uso de cabos de até 6 metros, contra apenas 1 metro no SATA. A maior distância é necessária ao conectar um grande número de extensores, já que eles são grandes e os últimos tendem a ficar fisicamente afastados do servidor.

As **controladoras SAS** incluem normalmente 4 ou 8 portas e são instaladas num slot PCI-X, ou PCI Express. Nada impede também que você instale duas ou até mesmo três controladoras no mesmo servidor caso precise de mais portas. Algumas placas-mãe destinadas a servidores já estão vindo com controladoras SAS onboard, reduzindo o custo.



Assim como a maioria das controladoras SAS, esta Adaptec da foto utiliza um único conector SFF 8484, ao invés de 4 conectores separados. Ele simplifica um pouco a instalação, mas na prática não muda muita coisa, pois o conector dá origem aos 4 cabos separados da mesma forma:



Um detalhe interessante é que o padrão SAS oferece compatibilidade retroativa com os HDs SATA, permitindo que você use HDs SATA convencionais como uma forma de cortar custos, sem ter que abrir mão da possibilidade de usar os extensores.

A relação, entretanto, não é recíproca: embora o conector seja o mesmo, HDs SAS não são reconhecidos caso instalados numa porta SATA convencional, pois eles utilizam comandos específicos, que vão bem além do conjunto suportado pelas controladoras SATA.

De qualquer forma, os HDs SAS são mais caros e não oferecem vantagens em termos de desempenho, de forma

que você dificilmente iria querer utilizar um em seu desktop, de qualquer forma. O principal diferencial é que eles são certificados para operação contínua e possuem garantias maiores, geralmente de 5 anos.



A maior parte dos HDs de alto desempenho, com rotação de 15.000 RPM, que antes só existiam em versão SCSI, estão sendo lançados também em versão SAS. Nos próximos anos é de se esperar que o SAS substitua gradualmente o SCSI, assim como o SATA já substituiu o IDE quase que completamente nos micros novos.

Não existe nada de fundamentalmente diferente, que impeça que estes drives de alto desempenho sejam lançados também em versão SATA, o problema reside unicamente na questão da demanda.

Por serem caros e possuírem capacidades reduzidas (devido ao uso de discos de 2.5"), os HDs de 15.000 RPM acabam não sendo muito adequados para o público doméstico. Você dificilmente pagaria R\$ 1500 por um HD de 73 GB (como Seagate Cheetah 15K.4), por mais rápido que ele fosse, quando pode comprar um HD SATA de 300 GB por menos de R\$ 250. Esta brutal diferença de custo acaba sendo justificável apenas no mercado de servidores de alto desempenho e workstations, onde, literalmente, "tempo é dinheiro".



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.

# APRENDA A LER EM INGLÊS

Aprender a ler em inglês é rápido e simples. Você não precisa sofrer de ansiedade ao encontrar um texto em inglês, necessário para sua profissão ou estudo. Ler em inglês é muito fácil, basta dominar as técnicas certas.

Conheça o método de aprendizado da leitura da língua inglesa desenvolvido pelo Instituto de Desenvolvimento do Potencial Humano.

Baixe gratuitamente uma aula de teste e comprove na prática como é simples aprender a ler em inglês

[http://www.idph.net/aula\\_inicial.zip](http://www.idph.net/aula_inicial.zip)

Saiba mais sobre o curso e faça sua inscrição:

<http://www.idph.net/matricula.shtml>

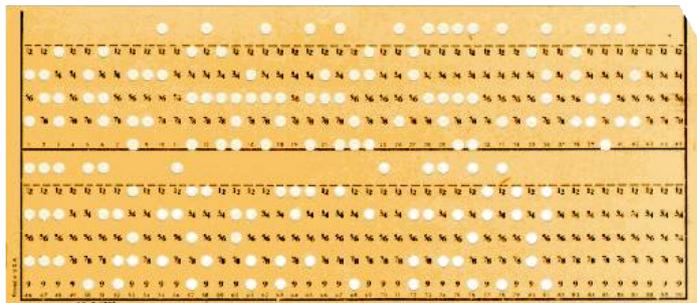
# *Especial* **Redes**

Por **Carlos E. Morimoto**

Inicialmente, as redes eram simplesmente uma forma de transmitir dados de um micro a outro, substituindo o famoso DPL/DPC (disquete pra lá, disquete pra cá), usado até então.

As primeiras redes de computadores foram criadas ainda durante a década de 60, como uma forma de transferir informações de um computador a outro. Na época, o meio mais usado para armazenamento externo de dados e transporte ainda eram os cartões perfurados, que armazenavam poucas dezenas de caracteres cada (o formato usado pela IBM, por exemplo, permitia armazenar 80 caracteres por cartão).

Eles são uma das formas mais lentas, trabalhosas e demoradas de transportar grandes quantidades de informação que se pode imaginar. São, literalmente, cartões de cartolina com furos, que representam os bits um e zero armazenados:



Cartão perfurado

De 1970 a 1973 foi criada a Arpanet, uma rede que interligava várias universidades e diversos órgãos militares. Nesta época surgiu o e-mail e o FTP, recursos que utilizamos até hoje. Ainda em 1973 foi feito o primeiro teste de transmissão de dados usando o padrão Ethernet, dentro do PARC (o laboratório de desenvolvimento da Xerox, em Palo Alto, EUA). Por sinal, foi no PARC onde várias outras tecnologias importantes, incluindo a interface gráfica e o mouse, foram originalmente desenvolvidas.

O padrão Ethernet é utilizado pela maioria das tecnologias de rede local em uso, das placas mais baratas às redes wireless. O padrão Ethernet define a forma como os dados são organizados e transmitidos. É graças a ele que placas de diferentes fabricantes funcionam perfeitamente em conjunto.

## Placas, cabos, conectores, hubs e switches

Os componentes básicos de uma rede são uma **placa de rede** para cada micro, os cabos e um **hub** ou **switch**, que serve como um ponto de encontro, permitindo que todos os micros se enxerguem e conversem entre si. Juntos, esses componentes fornecem a infra-estrutura básica da rede, incluindo o meio físico para a transmissão dos dados, modulação dos sinais e correção de erros.

## Placas de rede

As placas de rede já foram componentes caros. Mas, como elas são dispositivos relativamente simples e o funcionamento é baseado em padrões abertos, qualquer um com capital suficiente pode abrir uma fábrica de placas de rede. Isso faz com que exista uma concorrência acirrada, que obriga os fabricantes a produzirem placas cada vez mais baratas, trabalhando com margens de lucro cada vez mais estreitas.

As placas de rede mais baratas chegam a ser vendidas no atacado por menos de 3 dólares. O preço final é um pouco mais alto, naturalmente, mas não é difícil achar placas por 20 reais ou até menos, sem falar que, hoje em dia, praticamente todas as placas-mãe vendidas possuem rede onboard, muitas vezes duas, já pensando no público que precisa compartilhar a conexão.

No começo da década de 90, existiam três padrões de rede, as redes Arcnet, Token Ring e Ethernet. As redes Arcnet tinham problemas de desempenho e as Token Ring eram muito caras, o que fez com que as redes Ethernet se tornassem o padrão definitivo. Hoje em dia, "Ethernet" é quase um sinônimo de placa de rede. Até mesmo as placas wireless são placas Ethernet.

Lembre-se que **Ethernet** é o nome de um padrão que diz como os dados são transmitidos. Todas as placas que seguem este padrão são chamadas de placas Ethernet. Não estamos falando de uma marca ou de um fabricante específico.

Temos aqui alguns exemplos de **placas de rede**. O conector para o cabo é chamado de "RJ45" e o soquete vago permite instalar um chip de boot.



Existem três padrões de redes Ethernet (com fio): de 10 megabits, 100 megabits e 1000 megabits (também chamadas de Gigabit Ethernet). Já estão disponíveis também as redes de 10 gigabits, mas por enquanto elas ainda são muito caras, pois utilizam placas específicas e cabos de fibra óptica. Esses três padrões são intercompatíveis: você pode perfeitamente misturar placas de 100 e 1000 megabits na mesma rede, mas, ao usar placas de velocidades diferentes, a velocidade é sempre nivelada por baixo, ou seja, as placas Gigabit são obrigadas a respeitar a velocidade das placas mais lentas.

As redes de 10 megabits estão em desuso já a vários anos e tendem a se extinguir com o tempo. As de 100 megabits são o padrão (por enquanto), pois são muito baratas e propiciam uma velocidade suficiente para transmitir grandes arquivos e rodar aplicativos remotamente.

Tudo o que a placa de rede faz é transmitir os uns e zeros enviados pelo processador através do cabo de rede, de forma que a transmissão seja recebida pelos outros micros. Ao transferir um arquivo, o processador lê o arquivo gravado no HD e o envia à placa de rede para ser transmitido.

Os HDs atuais são capazes de ler dados a 30 ou 40 MB por segundo. Lembre-se que um byte tem 8 bits, logo 30 MB (megabytes, com o B maiúsculo) correspondem a 240 megabits (Mb, com o b minúsculo) e assim por diante.

Se você dividir 100 megabits por 8, terá 12.5 megabytes por segundo. É bem menos do que um HD atual é capaz, mas já é uma velocidade razoável. No que depender da rede, demora cerca de um minuto para copiar um CD inteiro, por exemplo.

A opção para quem precisa de mais velocidade são as redes Gigabit Ethernet, que transmitem a uma taxa de até 1000 megabits (125 megabytes) por segundo. As placas Gigabit atuais são compatíveis com os mesmos cabos de par trançado cat 5 usados pelas placas de 100 megabits (veja mais detalhes a seguir), por isso a diferença de custo fica por conta apenas das placas e do switch. Graças a isso elas estão caindo de preço e se popularizando rapidamente.

## Tipos de cabos de rede

Existem basicamente 3 tipos diferentes de cabos de rede: os cabos de par trançado (que são, de longe, os mais comuns), os cabos de fibra óptica (usados principalmente em links de longa distância) e os cabos coaxiais, ainda usados em algumas redes antigas.

Existem vários motivos para os cabos coaxiais não serem mais usados hoje em dia: eles são mais propensos a mal contato, os conectores são mais caros e os cabos são menos flexíveis que os de par trançado, o que torna mais difícil passá-los por dentro de tubulações.

No entanto, o principal motivo é o fato de que eles podem ser usados apenas em redes de 10 megabits: a partir do momento em que as redes 10/100 tornaram-se populares, eles entraram definitivamente em desuso, dando lugar aos cabos de par trançado. Entre eles, os que realmente usamos no dia-a-dia são os cabos "cat 5" ou "cat 5e", onde o "cat" é abreviação de "categoria" e o número indica a qualidade do cabo.

Fabricar cabos de rede é mais complicado do que parece. Diferente dos cabos de cobre comuns, usados em instalações elétricas, os cabos de rede precisam suportar frequências muito altas, causando um mínimo de atenuação do sinal. Para isso, é preciso minimizar ao máximo o aparecimento de bolhas e impurezas durante a fabricação dos cabos. No caso dos cabos de par trançado, é preciso, ainda, cuidar do entrançamento dos pares de cabos, que também é um fator crítico.

Existem cabos de cat 1 até cat 7. Como os cabos cat 5 são suficientes tanto para redes de 100 quanto de 1000 megabits, eles são os mais comuns e mais baratos; geralmente custam em torno de 1 real o metro. Os cabos cat5e (os mais comuns atualmente) seguem um padrão um pouco mais estrito, por isso dê preferência a eles na hora de comprar.

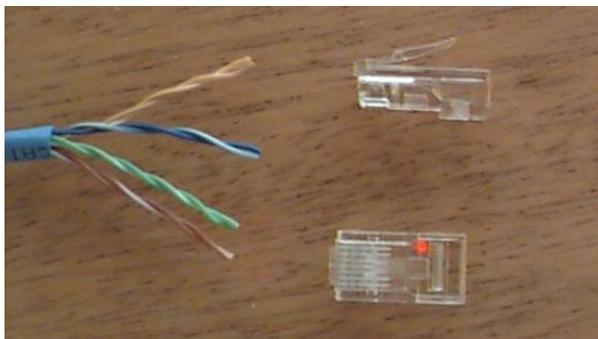
Em caso de dúvida, basta checar as inscrições decalcadas no cabo.



*Destaque para a inscrição no cabo indicando a categoria*

Você pode comprar alguns metros de cabo e alguns conectores e crimpar os cabos você mesmo, ou pode comprá-los já prontos. Em ambos os casos, os cabos devem ter no mínimo 30 centímetros e no máximo 100 metros, a distância máxima que o sinal elétrico percorre antes que comece a haver uma degradação que comprometa a comunicação.

Naturalmente, os 100 metros não são um número exato. A distância máxima que é possível atingir varia de acordo com a qualidade dos cabos e conectores e as interferências presentes no ambiente. Já vi casos de cabos de 180 metros que funcionavam perfeitamente, e casos de cabos de 150 que não. Ao trabalhar fora do padrão, os resultados variam muito de acordo com as placas de rede usadas e outros fatores. Ao invés de jogar com a sorte, é mais recomendável seguir o padrão, usando um hub/switch ou um repetidor a cada 100 metros, de forma a reforçar o sinal.



Comprar os cabos já prontos é muito mais prático, mas tem alguns inconvenientes. Muita gente (a maioria, acredito :) não acha muito legal ver cabos espalhados pelo chão da sala. Alguns desavisados chegam a tropeçar neles, derubando micros, quebrando os conectores das placas de rede, entre outros acidentes desagradáveis.

Para dar um acabamento mais profissional, você precisa passar os cabos por dentro das tubulações das paredes ou pelo teto e é mais fácil passar o cabo primeiro e crimpar o conector depois do que tentar fazer o contrário. Se preferir crimpar o cabo você mesmo, vai precisar comprar também um alicate de crimpagem. Ele "esmaga" os contatos do conector, fazendo com que eles entrem em contato com os fios do cabo de rede.

*Alicate de crimpagem*



Os cabos de rede transmitem sinais elétricos a uma frequência muito alta e a distâncias relativamente grandes, por isso são muito vulneráveis a interferências eletromagnéticas externas.

Além dos cabos sem blindagem, conhecidos como UTP (Unshielded Twisted Pair), existem os cabos blindados conhecidos como STP (Shielded Twisted Pair). A única diferença entre eles é que os cabos blindados, além de contarem com a proteção do entrelaçamento dos fios, possuem uma blindagem externa (assim como os cabos coaxiais) e por isso são mais adequados a ambientes com fortes fontes de interferências, como grandes motores elétricos ou grandes antenas de transmissão muito próximas.

Quanto maior for o nível de interferência, menor será o desempenho da rede, menor será a distância que poderá ser usada entre os micros e mais vantajosa será a instalação de cabos blindados. Em ambientes normais, porém, os cabos sem blindagem funcionam perfeitamente bem. Na ilustração temos um exemplo de *cabo com blindagem*, com proteção individual para cada par de cabos. Existem também cabos mais "populares", que utilizam apenas uma blindagem externa que envolve todos os cabos.



Outras fontes menores de interferências são as lâmpadas fluorescentes (principalmente lâmpadas cansadas, que ficam piscando), cabos elétricos, quando colocados lado a lado com os cabos de rede, e mesmo telefones celulares muito próximos dos cabos. Este tipo de interferência não chega a interromper o funcionamento da rede, mas pode causar perda de pacotes.

No final de cada pacote TCP são incluídos 32 bits de CRC, que permitem verificar a integridade dos dados. Ao receber cada pacote, a estação verifica se a soma dos bits "bate" com o valor do CRC. Sempre que a soma der errado, ela solicita a retransmissão do pacote, o que é repetido indefinidamente, até que ela receba uma cópia intacta. Graças a esse sistema é possível transmitir dados de forma confiável mesmo através de links ruins (como, por exemplo, uma conexão via modem). Porém, quanto mais intensas forem as interferências, mais pacotes precisam ser retransmitidos e pior é o desempenho da rede.

## Crimpando os cabos

Ao crimpar os cabos de rede, o primeiro passo é descascar os cabos, tomando cuidado para não ferir os fios internos, que são frágeis. Normalmente, o alicate inclui uma saliência no canto da guilhotina, que serve bem para isso. Existem também descascadores de cabos específicos para cabos de rede.

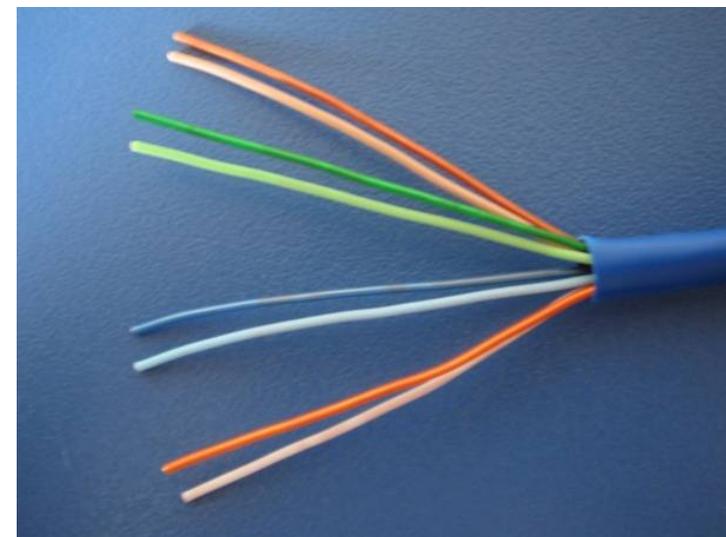


*Saliência no canto da guilhotina para descascar o cabo.*

Os quatro pares do cabo são diferenciados por cores. Um par é laranja, outro é azul, outro é verde e o último é marrom. Um dos cabos de cada par tem uma cor sólida e o outro é mais claro ou malhado, misturando a cor e pontos de branco. É pelas cores que diferenciamos os 8 fios.

O segundo passo é destrançar os cabos, deixando-os soltos. É preciso organizá-los em uma certa ordem para colocá-los dentro do conector e é meio complicado fazer isso se eles estiverem grudados entre si :-P.

Eu prefiro descascar um pedaço grande do cabo, uns 6 centímetros, para poder organizar os cabos com mais facilidade e depois cortar o excesso, deixando apenas os 2 centímetros que entrarão dentro do conector. O próprio alicate de crimpagem inclui uma guilhotina para cortar os cabos, mas você pode usar uma tesoura se preferir.



*Pares do cabo separados, prontos para serem organizados*

Existem dois padrões para a ordem dos fios dentro do conector, o EIA 568B (o mais comum) e o EIA 568A. A diferença entre os dois é que a posição dos pares de cabos laranja e verde são invertidos dentro do conector.

Existe muita discussão em relação com qual dos dois é "melhor", mas na prática não existe diferença de conectividade entre os dois padrões. A única observação é que você deve cabear toda a rede utilizando o mesmo padrão. Como o EIA 568B é de longe o mais comum, recomendo-o que você utilize-o ao crimpar seus próprios cabos.

No padrão EIA 568B, a ordem dos fios dentro do conector (em ambos os lados do cabo) é a seguinte:

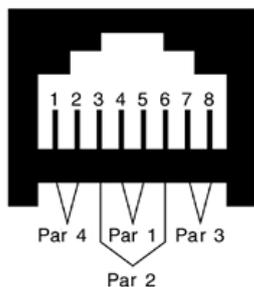
- ● 1- Branco com Laranja
- 2- Laranja
- ● 3- Branco com Verde
- 4- Azul
- ● 5- Branco com Azul
- 6- Verde
- ● 7- Branco com Marrom
- 8- Marrom

Os cabos são encaixados nesta ordem, com a trava do conector virada para baixo, como no diagrama.



Ou seja, se você olhar o conector "de cima", vendo a trava, o par de fios laranja estará à direita e, se olhar o conector "de baixo", vendo os contatos, eles estarão à esquerda.

Este outro diagrama mostra melhor como fica a posição dos cabos dentro do conector:



No caso de um cabo "reto" (straight), que vai ser usado para ligar o micro ao hub, você usa esta mesma disposição nas duas pontas do cabo. Existe ainda um outro tipo de cabo, chamado de "**cross-over**", que permite ligar diretamente dois micros, sem precisar do hub. Ele é uma opção mais barata quando você tem apenas dois micros. Neste tipo de cabo a posição dos fios é diferente nos dois conectores, de um dos lados a pinagem é a mesma de um cabo de rede normal, enquanto no outro a posição dos pares verde e laranja são trocados. Daí vem o nome cross-over, que significa, literalmente, "cruzado na ponta".

Para fazer um cabo cross-over, você crimpa uma das pontas seguindo o padrão EIA 568B que vimos acima e a outra utilizando o padrão EIA 568A, onde são trocadas as posições dos pares verde e laranja:

- ● 1- Branco com Verde
- 2- Verde
- ● 3- Branco com Laranja
- 4- Azul
- ● 5- Branco com Azul
- 6- Laranja
- ● 7- Branco com Marrom
- 8- Marrom

Esta mudança faz com que os fios usados para transmitir dados em um dos micros sejam conectados aos pinos conectores do outro, permitindo que eles conversem diretamente.

A maioria dos hub/switchs atuais é capaz de "descruzar" os cabos automaticamente quando necessário, permitindo que você misture cabos normais e cabos cross-over dentro do cabeamento da rede. Graças a isso, a rede vai funcionar mesmo que você use um cabo cross-over para conectar um dos micros ao hub por engano.

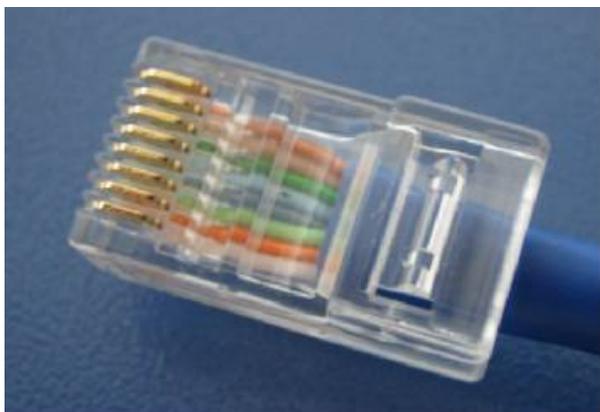
Na hora de crimpar é preciso fazer um pouco de força para que o conector fique firme. A qualidade do alicate é importante: evite comprar alicates muito baratos, pois eles precisam ser resistentes para aplicar a pressão necessária.

A função do alicate é fornecer pressão suficiente para que os pinos do conector RJ-45, que internamente possuem a forma de lâminas, esmaguem os fios do cabo, alcançando o fio de cobre e criando o contato. Você deve retirar apenas a capa externa do cabo e não descascar individualmente os fios, pois isso, ao invés de ajudar, serviria apenas para causar mau contato, deixando frouxo o encaixe com os pinos do conector.



É preciso um pouco de atenção ao cortar e encaixar os fios dentro do conector, pois eles precisam ficar perfeitamente retos. Isso demanda um pouco de prática. No começo, você vai sempre errar algumas vezes antes de conseguir.

Veja que o que protege os cabos contra as interferências externas são justamente as tranças. A parte destrançada que entra no conector é o ponto fraco do cabo, onde ele é mais vulnerável a todo tipo de interferência. Por isso, é recomendável deixar um espaço menor possível sem as tranças. Para crimpar cabos dentro do padrão, você precisa deixar menos de 2,5 centímetros destrançados. Você só vai conseguir isso *cortando o excesso de cabo solto* antes de encaixar o conector, como na foto:



O primeiro teste para ver se os cabos foram crimpados corretamente é conectar um dos micros (ligado) ao hub e ver se os LEDs da placas de rede e do hub acendem.

Isso mostra que os sinais elétricos enviados estão chegando até o hub e que ele foi capaz de abrir um canal de comunicação com a placa.

Se os LEDs nem acenderem, então não existe o que fazer. Corte os conectores e tente de novo. Infelizmente, os conectores são descartáveis: depois de crimpar errado uma vez, você precisa usar outro novo, aproveitando apenas o cabo. Mais um motivo para prestar atenção ;).

Existem também aparelhos testadores de cabos, que oferecem um diagnóstico muito mais sofisticado, dizendo, por exemplo, se os cabos são adequados para transmissões a 100 ou a 1000 megabits e avisando caso algum dos 8 fios do cabo esteja rompido. Os mais sofisticados avisam inclusive em que ponto o cabo está rompido, permitindo que você aproveite a parte boa.



Testador de cabos

Esses aparelhos serão bastante úteis se você for crimpar muitos cabos, mas são dispensáveis para trabalhos esporádicos, pois é muito raro que os cabos venham com fios rompidos de fábrica. Os cabos de rede apresentam também uma boa resistência mecânica e flexibilidade, para que possam passar por dentro de tubulações. Quase sempre os problemas de transmissão surgem por causa de conectores mal crimpados.

Existem ainda modelos mais simples de testadores de cabos, que chegam a custar em torno de 20 reais. Estes modelos mais simples realizam apenas um teste de continuidade do cabo, checando se o sinal elétrico chega até a outra ponta e, verificando o nível de atenuação, para certificar-se de que ele cumpre as especificações mínimas. Um conjunto de 8 leds se acende, mostrando o status de cada um dos 8 fios. Se algum fica apagado durante o teste, você sabe que o fio correspondente está partido. A limitação é que eles não são capazes de calcular em que ponto o cabo está partido, de forma que a sua única opção acaba sendo trocar e descartar o cabo inteiro.

Uma curiosidade é que algumas placas mãe da Asus, com rede Yukon Marvel (e, eventualmente, outros modelos lançados futuramente), incluem um software testador de cabos, que pode ser acessado pelo setup, ou através de uma interface dentro do Windows.

Ele funciona de uma forma bastante engenhosa. Quando o cabo está partido em algum ponto, o sinal elétrico percorre o cabo até o ponto onde ele está rompido e, por não ter para onde ir, retorna na forma de interferência. O software cronometra o tempo que o sinal demora para ir e voltar, apontando com uma certa precisão depois de quantos metros o cabo está rompido.

Além dos cabos em si, temos todo um conjunto de conduítes, painéis e tomadas, destinados a organizar a fiação. Você pode também passar cabos de rede através dos conduítes destinados aos fios de telefone e cabos de antena. Para isso, você vai precisar de uma guia para passar os cabos (você pode comprar em lojas de ferragens). Existem também lubrificantes específicos para cabos de rede, que ajudam o cabo a deslizar e podem ser usados para reduzir o stress mecânico sob o cabo ao passá-lo por conduítes apertados ou longas distâncias.

Uma boa opção ao cabear é usar tomadas para cabos de rede, ao invés de simplesmente deixar o cabos soltos. Elas dão um acabamento mais profissional e tornam o cabeamento mais flexível, já que você pode ligar cabos de diferentes tamanhos às tomadas e substituí-los conforme necessário (ao mudar os micros de lugar, por exemplo).

Uma observação importante é que você não deve passar cabos de rede pelas tubulações destinadas a cabos elétricos.

A rede elétrica utiliza corrente alternada, com ondas de 60 Hz, o que faz com que o ruído eletromagnético emitido pelos cabos elétricos prejudiquem a transmissão de sinais através do cabo de rede, aumentando o número de pacotes perdidos e assim por diante.

Em algumas situações, pode ser que você realmente não tenha outra saída, mas é uma coisa que você deve evitar ao máximo fazer. Verifique se não é possível passar os cabos por baixo do carpete, ou pelo forro do teto, por exemplo.

Para ser sincero, os padrões de cabeamento são definidos com uma boa margem de tolerância, para garantir que a rede funcione de forma confiável em qualquer situação. Já vi muitas redes com cabeamento completamente fora do padrão que continuavam funcionando, apesar dos abusos. Já vi casos de cabos com bem mais de 100 metros, cabos de rede passados lado a lado com fios elétricos e até mesmo um cabo cross-over feito com fios de telefone! Enfim, o simples caso da rede "funcionar" não significa que o cabeamento foi bem feito. Trabalhar próximo do limite vai fazer com que a velocidade de transmissão da rede fique abaixo do normal (por causa de colisões, pacotes perdidos e retransmissões) e pode causar problemas de conectividade diversos, que podem ser muito complicados de diagnosticar e corrigir.

Se você valoriza seu trabalho, procure seguir as regras e fazer um bom cabeamento. Redes bem cabeadas podem durar décadas. :)

## Hubs e switches

O **hub** ou **switch** é simplesmente o coração da rede. Ele serve como um ponto central, permitindo que todos os pontos se comuniquem entre si.

Todas as placas de rede são ligadas ao hub ou switch e é possível ligar vários hubs ou switches entre si (até um máximo de 7), caso necessário.



A diferença entre os hubs e switches é que o hub apenas retransmite tudo o que recebe para todos os micros conectados a ele, como se fosse um espelho. Isso significa que apenas um micro pode transmitir dados de cada vez e que todas as placas precisam operar na mesma velocidade, que é sempre nivelada por baixo. Caso você coloque um micro com uma placa de 10 megabits na rede, a rede toda passará a trabalhar a 10 megabits.

Os switches por sua vez são aparelhos muito mais inteligentes. Eles fecham canais exclusivos de comunicação entre o micro que está enviando dados e

o que está recebendo, permitindo que vários pares de micros troquem dados entre si ao mesmo tempo. Isso melhora bastante a velocidade em redes congestionadas, com muitos micros. Outra vantagem dos switches é que, em redes onde são misturadas placas 10/10 e 10/100, as comunicações podem ser feitas na velocidade das placas envolvidas, ou seja, quando duas placas 10/100 trocarem dados, a comunicação será feita a 100 megabits e quando uma das placas de 10 megabits estiver envolvida, será feita a 10 megabits.

Hoje em dia, os hubs "burros" caíram em desuso. Quase todos à venda atualmente são "**hub-switches**", modelos de switches mais baratos, que custam quase o mesmo que um hub antigo. Depois destes, temos os switches "de verdade", capazes de gerenciar um número muito maior de portas, sendo, por isso, adequados a redes de maior porte.

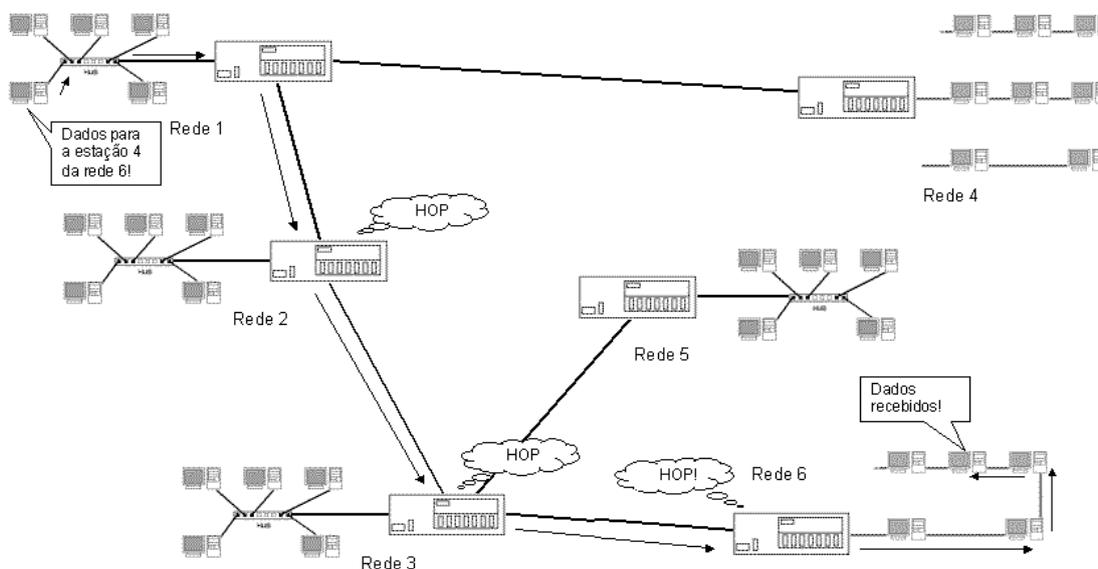
Tanto os "hub-switches", quanto os switches "de verdade" são dispositivos que trabalham no nível 2 do modelo OSI. O que muda entre as duas categorias é o número de portas e recursos. Os switches "de verdade" possuem interfaces de gerenciamento, que você acessa através do navegador em um dos micros da rede, que permitem visualizar diversos detalhes sobre o tráfego, descobrir problemas na rede e alterar diversas configurações, enquanto que os "hub-switches" são dispositivos burros.

Hoje em dia, existem ainda os "level 3 switches", uma categoria ainda mais inteligente de switches, que incorporam algumas características dos roteadores. Eles permitem definir rotas entre os diferentes micros da rede com base no endereço IP, criar "redes virtuais", onde os micros passam a se comportar como se estivessem ligados a dois switches diferentes, e assim por diante.

Finalmente, temos os **roteadores**, que são o topo da cadeia evolutiva. Os roteadores são ainda mais inteligentes, pois são capazes de interligar várias redes diferentes e sempre escolher a rota mais rápida para cada pacote de dados. Os roteadores operam no nível 3 do modelo OSI, procurando por endereços IP, ao invés de endereços MAC.

Usando roteadores, é possível interligar um número enorme de redes diferentes, mesmo que situadas em países ou mesmo continentes diferentes. Note que cada rede possui seu próprio roteador e os vários roteadores são interligados entre si. É possível interligar inúmeras redes diferentes usando roteadores, e não seria de se esperar que todos os roteadores tivessem acesso direto a todos os outros roteadores a que estivesse conectado.

Pode ser que, por exemplo, o roteador 4 esteja ligado apenas ao roteador 1, que esteja ligado ao roteador 2, que por sua vez seja ligado ao roteador 3, que esteja ligado aos roteadores 5 e 6. Se um micro da rede 1 precisar enviar dados para um dos micros da rede 6, então o pacote passará primeiro pelo roteador 2, será encaminhado ao roteador 3 e finalmente ao roteador 6. Cada vez que o dado é transmitido de um roteador para outro, temos um "**hop**".



Os roteadores são inteligentes o suficiente para determinar o melhor caminho a seguir. Inicialmente, o roteador procurará o caminho com o menor número de hops: o caminho mais curto. Mas se por acaso perceber que um dos roteadores desta rota está ocupado demais (o que pode ser medido pelo tempo de resposta), ele procurará caminhos alternativos para desviar deste roteador congestionado, mesmo que para isso o sinal tenha que passar por mais roteadores. No final, apesar do sinal ter percorrido o caminho mais longo, chegará mais rápido, pois não precisará ficar esperando na fila do roteador congestionado.

A internet é, na verdade, uma rede gigantesca, formada por várias sub-redes interligadas por roteadores. Todos os usuários de um pequeno provedor, por exemplo, podem ser conectados à internet por meio do mesmo roteador. Para baixar uma página do Yahoo, por exemplo, o sinal deverá passar por vários roteadores, várias dezenas em alguns casos. Se todos estiverem livres, a página será carregada rapidamente. Porém, se alguns estiverem congestionados, pode ser que a página demore vários segundos antes de começar a carregar.

Você pode medir o tempo que um pedido de conexão demora para ir até o destino e ser respondido usando o comando "**ping**", disponível tanto no Linux quanto no prompt do MS-DOS, no Windows.

Para verificar por quantos roteadores o pacote está passando até chegar ao destino, use o comando "**tracert**" (no Linux) ou "**tracert**" (no Windows).

Os roteadores podem ser desde PCs comuns, com duas ou mais placas de rede, até supercomputadores capazes de gerenciar centenas de links de alta velocidade. Eles formam a espinha dorsal



Roteador

Quando você usa um PC com duas placas de rede para compartilhar a conexão com os micros da rede local, você está configurando-o para funcionar como um roteador simples, que liga uma rede (a internet) a outra (a sua rede doméstica). O mesmo acontece ao configurar seu modem ADSL como roteador.

Pense que a diferença entre hubs e switches e os roteadores é justamente esta: os hubs e switches permitem que vários micros sejam ligados formando uma única rede, enquanto que os roteadores permitem interligar várias redes diferentes, criando redes ainda maiores, como a própria internet.

Dentro de uma mesma rede é possível enviar pacotes de broadcast, que são endereçados a todos os integrantes da rede simultaneamente. Ao usar um hub burro, todos os micros recebem todas as transmissões. Um roteador filtra tudo isso, fazendo com que apenas os pacotes especificamente endereçados a endereços de outras redes trafeguem entre elas. Lembre-se de que, ao contrário das redes locais, os links de internet são muito caros (muitas vezes se paga por gigabyte transferido), por isso é essencial que sejam bem aproveitados.

## Redes wireless

Usar algum tipo de cabo, seja um cabo de par trançado ou de fibra óptica, é a forma mais rápida e em geral a mais barata de transmitir dados. Os cabos de par trançado cat 5e podem transmitir dados a até 1 gigabit a uma distância de até 100 metros, enquanto os cabos de fibra óptica são usados em links de longa distância, quando é necessário atingir distâncias maiores. Usando 10G, é possível atingir distâncias de até 40 km, sem necessidade de usar repetidores.

Mas, em muitos casos não é viável usar cabos. Imagine que você precise ligar dois escritórios situados em dois prédios diferentes (porém próximos), ou que a sua mãe/esposa/marido não deixa você nem pensar em espalhar cabos pela casa.

A solução nesses casos são as redes sem fio, que estão caindo de preço e, por isso, tornando-se bastante populares. O padrão mais usado é o Wi-Fi (Wireless Fidelity), o nome comercial para os padrões 802.11b, 802.11a e 802.11g. A topologia deste tipo de rede é semelhante a das redes de par trançado, com o hub central substituído pelo ponto de acesso. A diferença no caso é que são usados transmissores e antenas ao invés de cabos. É possível encontrar tanto placas PCMCIA ou mini-PCI, para notebooks, quanto placas PCI, para micros desktop.

Quase todos os notebooks à venda atualmente, muitos modelos de palmtops e até mesmo smartphones já incluem transmissores wireless integrados. Muita gente já acha inconcebível comprar um notebook sem wireless, da mesma forma que ninguém mais imagina a idéia de um PC sem disco rígido, como os modelos vendidos no início da década de 80.



Na verdade, é bastante raro um notebook que venha com uma placa wireless "onboard". Quase sempre é usada uma placa mini-pci (uma versão miniaturizada de uma placa PCI tradicional, que usa um encaixe próprio), que pode ser substituída como qualquer outro componente. A antena não vai na própria placa, mas é montada na tampa do monitor, atrás do LCD e o sinal vai até a placa através de dois cabos, que correm dentro da carcaça do notebook.

Estas placas mini-pci levam uma vantagem muito grande sobre as placas wireless PCMCIA por causa da antena. As placas PCMCIA precisam ser muito compactas, por isso invariavelmente possuem uma antena muito pequena, com pouca sensibilidade. As antenas incluídas nos notebooks, por sua vez, são invariavelmente muito maiores, o que garante uma conexão muito mais estável, com um alcance muito maior e ajuda até mesmo na autonomia das baterias (já que é possível reduzir a potência do transmissor).

A maioria dos notebooks fabricados a partir do final de 2002 trazem o slot mini-pci e a antena, permitindo que você compre e instale uma placa mini-pci, ao invés de ficar brigando com o alcance reduzido das placas PCMCIA.

Existem vários modelos de placas mini-pci no mercado, mas elas não são um componente comum, de forma que você só vai encontrá-las em lojas especializadas. É possível também substituir a placa que acompanha o notebook por outro modelo, melhor ou mais bem suportado no Linux.



Slot mini-pci



Placa wireless mini-pci

Não se engane pela foto. As placas mini-pci são muito pequenas, quase do tamanho de uma caixa de fósforos e os conectores a antena são quase do tamanho de uma cabeça de alfinete. Eles são frágeis, por isso é preciso ter cuidado ao plugá-los na placa. O fio branco vai sempre no conector no canto da placa e o preto no conector mais ao centro, como na foto.

Quase sempre, o notebook tem uma chave ou um botão que permite ligar e desligar o transmissor wireless. Antes de testar, verifique se ele está ativado.

Embora as placas mini-pci sejam componentes tão padronizados quanto as placas PCMCIA, sempre existe a possibilidade de algumas placas específicas não serem compatíveis com seu notebook. O ideal é sempre testar antes de comprar, ou comprar em uma loja que aceite trocar a placa por outra em caso de problemas.

## O básico

Em uma rede wireless, o hub é substituído pelo **ponto de acesso** (access-point em inglês), que tem a mesma função central que o hub desempenha nas redes com fios: retransmitir os pacotes de dados, de forma que todos os micros da rede os recebam.

Os pontos de acesso possuem uma saída para serem conectados em um hub tradicional, permitindo que você "junte" os micros da rede com fios com os que estão acessando através da rede wireless, formando uma única rede, o que é justamente a configuração mais comum.

Existem poucas vantagens em utilizar uma rede wireless para interligar micros desktops, que afinal não precisam sair do lugar. O mais comum é utilizar uma rede cabeada normal para os desktops e utilizar uma rede wireless complementar para os notebooks, palmtops e outros dispositivos móveis.

Você utiliza um hub/switch tradicional para a parte cabeada, usando cabo também para interligar o ponto de acesso à rede.



O ponto de acesso serve apenas como a "última milha", levando o sinal da rede até os micros com placas wireless. Eles podem acessar os recursos da rede normalmente, acessar arquivos compartilhados, imprimir, acessar a internet, etc. A única limitação fica sendo a velocidade mais baixa e o tempo de acesso mais alto das redes wireless.

Isso é muito parecido com juntar uma rede de 10 megabits, que utiliza um hub "burro" a uma rede de 100 megabits, que utiliza um switch. Os micros da rede de 10 megabits continuam se comunicando entre si a 10 megabits, e os de 100 continuam trabalhando a 100 megabits, sem serem incomodados pelos vizinhos. Quando um dos micros da rede de 10 precisa transmitir para um da rede de 100, a transmissão é feita a 10 megabits, respeitando a velocidade do mais lento.

Para redes mais simples, onde você precise apenas compartilhar o acesso à internet entre poucos micros, todos com placas wireless, você pode ligar o modem ADSL (ou cabo) direto ao ponto de acesso. Alguns pontos de acesso trazem um switch de 4 ou 5 portas embutido, permitindo que você crie uma pequena rede cabeada sem precisar comprar um hub/switch adicional.



A principal diferença é que em uma rede wireless o meio de transmissão (o ar) é compartilhado por todos os clientes conectados ao ponto de acesso, como se todos estivessem ligados ao mesmo cabo coaxial. Isso significa que apenas uma estação pode transmitir de cada vez, e todas as estações recebem todos os pacotes transmitidos da rede, independentemente do destinatário. Isso faz com que a segurança dentro de uma rede wireless seja uma questão sempre bem mais delicada que em uma rede cabeada. Outra questão importante é que a velocidade da rede decai conforme aumenta o número de micros conectados, principalmente quando vários deles transmitem dados ao mesmo tempo.

Dependendo da potência dos transmissores nas placas e no pontos de acesso e do tipo de antenas usadas, é possível propagar o sinal da rede por 200, 300 ou até 500 metros de distância (desde que não existam obstáculos importantes pelo caminho). Usando antenas Yagi (que geram um sinal mais focalizado) e amplificadores é possível interligar dois pontos distantes a 2 km ou mais.

Isso traz mais um problema, que é a questão da interferência entre diferentes redes instaladas na mesma área. Imagine um grande prédio comercial, com muitos escritórios de empresas diferentes e cada uma com sua própria rede wireless. Os pontos de acesso podem ser configurados para utilizarem frequências diferentes, divididas em 16 canais. Devido à legislação de cada país, apenas 11, 13 ou 14 destes canais podem ser usados e destes, apenas 4 podem ser usados simultaneamente, sem que realmente não exista interferência. Ou seja, com várias redes instaladas próximas umas das outras, os canais disponíveis são rapidamente saturados, fazendo com que o tráfego de uma efetivamente reduza o desempenho da outra.

Existe ainda a questão das interferências e de materiais que atenuam o sinal. Em primeiro lugar temos as superfícies de metal em geral, como janelas, portas metálicas, lajes, vigas e até mesmo tintas com pigmentos metálicos. Depois temos concentrações de líquido, como aquários, piscinas, caixas d'água e até mesmo pessoas passeando pelo local (nosso corpo é composto de 70% de água).

Fornos de microondas operam na mesma frequência das redes wireless, fazendo com que, quando ligados, eles se transformem em uma forte fonte de interferência, prejudicando as transmissões num raio de alguns metros. Telefones sem fio, que operam na faixa dos 2.4 GHz, também interferem, embora em menor grau.

Os fabricantes falam em 150 ou até 300 metros de alcance máximo, mas essas distâncias são atingidas apenas em campo aberto, em condições ideais. Na prática, o alcance varia muito de acordo com o ambiente. Você pode conseguir pegar o sinal de um ponto de acesso instalado na janela de um prédio vizinho, distante 100 metros do seu (campo aberto), mas não conseguir acessar a rede do andar de cima (a armação de ferro e cimento da laje é um obstáculo difícil de transpor). Para compensar grandes distâncias, obstáculos ou interferências, o ponto de acesso reduz a velocidade de transmissão da rede, como um modem discado tentando se adaptar a uma linha ruidosa. Os 54 megabits originais podem se transformar rapidamente em 11, 5.5, 2 ou até mesmo 1 megabit.

Temos ainda a questão da segurança: se você morar em um sobrado e colocar o ponto de acesso próximo da janela da frente do quarto no primeiro andar, provavelmente um vizinho do quarteirão seguinte ainda vai conseguir se conectar à sua rede, desde que substitua a antena da placa por uma mais potente. Existe até uma velha receita que circula pela internet de como fazer uma antena caseira razoável usando um tubo de batata Pringles. Não é brincadeira: o tubo é forrado de papel alumínio e tem um formato adequado para atuar como uma antena.

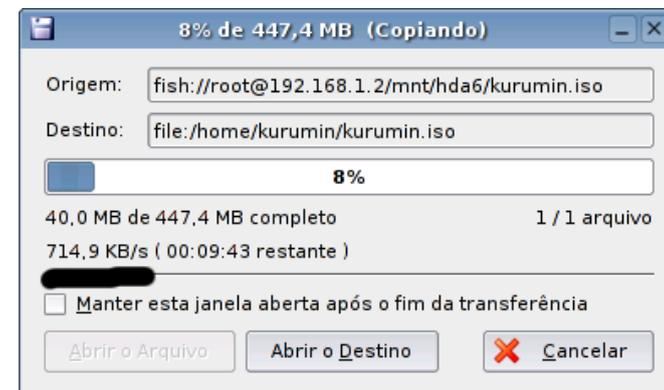
Caímos, então, em um outro problema. Você simplesmente não tem como controlar o alcance do sinal da rede. Qualquer vizinho próximo, com uma antena potente (ou um tubo de batata), pode conseguir captar o sinal da sua rede e se conectar a ela, tendo acesso à sua conexão com a web, além de arquivos e outros recursos que você tenha compartilhado entre os micros da rede, o que não é muito interessante.

Eis que surge o **WEP**, abreviação de "Wireless-Equivalent Privacy", que, como o nome sugere, traz como promessa um nível de segurança equivalente ao das redes cabeadas. Na prática, o WEP tem muitas falhas e é relativamente simples de quebrar, mas não deixa de ser uma camada de proteção básica que você sempre deve manter ativa. A opção de ativar o WEP aparece no painel de configuração do ponto de acesso.

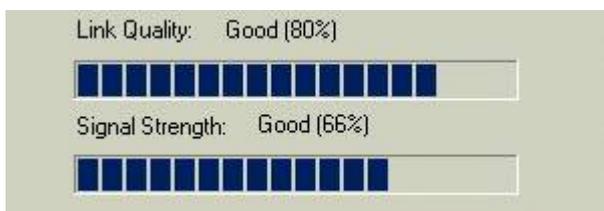
O WEP se encarrega de encriptar os dados transmitidos através da rede. Existem dois padrões WEP: de 64 e de 128 bits. O padrão de 64 bits é suportado por qualquer ponto de acesso ou interface que siga o padrão WI-FI, o que engloba todos os produtos comercializados atualmente. O padrão de 128 bits, por sua vez, não é suportado por todos os produtos, mas em compensação é bem menos inseguro. Para habilitá-lo será preciso que todos os componentes usados na sua rede suportem o padrão, caso contrário os nós que suportarem apenas o padrão de 64 bits ficarão fora da rede.

Existe ainda o WPA, um padrão mais seguro, que já é suportado pela grande maioria das placas e dos pontos de acesso. Existem várias variações do WPA, que utilizam diversos sistemas de encriptação diferentes, com a opção de usar um servidor Radius para centralizar os logins da rede, opção muito usada em empresas. No entanto, o mais comum em pequenas redes é usar o WPA-PSK (o padrão mais simples), onde é definida uma chave (uma espécie de senha), usada para autenticar os clientes da rede. PSK é abreviação de "Pre-Shared Key", ou "chave previamente compartilhada".

Temos, em seguida, a questão da velocidade. Nas redes 802.11b, o padrão original, a velocidade teórica é de apenas 11 megabits (ou 1.35 MB/s). Como as redes wireless possuem um overhead muito grande, por causa da modulação do sinal, checagem e retransmissão dos dados, as taxas de transferências, na prática, ficam em torno de 750 KB/s, menos de dois terços do máximo.



Conforme o cliente se distancia do ponto de acesso, a taxa de transmissão cai para 5 megabits, 2 megabits e 1 megabit, até que o sinal se perca definitivamente. No Windows você pode usar o utilitário que acompanha a placa de rede para verificar a qualidade do sinal em cada parte do ambiente onde a rede deverá estar disponível. No Linux isso é feito por programas como o Kwifimanager, que veremos a seguir.



Veja que tanto na questão da segurança, quanto na questão do desempenho, as redes wireless perdem para as redes cabeadas. A maior arma das redes wireless é a versatilidade. O simples fato de poder interligar os PCs sem precisar passar cabos pelas paredes já é o suficiente para convencer muitas pessoas, mas existem mais alguns recursos interessantes que podem ser explorados.

Sem dúvida, a possibilidade mais interessante é a mobilidade para os portáteis. Tanto os notebooks, quanto handhelds e as webpads podem ser movidos livremente dentro da área coberta pelos pontos de acesso sem que seja perdido o acesso à rede. Essa possibilidade lhe dá mobilidade dentro de casa para levar o notebook para onde quiser, sem perder

o acesso à web, mas é ainda mais interessante para empresas e escolas. No caso das empresas, a rede permite que os funcionários possam se deslocar pela empresa sem perder a conectividade com a rede (entrando e saindo de reuniões ou caminhando pela linha de produção, por exemplo), e basta se aproximar do prédio para que seja possível se conectar à rede e ter acesso aos recursos necessários.

No caso das escolas, a principal utilidade seria fornecer acesso à web aos alunos. Muitas lojas e a maior parte dos aeroportos pelo mundo já oferecem acesso à web através de redes sem fio como uma forma de serviço para seus clientes. Um exemplo famoso é o da rede de cafés Starbucks nos EUA e Europa, onde todas as lojas oferecem acesso gratuito à web para os clientes que possuem um notebook ou outro portátil com placa wireless.

## Padrões

O **802.11b** foi o primeiro padrão wireless usado em grande escala. Ele marcou a popularização da tecnologia. Naturalmente, existiram vários padrões anteriores, como o 802.11 (que trabalhava a 1 ou 2 megabits) e também alguns padrões proprietários, incompatíveis entre si, como o Arlan da Aironet e o WaveLAN, da NCR, que trabalhavam na faixa dos 900 MHz e transmitiam a respectivamente 860 kbits e 2 megabits.

O 802.11b permitiu que placas de diferentes fabricantes se tornassem compatíveis e os custos caíssem, graças ao aumento na demanda e à concorrência. O padrão seguinte foi o 802.11a (que na verdade começou a ser desenvolvido antes do 802.11b, mas foi finalizado depois), que utiliza uma faixa de frequência mais alta: 5 GHz e oferece uma velocidade teórica de 54 megabits, porém a distâncias menores, cerca de metade da distância atingida por uma placa 802.11b usando o mesmo tipo de antena.

Embora os dois padrões sejam incompatíveis, a maior parte das placas 802.11a incorporam chips capazes de trabalhar nas duas faixas de frequência, permitindo que sejam usadas nos dois tipos de redes. Uma observação importante é que, ao misturar placas 802.11a e 802.11b, a velocidade é nivelada por baixo e toda a rede passa a operar a 11 megabits. Lembre-se que uma rede wireless opera de forma similar às redes antigas, com cabos coaxiais: todos compartilham o mesmo "cabo".

Finalmente, temos o padrão atual, o **802.11g**. Ele utiliza a mesma faixa de frequência do 802.11b: 2.4 GHz. Isso permite que os dois padrões sejam interoperáveis. A idéia é que você possa adicionar placas e pontos de acesso 802.11g a uma rede 802.11b já existente, mantendo os componentes antigos, do mesmo modo como hoje em dia temos liberdade para adicionar placas e switches Gigabit Ethernet a uma rede já existente de 100 megabits.

Apesar disso, a velocidade de transmissão no 802.11g é de 54 megabits, como nas redes 802.11a. Na prática, em redes 802.11a é possível atingir taxas de transmissão (reais) em torno de 3,4 MB/s, enquanto que as redes 802.11g são um pouco mais lentas, atingindo cerca de 3,0 MB/s em condições ideais. Mas, fora esta pequena desvantagem no desempenho, as redes 802.11g juntam o melhor dos dois mundos.

Note que, para que a rede efetivamente trabalhe a 54 megabits, é necessário que o ponto de acesso e todas as placas sejam 802.11g. Ao incluir uma única placa 802.11b na rede (mesmo que seja seu vizinho roubando sinal), toda a rede passa a operar a 11 megabits. As placas 802.11g não são compatíveis com o padrão 802.11a, mas os dois tipos de placas podem conversar a 11 megabits, utilizando o padrão 801.11b, que vira um denominador comum.

Além dos padrões oficiais, existem as extensões proprietárias criadas pela Atheros e Broadcom para aumentar o desempenho das redes baseadas em seus produtos.

As placas e pontos de acesso 802.11g baseados em chips da Atheros utilizam o "**Atheros Super G**", um sistema dual-band, onde a placa passa a transmitir usando dois canais simultaneamente, dobrando a taxa de transmissão. Ele é encontrado nas placas e pontos de acesso D-Link AirPlus Xtreme G e nos produtos recentes da Netgear.

O efeito colateral é que, por transmitir usando dois canais simultâneos, ele acaba gerando bem mais interferência com outras redes próximas.

A Broadcom oferece o "**Broadcom Afterburner**", que mantém o uso de um único canal, mas utiliza uma série de otimizações, reduzindo o overhead das transmissões e conseguindo assim aumentar a porcentagem de bytes "úteis" transmitidos. Entre as técnicas utilizadas estão o frame-bursting (onde são enviados uma série de pacotes de dados dentro de um único frame, reduzindo o overhead da transmissão) e a compressão de dados, que ajuda ao transferir arquivos com baixo índice de compressão através da rede. O ponto fraco é que o ganho de velocidade depende muito do tipo de dados transmitidos (por causa da compressão).

O Afterburner promete até 125 megabits, contra os 108 megabits do Super G e os 54 megabits do 802.11g "regular". Na prática, as diferenças acabam não sendo tão grandes, pois o uso de dois canais do Super G aumenta o nível de interferência com redes próximas e a vulnerabilidade a interferências de uma forma geral e as otimizações utilizadas pelo Afterburner aumentam o número de pacotes perdidos ou corrompidos, reduzindo o ganho real de desempenho.

Outro problema é que as otimizações só funcionam caso você baseie toda a sua rede em placas e pontos de acesso compatíveis com um dos dois padrões, caso

contrário a rede passa a operar no modo 802.11g "padrão", para manter a compatibilidade com todos os clientes. Na prática isso é bem complicado, pois você raramente pode escolher qual placa virá instalada ao comprar um notebook ou um PC montado, por exemplo.

## Aumentando o alcance

Assim como em outras tecnologias de transmissão via rádio, a distância que o sinal é capaz de percorrer depende também da qualidade da antena usada. As antenas padrão utilizadas nos pontos de acesso (geralmente de 2 dBi) são pequenas, práticas e baratas, mas existe a opção de utilizar antenas mais sofisticadas para aumentar o alcance da rede.

Alguns fabricantes chegam a dizer que o alcance dos seus pontos de acesso chega a 300 metros, usando as pequenas antenas padrão. Isso está um pouco longe da realidade, pois só pode ser obtido em campos abertos, livres de qualquer obstáculo e, mesmo assim, com o sinal chegando muito fraco ao final dos 300 metros, já com a rede trabalhando na velocidade mínima, a 1 megabit e com um lag muito grande.



Apesar disso, a distância máxima e a qualidade do sinal (e, conseqüentemente, a velocidade de transmissão) podem variar bastante de um modelo de ponto de acesso para outro, de acordo com a qualidade e potência do transmissor e da antena usada pelo fabricante. Existem basicamente três tipos de antenas que podem ser utilizadas para aumentar o alcance da rede.

As antenas Yagi são as que oferecem um maior alcance, mas em compensação são capazes de cobrir apenas uma pequena área, para onde são apontadas. Estas antenas são mais úteis para cobrir alguma área específica, longe do ponto de acesso, ou interligar duas redes distantes.

Em ambos os casos, o alcance ao usar uma antena Yagi pode facilmente ultrapassar os 1000 metros. Usando uma antena de alto ganho em cada ponto, uma delas com um amplificador de 1 watt (o máximo permitido pela legislação), é possível atingir 5 km ou mais. As Yagi são também o melhor tipo de antena a usar quando é preciso concentrar o sinal para "furar" um obstáculo entre as duas redes, como, por exemplo, um prédio bem no meio do caminho. Nestes casos a distância atingida será sempre mais curta, naturalmente.

Uma solução muito adotada nestes casos é usar um repetidor instalado num ponto intermediário, permitindo que o sinal desvie do obstáculo.

Existem até mesmo pontos de acesso extremamente robustos, desenvolvidos para uso industrial, que além de uma gabinete reforçado, utilizam placas solares e baterias, que permitem a eles funcionar de forma inteiramente autônoma.



*Estação repetidora*

Outra solução comum é usar dois pares do cabo de rede (a rede funciona perfeitamente apenas com dois pares) para enviar energia ao ponto de acesso, eliminando o uso de um cabo de força separado. Esta solução é chamada de "Power Over Ethernet" (POE), veja mais detalhes no:

<http://www.poweroverethernet.com/>.

Voltando ao tema principal, a instalação das antenas Yagi é complicada, pois uma antena deve ficar apontada exatamente para a outra, cada uma no topo de um prédio ou morro, de forma que não exista nenhum obstáculo entre as duas.

No final da instalação é usado um laser para fazer um ajuste fino "mirando" as duas antenas.

As antenas feitas com tubos de batatas Pringles são justamente um tipo de antena Yagi de baixo ganho. Outra dica é que os pontos de acesso quase sempre possuem duas saídas de antena. Você pode usar uma antena convencional em uma delas, para manter o sinal em um raio circular, atendendo aos micros próximos e usar uma antena Yagi na segunda, para criar um link com um local específico, distante do ponto de acesso.



*Antena Yagi*

A segunda opção são as antenas **omnidirecionais**, que, assim como as antenas padrão dos pontos de acesso, cobrem uma área circular em torno da antena. Elas são boas irradiando o sinal na horizontal, mas não na vertical, por isso devem ser sempre instaladas "de pé", a menos que a intenção seja pegar sinal no andar de cima.

As antenas nos clientes devem sempre estar alinhadas (também de pé) com a antena do ponto de acesso, para uma melhor recepção. Caso o cliente use algum tipo de antena mini-yagi, então a antena deve ficar apontada para o ponto de acesso.

A vantagem de usar uma omnidirecional externa é a possibilidade de utilizar uma antena de maior ganho. Existem modelos de antenas omnidirecionais de 3 dBi, 5 dBi, 10 dBi ou até mesmo 15 dBi, um grande avanço sobre as antenas de 2 ou 3 dBi que acompanham a maioria dos pontos de acesso.



*Antena omnidirecional*

Assim como as Yagi, as antenas omnidirecionais podem ser usadas tanto para aumentar a área de cobertura do ponto de acesso, quanto serem instaladas em placas de rede wireless com antenas destacáveis, permitindo captar o sinal do ponto de acesso de uma distância maior.

Uma terceira opção de antena são as **parabólicas** ou **miniparabólicas**, que também captam o sinal em apenas uma direção, de forma ainda mais concentrada que as Yagi, permitindo que sejam atingidas distâncias maiores. As miniparabólicas mais "populares" possuem, geralmente, 24 ou 28 dBi de potência, enquanto as maiores e mais caras podem chegar a 124 dBi (ou mais).



*Antena miniparabólica*

Estas antenas podem custar de 30 a mais de 200 dólares, dependendo da potência. As antenas Yagi estão entre as mais caras, vendidas por 150 dólares ou mais. Além do problema do preço, existe um aumento no risco de uso indevido na rede, já que o sinal irá se propagar por uma distância maior, mais uma razão para reforçar a segurança.

Para ligar as antenas ao ponto de acesso ou à placa é usado um cabo especial chamado **pigtail**, um cabo fino, sempre relativamente curto, usado como um adaptador entre a minúscula saída usada nas placas e a entrada do cabo ou antena.

Os pigtails invariavelmente causam uma pequena perda de sinal, pois para ser flexível o cabo possui apenas uma fina camada de blindagem. Justamente por isso, eles devem ser o mais curto possíveis, tendo apenas o comprimento necessário para realizar a conexão.



*Cabo pigtail*

Ao cobrir distâncias maiores, o ideal é que o ponto de acesso seja instalado próximo à antena, com um cabo de rede ligando-o ao servidor ou switch. As redes 801.11x trabalham com sinais de baixa potência (em geral menos de 0.25 watt); por isso, qualquer tipo de cabo longo causa uma grande perda de sinal.

Para casos em que a antena do ponto de acesso não é suficiente, mas também não existe necessidade de uma antena cara, existe a opção de fazer um defletor caseiro, que concentra o sinal recebido pela antena padrão do ponto de acesso, fazendo com que ela cubra uma área mais focalizada, porém com um ganho maior.

Além de melhorar a qualidade do sinal na área desejada, ela reduz o alcance nas demais direções, fazendo com que seja muito mais difícil captar o sinal da sua rede de fora.

Esta é uma receita muito simples. Você precisa de alguma folha de metal ou fio (como uma malha de fios, papel alumínio, papel laminado ou um pedaço de lata) e papelão. Cobrindo um pedaço retangular do papelão com a folha metálica e dobrando-o num ângulo de 90 graus (formando um meio quadrado) você obtém um concentrador de sinal, que pode ser encaixado nas antenas do ponto de acesso:



Assim como em uma antena parabólica, os sinais são refletidos pela folha metálica e concentrados em direção à antena do ponto de acesso, aumentando o ganho. Por outro lado, o sinal torna-se muito mais fraco nas outras direções, dificultando as coisas para seu vizinho interessado em roubar sinal. Apesar de primitivos, estes defletores podem proporcionar um ganho de até 12 dBi, um upgrade respeitável.

Você pode baixar o modelo com os ângulos corretos no:

<http://www.freeantennas.com/projects/Ez-10/>

Várias fotos com exemplos estão disponíveis no:

<http://www.freeantennas.com/projects/template/gallery/>

Existe ainda a popular "cantenna", um tipo de antena Yagi feita usando uma lata de batata Pringles. Você encontra a receita no:

<http://www.oreillynet.com/cs/weblog/view/wlg/448>

Vamos então a um conjunto de respostas rápidas às dúvidas mais comuns relacionadas à antenas:

**Interferência:** Usar uma antena de alto ganho não ajuda muito com relação a interferências criadas por outras redes próximas, telefones sem fio ou aparelhos de microondas, já que junto com o sinal, a antena também amplifica todas as fontes de interferência na mesma proporção. Uma solução neste caso pode ser substituir a antena omnidirecional do ponto de acesso, ou do cliente afetado por uma antena yagi ou outro tipo de antena direcional. Isto permite concentrar o sinal, evitando as fontes de interferência.

**Uso de duas antenas:** A maioria dos pontos de acesso 802.11b e 802.11g utilizam duas antenas, mas (com exceção de alguns hacks que ativam esta função), elas não são usadas de forma independente, uma para enviar e outra para receber, por exemplo.

Ao invés disso, o ponto de acesso transmite o mesmo sinal usando ambas as antenas, simplesmente selecionando a que oferece um sinal de melhor qualidade com relação a cada cliente. Muitos pontos de acesso de baixo custo, estão passando a utilizar uma única antena, o que favorece o surgimento de pontos cegos.

Os pontos de acesso 802.11n (com três antenas), por sua vez, utilizam o mimo, um sistema mais sofisticado, onde cada uma das antenas transmite um sinal independente e o ponto de acesso se encarrega de remontar o sinal original combinando os sinais das três antenas, além de levar em conta fatores como a reflexão do sinal por paredes e outros objetos. Isso permite que o 802.11n ofereça uma taxa de transmissão e alcance maiores que os 802.11g.

**Comprimento do cabo:** O sinal de uma rede wireless é bastante fraco, por isso os cabos e conectores representam sempre um ponto importante de perda. O ideal é sempre utilizar cabos com 3 metros ou menos, de forma que a perda seja limitada. Caso precise de cabos mais longos, procure cabos blindados, que reduzem a perda. Leve em conta que por melhor que seja a qualidade do cabo e conectores usados, você quase sempre terá uma perda de 2 a 3 dBi. Leve isso em consideração ao escolher qual antena usar.

## Portas TCP e UDP

Ao conectar na internet, seu micro recebe um endereço IP válido. Mas, normalmente mantemos vários programas ou serviços abertos simultaneamente. Em um desktop é normal ter um programa de e-mail, um cliente de FTP ou SSH, o navegador, um cliente de ICQ ou MSN, dois ou três downloads via bittorrent e vários outros programas que enviam e recebem informações, enquanto um único servidor pode manter ativos servidores web, FTP, SSH, DNS, LDAP e muitos outros serviços.

Se temos apenas um endereço IP, como todos estes serviços podem funcionar ao mesmo tempo sem entrar em conflito?

Imagine que as duas partes do endereço IP (a parte referente à rede e a parte referente ao host) correspondem ao CEP da rua e ao número do prédio. Um carteiro só precisa destas duas informações para entregar uma carta. Mas, dentro do prédio moram várias pessoas. O CEP e número do prédio só vão fazer a carta chegar até a portaria. Daí em diante é preciso saber o número do apartamento. É aqui que entram as famosas **portas TCP**.

Existem 65.536 portas TCP, numeradas de 1 a 65536. Cada porta pode ser usada por um programa ou serviço diferente, de forma que em teoria poderíamos ter até 65536 serviços diferentes ativos simultaneamente

em um mesmo servidor, com um único endereço IP válido. O endereço IP contém o CEP da rua e o número do prédio, enquanto a **porta TCP** determina a que sala dentro do prédio a carta se destina.

As portas TCP mais usadas (também chamadas de "well known ports") são as portas de 0 a 1023, que são reservadas para serviços mais conhecidos e utilizados, como servidores web, FTP, servidores de e-mail, compartilhamento de arquivos, etc. A porta 80, por exemplo, é reservada para uso de servidores web, enquanto a porta 21 é a porta padrão para servidores FTP. A porta "0" é reservada, por isso não entra realmente na lista.

Além do endereço IP, qualquer pacote que circula na internet precisa conter também a porta TCP a que se destina. É isso que faz com que um pacote chegue até o servidor web e não ao servidor FTP instalado na mesma máquina.

Além das 65.536 portas TCP, temos o mesmo número de portas UDP, seu protocolo irmão. Embora seja um protocolo menos usado que o TCP, o UDP continua presente nas redes atuais pois oferece uma forma alternativa de envio de dados, onde ao invés da confiabilidade é privilegiada velocidade e simplicidade. Vale lembrar que, tanto o TCP, quanto o UDP, trabalham na camada 4 do modelo OSI. Ambos trabalham em conjunto com o IP, que cuida do endereçamento.



No TCP, os dados são transmitidos através de conexões. Tudo começa com o cliente enviando o pacote "SYN", que solicita a abertura da conexão. Caso a porta esteja fechada, o servidor responde com um pacote "RST" e a conversa para por aí. Caso, por outro lado, exista algum servidor disponível na porta solicitada (um servidor apache, por exemplo), então ele responde com outro pacote "SYN", seguido de um pacote "ACK", avisando que a porta está disponível e prosseguindo com a abertura da conexão.

O cliente responde então com outro pacote "ACK", o que abre oficialmente a conexão. Começa então a transferência dos dados, que são organizados em pacotes com até 1550 bytes cada um. Para cada pacote recebido, a estação envia um pacote de confirmação e, caso algum pacote se perca, ela solicita a retransmissão. Cada pacote inclui 4 bytes adicionais com um código de CRC, que permite verificar a integridade do pacote. É através dele que o cliente sabe quais pacotes chegaram danificados.

Depois que todos os dados são transmitidos, o servidor envia um pacote "FYN" que avisa que não tem mais nada a transmitir. O cliente responde com outro pacote "FYN" e a conexão é oficialmente encerrada.

Graças a tudo isso, a confiabilidade é muito boa. Quando a conexão está ruim, é normal ocorrerem mais perdas de pacotes e retransmissões, mas as corrupções são geralmente causadas pelo próprio programa que está baixando o arquivo e não pelo protocolo.

O problema é que toda esta formalidade torna as transferências um pouco mais lentas. Imagine que, para transmitir uma mensagem de texto com 300 bytes, via TCP, seria necessário transmitir um total de 9 pacotes!

Veja um exemplo de como a transmissão funcionaria:

Estação: **SYN** (*solicita a abertura da conexão*)

Servidor: **SYN** (*confirma o recebimento e avisa que a porta está disponível*)

Servidor: **ACK** (*inicia a conexão*)

Estação: **ACK** (*confirma*)

Estação: **DATA** (*é enviado o pacote com a mensagem de texto*)

Servidor: **OK** (*a confirmação, depois de verificar a integridade do pacote*)

Estação: **FYN** (*solicita o fechamento da conexão*)

Servidor: **FYN** (*confirma*)

Estação: **FYN** (*confirma que recebeu a confirmação*)

No UDP, as coisas são mais simples. Nele não existe abertura de conexão, os pacotes são transmitidos diretamente. A estação solicita alguma informação e o servidor envia a resposta.

Assim como no TCP, são usados pacotes de até 1550 bytes, contendo os bits adicionais de verificação. A estação pode verificar a integridade dos pacotes, mas não tem como perceber se algum pacote se perdeu, ou solicitar a retransmissão de um pacote corrompido. Se um pacote se perde, fica por isso mesmo.

Um exemplo típico do uso do UDP é o streaming de vídeo e áudio via web, uma situação onde o que vale é a velocidade e não a confiabilidade. Você não gostaria nada se o navegador parasse a exibição do vídeo para solicitar uma retransmissão cada vez que um pacote se perdesse ou chegasse corrompido. É preferível que ele pule o quadro e continue exibindo o restante do vídeo.

Outra aplicação comum são os servidores DNS. Sempre que você acessa um site, a solicitação do endereço IP referente ao domínio do site e a resposta do servidor são enviadas via UDP, para ganhar tempo.

Na prática, é bem raro encontrar algum programa que utilize unicamente pacotes UDP para qualquer coisa além do envio de mensagens curtas. Mesmo no caso do streaming de vídeo, é quase sempre usada uma porta TCP para estabelecer a conexão e enviar informações de controle, deixando o UDP apenas para o envio dos dados.

As portas TCP mais usadas são:

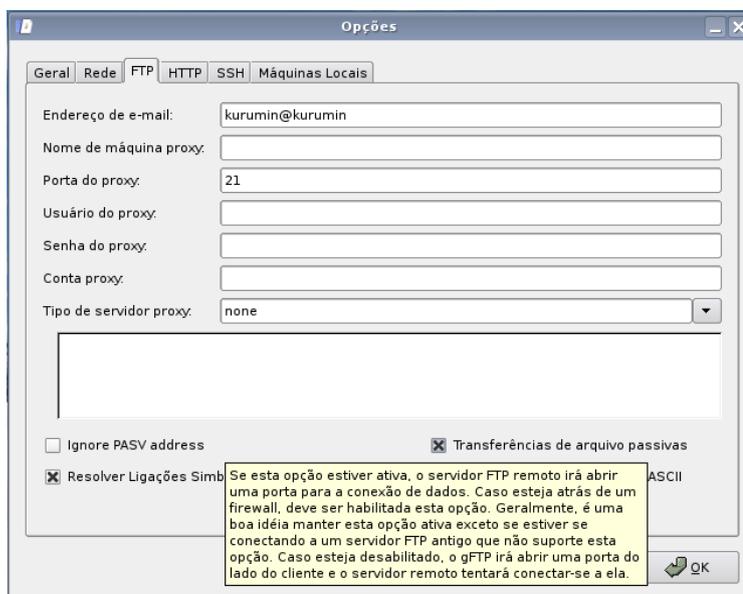
**21: FTP** – O FTP é um dos protocolos de transferência de arquivos mais antigos e ainda assim um dos mais usados. O ponto fraco do FTP é a questão da segurança: todas as informações, incluindo as senhas trafegam em texto puro e podem ser capturadas por qualquer um que tenha acesso à transmissão.

O FTP possui dois modos de operação: passivo e ativo. No modo ativo, o cliente contata o servidor usando uma porta vaga aleatória, como, por exemplo, a porta 1026, endereçando o pacote à porta 21 do servidor. O servidor imediatamente contata o cliente de volta, usando a porta seguinte (do cliente) para enviar os dados. Se o cliente usou a porta 1026 para abrir a conexão, então o servidor enviará os dados na porta 1027. O problema é que o modo ativo não funciona quando o cliente acessa através de uma conexão compartilhada. Ao tentar responder, o servidor cairia na porta 1027 do gateway da rede, sem conseguir chegar ao cliente.

No modo passivo, o cliente também abre a conexão contatando a porta 21 do servidor; entretanto, ao invés de iniciar a conexão imediatamente, o servidor responde avisando que o cliente pode contatá-lo numa segunda porta, escolhida aleatoriamente (a 2026, por exemplo). O cliente inicia, então, uma nova conexão na porta especificada e o servidor responde enviando os dados.

Esta porta fica reservada ao cliente durante o tempo que durar a transferência. Em teoria, isto seria um limite ao número de clientes que poderiam se conectar simultaneamente, mas, na prática, seriam necessárias mais de 64.000 conexões simultâneas ao mesmo servidor FTP para esgotar as portas disponíveis.

Praticamente todos os clientes de FTP atuais utilizam o modo passivo por padrão, mas isso pode ser modificado dentro da configuração. Alguns poucos servidores de FTP não podem ser acessados em modo passivo, pois para isso é necessário que o administrador faça uma configuração de firewall mais cuidadosa, mantendo abertas um conjunto de portas altas.



Em resumo, no modo ativo o servidor precisa ter aberta apenas a porta 21, mas em compensação o cliente precisa acessar a web diretamente e ter um conjunto de portas altas abertas no firewall. No modo passivo, os papéis se invertem: o cliente não precisa ter portas abertas, mas o servidor sim.

**22: SSH** – O SSH é o canivete suíço da administração remota em servidores Linux. Inicialmente o SSH permitia executar apenas comandos de texto remotamente; depois passou a permitir executar também aplicativos gráficos e, em seguida, ganhou também um módulo para transferência de arquivos, o SFTP. A vantagem do SSH sobre o Telnet e o FTP é que tudo é feito através de um canal encriptado, com uma excelente segurança.

O SSH pode ser usado também para encapsular outros protocolos, criando um túnel seguro para a passagem dos dados. Criando túneis, é possível acessar servidores de FTP, proxy, e-mail, rsync, etc. de forma segura. Graças a isso, o SSH é usado como meio de transporte por diversos programas, como o NX Server.

O sistema de encriptação utilizado pelo SSH, assim como os túneis encriptados trabalham no nível 6 do modelo OSI, acima da camada de sessão, do protocolo TCP/IP e de toda a parte física da rede.

Ao contrário do FTP, o SSH não precisa de portas adicionais: tudo é feito através da porta 22, que é a única que precisa ficar aberta no firewall do servidor. O cliente não precisa ter porta alguma aberta e pode acessar através de uma conexão compartilhada.

**23: Telnet** – O Telnet é provavelmente o protocolo de acesso remoto mais antigo. A primeira demonstração foi feita em 1969, com o acesso de um servidor Unix remoto, ainda através da antiga Arpanet, muito antes de ser inventado o padrão Ethernet e, antes mesmo da primeira versão do TCP/IP. O Telnet foi muito usado durante a década de 80 e 90, mas depois caiu em desuso, sendo rapidamente substituído pelo SSH. Além de não possuir nenhum dos recursos mais sofisticados suportados pelo SSH, o Telnet é um protocolo completamente aberto (no sentido pejorativo), que transmite login, senha e todos os comandos em texto puro. Isso torna ridiculamente simples capturar a transmissão (usando, por exemplo, o Wireshark) e assim "invadir" o servidor, usando a senha roubada.

Uma curiosidade, é que o sistema usado pelo Telnet para a transmissão de comandos é usado como base para diversos outros protocolos, como o SMTP e o HTTP. De fato, você pode usar um cliente Telnet para mandar um e-mail, ou mesmo acessar um servidor web, desde que consiga simular uma conexão HTTP válida, como faria um navegador.

**25: SMTP** – O SMTP é o protocolo padrão para o envio de e-mails. Ele é usado tanto para o envio da mensagem original, do seu micro até o servidor SMTP do provedor, quanto para transferir a mensagem para outros servidores, até que ela chegue ao servidor destino. Tradicionalmente, o Sendmail é o servidor de e-mails mais usado, mas, devido aos problemas de segurança, ele vem perdendo espaço para o Qmail e o Postfix.

**53 (UDP): DNS** – Os servidores DNS são contatados pelos clientes através da porta 53, UDP. Eles são responsáveis por converter nomes de domínios como "guiadohardware.net" nos endereços IP reais dos servidores.

Existem no mundo 13 servidores DNS principais, chamados "root servers". Cada um deles armazena uma cópia completa de toda a base de endereços. Estes servidores estão instalados em países diferentes e ligados a links independentes. A maior parte deles roda o Bind, mas pelo menos um deles roda um servidor diferente, de forma que, mesmo no caso de um gigantesco cyberataque, pelo menos um dos servidores continue no ar, mantendo a internet operacional.

Para acessar qualquer endereço, é preciso primeiro consultar um servidor DNS e obter o endereço IP real do servidor. Em geral, uma consulta a um dos root servers demora alguns segundos, por isso os provedores de acesso e

responsáveis por grandes redes sempre configuram servidores DNS locais, que criam um cache das consultas anteriores, de forma a agilizar o acesso. Você mesmo pode configurar um servidor DNS para a sua rede usando o Bind.

**67: Bootps, 68: Bootpc** – Estes dois protocolos são usados em sistemas de boot remoto (como no LTSP), onde os clientes não possuem HD nem CD-ROM e acessam todos os arquivos de que precisam a partir do servidor.

**69 (UDP): TFTP** – O TFTP é uma versão simplificada do FTP, que utiliza portas UDP para a transferência dos dados e não inclui suporte à correção de erros. Ele pode ser usado para transferência de arquivos em geral, mas é mais freqüentemente usado em sistemas de boot remoto.

**80: HTTP** – O HTTP é o principal protocolo da internet, por onde acessamos as páginas. Embora a porta 80 seja a porta padrão dos servidores web, é possível configurar um servidor web para usar qualquer outra porta TCP. Neste caso, você precisa especificar a porta ao acessar o site, como em: <http://200.234.34.12:8080>.

**110: POP3** – Servidores de e-mail, como o Postfix, armazenam os e-mails recebidos numa pasta local. Se você tiver acesso ao servidor via SSH, pode ler estes e-mails localmente, usando Mutt.

Entretanto, para transferir os e-mails para sua máquina, é necessário um servidor adicional. É aí que entra o protocolo POP3, representado pelo courier-pop e outros servidores.

Programas como o Thunderbird e o Outlook contatam o servidor POP3 através da porta 110 e baixam as mensagens utilizando um conjunto de comandos de texto, derivados do Telnet. Originalmente, o POP3 é um protocolo tão inseguro quanto o Telnet, mas os servidores atuais suportam encriptação via SSL (o mesmo sistema de encriptação usado para acessar páginas seguras, via HTTPSs), o que garante um bom nível de segurança.

137, 138 e 139: **Netbios** – Estas três portas são usadas pelo protocolo de compartilhamento de arquivos em redes Microsoft. Cada uma das portas tem uma função específica (nome, datagrama e sessão), mas é necessário que as três estejam abertas no firewall para que a visualização dos compartilhamentos e acesso aos arquivos funcione corretamente.

**143: IMAP** – O IMAP é mais um protocolo para recebimento de e-mails, assim como o POP3. A diferença entre os dois é que, ao receber os e-mails via POP3, eles são apagados do servidor assim que baixados, liberando o espaço usado na caixa postal. No IMAP, os e-mails continuam no servidor até serem deletados manualmente.

Embora oferecer contas de e-mail com acesso via IMAP seja muito mais oneroso do que via POP3 (já que o número de requisições é maior, e os usuários podem conservar mensagens antigas por muito tempo), ele vem "roubando a cena" com a popularização dos webmails, que são justamente clientes IMAP, que rodam no próprio servidor (através do Apache ou outro servidor web), e são acessados no cliente usando o navegador.

**177: XDMCP** – O XDMCP é um protocolo de acesso remoto, suportado nativamente pelo X. Ele permite rodar aplicativos remotamente e é a base para o LTSP e outros sistemas onde é usado um servidor central e terminais leves. Pode ser também usado no dia-a-dia, para simplesmente rodar programas instalados em outra máquina da rede.

A vantagem do XDMCP é que ele é um protocolo bastante simples e rápido, que oferece um bom desempenho via rede local e consome poucos recursos, tanto no servidor, quanto no cliente. Ele é também um recurso nativo do X, de forma que você não precisa instalar nenhum software adicional, basta ativar o recurso na configuração do KDM ou GDM (os gerenciadores de login usados nas distribuições atuais)

A desvantagem é que o XDMCP é um protocolo "da velha guarda", que não utiliza encriptação, e utiliza um conjunto de portas altas para enviar dados aos clientes.

Além da porta 177, onde o servidor recebe conexões, é necessário que estejam abertas as portas de 6010 à 6099 (no servidor) e as portas de 5000 a 5200 nos clientes, o que complica um pouco as coisas ao manter um firewall ativo.

**389: LDAP** – O LDAP é muito usado atualmente para criar servidores de autenticação e definir permissões de acesso para os diferentes usuários da rede. Existem vários padrões de LDAP, um dos mais usados é o OpenLDAP, suportado pela maioria das distribuições Linux atualmente em uso.

**443: HTTPS** – O HTTPS permite transmitir dados de forma segura, encriptados em SSL. Ele é usado por bancos e todo tipo de site de comércio eletrônico ou que armazene informações confidenciais.

Naturalmente, esta é uma lista rápida, contendo apenas as portas mais usadas. Você pode ver uma lista longa e completa, com todos os serviços conhecidos e as portas utilizadas por cada um no: <http://www.iana.org/assignments/port-numbers>

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.

ESPECIAL

# Montagem de micros

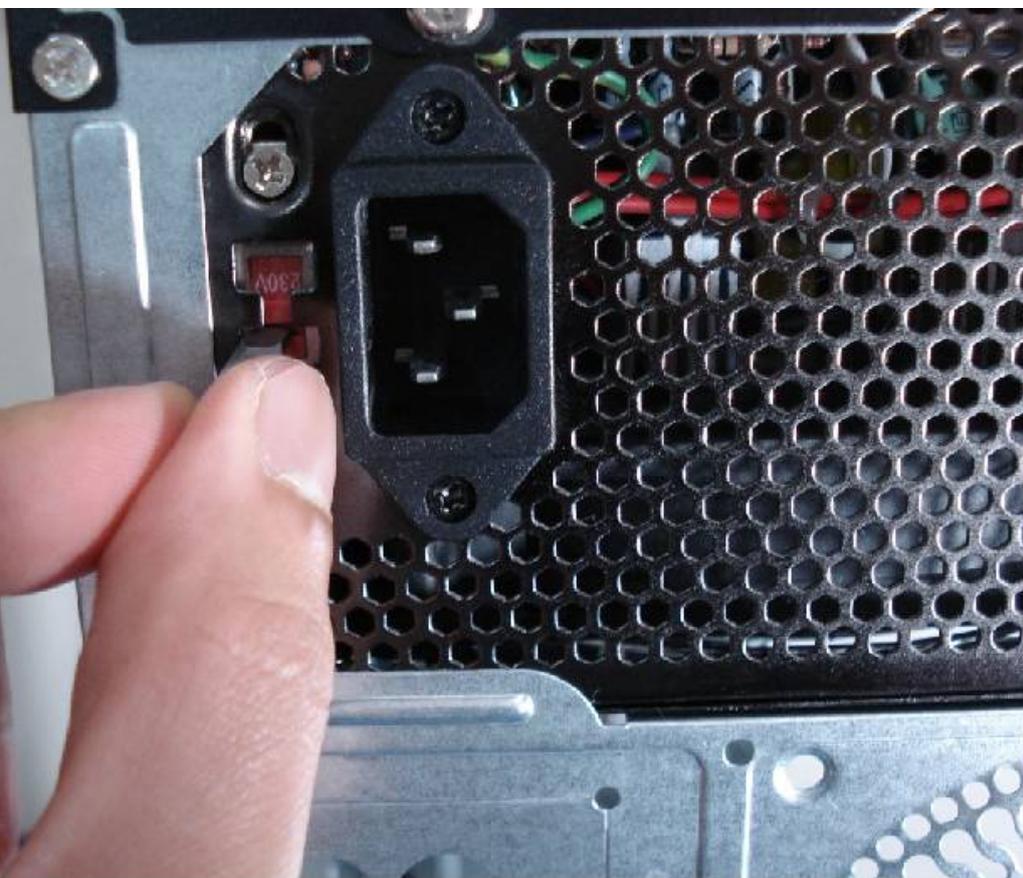
por Carlos E. Morimoto

Hoje em dia, graças ao programa do PC popular e outras formas de incentivos, PCs de marcas nacionais passaram a ser vendidos até mesmo em supermercados, com preços relativamente baixos e boas condições de pagamento.

O maior problema é que junto com o preço baixo, temos uma qualidade que muitas vezes deixa a desejar, de forma que o mercado para PCs montados, com configurações melhores continua existindo, talvez maior do que nunca. Além disso, a maior quantidade de micros vendidos, grande parte deles modelos com componentes de baixa qualidade, acabam aumentando as oportunidades de trabalhos para técnicos de manutenção.

Outra questão é que, qualquer micreiro que se preze, não compra PCs montados; compra as peças e monta ele mesmo :). Vamos então ao especial deste mês, sobre montagem de micros.

Depois de desempacotar as peças, a primeira coisa a fazer é mudar a posição da chave de tensão da fonte de alimentação. Por segurança, todas as fontes vem de fábrica com a posição no "220V", já que ligar a fonte chaveada para 220 em uma tomada de 110 não causa danos, bem diferente do que acontece ao fazer o contrário. O problema é que a lei de Murphy faz com que você sempre esqueça de trocar a chave de posição, fazendo com que mais adiante o micro simplesmente não ligue e você fique sem saber o por que. :)



Apesar de muitas vezes não parecer, o gabinete é um componente bastante barato e fácil de fabricar. A matéria prima básica são chapas de aço bastante finas, que são dobradas e prensadas até chegar à forma final. Este aço bruto é bastante barato e pouco resistente, ao contrário do aço temperado usado em aplicações mais nobres. Os gabinetes mais baratos chegam a custar menos de 100 reais e quase metade deste valor é referente à fonte de alimentação que vem de brinde.

O maior problema com os gabinetes baratos é a presença de rebarbas, que agem como lâminas, cortando os dedos dos descuidados. A presença de rebarbas é sinônimo de gabinete de baixa qualidade, uma dica para evitar o fabricante na próxima compra, para não cometer o mesmo erro duas vezes.

Além da questão do acabamento, existe uma tendência crescente de substituir o aço por alumínio nos modelos mais caros. Existem ainda gabinetes de materiais alternativos, voltados para quem gosta de casemod, feitos acrílico, resina, vidro ou até mesmo madeira.

Além do material usado, acabamento e da questão estética de uma forma geral, os gabinetes se diferenciam pela presença de portas USB ou conectores de áudio frontais (ou outros acessórios) e pela questão da ventilação.

De qualquer forma, a principal função do gabinete é servir como um suporte para os demais componentes. Você pode muito bem montar um micro dentro de um armário, de uma gaveta, ou até mesmo dentro de uma caixa de pizza, mas sem uma fonte de alimentação com um mínimo de qualidade, você corre o risco de ver pentes de memória queimados, HDs com badblocks, capacitores estufados na placa mãe e assim por diante em pouco tempo. De uma forma geral, as fontes que acompanham os gabinetes valem o que custam (muito pouco), por isso você deve procurar substituí-las por fontes melhores em qualquer micro com componentes mais caros, ou em micros de trabalho, que vão desempenhar um papel importante.

Como (com exceção de alguns modelos high-end) todas as fontes utilizam o mesmo tamanho padrão, é muito fácil substituir a fonte por outra.

Voltando à montagem, o próximo passo é tirar ambas as tampas do gabinete. Aproveite para remover também as tampas das baias dos drives de CD e DVD que for utilizar.



*A parte interna do gabinete possui um padrão de furação, destinado aos suportes e parafusos que prendem a placa mãe. Todos os parafusos necessários devem vir junto com o gabinete:*



Remova também a tampa do painel ATX, ao lado das aberturas dos exaustores. Cada placa mãe utiliza uma combinação própria de conectores, de forma que o que vem com o gabinete é inútil, já que nunca combina com os conectores da placa mãe. Por isso o substituímos pela tampa que acompanha a placa mãe, feita sob medida para ela.

Dependendo da marca e modelo, podem ser usados pinos plásticos, como os da esquerda, encaixes como os da direita ou (mais comum) espaçadores metálicos como os do centro. Existem ainda suportes plásticos como os dois na parte inferior da foto, que podem ser usados como apoio, inseridos nos furos na placa mãe que não possuem par no gabinete. Eles eram mais usados antigamente, na época dos gabinetes AT, mas é sempre bom ter alguns à mão.

*O conjunto com os parafusos e espaçadores necessários deve vir junto com o gabinete. Ele é chamado de "kit de montagem" pelos fabricantes. Normalmente o gabinete vem também com o cabo de força, com exceção dos modelos sem fonte, onde o cabo vem junto com a fonte avulsa.*

As placas ATX possuem normalmente 6 furos para parafusos e mais dois ou três pontos de apoio adicionais, que podem ser usados pelos suportes plásticos. A posição deles, entretanto, varia de acordo com a distribuição dos componentes na placa, de forma que o gabinete inclui um número muito maior de furos. Com o tempo, você acaba aprendendo a descobrir quais usar "de olho", mas no início você acaba perdendo tempo comparando as furações da placa e do gabinete para ver onde colocar os suportes.

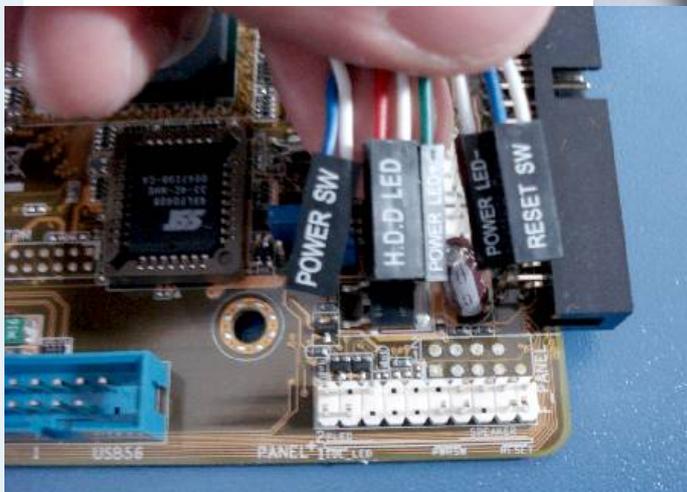
*Uma dica é que você pode usar uma folha de papel para achar mais facilmente as combinações entre a furação da placa mãe e a do gabinete. Coloque a placa mãe sobre o papel e use uma caneta para fazer pontos no papel, um para cada furo disponível. Depois, coloque o papel sobre a chapa do gabinete e vá colocando os parafusos onde os pontos coincidirem com a furação. Muito simples mas bastante prático. :)*



É importante apertar os parafusos de suporte usando uma chave torx, para que eles continuem no lugar depois de parafusar e desparafusar a placa mãe. Se não forem bem apertados, os parafusos de suporte acabam saindo junto com os usados para prender a placa mãe ao removê-la, o que não é muito agradável.



Antes de instalar a placa mãe, você pode aproveitar para encaixar os conectores do painel frontal do gabinete e das portas USB frontais, que são muito mais fáceis de encaixar com a placa mãe ainda sobre a mesa, do que com ela já instalada dentro do espaço apertado do gabinete, com pouca luz.



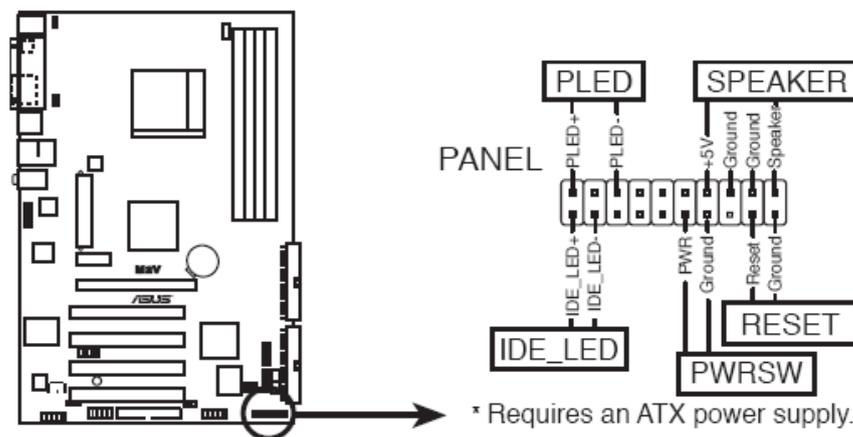
Tanto os dois botões, quanto o speaker (que usa um conector de 4 pinos, embora apenas 2 sejam usados) não possuem polaridade, de forma que podem ser ligados em qualquer sentido. Os LEDs por sua vez, precisam ser ligados na polaridade correta, caso contrário não funcionam.

Quase sempre, a própria placa traz uma indicação resumida decalcada, indicando inclusive as polaridades, mas em caso de dúvidas você pode dar uma olhada rápida no manual, que sempre traz um esquema mais visível:

Infelizmente, não existe muita padronização nos contatos do painel frontal, cada fabricante faz do seu jeito. Embora o mais comum seja que os pinos fiquem no canto inferior direito da placa, até mesmo a posição pode mudar de acordo com a placa. Em muitas ele fica mais para cima, quase no meio da placa.

Nos gabinetes ATX, temos basicamente 5 conectores: Power SW (o botão liga/desliga), Reset SW (o botão de reset), Power LED (o led que indica que o micro está ligado), HD LED (o led que mostra a atividade do HD) e o speaker.

Cada um dos contatos é formado por dois pinos, um positivo e um neutro. Nos conectores, o fio colorido corresponde ao positivo e o branco ao neutro.



\* Requires an ATX power supply.

Em micros antigos, ainda na época dos gabinetes AT, existiam também os conectores Keylock (uma chave no gabinete que permitia travar o teclado), Turbo SW (a chave do botão "turbo") e o Turbo LED (o LED correspondente).

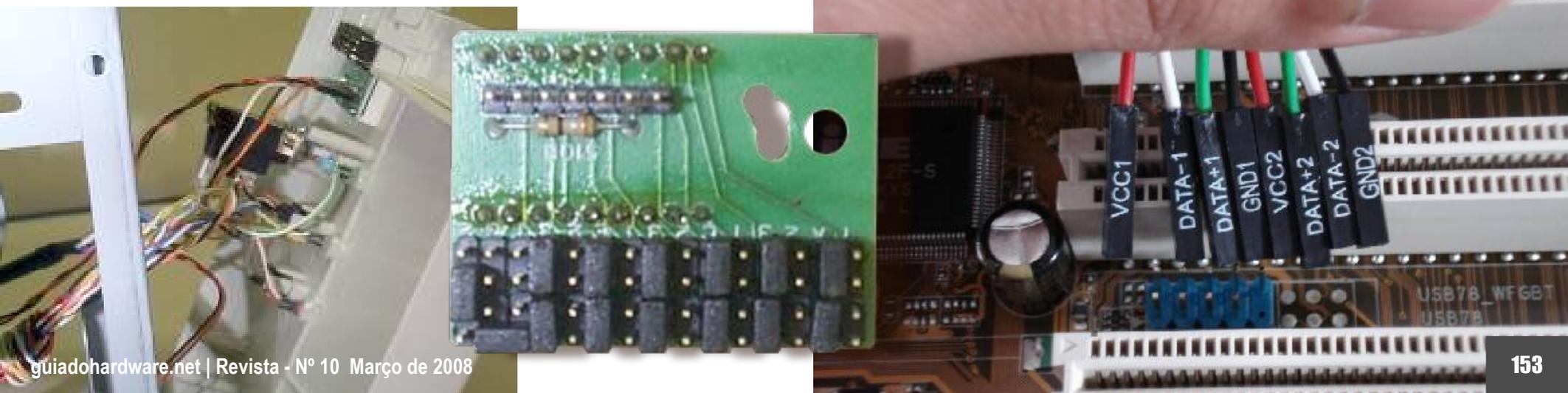
O botão "turbo" é uma história curiosa. Ele surgiu com o lançamento dos primeiros micros 286 e tinha a função de reduzir a frequência de operação do processador, fazendo com que o micro ficasse com um desempenho similar ao de um XT (o micro operava à frequência normal apenas enquanto o botão estivesse pressionado).

Isso permitia rodar alguns jogos e outros programas que ficavam rápidos demais se executados no 286. Por algum motivo, o botão "turbo" continuou presente nos gabinetes AT até a época dos micros Pentium, embora não fosse mais usado.

Outra curiosidade era o mostrador do clock, também usado na época dos micros Pentium 1. Ele tinha uma função puramente decorativa, mostrando a frequência de operação do processador. O engraçado era que ele não tinha relação nenhuma com a frequência real. Era simplesmente um painel digital, configurado através de jumpers, onde você podia colocar a frequência que quisesse. Felizmente ele também saiu de moda e não é mais usado nos gabinetes atuais.

Em seguida, temos os conectores das portas USB frontais, também conectados diretamente na placa mãe. Eles precisam ser encaixados com atenção, pois inverter os contatos das portas USB (colocando o polo positivo de alimentação na posição do negativo de dados, por exemplo) vai fazer com que pendrives, mp3players e outros dispositivos eletrônicos conectados nas portas USB sejam queimados, um problema muito mais grave do que deixar parafusos soltos ou inverter a polaridade de um LED, por exemplo.

*Os conectores USB (ou headers USB) na placa mãe são conectores de 9 pinos, facilmente reconhecíveis. Cada porta USB utiliza 4 pinos, dois para a alimentação e dois para dados, sendo que dentro de cada par, um é o positivo e o outro o negativo. O nono pino do conector serve apenas como orientação, indicando o lado referente aos dois fios pretos, referentes ao polo neutro do par de alimentação:*

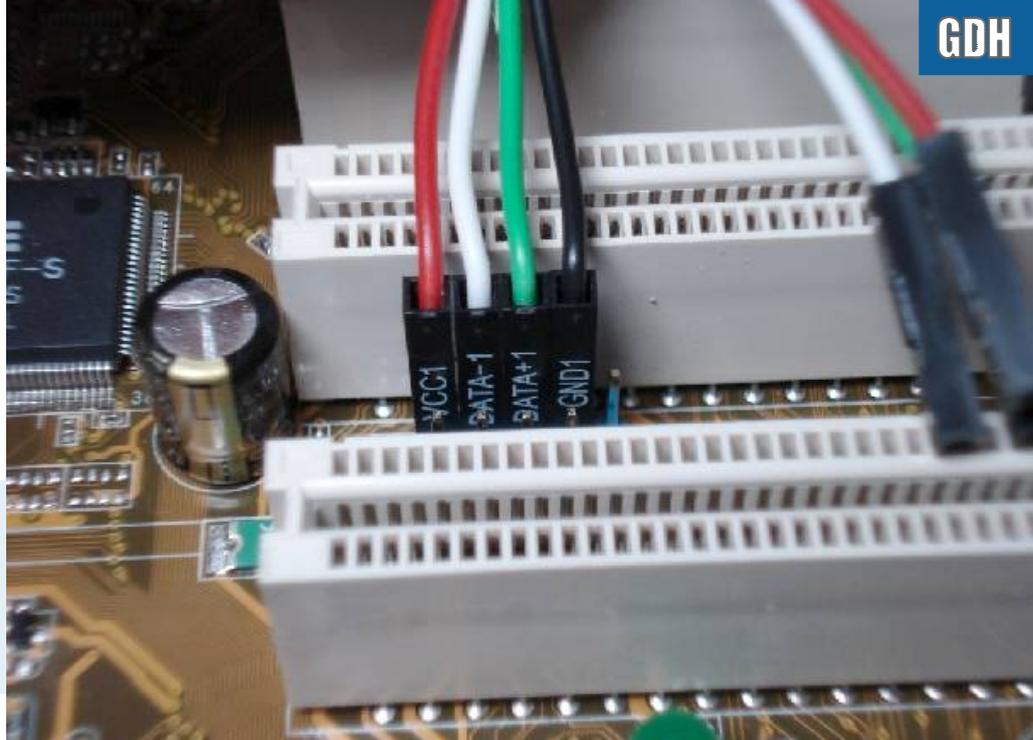


# Especial Montagem de micros

Cada header USB inclui duas portas. Uma placa mãe com "12 portas USB" normalmente inclui 4 portas no painel traseiro e mais 4 headers para a conexão das portas frontais do gabinete. Alguns gabinetes possuem 4 portas frontais, mas a maioria inclui apenas duas. Existem ainda diversos tipos de suportes com portas adicionais, leitores de cartões e outras bugigangas instaladas na baía do drive de disquetes, em uma das baias dos drives ópticos ou em uma das aberturas traseiras. Assim como as portas frontais, eles também são ligados nos headers USB da placa mãe.

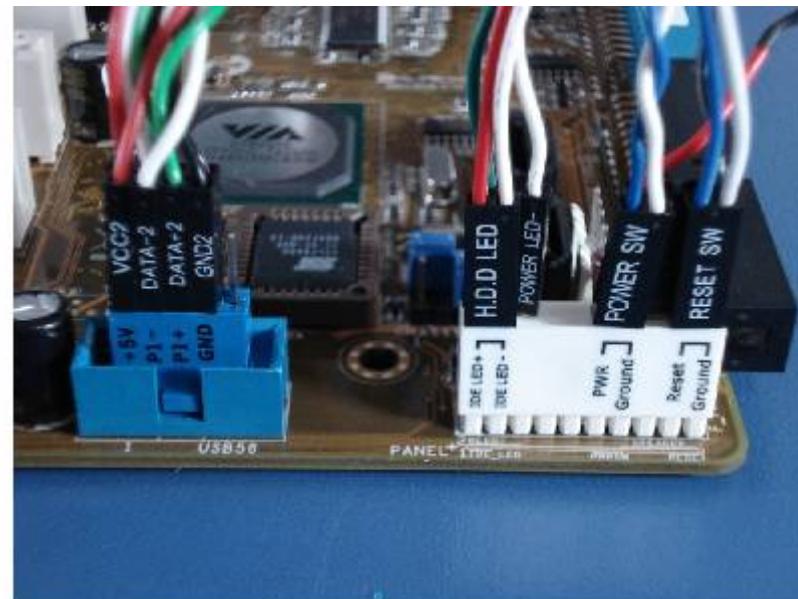
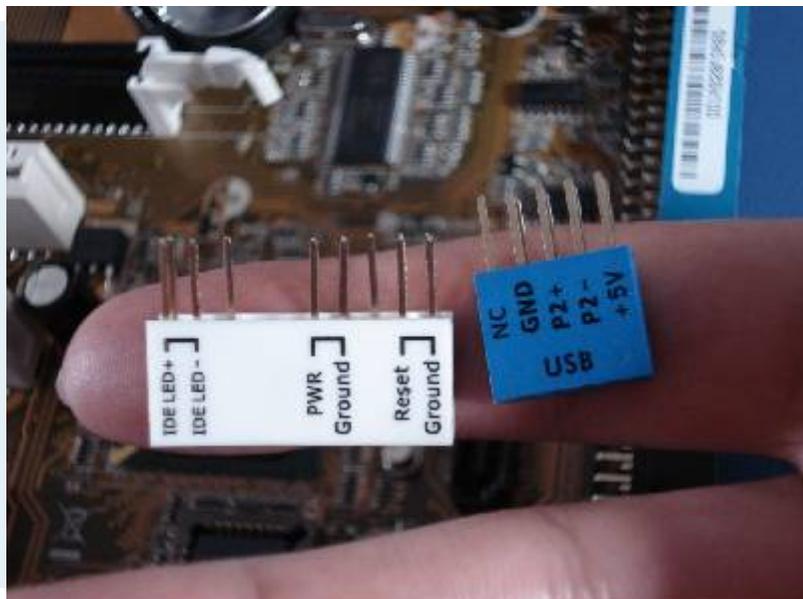
*Dentro de cada header a ordem os fios é a seguinte: VCC (vermelho), DATA - (branco), DATA + (verde) e GND (preto), onde o GND fica sempre do lado do nono pino, que serve como guia. Ligue primeiro os pinos da porta 1, para não arriscar misturá-los com os da segunda porta. :)*

Fazendo isso com a atenção, não existe muito o que errar; o problema é que se você precisa montar vários micros, acaba tendo que fazer tudo rápido, o que abre espaço para erros.



Instalação dos conectores das portas USB frontais do gabinete

*A partir de 2007, a Asus passou a fornecer "agrupadores" para os conectores do painel e das portas USB frontais junto com as placas. Eles são práticos, pois ao invés de ficar tentando enxergar as marcações na placa mãe você pode encaixar os conectores no suporte e depois encaixá-lo de uma vez na placa mãe:*



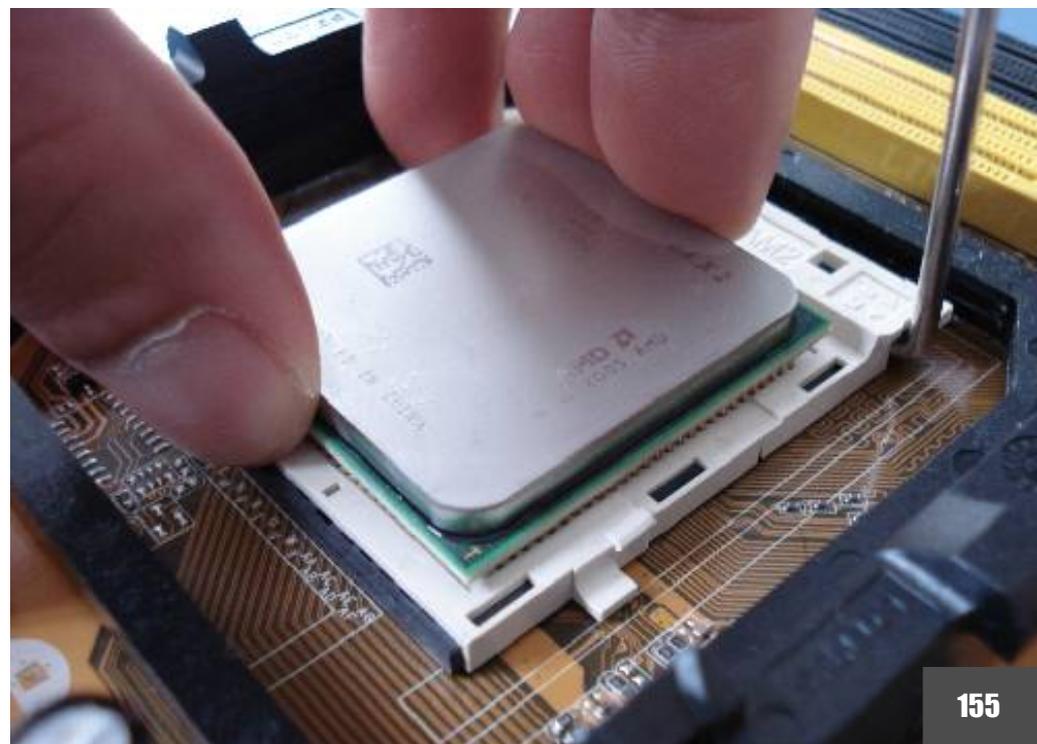
Antes de instalar a placa-mãe dentro do gabinete, você pode aproveitar também para instalar o processador, o cooler e os módulos de memória.

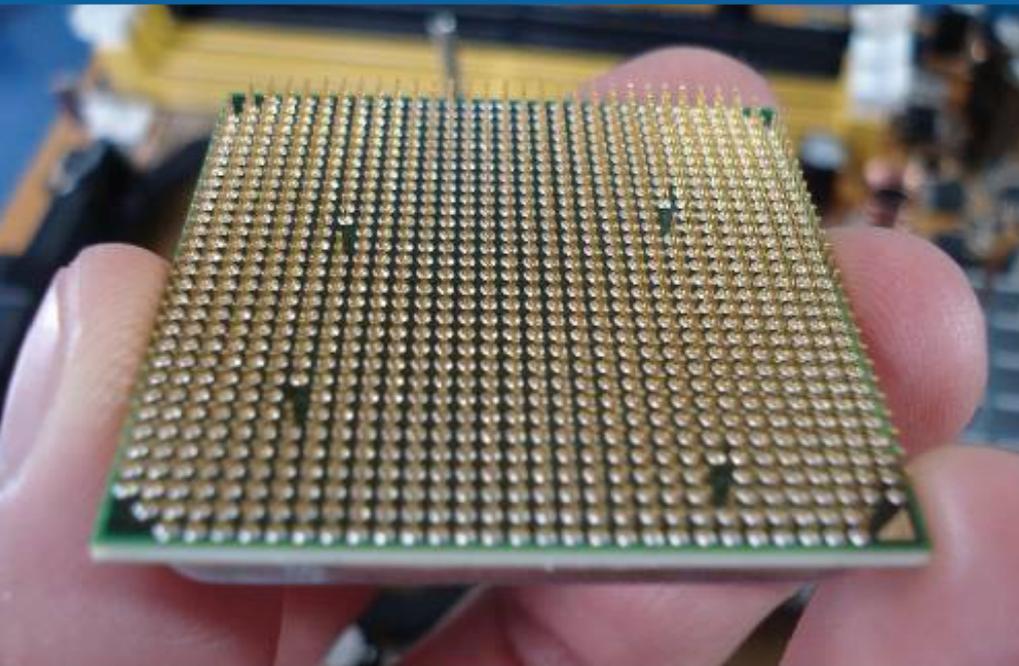
Com exceção dos antigos Pentiums e Athlons em formato de cartucho, todos os processadores são ligados ao chipset e demais componentes da placa mãe através de um grande número de pinos de contato. Como o encapsulamento do processador é quadrado, seria muito fácil inverter a posição de contato (como era possível nos 486), o que poderia inutilizar o processador quando o micro fosse ligado e a alimentação elétrica fornecida pela placa mãe atingisse os pinos errados.

*Para evitar isso, todos os processadores atuais possuem uma distribuição de pinos que coincide com a do soquete em apenas uma posição. Você pode notar que existe uma seta no canto inferior esquerdo deste Athlon X2, que coincide com uma pequena seta no soquete:*

O encaixe do processador é genericamente chamado de "ZIF" (zero insertion force), nome que indica justamente que você não precisa fazer nenhuma pressão para encaixar o processador. A própria ação da gravidade é suficiente para encaixá-lo no soquete. O ideal é simplesmente segurar o processador alguns milímetros acima do soquete e simplesmente soltá-lo, deixando que a lei da gravidade faça seu trabalho. Isso evita que você entorte os pinos se estiver sonolento e tentar encaixar o processador no sentido errado.

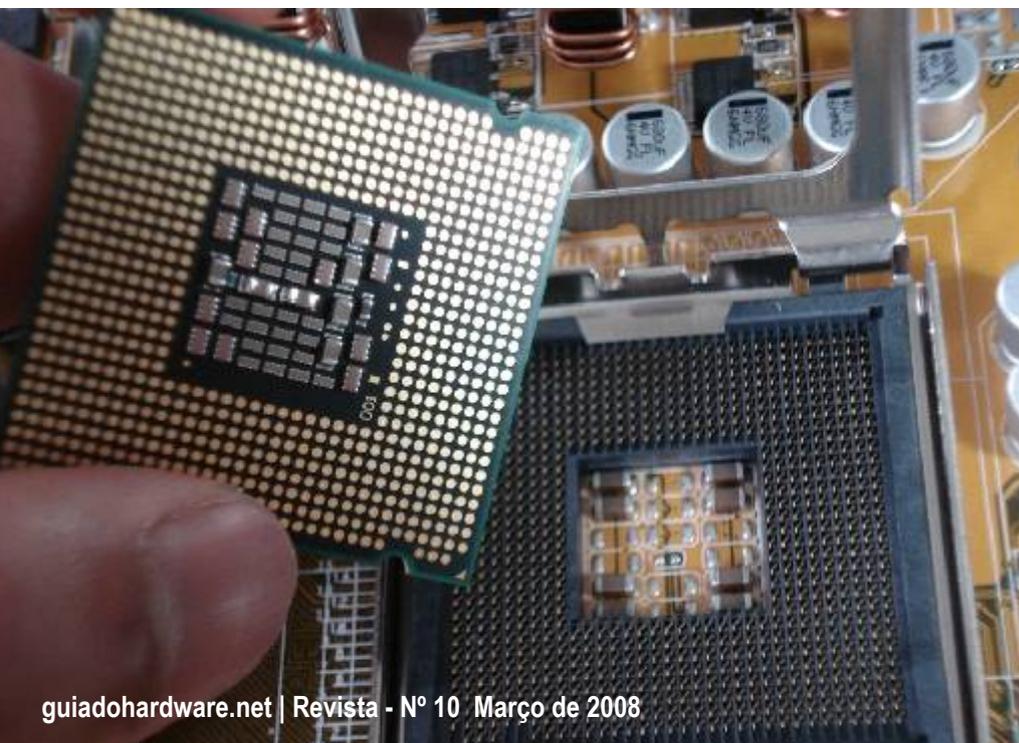
Danos aos pinos do processador são desesperadores, pois é muito difícil desentortar os pinos. Se alguns poucos pinos forem entortados, sobretudo pinos nos cantos, você pode tentar desentortá-los usando uma lâmina, tentando deixá-los alinhados com os outros da fileira. Em alguns casos, um alicate de precisão também pode ajudar. O trabalho nunca vai ficar perfeito, mas você tem a chance de deixar os pinos retos o suficiente para que eles entrem no soquete, mesmo que seja necessário aplicar um pouco de pressão.





O Athlon X2 e o Phenom X4 serão possivelmente os últimos processadores Intel/AMD para micros PCs a utilizarem o formato tradicional, com pinos. Desde o Pentium 4 com Core Prescott a Intel adotou o formato LGA, onde os pinos são movidos do processador para o soquete. A AMD utiliza um sistema semelhante no soquete-F utilizado pelos Opterons, Athlon Quad FX e Phenom FX e a tendência é que ele substitua as placas AM2, AM2+ e AM3 nos próximos anos.

A boa notícia é que no sistema LGA não existem mais pinos para serem entortados no processador, de forma que ele torna-se um componente muito resistente mecanicamente. A má é que agora temos um grande número de pinos ainda mais frágeis no soquete da placa mãe, o que demanda ainda mais cuidado ao instalar o processador. Diferentemente dos pinos dos processadores tradicionais, os pinos do soquete LGA são praticamente impossíveis de desentortar. Ao danificar um grande número deles, você simplesmente condena a placa mãe.



*A melhor estratégia continua sendo suspender o processador apenas alguns milímetros acima dos pinos de contato e simplesmente soltá-lo, deixando o resto por conta da gravidade. Assim você minimiza a possibilidade de danificar os pinos. No caso dos processadores soquete 775, duas guias de um dos lados do soquete impedem que o processador seja encaixado na direção errada. Olhando com atenção, você verá também uma seta em baixo relevo no canto inferior esquerdo do soquete, que faz par com a seta decalcada em um dos cantos do processador.*



*Outra mudança trazida pelo sistema LGA é que a pressão necessária para manter o processador no lugar é feita pelo próprio soquete, e não mais pelo cooler. Isso faz com que a força necessária para fechar a alavanca do soquete nas placas soquete 775 seja muito maior.*

Com o processador instalado, o próximo passo é usar a boa e velha pasta térmica para melhorar a condutividade térmica com o cooler. Hoje em dia, existe diversos tipos de pasta térmica, que vão desde a boa e velha pasta térmica branca, à base de óxido de zinco, que é bem barata e muitas vezes vendida em tubos de 50 gramas ou mais até diversos tipos de pasta térmica "premium" com diferentes compostos, vendidas em seringas ou vidros. Os próprios coolers muitas vezes acompanham envelopes de pasta térmica branca.

Usar uma pasta "premium", baseada em algum composto metálico normalmente reduz a temperatura de operação do processador em dois ou até três graus em relação a usar alguma pasta branca genérica. A diferença é maior em over-clocks mais extremos, onde a dissipação térmica do processador (e conseqüentemente a temperatura de funcionamento) é mais elevada.

Se você já está gastando mais no cooler e na placa mãe, pensando justamente em recuperar o investimento com um overclock agressivo, então gastar 20 reais em uma seringa de pasta Arctic Silver, para ganhar mais dois ou três graus faz sentido. Mas, ao montar um micro de baixo custo, onde você conta os trocados para conseguir colocar 512 MB de memória, vale mais à pena aproveitar a dose de pasta branca que veio de brinde com o cooler ou usar pasta branca genérica. O mais importante é não cair em modismos e deixar alguém te passar a perna tentando cobrar 40 ou 50 reais por um vidro de pasta térmica que não vai fazer milagres.



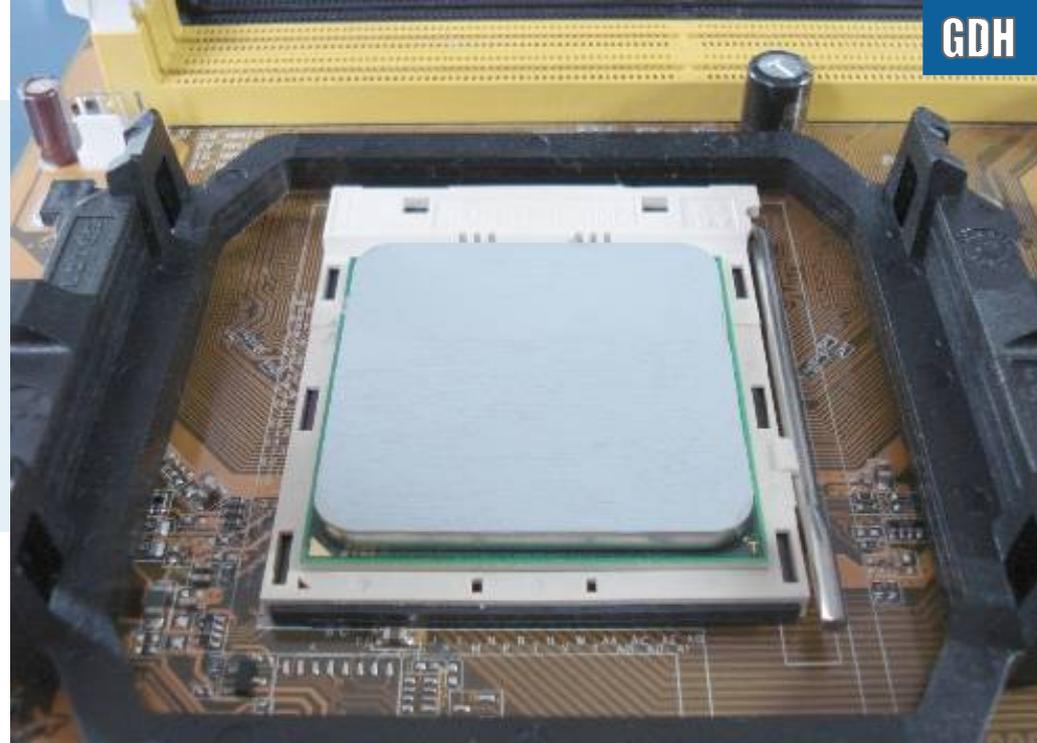
*Independentemente do tipo escolhido, a idéia básica é passar uma fina camada de pasta térmica cobrindo todo o dissipador do processador. Se você simplesmente esparar um montinho de pasta sobre o processador, a pressão exercida pelo cooler vai se encarregar de espalhá-la cobrindo a maior parte do dissipador de qualquer forma, mas a aplicação nunca fica perfeita, de forma que se você tiver tempo para espalhar a pasta uniformemente, antes de instalar o cooler, o resultado será sempre um pouco melhor. Aplicar uma camada de pasta é especialmente importante nos processadores LGA, pois neles o cooler não exerce uma pressão tão forte sobre o processador.*

Muitos coolers, sobretudo os coolers dos processadores boxed vem com uma camada de pasta térmica (quase sempre cinza) pré-aplicada. O principal objetivo é a praticidade, já que elimina uma das etapas da instalação do cooler.

Caso prefira utilizar sua própria pasta térmica, remova a camada pré-aplicada no cooler usando uma flanela e álcool isopropílico. Não use espátulas ou qualquer outro objeto metálico, pois você vai arranhar a base do cooler, o que também prejudica a dissipação de calor.

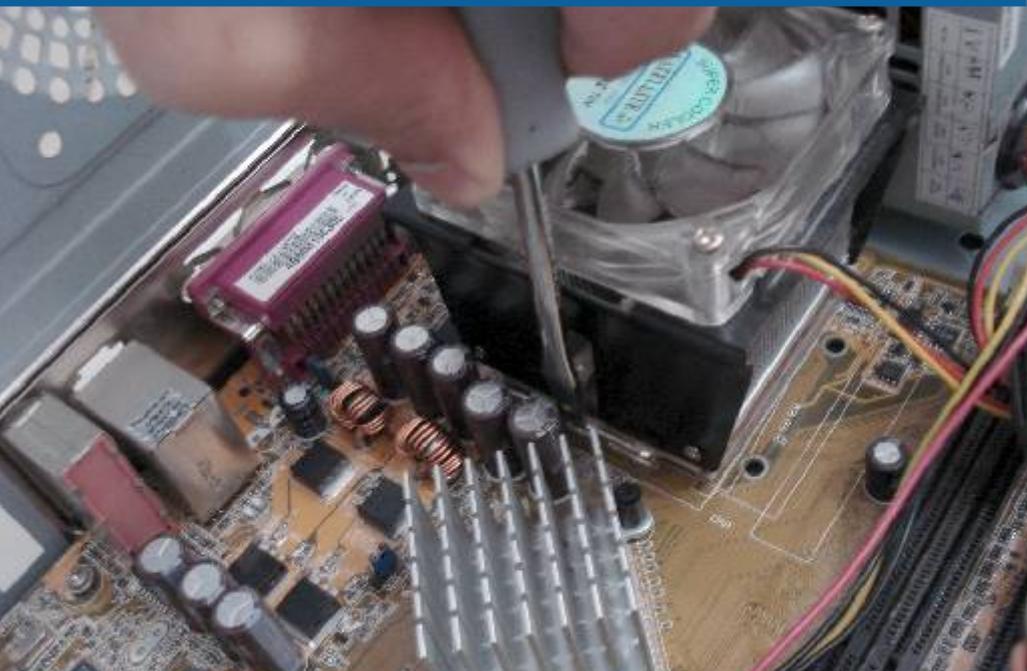
O maior problema é que muitos coolers (em sua maioria fabricadas entre 2001 e 2005) utilizavam uma camada de elastômero (um tipo de borracha, que podia ser rosa, cinza, ou mesmo branca), no lugar da pasta térmica. Ele é um material que derrete se aquecido a temperaturas superiores a 60 graus, de forma que a pressão do cooler acaba moldando-o ao processador.

O elastômero não é tão eficiente quanto a pasta térmica (mesmo se comparado à pasta branca comum) e tem a desvantagem de ser descartável, precisando ser substituído depois da primeira remoção do cooler. Ele era usado por que era barato e era considerado "bom o bastante" pelos integradores e não por ser realmente eficiente.



É fácil reconhecer o elastômero, pois ele tem aspecto e consistência de chiclete. É sempre recomendável removê-lo e substituí-lo por pasta térmica antes de instalar o cooler. Ao se deparar com um cooler com a camada de elastômero ao dar manutenção, remova sempre toda a camada antiga antes de aplicar a pasta e reinstalar o cooler. Misturar os dois materiais acaba resultando em uma camada ainda mais ineficiente.

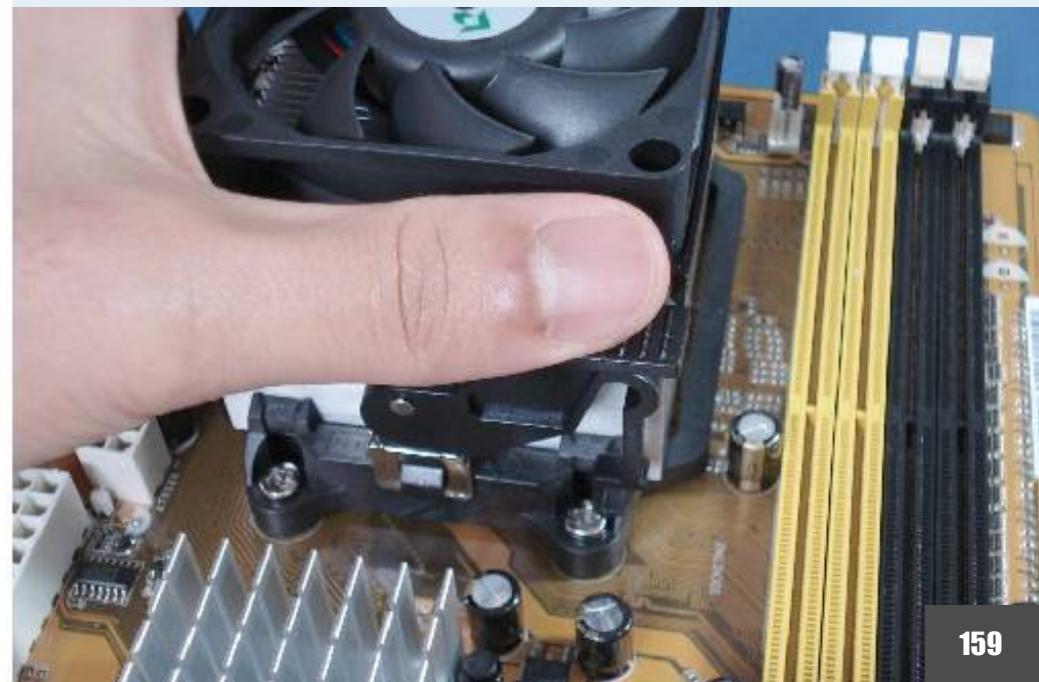
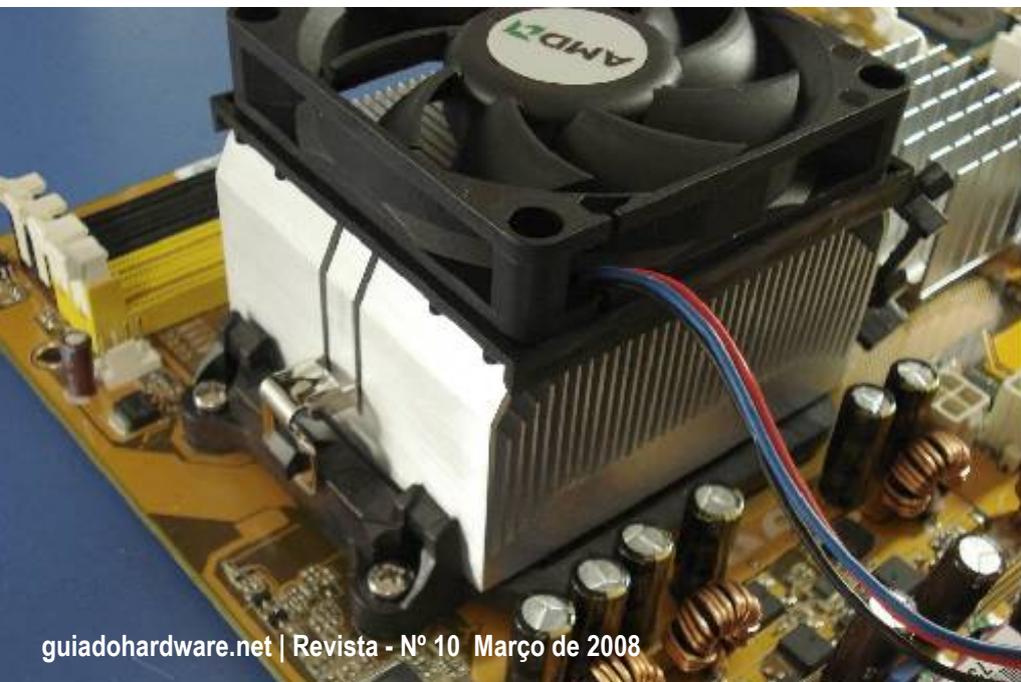
Para manter o processador firme no lugar (evitando mal contatos nos pinos) e eliminar o excesso de pasta térmica o cooler precisa pressionar o processador com uma certa pressão. Na maioria dos coolers antigos, você precisava da ajuda de uma chave de fenda para instalar e remover o cooler. A ponta era presa em um pequeno encaixe na presilha do cooler e você precisava de uma boa dose de força para encaixá-la no soquete.



Este sistema levava a acidentes, pois com frequência a chave de fenda escapava, muitas vezes destruindo trilhas e inutilizando a placa mãe. Como a pressão era exercida sobre os pinos laterais do soquete, também às vezes acontecia deles quebrarem. Para não ter que descartar a placa mãe, você acabava sendo obrigado a fazer algum "chunxo" para prender ou colar o cooler no soquete.

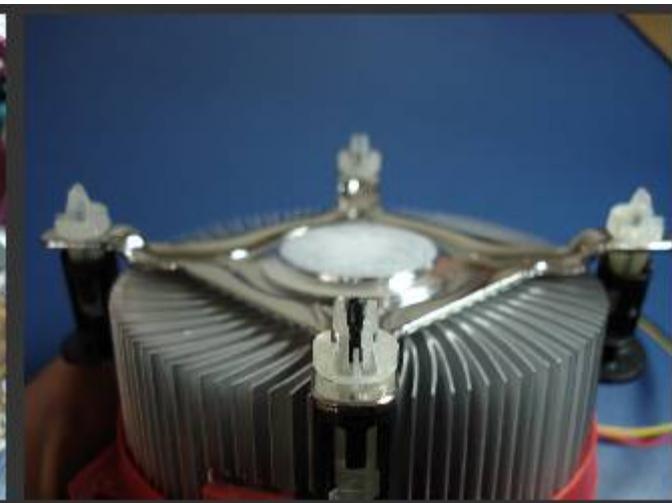
Para solucionar estes dois problemas, tanto a Intel quanto a AMD desenvolveram novos sistemas de encaixe.

*A AMD passou a usar uma "gaiola" plástica em torno do processador. Os pinos de encaixe ficam na gaiola, que é presa à placa por dois ou quatro parafusos e pode ser substituída em caso de quebra. O cooler é encaixado através de um sistema de alavanca, onde você encaixa a presilha dos dois lados e usa a alavanca presente no cooler para prendê-lo ao soquete:*



Nas placas soquete 775, a pressão necessária para manter o processador preso é exercida pelo encaixe metálico incluído no próprio soquete. A Intel se aproveitou disso para desenvolver um sistema de encaixe bastante engenhoso, onde o cooler exerce menos pressão sobre a placa mãe e é preso por 4 presilhas.

As presilhas utilizam um sistema de retenção peculiar. Girando o prendedor no sentido horário (o sentido oposto à seta em baixo relevo) você o deixa na posição de encaixe, pronto para ser instalado. Girando no sentido anti-horário, o prendedor se solta, permitindo que o cooler seja removido:



Ao instalar o cooler, você só precisa deixar as presilhas na posição de instalação e pressioná-la em direção a placa. Ao contrário dos coolers para placas soquete 754, 939 e AM2, você pode encaixar o cooler em qualquer sentido.

Eu pessoalmente prefiro instalar o cooler pressionando uma das presilhas de cada vez, antes de instalar a placa mãe dentro do gabinete:

Outra forma de instalar o cooler seria pressionar as 4 presilhas de uma vez, usando as duas mãos, com a placa já instalada dentro do gabinete. Esta segunda opção faz com que seja exercida uma grande pressão sobre a placa mãe, o que é sempre bom evitar.

Com o cooler instalado, não se esqueça de instalar o conector de alimentação do cooler. As placas atuais oferecem pelo menos dois conectores de alimentação; uma para o cooler do processador e outro para a instalação de um exaustor frontal ou traseiro.

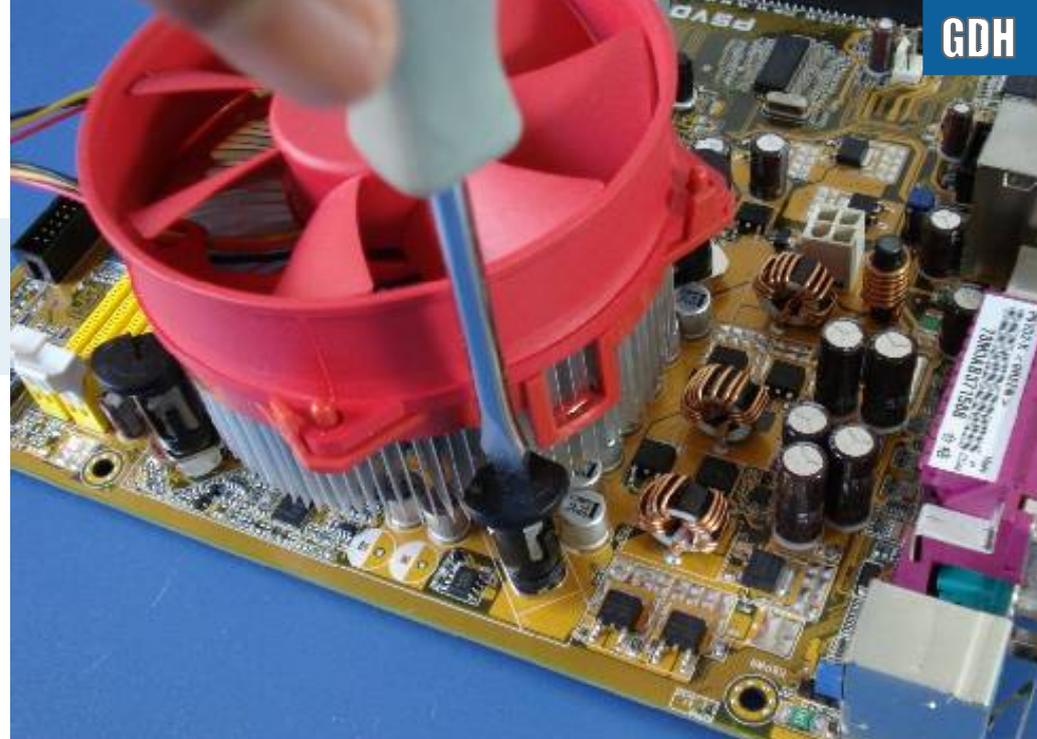
Muitas placas oferecem 3 ou 4 conectores, facilitando a instalação de exaustores adicionais.

*Para remover o cooler, basta girar as presilhas no sentido anti-horário, destravando o mecanismo. É mais fácil fazer isso usando uma chave de fenda:*

Um problema que temos no Brasil é o uso dos famigerados (para não usar um adjetivo pior) adesivos de garantia, usados por muitos distribuidores. Antigamente, eles costumavam ser colados na parte inferior do processador, mas com o lançamento dos processadores soquete 939, AM2 e LGA 775, onde não existe espaço na parte inferior, muitos distribuidores e lojas passaram a colar adesivos sobre o spreader do processador, o que prejudica brutalmente o contato entre o processador e o cooler, causando problemas de superaquecimento.

*Como você pode ver na foto, os adesivos formam uma "cratera" de área sem contato com o cooler em torno deles. Para amenizar o problema, você acaba tendo que usar mais pasta térmica, o que também é ruim, já que para ser eficiente, a camada de pasta térmica deve ser o mais fina possível. Por serem feitos de material plástico, os próprios adesivos não conduzem bem o calor, agravando ainda mais o problema:*

Na maioria dos casos, fornecedores com conhecimento de causa e preocupados com a qualidade não fazem esse tipo de coisa, até por que, é perfeitamente possível controlar as trocas dos processadores utilizando a numeração usada tanto pela Intel, quanto pela AMD. Em casos onde o fornecedor for irreduzível com relação ao uso dos adesivos, recomendo que procure outro.



*Os famigerados adesivos de garantia, que prejudicam o contato com o cooler*



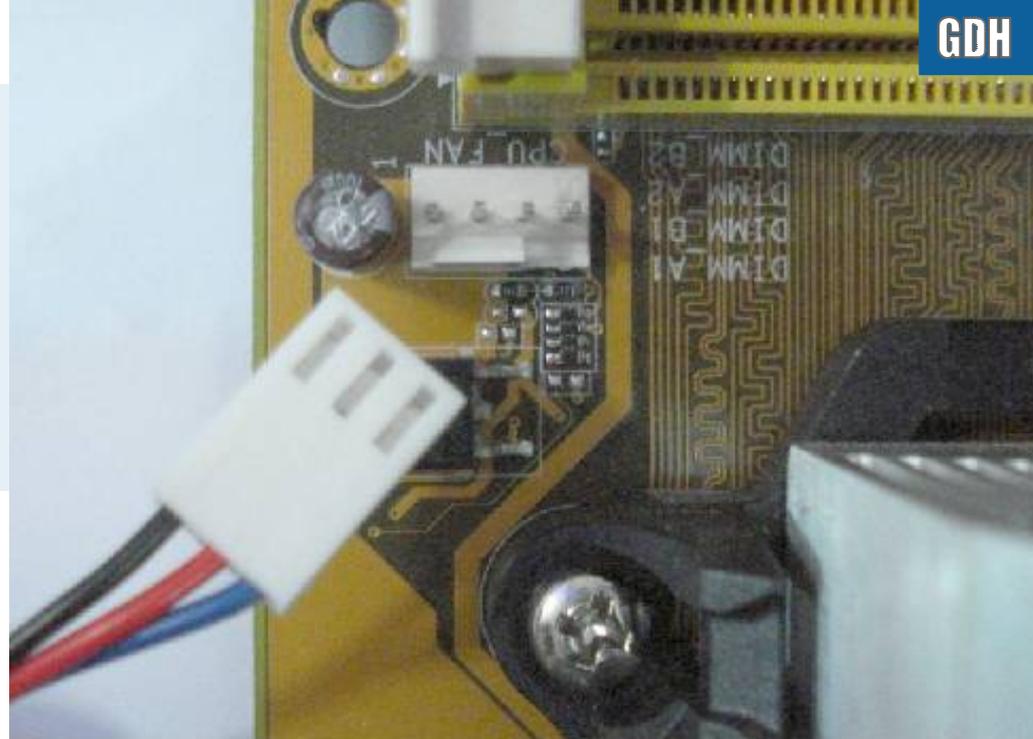
Com relação à alimentação, existem dois tipos de conectores para o cooler. Além do conector tradicional, com 3 pinos, existe o conector PWM, que possui 4 pinos. Ele foi introduzido pela Intel em 2004 e é usado na maioria das placas atuais (tanto para processadores Intel quanto AMD). O conector de 4 pinos é perfeitamente compatível com coolers que utilizam o conector antigo de 3 e você também pode conectar coolers que utilizam o conector de 4 pinos em placas com o conector de 3 pinos sem risco. A guia presente em um dos lados do conector impede que você encaixe o conector invertido ou ocupando os pinos errados, por isso não existe o que errar:

No conector de 3 pinos, dois deles são responsáveis pela alimentação elétrica (+12V e GND), enquanto o terceiro é usado pela placa mãe para monitorar a velocidade de rotação do cooler (speed sensor). O quarto pino permite que o BIOS da placa mãe controle a velocidade de rotação do cooler (PWM pulse), baseado na temperatura do processador. Com isso o cooler não precisa ficar o tempo todo girando na rotação máxima, o que além de reduzir o nível de ruído do micro, ajuda a economizar energia.

Ao conectar um cooler com o conector de 4 pinos em uma placa com o conector de 3, você perde o ajuste da rotação, de forma que o cooler simplesmente passa a girar continuamente na velocidade máxima, mas com exceção disso não existe problema algum.

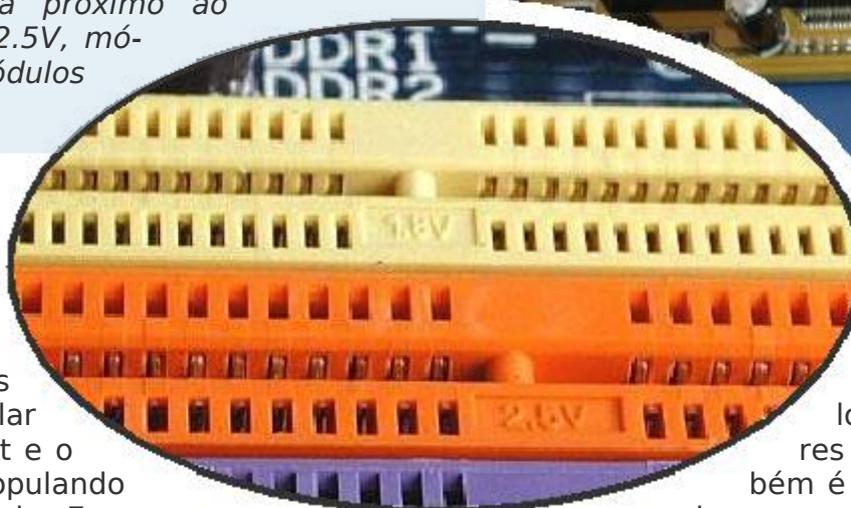
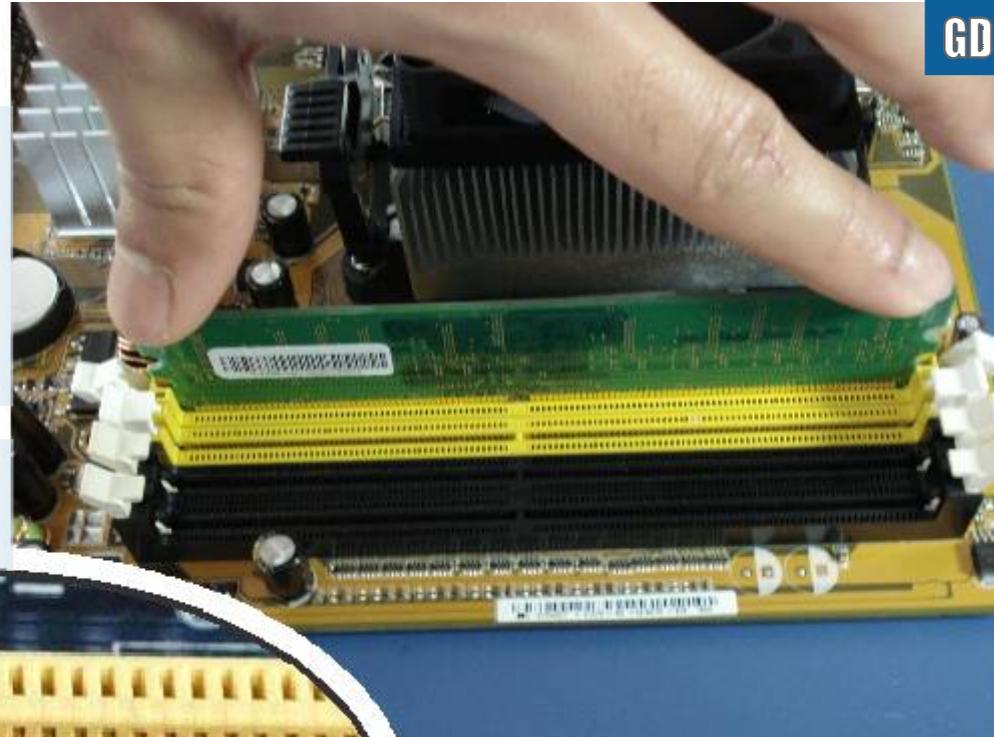
Além do cooler principal, temos a questão dos exaustores extra, que são um item cada vez mais importante nos PCs atuais. Alguns exaustores ainda utilizam conectores molex, como os utilizados pelo HD, mas a grande maioria dos de fabricação recente podem ser ligados aos conectores oferecidos pela placa mãe. A vantagem de utilizá-los é que a placa mãe pode monitorar as velocidades de rotação dos exaustores, permitindo que você os monitore via software.

*Esta placa da foto, por exemplo, possui 4 conectores, sendo que dois foram posicionados próximos às portas SATA:*



*Continuando, você pode aproveitar também para instalar os módulos de memória com a placa ainda fora do gabinete. O chanfro do conector impede que você encaixe um módulo DDR2 (ou DDR3) em uma placa que suporte apenas módulos DDR ou vice-versa, de forma que a principal dica é segurar sempre os módulos pelas bordas, evitando assim qualquer possibilidade de danificá-los com estática:*

*Além da posição do chanfro, outra forma de verificar rapidamente qual o tipo de memória utilizado pela placa, é verificar a tensão, decalcada próximo ao chanfro. Módulos DDR utilizam 2.5V, módulos DDR2 utilizam 1.8V e módulos DDR3 utilizam 1.5V:*



Em placas com 4 slots de memória, o primeiro e o terceiro slots formam o canal A, enquanto o segundo e o quarto formam o canal B. Para usar dois módulos em dual-channel, você deve instalar o primeiro módulo no primeiro slot e o segundo módulo no segundo, populando simultaneamente ambos os canais. Em caso de dúvidas sobre a instalação em alguma placa específica, você pode confirmar a posição correta na seção "Memory" ou "System Memory" do manual.

Outra observação é que não é obrigatório usar dois módulos em placas dual-channel. O uso de dois módulos é desejável do ponto de vista do desempenho, mas a placa funciona perfeitamente com apenas um.

As excessões ficam por conta das antigas placas para Pentium 4 que utilizavam módulos de memórias Rambus. Nelas era realmente obrigatório instalar módulos RIMM em pares e usar terminadores nos soquetes não utilizados. Também é preciso usar módulos em pares em placas soquete 7 antigas, que utilizam módulos de 72 vias.

Depois de tudo isso, podemos finalmente instalar a placa dentro do gabinete, prendendo-a nos suportes usando parafusos. Na verdade, você pode instalar a placa logo no início da montagem, e encaixar o processador, cooler, memória e os conectores do painel frontal com ela já dentro do gabinete. A questão é que é bem mais fácil instalar estes componentes com a placa "livre" sobre a bancada do que dentro do espaço apertado no gabinete.

# Especial Montagem de micros

Uma chave magnética ajuda bastante na hora de posicionar os parafusos. Lembre-se que você pode transformar qualquer chave de fenda em uma chave magnética usando um pequeno ímã de neodímio, como os encontrados dentro do mecanismo que movimenta a cabeça de leitura do HD. Cuide apenas para não largá-los sobre mídias magnéticas, como disquetes ou o próprio HD.

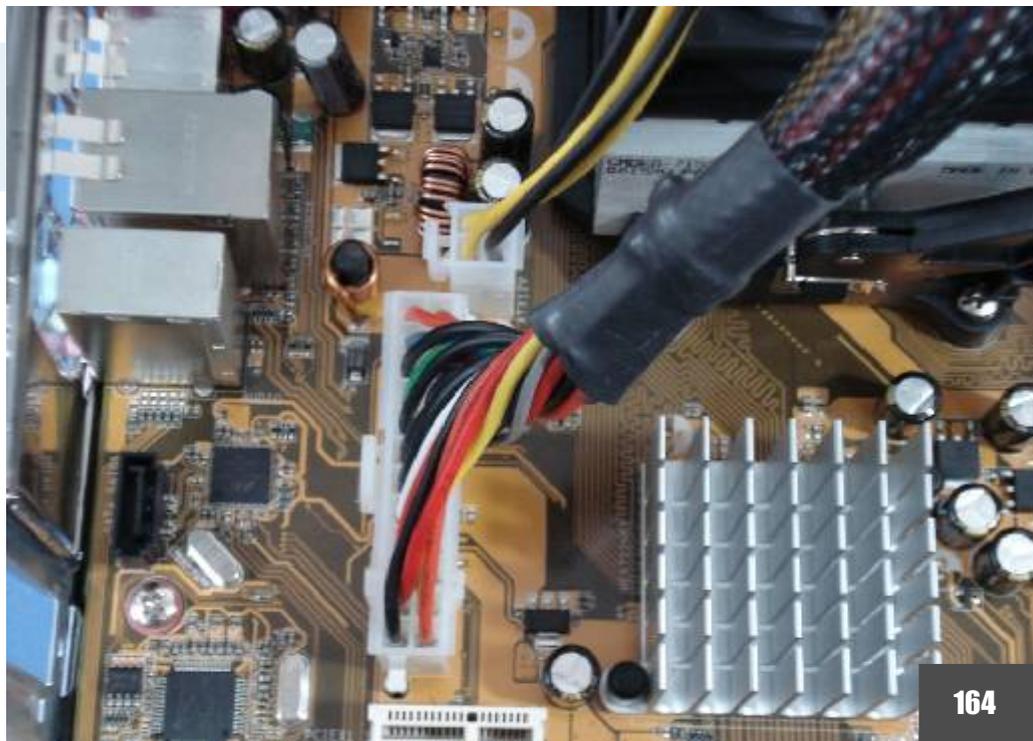
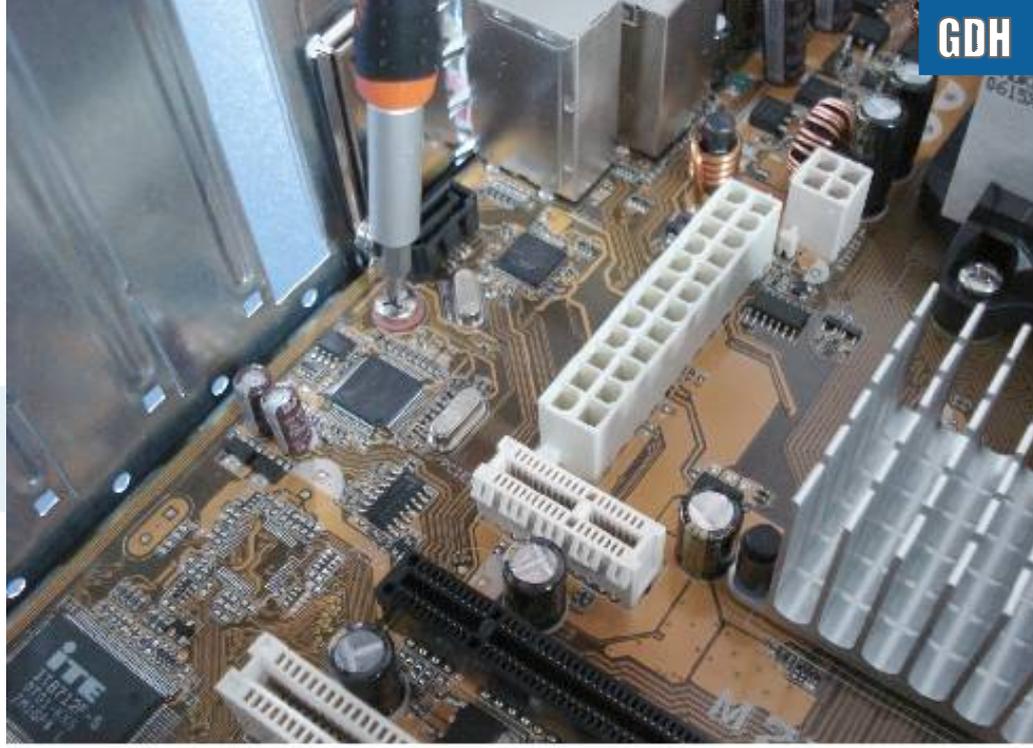
*Não se esqueça também de encaixar a tampa do painel ATX que acompanha a placa antes de instalá-la:*

O próximo passo é ligar os conectores de força na placa mãe. Praticamente todas as placas atuais utilizam tanto o conector ATX de 24 pinos e o conector P4, de 4 pinos, que fornece anergia adicional, reforçando o fornecimento elétrico para o processador e também para o slot PCI Express x16.

*Ao montar qualquer PC atual, você deve utilizar uma fonte de pelo menos 450 watts, que ofereça ambos os conectores:*

Lembre-se de que 90% das fontes vendidas no Brasil são produtos de baixa qualidade. Mesmo que a etiqueta diga que a fonte é capaz de fornecer 450 watts, é bem provável que ela na verdade ofereça apenas 350 watts ou menos, por isso é importante manter uma boa margem de segurança.

Voltamos então à velha pergunta: o que fazer com fontes antigas, que oferecem apenas 300 ou 350 watts e ainda utilizam o conector ATX de 20 pinos? A resposta curta é que você não deve usá-las ao montar um PC novo, pois não vale à pena arriscar a saúde dos demais componentes para economizar os 50 ou 70 reais de uma fonte nova.



A resposta longa é que a maioria das placas funciona usando um conector ATX de 20 pinos, desde que o conector P4 auxiliar esteja conectado. Entretanto, isto reduz o fornecimento elétrico da placa mãe, o que pode causar problemas ao utilizar processadores e/ou placas 3D com um consumo elétrico mais elevado.

*Algumas placas possuem um conector molex ao lado do conector P4 auxiliar. Esta combinação era comum por volta de 2001 a 2002, quando as fontes com o conector extra ainda não eram comuns. Neste caso, você pode escolher qual dos dois usar:*

O próximo passo é instalar os drives. Alguns gabinetes são espaçosos o suficiente para que você instale os HDs antes mesmo de prender a placa mãe, mas na maioria dos casos eles ficam parcialmente sobre a placa, de forma que você precisa deixar para instalá-los depois.

Ao usar drives IDE, você precisa se preocupar também com a configuração de master/slave. No caso do drive óptico (vou adotar este termo daqui em diante, já que você pode usar tanto um drive de CD quanto de DVD), o jumper está disponível bem ao lado do conector IDE. Colocá-lo na posição central configura o drive como slave, enquanto colocá-lo à direita configura o drive como master. Para o HD, a configuração do jumper varia de acordo com o fabricante, mas você encontra o esquema de configuração na etiqueta de informação do drive. Quase sempre, o HD vem configurado de fábrica como master e ao retirar o jumper ele é configurado como slave.

HDs SATA não utilizam jumpers de configuração de master/slave, pois cada porta permite a instalação de um único HD.



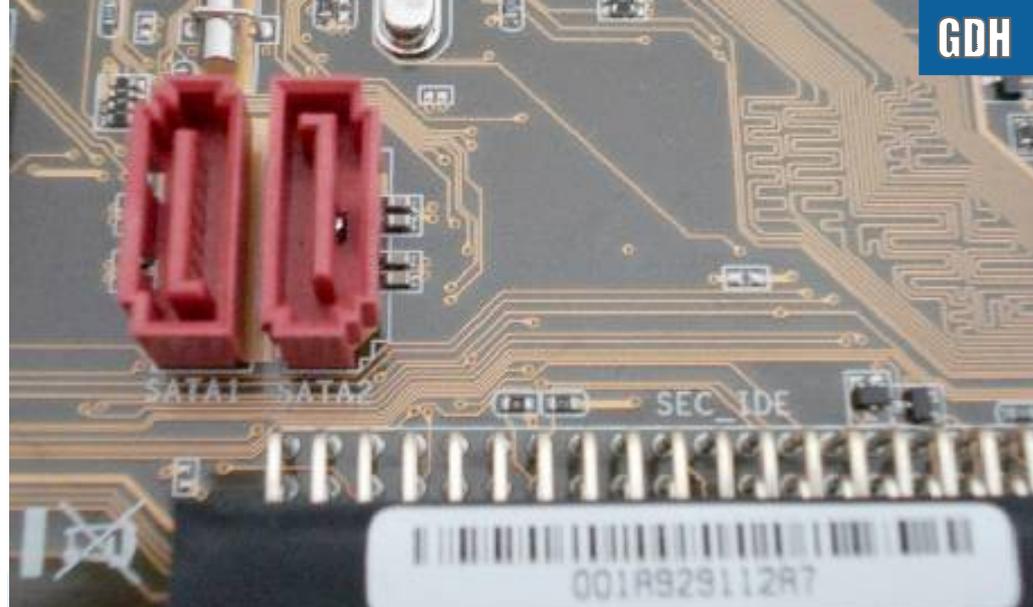
Jumpers em um HD IDE, HD SATA e drive de DVD IDE

Apesar disso, a maioria dos drives incluem um jumper que permite forçar o HD a operar em modo SATA/150 (evitando problemas de compatibilidade com algumas placas antigas). Em muitos HDs (como na maioria dos modelos da Seagate) ele vem ativado por padrão, fazendo com que o drive opere em modo SATA/150 por default. Ao usar uma placa equipada com portas SATA/300, não se esqueça de verificar a posição do jumper, para que a taxa de transferência da interface não seja artificialmente limitada.

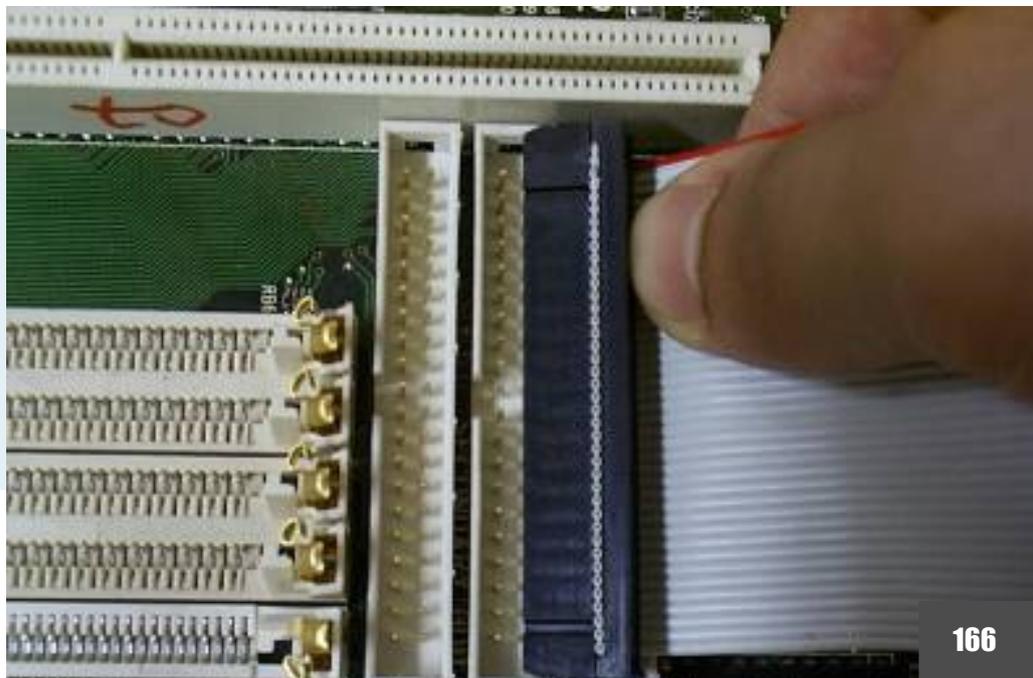
Ao instalar o HD e o drive óptico em portas separadas, você pode configurar ambos como master. Atualmente é cada vez mais comum que placas novas venham com apenas uma porta IDE, o que o obriga a instalar um como master e o outro como slave. É comum também que o drive óptico seja instalado como slave mesmo ao ficar sozinho na segunda porta, já deixando o caminho pronto para instalar um segundo HD como master futuramente.

Ao usar dois (ou mais) HDs SATA, é importante que o HD de boot, onde você pretende instalar o sistema operacional, seja instalado na porta SATA 1. É possível mudar a configuração de boot através do setup, dando boot através dos outros HDs, mas o default é que o primeiro seja usado.

*Nas placas e cabos atuais, é usada uma guia e um pino de controle, que impedem que você inverta a posição da cabos IDE. Em placas e cabos antigos era comum que estas proteções não estejam presentes. Nestes casos, procure um número "1" decalcado em um dos lados do conector. A posição do "1" deve coincidir com a tarja vermelha no cabo e, do lado do drive, a tarja vermelha fica sempre virada na direção do conector de força:*



*A identificação de cada porta vem decalcada sobre a própria placa mãe. Na foto temos "SATA1" e "SATA2" indicando as duas portas SATA e "SEC\_IDE", indicando a porta IDE secundária. Ao lado dela estaria a "PRI\_IDE", a porta primária:*



Os cabos IDE possuem três conectores. Normalmente dois estão próximos e o terceiro mais afastado. O conector mais distante é o que deve ser ligado na placa mãe, enquanto os dois mais próximos são destinados a serem encaixados nos drives. Ao instalar apenas um drive no cabo, você deve usar sempre as duas pontas do conector, deixando o conector do meio vago (nunca o contrário).

Você deve utilizar sempre cabos de 80 vias em conjunto com os HDs IDE atuais, pois eles oferecem suporte aos modos ATA-66, ATA-100 e ATA-133. Os drives ópticos podem utilizar cabos comuns, de 40 vias, pois eles trabalham sempre em modo ATA-33.

Você deve receber os cabos IDE e SATA juntamente com a placa mãe. Normalmente o pacote inclui também o cabo do disquete (embora hoje em dia seja cada vez mais raro usá-lo) e também um adaptador para converter um conector molex da fonte no conector de força SATA. A maioria das fontes oferece apenas um único conector de força SATA, de forma que você acaba precisando do adaptador ao instalar um segundo HD. Em placas que não possuem portas IDE, o cabo é substituído por um segundo cabo SATA.

O drive óptico acompanha um segundo cabo IDE (quase sempre um cabo de 40 vias), permitindo que, ao usar um drive óptico e HD IDE, você os instale em portas separadas.



"Kit" com cabos e manuais que acompanha a placa mãe

*Aqui temos os cabos IDE e SATA instalados. O cabo IDE preto está instalado na IDE primária e vai ser usado pelo HD, enquanto o cinza, instalado na IDE secundária, vai ser usado pelo drive óptico:*



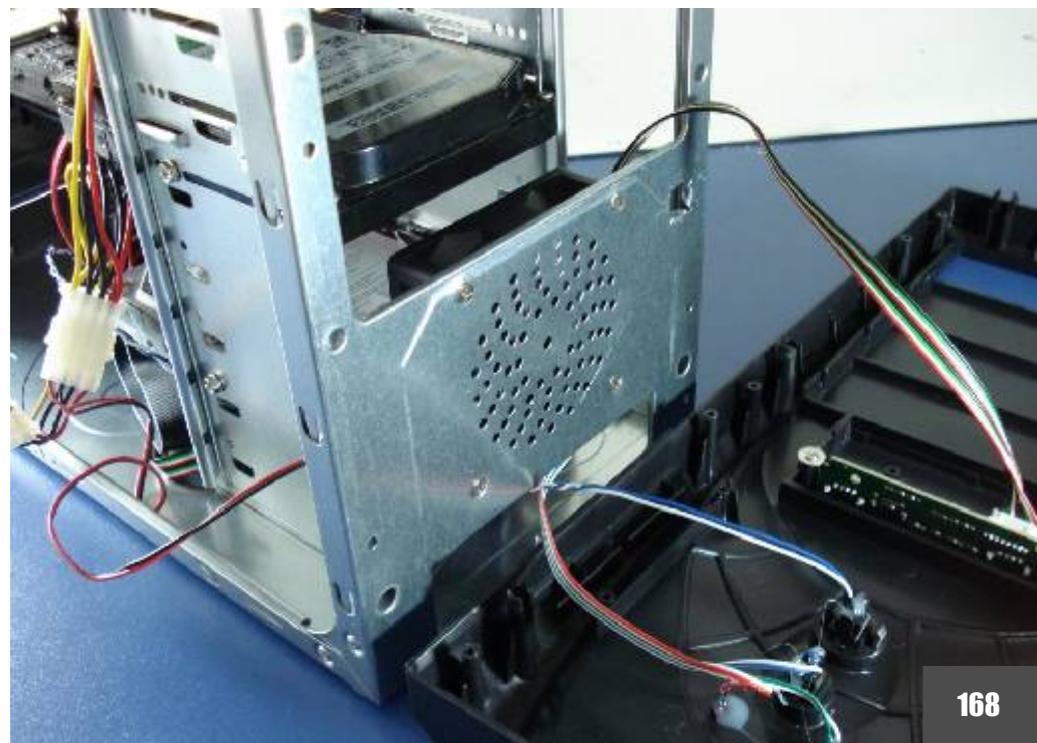
*Ao instalar dois ou mais HDs na mesma máquina, deixe sempre que possível um espaço de uma ou duas baias entre eles, o que ajuda bastante na questão da refrigeração:*

Assim como em outros componentes, a temperatura de funcionamento dos HDs tem um impacto direto sob a sua vida útil. O ideal é que a temperatura de operação do HD não ultrapasse os 45 graus (você pode monitorá-la usando o programa de monitoramento incluído no CD de drivers da placa, ou usando o lm-sensors no Linux), mas, quanto mais baixa a temperatura de funcionamento, melhor.

Caso tenha alguns trocados disponíveis, uma medida saudável é instalar um exaustor na entrada frontal do gabinete, puxando o ar para dentro. O fluxo de ar vai não apenas reduzir a temperatura de operação dos HDs (muitas vezes em 10 graus, ou mais) mas também dos demais componentes do micro, incluindo o processador. Para melhores resultados, o exaustor frontal deve ser combinado com outro na parte traseira, na abertura ao lado do processador, desta vez soprando o ar para fora.

Para instalar o exaustor frontal, você precisa remover a frente do gabinete. Em muitos dos modelos atuais, ela é apenas encaixada, de forma que basta puxar com cuidado. Em outros ela é presa com parafusos, escondidos nas laterais.

É sempre chato ficar colocando parafusos dos dois lados, tanto para os HDs, quanto para o drive óptico, mas é importante que você resista à tentação de instalar os drives "nas coxas", sem usar todos os parafusos. A questão fundamental aqui é a vibração. Colocando parafusos apenas de um lado, ou colocando apenas um de cada lado, a movimentação da cabeça de leitura dos HDs e do drive óptico vão fazer com que o drive vibre dentro da baia,

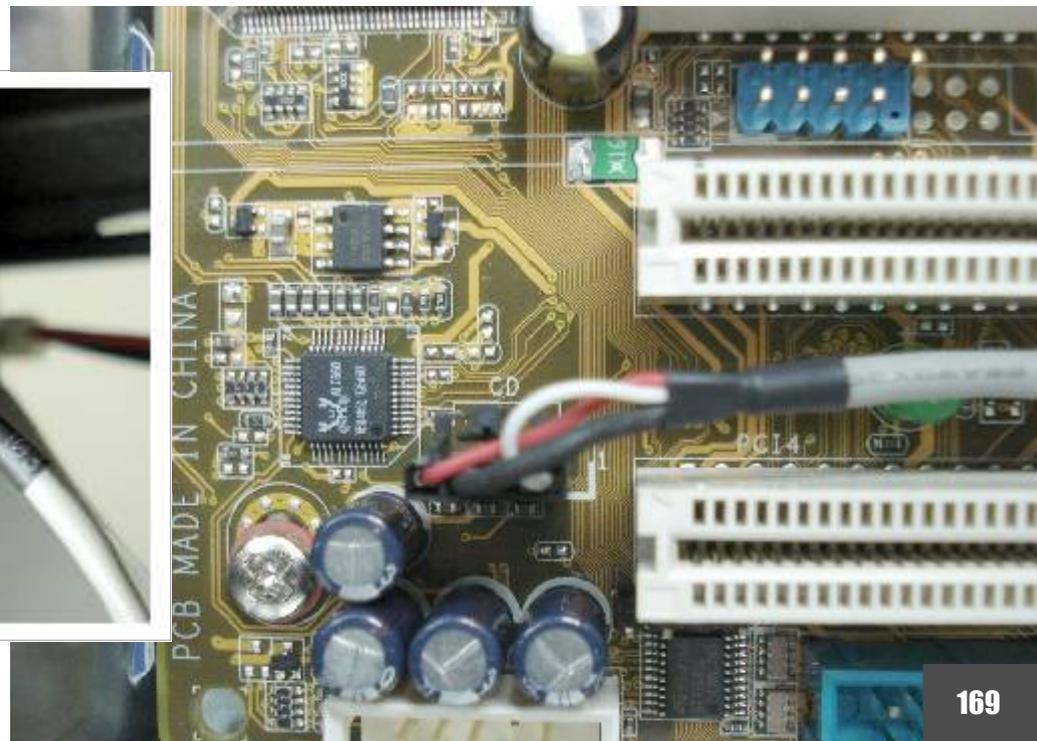
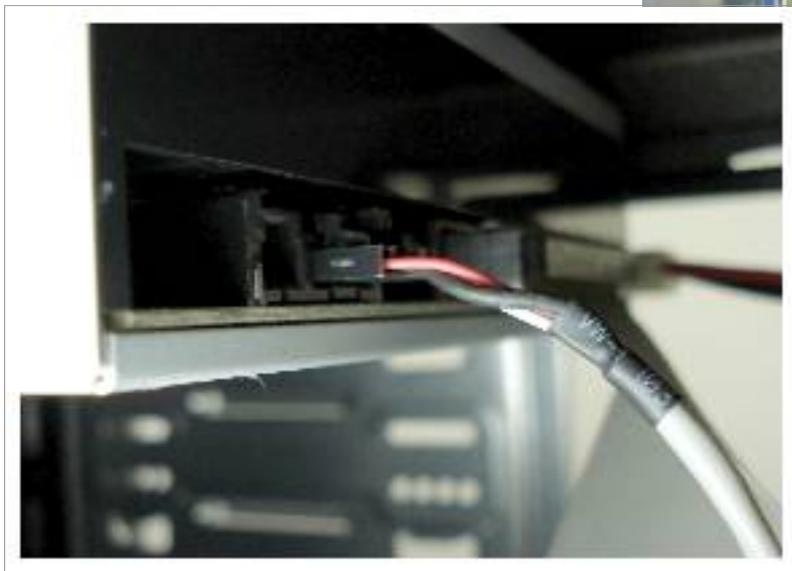


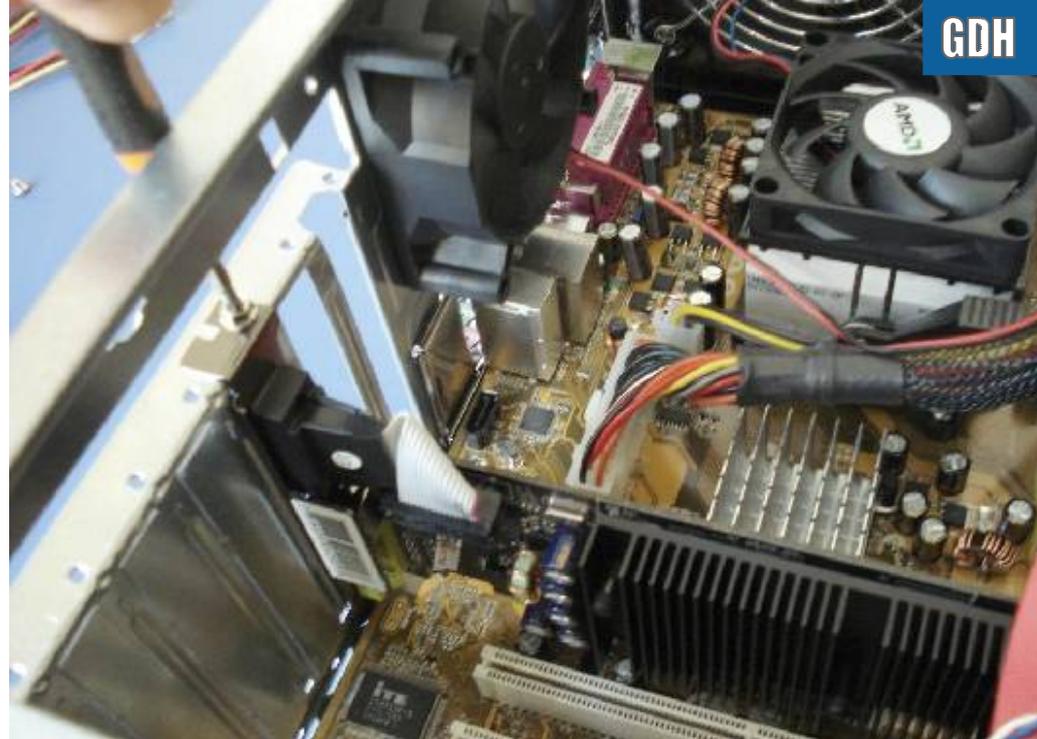
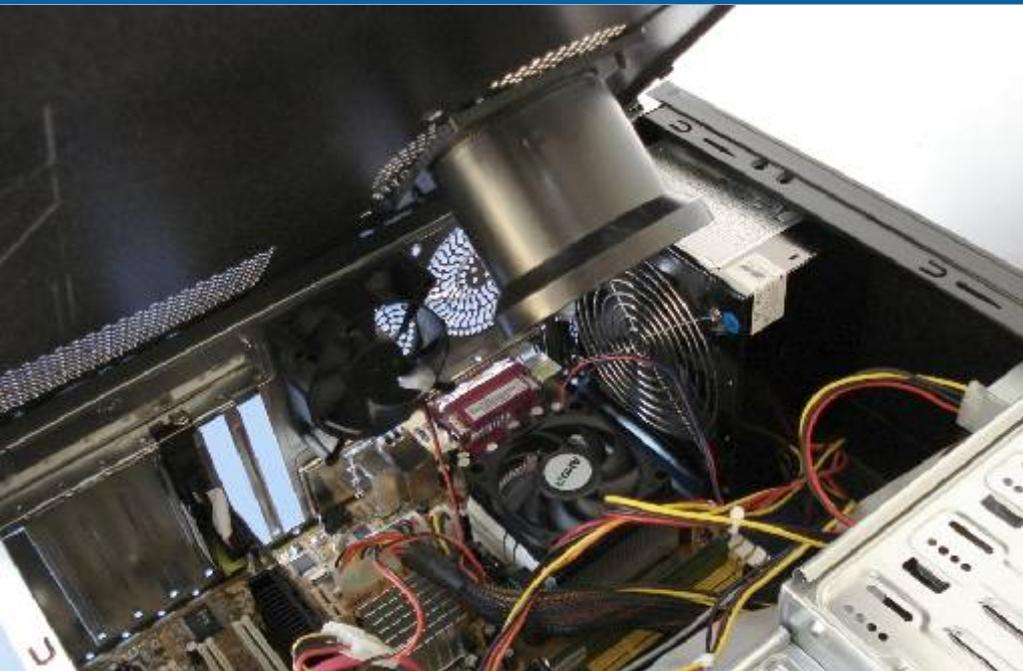
aumentando o nível de ruído do micro, sem falar de possíveis problemas relacionados ao desempenho ou mesmo à vida útil dos drives.

O toque final é instalar o cabo de áudio do drive de CD, usado para tocar CDs de áudio. Hoje em dia ele não é mais tão usado, pois a maioria dos programas é capaz de reproduzir CDs obtendo as faixas digitalmente, a partir do próprio cabo de dados do drive (o mesmo processo usado para ripar CDs), mas é sempre bom ter o cabo instalado, já que você nunca sabe que programas o dono do micro vai utilizar. O cabo é fornecido junto com o drive e é encaixado na entrada "CD" da placa mãe, um conector de 4 pinos.

Como disse a pouco, é importante instalar um exaustor na abertura traseira do micro, soprando o ar para fora. O exaustor dentro da fonte de alimentação também faz este trabalho, mas a principal função dele é resfriar a própria fonte. O exaustor traseiro age mais diretamente, empurrando pra fora rapidamente o ar quente que já passou pelo cooler do processador.

A maioria dos gabinetes atuais inclui um tubo (chamado de "túnel de vento" pelos fabricantes) que vai sobre o processador. O tubo canaliza o ar externo, fazendo com que o cooler do processador utilize o ar frio vindo de fora, ao invés de ficar simplesmente circulando o ar quente dentro do gabinete.

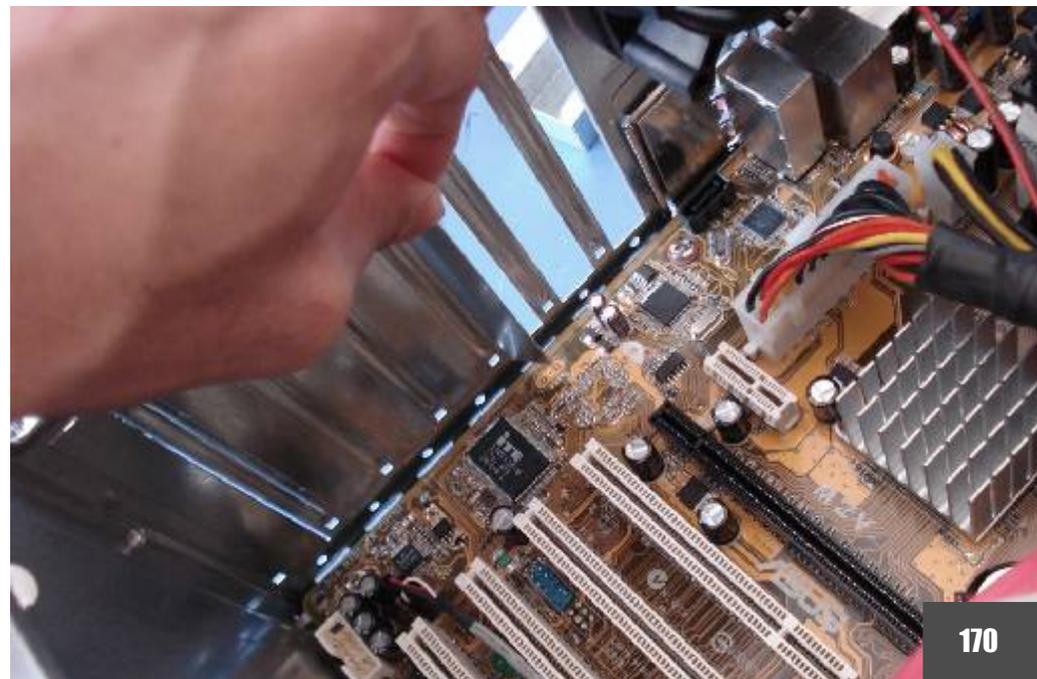


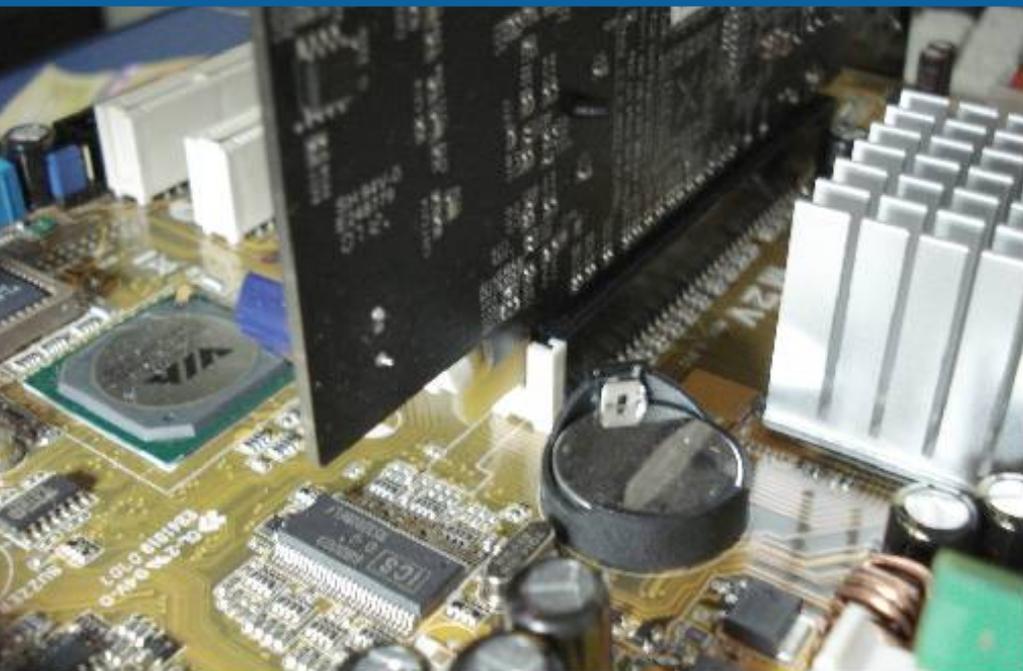


Concluindo, falta apenas instalar a placa de vídeo e outras placas de expansão (como uma segunda placa de rede, modem ou uma placa de captura) e a montagem está completa.

Alguns poucos gabinetes utilizam protetores independentes para as aberturas dos slots, mas na maioria é usada uma simples chapa cortada, onde você precisa remover as tampas dos slots que serão usados. Algumas sempre esbarram em capacitores da placa mãe, por isso precisam ser removidas com mais cuidado. O aço cortado é praticamente uma lâmina, é bem fácil se cortar.

Tanto os slots PCI Express x16, quanto os slots AGP, utilizam um sistema de retenção para tornar o encaixe da placa de vídeo mais firme. Ao remover a placa, não se esqueça de puxar o pino do lado direito do slot, senão você acaba quebrando-o.





Toda placa mãe inclui pelo menos um jumper, o jumper responsável por limpar o CMOS (CLR\_CMOS ou CLRRTC). Em muitas placas, ele vem de fábrica na posição discharge (com o jumper entre os pinos 2 e 3), para evitar que a bateria seja consumida enquanto a placa fica em estoque. A maioria das placas não dão boot enquanto o jumper estiver nesta posição, o que pode ser confundido com defeitos na placa.

Antes de ligar o micro, certifique-se que o jumper está na posição 1-2 (indicada no manual como "Normal" ou "Default").

**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.

## Já visitou o Guiadohardware.NET hoje?

acesse:

<http://guiadohardware.net>



# *Especial* NOTEBOOKS

por Carlos E. Morimoto



**Antigamente, ter um notebook era um luxo reservado apenas aos que realmente precisavam de portabilidade e podiam gastar o triplo ou o quádruplo do valor que pagariam por um desktop de configuração equivalente.**

**Felizmente, este tempo já passou e hoje em dia os notebooks mais populares custam apenas um pouco mais do que um desktop equivalente, com monitor de LCD e nobreak. Em alguns casos, promoções e condições de parcelamento podem fazer com que o note chegue até a custar mais barato.**

Outra área em que houve avanço foi a questão do desempenho. Antigamente, notebook era sinônimo de desempenho inferior. Os modelos antigos utilizavam HDs muito lentos, processadores de baixo clock, menos memória e antiquados monitores LCD de matiz passiva, que arruinavam o desempenho e tornavam o equipamento praticamente inutilizável para jogos e multimídia.

Embora os notebooks atuais ainda continuem perdendo em certas áreas, como no caso do desempenho do HD e da placa de vídeo, na maioria dos demais quesitos as coisas já estão equilibradas. Você pode comprar um notebook com 2 GB ou mais de RAM, com um processador dual core, com gravador de DVD, com uma placa 3D razoável ou até mesmo com uma tela de 17", depende apenas de quanto você está disposto a gastar.

Os notebooks também ficam atrás na questão do upgrade, já que (com exceção de modelos específicos) você não tem como instalar mais de um HD ou espetar mais do que dois pentes de memória. Atualizar o processador também é complicado, pois usar um modelo de maior clock (e maior dissipação térmica) exigiria também a substituição do cooler, o que é raramente possível num notebook. Em geral, você fica restrito a adicionar mais memória ou substituir o HD por um de maior capacidade.

A partir daí a única forma de upgrade acaba sendo usar periféricos externos, ligados às portas USB ou firewire.

Apesar disso, a portabilidade e o "cool factor" dos notebooks acabam superando suas desvantagens e fazendo com que cada vez mais gente acabe optando por um. Segundo o IDC, as vendas de notebooks devem superar as de desktops (em número de unidades) em 2011, uma tendência que deve ser percebida também aqui no Brasil.

Com a crescente redução na diferença de preço, não é difícil de imaginar que no futuro os notebooks se tornem padrão, com os desktops cada vez mais restritos a nichos específicos, como no caso dos gamers mais inveterados e nas estações de trabalho.

Muitos acham que a popularização dos notebooks vai reduzir o campo de trabalho para os técnicos de manutenção, mas eu vejo justamente o cenário oposto. Notebooks precisam de tanta manutenção quanto os desktops (ou até mais, já que acidentes e quedas são comuns), o que vai na verdade aumentar a oferta de trabalho. A questão fundamental é que serão exigidos profissionais com mais conhecimento técnico, que sejam capazes não apenas de identificar os defeitos e substituir as peças necessárias, mas também de obter as peças de reposição a um preço aceitável.

Se um técnico pode solucionar um problema na tela trocando todo o LCD e um segundo pode corrigir o mesmo problema trocando apenas o FL Inverter, sem dúvida o segundo teria como cobrar um preço muito mais competitivo pelo conserto.

Agora vou começar com uma explicação teórica sobre as diferenças entre os componentes dos desktops e notebooks, incluindo os processadores, chipsets, aceleradoras 3D e baterias.



## Categorias

Como tudo na informática, os portáteis podem ser divididos em categorias, que definem as combinações de tamanho e recursos mais usadas pelos fabricantes.

Antigamente, era comum que os portáteis fossem classificados em três categorias: laptops, notebooks e subnotebooks. Os laptops eram os modelos maiores, basicamente qualquer computador portátil o suficiente para que você pudesse colocá-lo no colo ("laptop" significa, literalmente, "no colo" ou

"sobre o colo") e usá-lo com relativo conforto. O notebook seria um aparelho menor, aproximadamente do tamanho de um caderno universitário (os IBM Thinkpad antigos são um bom exemplo), enquanto os subnotebooks eram os portáteis ainda mais compactos, que frequentemente sacrificavam o drive óptico e utilizavam processadores de baixo consumo para atingir o objetivo.

Um exemplo de subnotebook é o Toshiba Libretto, que foi relativamente popular durante a década de 1990. A configuração era fraca, mesmo se comparado com outros portáteis da época, mas em compensação ele era pouco maior que uma fita VHS e pesava apenas 850 gramas. O modelo mais rápido da safra inicial foi o Libretto 70, lançado em 1997. Ele era baseado em uma versão de baixo consumo do Pentium MMX, que operava a 120 MHz e suportava o uso de até 32 MB de RAM. Uma das maiores limitações era a tela, de 640x480:



*Toshiba Libretto 70CT*

A partir de um certo ponto, entretanto, cada vez mais fabricantes passaram a chamar seus portáteis de "notebooks", independentemente do tamanho. Com isso a designação tradicional deixou de fazer sentido, de forma que hoje em dia podemos dizer que os termos "laptop" e "notebook" tornaram-se sinônimos.

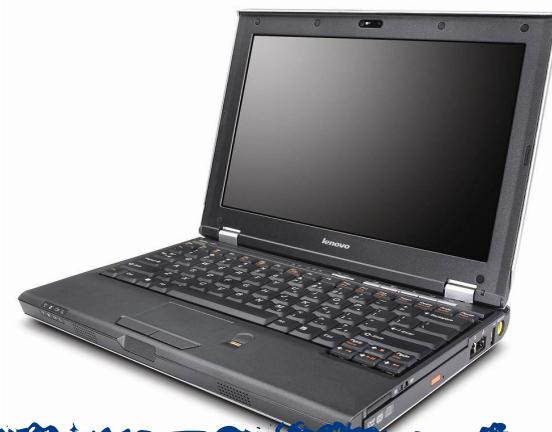
No lugar da classificação tradicional, os fabricantes passaram a usar os termos "Desktop replacement" (substituto para o desktop), "Thin-and-light" (leve e fino) e "Ultraportable" (ultraportátil).

Os desktop replacement são notebooks maiores, com tela de 15" ou mais (alguns chegam a usar telas de 21"!), desenvolvidos com o objetivo de serem confortáveis de usar sobre uma mesa e serem relativamente poderosos, sem tanta preocupação com a portabilidade ou com a autonomia das baterias.

Os thin-and-light já são modelos menores, com telas de 14" ou menos, desenvolvidos com o objetivo de serem mais portáteis. Em geral, os thin-and-light preservam um teclado de dimensões normais e continuam sendo confortáveis de usar, apesar da tela um pouco menor. O menor espaço interno limita um pouco a seleção de processadores, mas isto não chega a ser ruim, pois deixa de fora processadores muito gastadores como os mobile Pentium 4 e mobile Athlon 64.

Finalmente, temos os ultraportáteis, modelos com tela de 12" ou menos, que pesam menos de 1.7 kg. Para atingir esta marca, eles utilizam processadores de baixo consumo (e, conseqüentemente, de baixa frequência), teclados de dimensões reduzidas, drives ópticos miniaturizados (mais caros e difíceis de substituir em caso de defeito) ou drives externos e, em muitos casos, HDs de 1.8" ou drives de estado sólido, de memória Flash.

A questão do peso não é necessariamente uma regra. Por exemplo, a Lenovo classifica o A100, como um ultraportátil por causa da tela de 12", muito embora ele pese exatos 2 kg e tenha 3 centímetros de espessura, mais do que a maioria dos ultraportáteis, que ficam entre os 2 e 2.5 cm. Ele fica no limite entre o que seria considerado um thin-and-light e um ultraportátil:



*Lenovo A100*

Na foto a seguir temos uma comparação entre um Acer 5043WLMi e um Sony Vaio GN-TX670P. O 5043WLMi está no limite entre a classificação de desktop replacement e thin-and-light (ele possui uma tela de 15.4", mas é relativamente leve, pesando 2.85 kg). O GN-TX670P, por sua vez, é indiscutivelmente um ultraportátil, com tela de 11.1" e apenas 1.26 kg. Ele utiliza um processador Pentium M ULV de 1.2 GHz e um HD de 1.8", da Fujitsu:



Acer 5043WLI e Sony Vaio GN-TX670P

Muitos fabricantes ainda usam o termo "subnotebook" para aparelhos ainda menores, com menos de 1.2 kg, mas atualmente a designação está caindo em desuso, cedendo espaço para os termos "UMPC" e "MID" (veja a seguir) que são as novas categorias de pesos leves.

Outro termo conhecido é o "desknote", que tem duplo sentido. Atualmente, o uso mais comum é em relação aos notebooks maiores, no lugar do termo "desktop replacement". Nesta conotação, um desknote é um notebook grande, pouco portátil, feito para ser usado sobre a mesa.

Outro uso é em relação aos "desknotes" fabricados pela PC-Chips/ECS entre 2001 e 2005, uma classe de portáteis de baixo custo, sem bateria, que aproveitavam componentes usados em micros de mesa. A idéia era até boa, afinal, removendo a bateria, utilizando processadores de baixo custo e aproveitando projetos de placas-mãe para micros desktop, seria possível produzir notebooks relativamente baratos. Na prática, entretanto, acabou não dando muito certo, pois a pequena redução de preço acabava não sendo suficiente para compensar a perda de portabilidade. Os desknotes eram famosos também pela baixa qualidade e pelo uso de processadores de baixo desempenho, como os C3 e os Crusoe, vendidos sob a marca "GigaPro". Existiram também modelos baseados no Celeron P6 e no Athlon (as séries A901, A927 e A929), mas eles esquentavam bastante, resultando em problemas de estabilidade.

Outra categoria é a dos tablet-PCs, onde o uso de uma tela touch-screen permite que você use o notebook como

uma espécie de bloco de anotações, navegando entre as funções usando uma stylus e utilizando um teclado onscreen ou um sistema de reconhecimento de escrita para a entrada de informações.

A maioria dos modelos atuais são "conversíveis", ou seja, notebooks normais, onde você pode girar a tela touch-screen, fazendo com que ela se feche sobre o teclado. Desta forma, é possível usá-lo tanto como um notebook normal, como um tablet, de acordo com a situação:



Somados aos recursos de reconhecimento de escrita e anotações, a solução acaba se tornando relativamente prática. O problema é que o baixo volume de produção faz com que os tablets tornem-se mais caros que um notebook "normal", de configuração equivalente, o que acaba fazendo com que eles fiquem restritos a nichos muito específicos.

Mais uma observação é que nas especificações de notebooks é comum que o

peso seja informado em libras (pounds) ao invés de quilogramas, já que este é o padrão mais usado nos EUA. Neste caso é só fazer as contas. Uma libra internacional equivale a 454 gramas (453.59237g se quiser o número exato), de forma que um "6 pounds notebook" é um notebook que pesa aproximadamente 2.7 kg.

## Barebones

Um barebone é um notebook personalizável, onde você compra apenas a carcaça, com a placa-mãe, tela, bateria, teclado e touchpad. Desta forma, você pode personalizar o equipamento, escolhendo o processador, módulos de memória, HD, drive óptico e, em alguns modelos, também a placa de vídeo (neste caso escolhendo entre um pequeno conjunto de modelos mobile).

Dois exemplos de barebones são o Asus Z84JP e o MSI MS-1058. À primeira vista eles parecem notebooks normais, e realmente são. Um barebone nada mais é do que um notebook desenvolvido com o objetivo de permitir uma certa flexibilidade na escolha dos componentes, que é vendido "incompleto", permitindo que você escolha o que usar. Não é incomum que o fabricante ofereça também opções de notebooks "completos" baseados nos mesmos.



*Asus Z84JP e MSI MS-1058*

O Asus Z84JP é um desktop replacement, grande e equipado com uma tela de 17", enquanto o MSI MS-1058 é um thin-and-light, compacto e com tela de 12". Escolhi os dois como exemplo justamente devido às diferenças entre eles.

O Z84JP usa uma placa mãe baseada no chipset Intel 945PM. Ao contrário do 945GM, este chipset não tem vídeo onboard, por isso o notebook inclui um slot mobile PCI Express interno, onde você pode instalar uma placa de vídeo MXM, da série GeForce Go, o que permite montar um notebook com um desempenho 3D respeitável, usando uma GeForce Go 7900 GTX com 512MB, por exemplo.

O chipset suporta processadores Intel baseados no soquete M, incluindo os Core 2 Duo (Meron), Core Duo (Yonah) e Pentium M (Dothan) e a placa possui dois slots para módulos SODIMM de memória, de até 2 GB cada, o que permite instalar até 4 GB. O barebone já vem com um drive DVD-RW, leitor de cartões, som, rede e modem onboard, transmissor Bluetooth e uma webcam integrada na tela, mas (além do processador, memória e placa de vídeo), faltam o HD e a placa wireless.

Não existem limitações com relação à escolha do HD, você pode escolher qualquer modelo SATA de 2.5", incluindo os modelos de 7200 RPM. Apesar da placa-mãe ser baseada em um chipset

Intel, você pode utilizar uma placa wireless de qualquer fabricante. A única observação é que o barebone utiliza um slot Express Mini, o que deixa de fora as placas wireless no formato mini-PCI.

O MS-1058 é numa plataforma bem diferente. Em primeiro lugar, ele é baseado no chipset ATI Radeon Xpress 1100 (composto pelos RS485M e SB460), que inclui um chipset de vídeo onboard relativamente poderoso (pelo menos se comparado com os chipsets de vídeo integrado da Intel), mas em compensação não suporta o uso de uma placa MXM externa.

Ele suporta processadores AMD soquete S1, incluindo, naturalmente, todos os modelos do Turion X2 e oferece suporte a módulos SODIMM de até 1 GB, permitindo um máximo de 2 GB de memória instalada.

Assim como no caso do Asus Z84JP, o MSI MS-1058 inclui um drive DVD-RW, som, rede, modem, leitor de cartões, Bluetooth e um slot Express Mini para a conexão da placa wireless, mas, ao invés de uma interface SATA, ele ainda utiliza a boa e velha porta IDE para a conexão do HD, o que limita um pouco as escolhas, pois mesmo entre os HDs de 2.5", os modelos IDE são cada vez mais raros.

Em agosto de 2007, o Z84JP custava US \$ 973 e o MS-1058 custava US\$ 719.

Os preços podem parecer baixos a princípio, mas lembre-se de que estes são os preços nos EUA. Para comprá-los aqui no Brasil você teria que pagar o envio e os impostos de importação, o que dobraria o valor.

Somando o valor do barebone, com o valor individual dos componentes, você logo percebe que o custo excede bastante o da maioria dos notebooks de baixo e médio custo. Este é o primeiro segredo dos barebones: eles não são vantajosos quando você quer montar um notebook de baixo custo, mas sim em casos onde você quer uma configuração mais parruda ou incomum.

O principal motivo desta discrepância é que notebooks de baixo custo tem seu custo amortizado pelo grande volume de produção e são parcialmente subsidiados pelos modelos mais caros da mesma linha. Os barebones são quase sempre produzidos em pequena quantidade e por isso (depois de somados os custos individuais dos componentes) acabam saindo mais caro.

A situação muda um pouco, entretanto, quando você quer comprar um notebook high-end. Como os fabricantes trabalham com margens de lucro muito maiores nos modelos topo de linha (justamente o oposto do que temos nos modelos de baixo custo), comprar os componentes separadamente pode sair mais barato, além de que um barebone

lhe oferecerá melhores possibilidades de upgrade.

Por exemplo, imagine o caso de alguém que quer desesperadamente um notebook para games. Todos sabemos que as opções de vídeo integrado atendem apenas jogadores ocasionais, nenhum aficionado ficaria satisfeito jogando o F.E.A.R a 20 FPS, usando as configurações mínimas, em uma ATI X200M, por exemplo.

Presumindo que dinheiro não seja problema, esta seria uma situação onde um barebone poderia prestar bons serviços. Veja o caso do Clevo M590KE, por exemplo. Ele inclui uma tela de 20.1", suporta o uso de dois HDs SATA de 2.5" em RAID e suporta o uso de duas placas GeForce Go em SLI, com a possibilidade de usar duas GeForce Go 7950 GTX com 512 MB cada, ou (no caso de um notebook destinado a uso profissional) duas Quadro FX Go 2500M. Com relação ao processador, você pode utilizar um Turion X2 TL-66 (2.3 GHz), combinado com até 2 GB de memória. Ou seja, utilizando este barebone como base, você poderia montar um laptop com uma configuração superior à de muitos desktops.





Clevo M590KE

Naturalmente, isso tem um preço em termos de portabilidade. O M590KE pesa quase 7 kg e a bateria de 12 células dura pouco mais de meia hora rodando jogos pesados com duas placas em SLI.

O barebone em si custa US\$ 2259 (em Agosto de 2007), mas você poderia gastar mais de US\$ 4000 no total (preço dos EUA), já que cada GeForce Go 7950 GTX custa US\$ 390 e ainda temos o custo do processador, memória, placa wireless e HDs. Naturalmente, um desktop com um desempenho similar sairia brutalmente mais barato, mas o preço ainda é baixo se comparado com o de outros laptops de configuração similar.

O maior problema com relação aos barebones, é a dificuldade em comprar os componentes aqui no Brasil. Uma coisa é comprar um notebook montado dentro de uma determinada configuração ou faixa de preço, outra é conseguir encontrar modelos específicos de barebones, processadores mobile e placas MXM à venda. Você pode perfeitamente comprar tudo online, em lojas do exterior, mas os gastos com o transporte e impostos acabam elevando muito os valores.

## Fabricantes

A primeira questão a ter em mente com relação aos notebooks é que existe um número relativamente pequeno de fabricantes, como a:

**Quanta** (<http://www.quantatw.com>)

**Compal** (<http://www.compal.com>)

**Clevo** (<http://www.clevo.com.tw>)

Que embora não vendam notebooks sob sua marca, desenvolvem e fabricam equipamentos para inúmeros outros fabricantes que os vendem sob sua marca, tais como a HP, Dell, etc. Estes fabricantes são chamados genericamente de "Contract manufacturer" (CM).

O mesmo acontece com os integradores nacionais, como a Positivo e Amazon.

Com exceção de algumas das etapas finais da montagem, empacotamento e venda, praticamente tudo é terceirizado. A grande maioria destes modelos populares são na verdade fabricados pela Compal, Clevo ou a ECS.

A Toshiba já foi uma das maiores fabricantes de notebooks, mas atualmente terceiriza a maior parte da produção para a Compal. Outro caso interessante é o da IBM, que vendeu sua divisão de notebooks para a Lenovo.

Temos em seguida o time de fabricantes que vendem equipamentos sob sua própria marca, como a Asus, a ECS e a Acer. A Asus por exemplo, fabrica desde os Apple Mac Book até algumas séries do Playstation 3, enquanto a Acer fabrica alguns dos notebooks da Dell e de diversos outros integradores espalhados pelo mundo.

O fato de fabricar ou terceirizar acaba influenciando pouco no preço final dos produtos, pois devido à concorrência, os fabricantes trabalham com margens muito apertadas de lucro. Se a Acer e a HP resolvessem produzir um notebook com a mesma configuração, onde a Acer o fabrica diretamente e a HP o terceiriza para a Quanta (por exemplo), o custo inicial acabaria sendo praticamente o mesmo. As diferenças de preço são geralmente introduzidas mais adiante no processo, quando são incluídos os gastos com distribuição,

marketing, substituição de aparelhos na garantia e a margem de lucro de cada fabricante. Quem consegue ser mais eficiente na combinação de todas estas etapas, acaba levando a melhor.

Em seguida, temos a questão da variação de preço entre diferentes modelos do mesmo fabricante. Enquanto os modelos mais básicos são vendidos no exterior por 600 dólares ou menos, modelos apenas um pouco mais parrudos podem custar o dobro ou o triplo deste valor. Mesmo aqui no Brasil temos esta variação.

O que ocorre neste caso é que os fabricantes trabalham com margens de lucro maiores nos lançamentos, aproveitando-se do público que quer "o melhor" e está disposto a pagar caro por isto, e margens progressivamente menores nos modelos mais baratos, chegando a vender os modelos antigos com prejuízo, para se livrar dos estoques. Muita gente acha que os fabricantes nunca vendem equipamentos com prejuízo, mas devido à rápida evolução dos componentes, isso é extremamente comum. Ou seja, do ponto de vista do custo-benefício, os notebooks mais baratos são geralmente melhores, pois são "subsidiados" pelos lucros obtidos nos modelos mais caros.

Finalmente, temos a "terceira linha", representada pelos equipamentos remanufaturados (refurbished).

Estes são equipamentos que deram defeito, foram trocados dentro do período de garantia e estão sendo vendidos novamente depois de consertados. Embora sejam mais baratos, os remanufaturados nem sempre são uma boa opção, pois além de serem equipamentos já com algum uso (muitas vezes com riscos e outros detalhes), são invariavelmente vendidos com uma garantia menor. Preste atenção na hora de comprar.

## Drivers

Infelizmente, cada vez mais fabricantes optam por não fornecer um CD de quick restore e muitas vezes nem mesmo um CD de drivers junto com os notebooks vendidos, uma pequena economia, que acaba representando dor de cabeça para quem compra.

Nesses casos, para gerar o quick-restore, você deve usar o utilitário incluído no sistema e fornecer um DVD virgem. O maior problema é que ele inclui apenas uma imagem do sistema, sem uma pasta com os drivers.

Ao instalar uma cópia limpa do Windows, você precisa baixar os drivers e utilitários necessários no site do fabricante. Em muitos casos, além dos drivers, você precisa também de um software de gerenciamento para ativar

os botões controlados via software, como o "Launch Manager", que controla o botão do transmissor da placa wireless e outras funções em notebooks da Acer.

Vamos então a uma lista dos principais fabricantes:

### Acer:

<http://support.acer-euro.com/drivers/downloads.html>

ou

<http://www.cpsy.com.br/>.

### HP:

<http://welcome.hp.com/country/br/pt/support.html>

### Lenovo e IBM:

<http://www.lenovo.com/support/br/> (a página inclui também manuais técnicos detalhados para a maioria dos modelos).

**Toshiba:** <http://support.toshiba.com/>

### Asus:

<http://support.asus.com/download/>

### Averatec:

<http://www.averatec.com/customercare/downloads.asp>

### Dell:

<http://support.dell.com/support/downloads/>

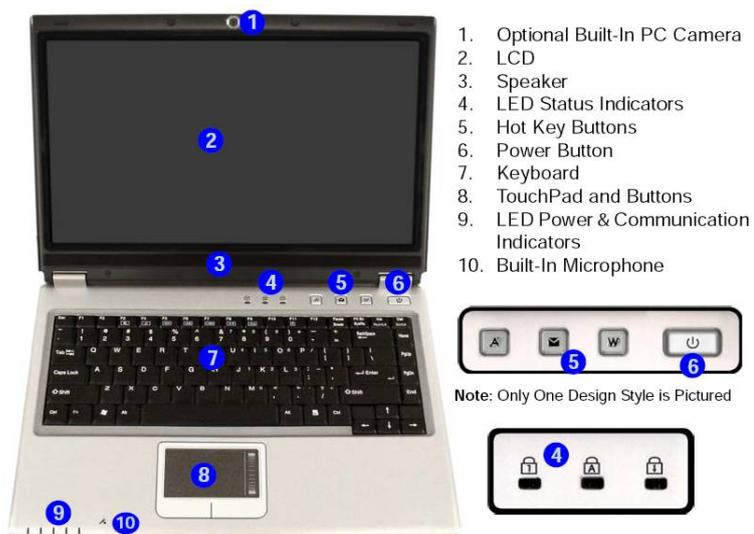
**Compal:** Embora a Compal venda um número muito pequeno de notebooks

sob sua própria marca, ela é a verdadeira fabricante de uma grande parcela dos notebooks vendidos pela Amazon, Toshiba e outros. O Amazon L81 é na verdade um Compal EL81, o Amazon L71 é um Compal DL71 e assim por diante. Os drivers para notebooks fabricados pela Compal podem ser encontrados no:

[http://www.compal.com/asp/driver\\_dnd/](http://www.compal.com/asp/driver_dnd/).

**Clevo:** A Clevo é a verdadeira fabricante da maior parte (senão todos) os notebooks da Positivo, entre outros integradores nacionais. O Positivo V41, por exemplo, é na verdade um Clevo M540S. Os drivers estão disponíveis no: <http://www.clevo.com.tw/download/>

Além de drivers, as páginas de download da Compal e Clevo incluem manuais e até mesmo guias de desmontagem (para alguns modelos). As fotos incluídas nos manuais também ajudam a identificar os modelos vendidos pelos integradores nacionais em caso de dúvidas.



1. Optional Built-In PC Camera
2. LCD
3. Speaker
4. LED Status Indicators
5. Hot Key Buttons
6. Power Button
7. Keyboard
8. TouchPad and Buttons
9. LED Power & Communication Indicators
10. Built-In Microphone

Note: Only One Design Style is Pictured



## Processadores

Dois componentes que são consideravelmente diferentes nos portáteis são os processadores e as aceleradoras 3D. Você pode perfeitamente comprar um notebook com um processador dual-core ou até mesmo com duas aceleradoras 3D em SLI, mas naturalmente ele será equipado com componentes de baixo consumo, diferentes dos que seriam usados em um desktop.

Antigamente, o mercado de micros portáteis era muito menor, de forma que os fabricantes não tinham o hábito de investir muito no desenvolvimento de processadores para portáteis. Até pouco tempo atrás, os processadores mobile eram quase sempre versões de baixo consumo dos mesmos processadores destinados a desktops, apenas operando a frequências um pouco mais baixas e incluindo algum sistema rudimentar de gerenciamento de energia.

Uma das primeiras empresas a tentar produzir processadores de baixo consumo, destinado especificamente a notebooks e outros equipamentos portáteis foi a Transmeta.

Foto do manual do Clevo M540S e foto de divulgação do Positivo V41.  
Nota alguma semelhança?

Ela produziu duas linhas de processadores, o Crusoe e o Efficeon. Ambos eram incrivelmente eficientes, mas ofereciam um baixo desempenho e o preço não era competitivo com relação aos concorrentes diretos da Intel e da AMD. A Transmeta conseguiu ganhar algumas batalhas entre 2002 e 2005, mas acabou naufragando com o lançamento do Pentium M e do Turion, processadores que conseguem combinar um bom desempenho com um sistema eficiente de gerenciamento de energia, além de ter sofrido com a concorrência direta do VIA C3 (veja mais detalhes a seguir), também desenvolvido para ser um processador de baixo consumo.

Antigamente, os processadores AMD não eram exatamente uma boa opção para portáteis, pois a AMD não possuía um sistema eficiente de gerenciamento de energia. Antes do Turion, os processadores "Mobile" da AMD eram basicamente versões de baixo consumo dos chips para desktops, fazendo com que o aquecimento e o consumo elétrico ficassem longe do ideal.

A Intel também cometeu suas gafes, com os Pentium 4 Mobile e os modelos equivalentes do Celeron Mobile, ambos baseados na arquitetura NetBurst. Um Mobile Pentium 4 de 3.06 GHz (baseado no core Northwood), por exemplo, possui um TDP de absurdos 70 watts, o que resultava em notebooks volumosos, pesados e com pouca autonomia

de baterias. Para completar, além de todas as desvantagens, o Mobile Pentium 4 de 3.06 GHz baseado no core Northwood perde em processamento para um simples Turion MT-34, lançado menos de dois anos depois.

A situação mudou com o lançamento do Banias, a primeira geração do Pentium M. Ele foi um processador desenvolvido sob medida para o uso em notebooks e acabou se revelando tão eficiente que acabou dando origem ao Core 2 Duo. A AMD respondeu lançando as versões Mobile do Athlon 64 e do Sempron e em seguida lançando o Turion e o Turion X2, também processadores bastante eficientes. Vamos então aos detalhes.

## Pentium M

A primeira geração do Pentium M era baseada no Banias, um descendente indireto do Pentium III Tualatin, em vez de ser uma versão de baixo consumo do Pentium 4. Sob diversos pontos de vista, ele era um processador mobile desenvolvido especificamente para ser usado em portáteis, ao invés de ser um processador para desktops adaptado para uso em notebooks.

Apesar do baixo consumo, o Banias se revelou relativamente poderoso.

Um Pentium M de 1.6 GHz baseado no core Banias supera facilmente um Pentium 4 de 2.4 GHz na maior parte dos aplicativos. Com a decadência da plataforma NetBurst, o Banias foi sucessivamente atualizado, dando origem à plataforma Core.

A segunda geração do Pentium M é baseada no core Dothan, sucessor do Banias. A primeira diferença entre eles é que o Banias ainda era produzido usando uma técnica de 0.13 micron, enquanto o Dothan inaugurou o uso da técnica de 0.09 micron. A redução possibilitou o aumento no cache, que passou a ser de 2 MB, além de reduzir o consumo do processador que, apesar do aumento na frequência de operação, caiu de 24.5W para apenas 21W. O Dothan trouxe também melhorias no circuito de branch-prediction, que é basicamente o mesmo usado no Conroe, a primeira versão do Core 2 Duo.



*Pentium M com core Dothan*

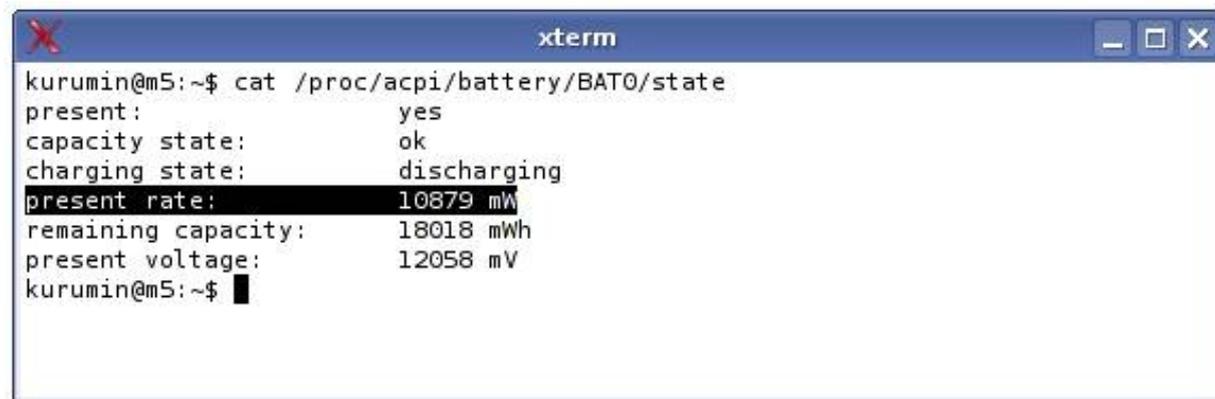
Com o lançamento da plataforma Sonoma, a segunda geração da plataforma Centrino, o Pentium M com core Dothan recebeu mais um upgrade, passando a utilizar memórias DDR2 e bus de 533 MHz. Estas melhorias, entretanto, são relacionadas ao controlador de memória incluído no chipset e não ao processador em si.

Todos os processadores Pentium M e Core oferecem suporte ao SpeedStep III (o SpeedStep original era usado nos processadores baseados na arquitetura P6 e o SpeedStep II era usado pelos processadores Mobile Pentium 4), que permite que a frequência e tensão usadas pelo processador sejam ajustadas dinamicamente, de acordo com a carga de processamento.

O ajuste é feito em "degraus" (steps, daí o nome). Em um Pentium M de 1.73 GHz, com core Dothan, por exemplo, os "degraus" disponíveis são 1.73 GHz, 1.33 GHz, 1.06 GHz e 800 MHz (a frequência mínima) e a tensão pode oscilar entre 0.988 e 1.356V.

No screenshot a seguir temos uma medição de consumo feita em um Asus M5200, baseado no Pentium M com core Dothan de 1.73 GHz. Ao executar tarefas pesadas, com o processador operando na frequência máxima, com o HD sendo acessado, o transmissor wireless ativado e o brilho da tela no máximo, o consumo do notebook chega a

atingir os 36 watts. Ao executar tarefas leves, por outro lado, com o processador operando na frequência mínima (800 MHz), o transmissor da placa wireless desativado e o brilho da tela reduzido em 6 níveis, o consumo do notebook chega a se aproximar da marca dos 10 watts. No screenshot ele está consumindo 10.879W:



```

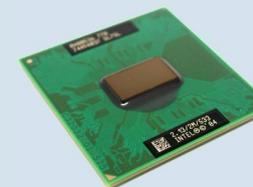
xterm
kurumin@m5:~$ cat /proc/acpi/battery/BAT0/state
present:          yes
capacity state:   ok
charging state:   discharging
present rate:     10879 mW
remaining capacity: 18018 mWh
present voltage:  12058 mV
kurumin@m5:~$
  
```

Continuando, o Dothan foi lançado em versões de até 2.26 GHz, mas estas últimas trabalhavam com um TDP muito mais elevado, de 27 watts. Como a Intel calcula seu TDP com base em uma estimativa de 75% da dissipação máxima do processador, é seguro dizer que um Dothan de 2.26 pode dissipar até 36 watts (durante curtos espaços de tempo) em determinadas aplicações.

A partir do Dothan, a Intel passou a vender os processadores sob um sistema de numeração, similar ao usado nos processadores para desktop. A lista de processadores baseados no core Dothan inclui o Pentium M **710** (1.4 GHz, 2 MB, 21W, 400 MHz), **715** (1.5 GHz, 2 MB, 21W, 400 MHz), **725** (1.6 GHz, 2 MB, 21W, 400 MHz), **730** (1.6

2 MB, 27W, 533 MHz), **735** (1.7 GHz, 2 MB, 21W, 400 MHz), **740** (1.73 GHz, 2 MB, 27W, 533 MHz), **745** (1.8 GHz, 2 MB, 21W, 400 MHz), **750** (1.86 GHz, 2 MB, 27W, 533 MHz), **755** (2.0 GHz, 2 MB, 21W, 400 MHz), **760** (2.0 GHz, 2 MB, 27W, 533 MHz), **765** (2.1 GHz, 2 MB, 21W, 400 MHz), **770** (2.13 GHz, 2 MB, 27W, 533 MHz) e **780** (2.26 GHz, 2 MB, 27W, 533 MHz).

Todos estes modelos utilizam o soquete 479 e não oferecem suporte às instruções SSE3 (apenas ao SSE2), que, ironicamente, são suportadas pelo AMD Turion, que seria seu concorrente direto.



Pentium M  
Dothan

## Core Duo e Core 2 Duo

O **Yonah**, lançado em 2006, foi um projeto ambicioso e o primeiro processador mobile dual-core desenvolvido pela Intel.

A partir do Yonah, a Intel abandonou o uso da marca "Pentium M" e passou a usar as marcas "Core Solo" e "Core Duo", para (respectivamente) as versões single-core e dual-core do processador. O Core 2 Duo para desktops se chama "Core 2 Duo" justamente por ser a segunda versão da plataforma iniciada com o Yonah.

O uso de dois núcleos é mais problemático em um processador mobile do que em um processador desktop devido à questão do consumo. Dois núcleos dissipam mais calor do que apenas um e consomem o dobro de energia. Apesar disso, o TDP dos Core Duo é de 31 watts, contra os 27 watts dos Pentium M single-core baseados no Dothan, um aumento de apenas 4 watts.

Isso ocorre por dois motivos. O primeiro é que o Yonah é produzido usando a técnica de 0.065 micron, o que faz com que, mesmo com dois núcleos, ele ocupe uma área de apenas 90 mm<sup>2</sup>, contra 84 mm<sup>2</sup> do Dothan e tenha um consumo elétrico apenas um pouco mais alto.

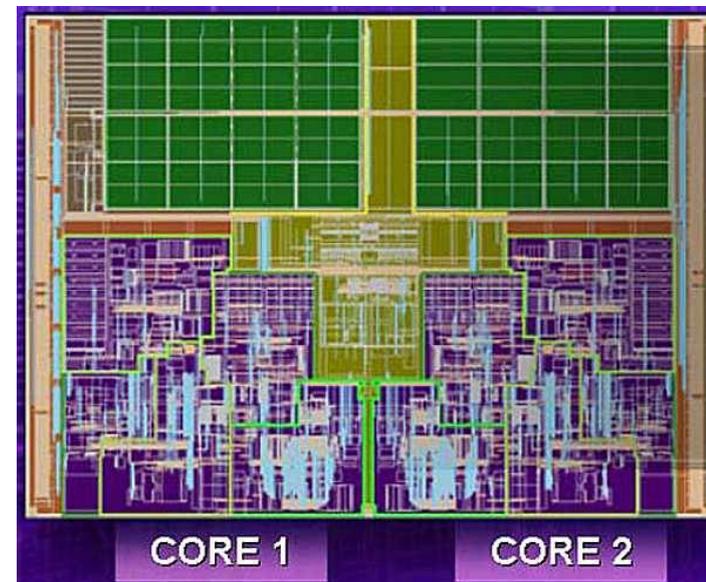
O segundo é que o Yonah inclui uma versão aprimorada do SpeedStep, que é capaz de ajustar a frequência de operação dos dois núcleos de forma independente.

Isso faz com que o segundo núcleo passe a maior parte do tempo operando na frequência mínima, enquanto o primeiro tem sua frequência ajustada de acordo com a demanda. Em situações onde os dois núcleos sejam muito exigidos durante longos períodos e o processador aqueça de forma exagerada, o SpeedStep pode também reduzir a frequência de um dos núcleos "na marra", reduzindo assim a dissipação até que a temperatura do processador volte a níveis seguros.

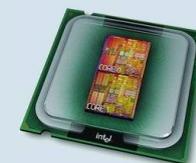
Além de ser um processador dual-core, o Yonah incorporou o suporte a SSE3, o que melhorou o desempenho em diversas aplicações. Entretanto, a latência do cache L2 aumentou de 10 para 14 ciclos, o que prejudica o desempenho em diversas áreas, sobretudo em aplicativos de produtividade, que usam predominantemente processamento de números inteiros. Isso faz com que um Core Solo baseado no Yonah realmente perca para um Dothan do mesmo clock em algumas aplicações, embora ganhe em outras devido ao suporte ao SSE3.

O cache L2 de 2 MB é compartilhado entre os dois núcleos, essencialmente o mesmo sistema utilizado no Core 2 Duo.

Isso permite que o espaço utilizado por cada núcleo seja ajustado dinamicamente, conforme o uso. Temos aqui uma foto de divulgação da Intel que mostra os componentes internos do processador:



Uma curiosidade é que os processadores "Core Solo" na verdade não constituem uma linha separada. Eles são simplesmente versões do Yonah com um dos núcleos desativados, o que os transforma em processadores single-core com 2 MB de cache, muito similares ao Pentium M com core Dothan.



Intel dual core  
Yonah

Mesmo os Celerons baseados no "Yonah-1024" começam a vida como processadores dual-core e são sucessivamente castrados, até que reste apenas um dos núcleos, com metade do cache e sem suporte ao SpeedStep. Por motivos diversos, a Intel acredita ser mais vantajoso manter uma linha unificada de produção e castrar os processadores conforme a aplicação (aproveitando assim processadores defeituosos que de outra forma precisariam ser descartados) do que manter linhas separadas para os processadores single-core, dual-core e para os Celerons.

O Yonah foi usado nos Core Duo **T2050** (1.6 GHz, 2 MB, 31W, 533 MHz), **T2250** (1.73 GHz, 2 MB, 31W, 533 MHz), **T2300** (1.66 GHz, 2 MB, 31W, 667 MHz), **T2350** (1.86 GHz, 2 MB, 31W, 533 MHz), **T2400** (1.83 GHz, 2 MB, 31W, 667 MHz), **T2450** (2.0 GHz, 2 MB, 31W, 533 MHz), **T2500** (2.0 GHz, 2 MB, 31W, 667 MHz), **T2600** (2.16 GHz, 2 MB, 31W, 667 MHz) e **T2700** (2.33 GHz, 2 MB, 31W, 667 MHz).

Temos ainda três modelos de baixo consumo, o Core Duo LV **L2300** (1.5 GHz, 2 MB, 15W, 667 MHz), LV **L2400** (1.66 GHz, 2 MB, 15W, 667 MHz) e o LV **L2500** (1.83 GHz, 2 MB, 15W, 667 MHz). Como pode ver, eles possuem um TDP de apenas 15 watts, inferior até mesmo que o dos Celerons baseados no Dothan-1024.

O Yonah foi utilizado também nos Core Solo **T1200** (1.5 GHz, 2 MB, 27W, 667 MHz), **T1300** (1.66 GHz, 2 MB, 27W, 667 MHz), **T1350** (1.86 GHz, 2 MB, 27W, 533 MHz), **T1400** (1.86 GHz, 2 MB, 27W, 667 MHz), **T1500** (2.0 GHz, 2 MB, 27W, 667 MHz) e **T1600** (2.16 GHz, 2 MB, 27W, 667 Mhz).

Todos os processadores baseados no Yonah utilizam o soquete M e são por isso incompatíveis com as placas anteriores. É importante enfatizar que tanto os processadores baseados no core Dothan quanto no Yonah não possuem suporte a instruções de 64 bits (o EM64 está disponível apenas nos Core 2 Duo e superiores). Este é um dos pontos em que os processadores Intel desta safra inicial perdem para os processadores da AMD.

Em seguida temos o Meron, que é a versão mobile do Core 2 Duo. Se comparado com o Yonah, ele inclui diversas melhorias, com destaque para o cache de 4 MB, adição de um decodificador adicional de instruções simples (o Meron possui três, contra dois do Yonah), reforço nas unidades de processamento de instruções SSE, aumento do comprimento do buffer do circuito ordenador de instruções e melhorias no circuito de branch prediction. Combinadas, estas melhorias justificam o brutal aumento no número de transistores, que saltou de 151 milhões no Yonah para nada menos do que 291 milhões no Meron, distribuídos sob uma área de 144 mm<sup>2</sup>.

Quase tão importante quanto as melhorias relacionadas ao desempenho é o suporte a instruções de 64 bits, incluído apenas a partir do Meron. Nenhum dos processadores baseados no Banias, Dothan ou Yonah oferece suporte ao EM64, o que acaba sendo uma limitação grave.

A questão é que o Meron não é diferente do Conroe, usado nos Core 2 Duo para desktops.



Core Duo baseado no core Yonah (à esquerda) e Pentium M com core Dothan

Naturalmente ele possui otimizações relacionadas à redução do consumo elétrico, mas a arquitetura dos dois é idêntica. Comparando dois processadores do mesmo clock, o Meron consegue ser de 10 a 20% mais rápido que o Yonah em quase todos os aplicativos. Apesar do brutal aumento no número de transístores, o Meron também não consome consideravelmente mais energia que um Yonah do mesmo clock, de forma que realmente não existem pontos negativos no Core 2 Duo em relação ao Core Duo.

Apesar das melhorias, a Intel optou por não fazer nenhum grande esforço de marketing com relação a ele. Notebooks vendidos sob a marca "Centrino Duo" podem tanto ser baseados no Core Duo baseado no Yonah quanto no Core 2 Duo baseado no Meron, sem distinção. É inclusive possível substituir um Yonah soquete M, por um Meron (também soquete M) diretamente, na maioria dos casos precisando apenas de um upgrade de BIOS. O preço dos processadores também não mudou, de forma que durante muito tempo, era possível encontrar processadores baseados nas duas arquiteturas pelo mesmo preço.

A lista de processadores baseados no Meron inclui o Core 2 Duo **T5200** (1.6 GHz, 2 MB, 34W, 533 MHz), **T5500** (1.66 GHz, 2 MB, 34W, 667 MHz), **T5600** (1.83 GHz, 2 MB, 34W, 667 MHz), **T7200** (2.0

GHz, 4 MB, 34W, 667 MHz), **T7300** (2.0 GHz, 4 MB, 35W, 800 MHz), **T7400** (2.16 GHz, 4 MB, 34W, 667 MHz), **T7500** (2.2 GHz, 4 MB, 35W, 800 MHz), **T7600** (2.33 GHz, 4 MB, 34W, 667 MHz) e **T7700** (2.4 GHz, 4 MB, 35W, 800 MHz).

Existe também o Meron-2M, uma versão reduzida do Meron, com apenas 2 MB de cache, em vez de 4 MB. Ele é diferente dos T5200, T5500 e T5600, modelos baseados no Meron que possuem apenas 2 MB do cache (com os outros 2 MB desativados), pois ele é realmente um processador diferente, com menos cache, e não um Meron regular com metade do cache desativado.

O Meron-2M é usado no Core 2 Duo **T5250** (1.5 GHz, 2 MB, 34W, 667 MHz), **T5300** (1.73 GHz, 2 MB, 34W, 533 MHz), **T5450** (1.66 GHz, 2 MB, 34W, 667 MHz), **T5470** (1.6 GHz, 2 MB, 34W, 800 MHz), **T5500** (1.66 GHz, 2 MB, 34W, 667 MHz), **T5600** (1.83 GHz, 2 MB, 34W, 667 MHz) e **T7100** (1.8 GHz, 2 MB, 35W, 800 MHz).

Os Core 2 Duo baseados no Meron e Meron-2M são produzidos tanto em formato soquete M quanto soquete P. Isso permite que eles sejam usados tanto em placas novas quanto como substitutos diretos para os Core Duo baseados no Yonah em projetos já existentes.



Core 2 Duo T5600, baseado no Meron-2M

## Celeron M

O Celeron M é possivelmente o processador mobile da Intel mais vendido, usado em uma infinidade de notebooks de baixo custo. Embora todo o marketing seja concentrado em torno da plataforma Centrino, os notebooks baseados no Celeron acabam sendo vendidos em maior número, já que são mais baratos.



Celeron M 350, baseado no core Dothan-1024

Como de praxe, o Celeron M possui metade do cache do Pentium M. No caso do Celeron com core Dothan, por exemplo, temos 1 MB contra 2 MB do Pentium M. Isto não chega a ser uma desvantagem tão grande, já que reduz o desempenho em apenas 10%, em média. A principal fraqueza do Celeron M reside na falta de gerenciamento avançado de energia. Ao contrário do Pentium M, ele trabalha sempre na frequência máxima, sem a opção de usar o speedstep, o que significa mais consumo e uma menor autonomia das baterias, sobretudo ao rodar aplicativos leves, situação onde a diferença de consumo entre o Celeron e outros processadores (com suporte a sistemas de gerenciamento de energia) é maior.

As primeiras versões do Celeron eram baseadas no "Banias-512" que, como o nome sugere, é uma versão do Pentium M com core Banias com metade do cache L2. Esta série inicial engloba o Celeron **M 310** (1.2 GHz, 512 KB, 24.5W), **320** (1.3 GHz, 512 KB, 24.5W), **330** (1.4 GHz, 512 KB, 24.5W) e **340** (1.5 GHz, 512 KB, 24.5W).

A série seguinte é baseada no "Dothan-1024", novamente uma versão do Pentium M com metade do cache. Ela inclui o Celeron **M 350** (1.3 GHz, 1 MB, 21W), **360** (1.4 GHz, 1 MB, 21W), **360J** (1.4 GHz, 1 MB, 21W), **370** (1.5 GHz, 1 MB, 21W), **380** (1.6 GHz, 1 MB, 21W) e **390** (1.7 GHz, 1 MB, 21W).

Todos estes modelos utilizam placas soquete 479 e suportam apenas instruções MMX, SSE e SSE2, sem suporte a SSE3. Os modelos **360J**, **370**, **380**, **390** suportam XD bit (o recurso de segurança, que protege áreas de memória, dificultando a ação dos vírus). Todos os modelos utilizam também bus de 400 Mhz.

Em seguida temos os modelos mais atuais, baseados no Yonah e no Meron. Ao contrário do Pentium M, os Celerons baseados em ambas as séries são single-core e possuem apenas 1 MB de cache L2. Os núcleos são chamados respectivamente de "Yonah-1024" e "Meron-1024".

Até o momento (outubro de 2007) não existem Celerons dual-core, nem notícias sobre um possível lançamento. Mesmo que a Intel decidisse lançar versões de baixo custo do Meron dual-core, possivelmente venderia os processadores sob a marca "Pentium", assim como no caso do Pentium E para desktops (baseados no Allendale) e não sob a marca "Celeron".

Voltando ao que interessa, a série baseada no **Yonah-1024** inclui o Celeron **M 410** (1.46 GHz, 1 MB, 27W), **420** (1.6 GHz, 1 MB, 27W), **430** (1.73 GHz, 1 MB, 27W), **440** (1.86 GHz, 1 MB, 27W) e **450** (2.0 GHz, 1 MB, 27W).

A série baseada no Meron-1024 inclui o Celeron M 520 (1.6 GHz, 1 MB, 30W) e

**530** (1.73 GHz, 1 MB, 30W). Existe ainda a série baseada no "Meron-L", uma versão de baixo consumo do **Meron-1024** que inclui o Celeron **520** (1.6 GHz, 1 MB, 26W), **540** (1.86 GHz, 1 MB, 30W) e **550** (2.0 GHz, 1 MB, 30W). O Celeron 520 baseado no Meron-L diferencia-se do 520 baseado no Meron-1024 apenas pelo TDP, que é um pouco mais baixo. As demais características dos dois processadores são idênticas.

Todos estes modelos utilizam o soquete M e bus de 533 MHz. A exceção fica por conta do Celeron M 540, que utiliza o soquete P. Todos eles oferecem suporte também ao SSE3 e ao XD bit.

## A plataforma Centrino

Sem dúvida, vender um pacote com processador, chipset e placa wireless é muito mais lucrativo do que vender apenas o processador. Controlar o desenvolvimento dos três componentes facilita também o desenvolvimento de sistemas mais eficientes de gerenciamento de energia e otimizações em geral.

A idéia por trás de todo o marketing em torno da marca "Centrino" é convencer os compradores de que os notebooks com o selo são uma escolha segura.

A partir do momento em que os compradores passam a preferir notebooks baseados na plataforma, a Intel pode trabalhar com margens de lucro maiores e assim maximizar os lucros, ao mesmo tempo em que mantém o controle sobre toda a plataforma.



Processador, placa wireless e chipset: o "kit" Centrino

Fabricantes, como a nVidia, Broadcom, Ralink, Atheros, VIA, SiS, Realtek e outros vendem chipsets e placas wireless a preços mais competitivos que a Intel. Em muitos casos os produtos alternativos são inclusive claramente superiores, como no caso das soluções de vídeo onboard da nVidia e ATI, que superaram em muito os chipsets de vídeo da série GMA900 utilizados nos chipsets Intel. Apesar disso, o marketing em torno da marca Centrino faz com que os fabricantes prefiram utilizar os chipsets e placas wireless da Intel, relegando os produtos de outros fabricantes aos modelos de baixo custo ou a configurações específicas.

A única brecha importante para o uso de componentes "não-Intel" em notebooks Centrino é no caso de chipsets de vídeo alternativos. Embora os chipsets Intel mais usados sejam os chipsets da linha "GM", que incluem vídeo onboard, estão disponíveis também os chipsets da linha "PM", que são idênticos, mas não incluem o chipset de vídeo. Eles permitem que os fabricantes de notebooks incluam chips ou placas MXM ou AXION da nVidia ou ATI sem com isso abandonar o uso da marca "Centrino". Entre os notebooks mais caros, é comum o uso de placas nVidia offboard no lugar do fraco vídeo Intel onboard.

A primeira encarnação da plataforma Centrino foi lançada em 2003 e responde pelo codenome Carmel. Ela consiste na combinação de um Pentium M com core Banias ou Dothan, um chipset i855 e uma placa wireless Intel 2100 ou 2200.

O chipset Intel 855 oferece suporte apenas a memórias DDR e as placas wireless Intel PRO/Wireless 2100 ainda utilizam o padrão 802.11b (11 megabits), sem suporte ao 802.11g, o que hoje em dia são duas pesadas limitações. A questão da placa wireless foi solucionada em 2004, com o lançamento da Intel PRO/Wireless 2200, que

pode operar tanto em modo b quanto g. Quase todos os notebooks Centrino produzidos a partir do segundo trimestre de 2004 já são equipados com a placa wireless atualizada.

Em 2005 foi lançada a segunda geração, sob o codenome Sonoma. Nessa época, o Banias já havia sido descontinuado, de forma que passou a ser usado exclusivamente o Pentium M com core Dothan. O limitado 855 foi substituído pelo Intel 915, que trouxe o suporte a memórias DDR2, SATA, slots Express Card, áudio HDA e bus de 533 MHz.

O uso de memórias DDR2 ajuda a reduzir o consumo elétrico, já que os módulos utilizam uma tensão mais baixa. Existe também um pequeno ganho com relação à compatibilidade com módulos de diferentes fabricantes, já que os módulos DDR2 possuem um terminador resistivo dentro de cada chip, o que garante uma melhor qualidade de sinal e reduz o número de casos de incompatibilidade, sobretudo ao utilizar dois módulos diferentes.

A partir do Intel 915, todos os chipsets mobile da Intel oferecem suporte a dual-channel (com exceção dos chipsets da linha "GMS", como o 915GMS, que são uma linha de baixo custo, com um controlador single-channel). Apesar disso, a maior parte dos notebooks não vem com o recurso ativado, simplesmente porque o uso de dual-channel

exige o uso de dois módulos de memória, o que aumenta os custos. Nada impede, entretanto, que você instale um segundo módulo de memória, ativando assim o dual-channel.

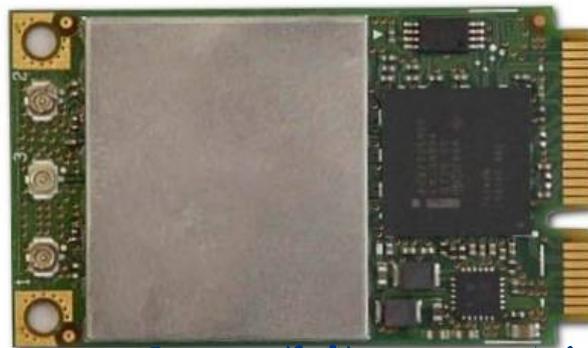
Com relação à placa wireless, os notebooks baseados na plataforma Sonoma podem utilizar tanto a PRO/Wireless 2200 (BG) quanto a 2915ABG, que, como o nome sugere, oferece como diferencial o suporte também ao padrão 802.11a.



*Placa wireless Intel 2915ABG mini-PCI instalada*

A terceira geração é a plataforma Napa, lançada em 2006. Ela consiste no uso de um processador Core Solo, Core Duo ou Core 2 Duo em versão soquete M, combinado com o chipset Intel 945 (que inclui suporte a bus de 667 MHz) e uma placa wireless Intel 3945ABG ou 4965AGN.

A 3945ABG é uma versão atualizada da 2915ABG, que mantém o suporte aos padrões a, b e g, enquanto a 4965AGN (a mesma utilizada no Santa Rosa) é uma placa "draft-n", com suporte a uma versão prévia no padrão 802.11n, que oferece uma taxa de transmissão teórica de 300 megabits e um alcance até duas vezes maior que uma placa 802.11g equipada com um transmissor de potência similar. É fácil diferenciar as placas 4965AGN das antigas, pois além de serem produzidas exclusivamente no formato Express Mini, elas possuem três conectores de antena, em vez de apenas dois:



*Placa wireless 4965AGN, no formato Express Mini*

Os notebooks baseados no Napa, equipados com processadores Core Solo ainda são vendidos sob a marca "Centrino", enquanto os baseados em processadores Core Duo ou Core 2 Duo passaram a utilizar a marca "Centrino

Duo". Não existem notebooks "Centrino Duo" baseados na plataforma Sonoma, já que ela não prevê o uso de processadores dual-core.

Temos ainda a plataforma **Santa Rosa**, lançada em 2007. Ela prevê o uso de um processador Core 2 Duo soquete P (bus de 800 MHz), combinado com um chipset Intel 965 e uma placa wireless Intel 4965AGN.

O sistema de gerenciamento de energia foi atualizado, de forma a (além de ajustar a frequência do processador) permitir a redução da frequência do FSB de 800 para 400 MHz nos momentos de baixa atividade, ajudando a compensar o aumento de consumo trazido pelo uso do bus de 800 MHz (recurso batizado de "Dynamic Front Side Bus Switching").

A redução da frequência do FSB resulta em uma economia relativamente pequena, de menos de 1 watt. Entretanto, muitos notebooks ultra-compactos baseados na plataforma Centrino chegam a consumir apenas 12 watts ou menos (incluindo a tela) quando o processador está operando na frequência mais baixa, de forma que qualquer nova redução acaba tendo um impacto significativo. Outro recurso divulgado ao extremo pela Intel é o "Turbo Memory" (tecnologia Robson), onde é utilizado um módulo de memória Flash ligado ao barramento PCI Express em conjunto

com o ReadyDrive do Vista, de forma a melhorar o desempenho do acesso ao HD e aumentar a autonomia da bateria.

É importante ressaltar que a economia de energia trazida pelo Turbo Memory é apenas incremental, pois os HDs de 2.5" e 5400 RPM atuais são bastante econômicos, consumindo em torno de 3 watts ou menos. Além disso, o HD não fica todo o tempo girando, o que reduz o consumo prático a níveis bem mais baixos. O Turbo Memory evita um certo número de acessos ao HD, mas em troca os próprios chips de memória Flash e a linha PCI Express a que eles estão ligados consomem uma certa quantidade de energia, o que reduz o ganho na prática. Não espere mais do que 2 ou 4% de ganho de autonomia de bateria em um notebook com o Turbo Memory em relação a outro de configuração similar sem o recurso. O ganho é tão pequeno que fica até mesmo difícil de medir, pois a diferença nos resultados fica muitas vezes dentro da variação normal de um teste para outro.

O Turbo Memory é desejável se for um "brinde", incluído em uma certa configuração sem um aumento considerável no custo. Não aceite pagar mais caro por um notebook equipado com ele, pois o ganho realmente não compensa qualquer aumento expressivo no preço.

É esperada uma atualização da plataforma Santa Rosa para 2008, com a inclusão

de processadores baseados no Penryn. Eles incluirão suporte ao EDAT (Enhanced Dynamic Acceleration Technology), onde o processador pode desativar o segundo núcleo e usar a redução no consumo para fazer um overclock temporário do núcleo ativo, melhorando assim o desempenho em aplicativos single threaded, que não são capazes de se beneficiar de um processador dual-core.

## Turion 64

Apesar dos processadores Pentium M, baseados no core Dothan, e os Core Duo, baseados no core Yonah não serem capazes de operar a frequências muito altas se comparados aos processadores para desktop, eles ofereciam um desempenho por ciclo de clock muito bom, o que permitia que competissem com processadores para desktop, baseados na arquitetura NetBurst, que operavam a frequências muito mais elevadas, mesmo mantendo um consumo elétrico relativamente baixo.

A arquitetura equilibrada, combinada com o esforço de marketing da Intel em torno da marca Centrino fizeram com que eles fossem um sucesso de vendas nos notebooks de médio e alto custo. Apesar disso, os notebooks baseados na plataforma Centrino sempre

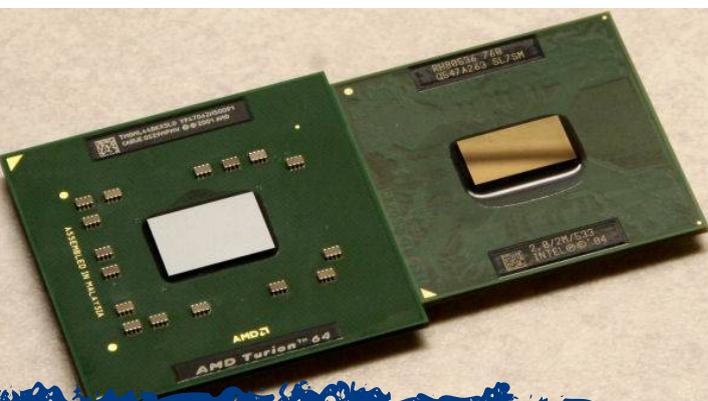
foram mais caros e o processador Intel para notebooks de baixo custo era o Celeron M, um chip muito menos competitivo, devido ao baixo clock, menos cache e à ausência do suporte ao SpeedStep.

A resposta da AMD veio na forma do Turion 64, uma versão mobile do Athlon 64 soquete 754, que operava a frequências mais baixas e incluía o PowerNow. Além de ser relativamente barato, o Turion oferecia um desempenho muito bom, chegando a superar o Pentium M baseado no core Dothan em diversas aplicações, e era relativamente econômico em termos de consumo elétrico.

Levando em conta apenas a questão do processador, o Turion era competitivo com o Pentium M e era uma escolha muito melhor que o Celeron M. A questão é que um notebook é um conjunto, onde o resultado não depende apenas das características do processador, mas sim da combinação de todos os componentes. A Intel possuía uma plataforma mais sólida, já que desenvolvia seus próprios chipsets e placas wireless, enquanto os notebooks baseados no Turion utilizavam chipsets da ATI, VIA ou SiS, criando plataformas mais problemáticas.

Isso fazia com que os Turion acabasse sendo usado predominantemente nos notebooks de baixo custo, o que não

era necessariamente ruim, pois permitia que você comprasse notebooks relativamente potentes, baseados nas diferentes versões do Turion a preços substancialmente mais baixos que os baseados na plataforma Centrino.



*Turion 64 e Pentium M com core Dothan*

O Turion 64 também surgiu como uma derivação do Athlon 64 com core Venice (assim como o Mobile Athlon 64 com core Newark), mas desta vez a AMD tomou cuidados adicionais. Para reduzir o consumo elétrico, a AMD adotou o uso de transistores de chaveamento mais lento, que reduzem a frequência máxima de operação do processador, mas em troca oferecem um consumo elétrico muito mais baixo. Como um processador mobile como o Turion nunca opera a frequências muito elevadas, devido à questão do consumo e dissipação térmica, a troca acabou fazendo sentido.

Foi incluído ainda um estágio adicional no sistema de gerenciamento de energia (o C3), que colocou o Turion um passo à frente dos Athlon 64 para desktops na questão do gerenciamento de energia.

Os Turions da série **MT** trabalham dentro de um TDP de 25 watts, enquanto os **ML** são um pouco mais gastadores e adotam um TDP de 35 watts. Existem ainda versões com 512 KB e 1 MB de cache L2, mas todas as versões utilizam um controlador de memória single-channel (justamente por isso é utilizado o soquete 754), uma restrição adicionada para reduzir o consumo.

Assim como os processadores soquete 754 para desktop, eles suportam apenas módulos DDR (sem suporte a DDR2) o que, combinado com o acesso single-channel, estrangula o barramento com a memória, prejudicando o desempenho do processador em diversas aplicações.

A diferença no consumo é justificada pelo tipo de transistores e pelas otimizações utilizados em cada série. Apesar disso, não existe diferença de desempenho entre um MT e um ML de mesmo clock (e com a mesma quantidade de cache).

Esta versão inicial do Turion 64 foi chamada de **Lancaster** e, assim como o Venice, era produzida usando a técnica de 0.09 micron e oferecia suporte a SSE3 e instruções de 64 bits, além do suporte ao PowerNow.

O core Lancaster foi usado no **MT-28** (1.6 GHz, 512 KB, 25W), **MT-30** (1.6 GHz, 1 MB, 25W), **MT-32** (1.8 GHz, 512 KB, 25W), **MT-34** (1.8 GHz, 1 MB, 25W), **MT-37** (2.0 GHz, 1 MB, 25W), **MT-40** (2.2 GHz, 1 MB, 25W), **ML-28** (1.6 GHz, 512 KB, 35W), **ML-30** (1.6 GHz, 1 MB, 35W), **ML-32** (1.8 GHz, 512 KB, 35W), **ML-34** (1.8 GHz, 1 MB, 35W), **ML-37** (2.0 GHz, 1 MB, 35W), **ML-40** (2.2 GHz, 1 MB, 35W), **ML-42** (2.4 GHz, 512 KB, 35W) e **ML-44** (2.4 GHz, 1 MB, 35W).

Mais adiante foi lançado o core **Richmond**, que utiliza o soquete S1, o mesmo utilizado pelo Turion X2. Ele tem a função de ser um processador de transição, permitindo que os fabricantes possam construir notebooks baseados em placas S1, sem necessariamente migrar para o Turion X2, que é mais caro. O Richmond foi vendido em apenas duas versões: **MK-36** (2.0 GHz, 512 KB, 31W) e **MK-38** (2.2 GHz, 512 KB, 31W).

Um dos principais pontos fortes do Turion é o sistema de gerenciamento de energia. Tanto o PowerNow, usado nos processadores mobile, quanto o Cool'n'Quiet, usado nos processadores para desktop, trabalham dentro do mesmo princípio, ajustando o multiplicador (e conseqüentemente a frequência) do processador, juntamente com a tensão, conforme mais ou menos poder de processamento é exigido.

Cada "degrau" corresponde a uma redução de 1x no multiplicador o que (no caso do Turion) corresponde a uma redução de 200 MHz na frequência do processador. A frequência é reduzida sequencialmente até atingir 1.6 GHz, ponto no qual o processador entra no estágio mais baixo, onde passa a operar a apenas 800 MHz (independentemente do modelo).

Juntamente com a frequência, a tensão também é reduzida, já que o processador é projetado para funcionar estavelmente usando tensões mais baixas, desde que operando a frequências também mais baixas. Quando mais processamento é exigido, a tensão volta ao valor inicial, seguida do aumento correspondente na frequência.

Um Turion ML-40, por exemplo, opera a 2.2 GHz e tem um TDP de 35 watts, mas consome este valor apenas quando está operando na frequência máxima. Conforme o PowerNow reduz a frequência de operação do processador, o consumo cai para 29 watts (2.0 GHz), 24.5 watts (1.8 GHz), 20 watts (1.6 GHz) e, finalmente, 7.9 watts (800 MHz).

Como a mesma frequência de 800 MHz é usada como estágio final em todos os Turions, independentemente da frequência original, todos os modelos acabam consumindo apenas 7.9 watts quando ociosos.

## Turion X2

O próximo passo da AMD foi o lançamento do Turion X2, com o objetivo de competir com o Core 2 Duo baseado no core Merom.

O Turion X2 é, essencialmente, uma versão de baixo consumo do Athlon X2, que preserva as mesmas características básicas, como o uso do Crossbar Switch, o controlador de memória DDR2 dual-channel (compartilhado entre os dois núcleos) e o uso do barramento HyperTransport.

Naturalmente, o fato de ser um processador mobile torna necessário o uso de um gerenciamento mais agressivo de energia. Entra em cena então uma versão atualizada do PowerNow, com suporte a um sistema de gerenciamento independente de cada núcleo.

O nível de utilização de cada núcleo é continuamente monitorado e os níveis apropriados de economia de energia são aplicados. Isso permite que um dos núcleos continue operacional, enquanto o outro "hiberna", em um dos estágios de baixo consumo.

O primeiro nível é o ajuste de frequência e tensão de operação do processador, essencialmente o mesmo sistema já usado nos processadores anteriores.

A frequência de operação é reduzida para 800 MHz e a tensão para 1.075V (ou menos, de acordo com a revisão do processador). O próximo nível é o estágio C1 (Halt) onde o núcleo é essencialmente desativado, mas os caches e registradores continuam ativos. A partir daí, existem mais três estágios adicionais (C2, C3 e C4), onde mais componentes internos são desativados, reduzindo o consumo, mas aumentando o tempo de reativação. No estágio C1 o consumo do processador cai para apenas 4.5 watts e no estágio C4 para apenas 0.085 watts.

Apesar de ser um processador dual-core, o sistema de gerenciamento faz com que o consumo acabe sendo similar, ou pouca coisa superior ao dos Turions single-core, muito embora o desempenho seja superior.

Com os dois núcleos operando à 800 MHz, o Turion X2 (de 0.09 micron) consome 10 watts, o que dá apenas 5 watts por núcleo, menos que os 7 watts do Turion single-core. O principal motivo da redução é o uso de uma versão aprimorada do SOI (silicon on insulator), a técnica que permite adicionar uma fina camada de material isolante entre os transistores do processador e o wafer, reduzindo a perda de elétrons e conseqüentemente o consumo do processador. O SOI é utilizado também em outros processadores AMD produzidos usando a técnica de 0.09 micron em diante, mas neste caso é utilizada

uma revisão anterior, menos eficiente que a usada no Turion X2.

O Turion X2 adotou o uso do soquete S1, uma versão mobile do soquete AM2, onde também são utilizados módulos de memória DDR2, com suporte a dual-channel. Como nem todos os fabricantes utilizam dois módulos de memória (sobretudo nos modelos de baixo custo), você nota um certo ganho de desempenho ao instalar um segundo módulo, não apenas pelo aumento no volume de memória instalada, mas também pela ativação do dual-channel.

Um dos destaques do **soquete S1** é o número reduzido de pinos. Ele possui apenas 638 pinos, contra 940 pinos do soquete AM2 para desktops (menos até mesmo que o soquete 754, que é single-channel).



*Turion X2 instalado no soquete S1*

O número reduzido de pinos ajudou a reduzir o tamanho do encapsulamento do processador, que passou a ser um pouco menor que o dos processadores soquete 754. O soquete S1 tem uma aparência muito similar ao soquete M usado pelos processadores Intel.

O soquete S1 não é uma exclusividade do Turion X2. Ele é utilizado também pelo Turion single-core com core Richmond e também pelo Mobile Sempron baseado no core Keene. Uma forma muito simples de verificar se o seu notebook utiliza um destes processadores (e não um dos modelos antigos, soquete 754) é verificar o tipo de memória usada. Todos os notebooks soquete 754 utilizam memórias DDR, enquanto os soquete S1 utilizam memórias DDR2.

Na ilustração a seguir temos um Core Duo baseado no Yonah, um Core 2 Duo baseado no Meron e um Turion X2. Note que apesar de possuir menos cache, o die do Turion X2 é maior, pois ele é ainda produzido usando uma técnica de 0.09 micron.



*Core Duo (à esquerda), Core 2 Duo e o Turion X2*

Do ponto de vista do upgrade, as memórias DDR2 são muito mais desejáveis, não apenas pela questão do desempenho, mas porque (assim como nos desktops) os módulos SODIMM DDR2 já são mais baratos que os módulos DDR. Outra questão é que usando módulos DDR2 você vai, muito provavelmente, poder aproveitar os mesmos módulos ao trocar de notebook, enquanto as DDR são um beco sem saída.

Voltando ao processador, a safra inicial, lançada em 2006, inclui os Turions baseados nos cores **Taylor** e **Trinidad**. Ambos são fabricados utilizando a mesma técnica de fabricação de 0.09 micron, com SOI e o uso de transistores de baixo consumo usados nos Turions single-core da série ML. Eles incluem suporte a SS3, instruções de 64 bits, NX Bit e suporte ao AMD-V, a única diferença é que o Taylor inclui apenas 256 KB de cache L2 por núcleo, enquanto o Trinidad (usado na maior parte dos modelos) inclui 512 KB por núcleo.

A lista de modelos é composta pelo Turion X2 TL-50 (1.6 GHz, 2x 256 KB, 31W), TL-52 (1.6 GHz, 2x 512 KB, 31W), TL-56 (1.8 GHz, 2x 512 KB, 33W), TL-60 (2.0 GHz, 2x 512 KB, 35W) e TL-64 (2.2 GHz, 2x 512 KB, 35W).

Em 2007 foi lançado o Turion X2 baseado no core **Tyler**, produzido usando a nova técnica de 0.065 micron, usada também no Athlon X2 Brisbane e outros processadores recentes. Ele é usado nos modelos TL-56 (1.8 GHz, 2x 512 KB, 31W), TL-58 (1.9 GHz, 2x 512 KB, 31W), TL-60 (2.0 GHz, 2x 512 KB, 31W), TL-64 (2.2 GHz, 2x 512 KB, 35W) e TL-66 (2.3 GHz, 2x 512 KB, 35W).

Você pode notar que não existe diferenciação entre os modelos baseados no Tyler e no Trinidad. Com exceção do TL-56 e do TL-60 (que possuem um TDP ligeiramente inferior), até mesmo o

TDP dos processadores baseados nos dois cores é o mesmo. Na verdade, o Tyler é mais econômico que o Trinidad (embora a diferença não seja tão grande), o que a AMD passou a fazer foi simplesmente passar a usar uma margem mais folgada para cálculo do TDP. Lembre-se de que o TDP é simplesmente um teto de consumo que não deve ser excedido pelo processador. O fabricante pode usar o mesmo TDP para vários processadores, muito embora o consumo real de alguns deles seja mais baixo.

## Localizando defeitos

Diferentemente de um desktop, onde você pode solucionar problemas simplesmente na base da tentativa e erro, trocando peças até descobrir qual é o culpado, nos notebooks isso nem sempre é possível, já que a maior parte dos componentes são diretamente integrados à placa-mãe e você dificilmente terá um LCD ou outra fonte de alimentação compatível (por exemplo) para testar.

Na maior parte dos casos, você precisa identificar o problema e certificar-se de que o componente X precisa mesmo ser substituído, para só então encomendar a peça de reposição e poder fazer o conserto. Um erro de diagnóstico pode sair muito caro, levando-o à compra de uma nova placa-mãe quando

o defeito era em um dos cabos flat, por exemplo.

Vamos então a um guia rápido de como localizar defeitos ao dar manutenção em notebooks e, assim, descobrir qual componente precisa ser reparado ou substituído.



### Não liga

Se o notebook simplesmente não dá nenhum sinal de vida, a primeira coisa a se verificar é a fonte de alimentação.

Não confie que a fonte está funcionando só porque o led está aceso, use um multímetro para medir a tensão de saída da fonte, ajustando-o para medir a tensão em uma escala de 20V. Compare a medição do multímetro com os valores informados na etiqueta da fonte. A fonte deve fornecer a tensão de saída especificada na etiqueta, com uma margem de tolerância de 5% para mais ou para menos:



Além da possibilidade da fonte ter queimado devido a alguma intempérie, também é comum que fontes de baixa qualidade apresentem capacitores estufados e outros defeitos após alguns meses de uso. A fonte pode então passar a oferecer uma tensão cada vez mais baixa, até que o notebook simplesmente não ligue e não inicie a carga da bateria.

Na maioria dos casos, a solução mais barata é reparar a fonte. As fontes usadas em notebooks não são diferentes das usadas em monitores de LCD e outros periféricos, por isso um técnico especializado em manutenção de fontes pode resolver a maioria dos defeitos sem muita dificuldade. Ao optar por trocar a fonte, você pode tanto procurar outra fonte idêntica, do mesmo fabricante, ou comprar um fonte "curinga", onde são fornecidos diversos encaixes, de forma que a fonte possa ser usada em conjunto com um grande número de modelos:



Essas fontes são muito comuns no Ebay (cheque a categoria "Computers & Networking > Desktop & Laptop Accessories > Adapters, Chargers for Laptops"), embora a qualidade nem sempre seja das melhores. Em qualquer um dos casos, não se esqueça de verificar a tensão de saída (novamente usando o multímetro) antes de ligar a nova fonte no note.

Se a fonte está fornecendo a tensão correta, a próxima possibilidade é que as soldas do encaixe para o conector da fonte na placa-mãe do note estejam quebradas. Este defeito é muito comum nos Toshiba M35X, A65, A70, A75 e outros modelos, onde o conector é diretamente soldado na placa-mãe e fica preso apenas pelas soldas, sem um bom suporte na carcaça. Embora isso exija uma certa dose de habilidade manual, é sempre possível desmontar o note, remover a placa-mãe e refazer as soldas do conector.

Assim como nos desktops, problemas de mal contato e oxidação são muito comuns. Experimente sempre desconectar o HD, o drive óptico, a placa wireless e outros componentes não-essenciais e limpar os contatos dos módulos de memória. Se o notebook tiver dois slots de memória, experimente instalar o módulo no segundo slot. Se tiver dois módulos, experimente ligá-lo com apenas um e assim por diante. Outra observação é que alguns notebooks (sobretudo modelos antigos) não ligam se a bateria não estiver instalada.

## Instabilidade

Ainda mais comuns do que os casos em que o notebook simplesmente "morre" são os casos de instabilidade, onde o notebook trava, apresenta erros diversos ou simplesmente desliga sozinho de tempos em tempos, ou nos momentos de maior atividade.

Assim como em um desktop, problemas de estabilidade podem ser causados pelos mais diversos fatores, incluindo problemas de software, problemas de superaquecimento causado pelo acúmulo de pó nos dissipadores, defeitos nos módulos de memória, entre outras possibilidades, por isso é quase sempre necessário fazer uma checagem geral, verificando diversas possibilidades até finalmente conseguir descobrir a fonte do problema.

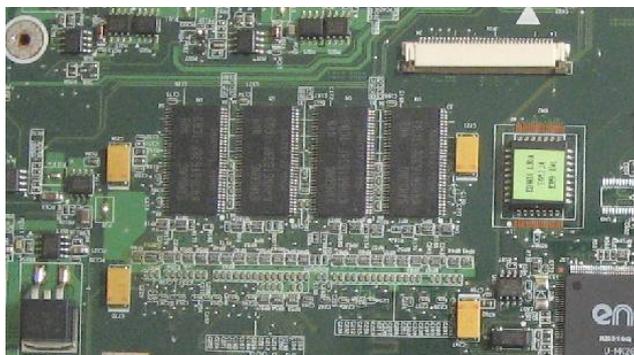
A primeira coisa a fazer é dar boot com uma distribuição Linux live-cd com que você tenha familiaridade. Use o sistema por algum tempo, execute algumas tarefas pesadas (como compactar e descompactar grandes quantidades de arquivos, por exemplo) e monitore as respostas do sistema. Se os problemas de estabilidade se manifestam apenas no Windows, muito provavelmente o problema se restringe aos softwares e drivers instalados e pode ser resolvido com uma simples reinstalação do sistema.

Se os problemas continuam mesmo depois de descartar o fator software, o próximo passo é fazer um teste completo da memória usando o memtest, já que a memória passa a ser a próxima suspeita. O mais comum é que o note possua dois slots de memória, um externo, acessível através das tampas inferiores, e outro interno, acessível depois de remover o teclado (como no HP6110). Nesses casos, você só precisa identificar qual dos módulos apresentou o defeito e substituí-lo. Se o defeito for nos últimos endereços, é possível também usar as dicas que vimos no capítulo sobre memória para limitar a quantidade de memória usada pelo sistema e, assim, evitar o uso da parte onde estão as células defeituosas.

Como a grande maioria dos notebooks utilizam memória compartilhada para o vídeo, defeitos na memória podem causar também o aparecimento de falhas na imagem, incluindo o aparecimento de linhas horizontais ou verticais.

Se o defeito se restringir à área de memória utilizada pelo chipset de vídeo (normalmente os primeiros endereços do módulo), o sistema pode funcionar de forma perfeitamente estável (com os problemas se restringindo ao vídeo), por isso é importante sempre checar a memória antes de colocar a culpa no LCD ou na controladora de vídeo.

Em casos em que os chips referentes ao módulo interno vêm soldados à placa-mãe do notebook, a situação fica mais complicada, já que você não tem como substituir os chips de memória diretamente:



A solução "correta" nesse caso seria substituir a placa-mãe. Algumas autorizadas possuem câmaras de vapor e são capazes de substituir os módulos, mas não é o tipo de coisa que você pode fazer usando um ferro de solda. Se o reparo não for possível e você chegar ao ponto de decidir descartar a placa, uma última solução desesperada que você pode tentar é remover os chips de memória (com muito cuidado, para evitar danificar outros componentes) e passar a usar um módulo instalado no slot de expansão. Se não houver nenhuma trava relacionada ao software, o BIOS vai detectar a remoção da memória integrada e passará a usar o módulo instalado no slot.

Se o notebook funciona de forma aparentemente normal por algum tempo, mas trava, reinicia ou desliga ao executar tarefas pesadas, muito provavelmente temos um problema de superaquecimento. A solução neste caso é remover o cooler do processador, fazer uma boa limpeza e substituir a pasta térmica do processador. Em alguns notes o cooler fica bem acessível através das tampas inferiores, mas em outros é preciso desmontar o note para chegar até ele.

Uma opção rápida para desobstruir o exaustor sem precisar desmontar o notebook é usar um jato de ar comprimido na saída de ar. O problema neste caso é que você apenas espalha a sujeira dentro do note, ao invés de removê-la. Isso faz com que o fluxo de ar gerado pelo cooler acabe movendo o pó novamente para a saída do cooler, fazendo com que o problema de superaquecimento reapareça mais rápido do que demoraria se você tivesse feito uma limpeza completa.

Além do acúmulo de sujeira nos dissipadores, é comum a entrada de pó dentro do próprio motor de rotação do cooler, o que causa o aparecimento de um ruído irritante e faz com que o exaustor gire cada vez mais devagar (ou até pare completamente). A solução nesse caso é desmontar o exaustor e fazer uma boa limpeza interna.

Na maioria dos coolers para notebook, o motor e a hélice são presos ao corpo principal apenas pelo conjunto de ímãs, de forma que basta puxar. Em alguns casos as duas partes são presas por uma presilha, escondida sob uma etiqueta.

Limpe bem as partes internas do motor, usando um cotonete embebido em álcool isopropílico e coloque um pouco (pouco!) de pó de grafite antes de fechar. Ele funciona como um lubrificante seco, que faz seu papel sem o risco de ressecar ou se misturar à sujeira com o tempo. O pó de grafite é usado para desemperrar fechaduras e pode ser encontrado facilmente em lojas de ferragens ou lojas de 1.99.



Assim como nos desktops, os HDs de notebook também apresentam defeitos mecânicos e muitas vezes precisam ser substituídos. Por sorte, os HDs são um componente padronizado, de forma que você pode substituir o drive em caso de defeito ou ao fazer upgrade sem muitas dificuldades. A principal cuidado ao comprar é verificar se o HD usa interface IDE ou SATA.

Em casos de perda de dados, os procedimentos de recuperação são os mesmos de um desktop.

Você pode remover o HD do notebook e plugá-lo em outro micro usando uma gaveta USB ou um adaptador para instalá-lo diretamente nas portas IDE ou SATA do desktop. É possível encontrar tanto adaptadores para drives de 2.5" IDE (velhos conhecidos de quem trabalha com manutenção) quanto adaptadores para os novos drives de 2.5" SATA. Por serem relativamente raros, estes adaptadores podem custar muitas vezes R\$ 50 ou mais em lojas do Brasil, mas são muito baratos se comprados no Ebay.

Outra opção é dar boot no próprio notebook, usando uma distribuição Linux live-CD, e copiar os dados para um compartilhamento de rede, ou um HD externo. Isso permite acessar os dados e fazer a recuperação muito mais facilmente, sem precisar abrir o note e sem precisar de um segundo micro. Carregando um CD de boot e um HD externo (ou pelo menos um pendrive), você poderia recuperar os dados diretamente no local, sem precisar de ferramentas extras.

Em casos em que os dados foram apagados e você precisa usar um programa de recuperação, como o Easy Recovery ou o PC Inspector para recuperá-los, um opção é fazer uma imagem do HD usando o dd (como vimos no tópico de recuperação de dados do capítulo sobre HDs) a partir do próprio live-CD (salvando a imagem em um HD externo),

restaurar a imagem em outro HD, instalado em um desktop (obtendo, assim, um clone do HD original) e rodar o programa de recuperação no HD clonado. Dessa forma, você não corre o risco de piorar as coisas manipulando os dados salvos no HD original.

Assim como no caso dos HDs, os drives ópticos são padronizados e podem ser substituídos, inclusive usando um drive removido de outro notebook. Em muitos casos, os problemas de leitura podem ser causados pelo acúmulo de sujeira na lente ou no mecanismo de leitura do drive. Nesse caso, uma boa limpeza e lubrificação pode resolver.



Defeitos na tela são possivelmente os mais comuns. Como vimos, problemas de "tela apagada" são quase sempre relacionados ao FL inverter e podem ser reparados com a substituição deste.

Os casos de defeitos relacionados às lâmpadas de catodo frio são mais raros, mas a possibilidade nunca deve ser descartada, sobretudo em notebooks com mais tempo de uso. As lâmpadas de catodo frio são um componente do LCD, mas podem ser substituídas separadamente, caso você encontre peças de reposição.

É preciso extremo cuidado ao substituí-las, pois elas são bastante frágeis (imagine uma lâmpada fluorescente comprida e muito fina).



*Lâmpada de cátodo frio testada fora do LCD*

Em notebooks que utilizam LEDs para a iluminação da tela, os problemas são muito mais raros, já que eles não utilizam o FL inverter (os LEDs utilizam tensão de 5V ou 3.3V, fornecida diretamente pelos circuitos reguladores da placa-mãe) e são utilizados um número muito grande de LEDs, de forma que a queima de um ou alguns deles não causa uma perda considerável de luminosidade.

Em alguns casos mais raros, o problema pode ser com os circuitos de alimentação na placa-mãe (causado por capacitores estufados, por exemplo). Esses defeitos são relativamente raros nos notebooks atuais, pois os fabricantes têm adotado o uso de capacitores de estado sólido, que são muito mais duráveis que os capacitores eletrolíticos, como você pode ver na foto a seguir:



Cada pixel do LCD é formado por um conjunto de três transistores, um para cada cor. Quando alguns dos transistores passam a queimar, seja qual for o motivo, a tela passa a apresentar um número crescente de bad-pixels. Eles são relativamente benignos, já que não impedem o uso da tela, apenas têm um aspecto visual desagradável.

O aparecimento de bad-pixels causados pelo "envelhecimento" da tela era comum em notebooks antigos, mas as telas atuais possuem uma qualidade de fabricação muito superior e por isso eles se tornaram um problema mais raro. Em um notebook atual, se a tela não vem com bad-pixels de fábrica, é muito difícil que eles apareçam posteriormente.

Um problema mais grave é a corrupção da tela, causada por defeitos na placa controladora. Nesse caso, o LCD pode apresentar desde linhas horizontais ou verticais ou defeitos de atualização em imagens em movimento, até um borrão completo em vez da imagem:



Você pode diferenciar os defeitos no LCD de defeitos causados pela memória ou defeitos no controlador de vídeo (que faz parte da placa-mãe) usando um monitor externo. Na maioria dos notebooks, você precisa pressionar uma combinação de teclas, como Fn+F4, Fn+F5, Fn+F7 ou Fn+F8, para ativar o monitor externo. Se a imagem aparece corrompida no LCD, mas fica perfeita no monitor externo, então o problema é ou nos cabos ou no próprio LCD.

Esse tipo de defeito é causado por defeitos na placa controladora incluída na tela, e não pelo LCD em si. Em muitos casos a controladora é presa à tela por um cabo flat e pode ser substituída, mas em muitos ela é soldada, de forma que só é possível trocá-la em conjunto com o LCD. Também não é comum encontrar apenas a controladora à venda, de forma que você acaba tendo que substituir todo o LCD.



*LCD desmontado, mostrando a placa controladora*

Nem sempre vale a pena substituir o LCD em caso de defeito, pois nos notebooks mais baratos, a tela custa mais de metade do valor de um note novo. Em muitos casos vale mais à pena comprar outro note e passar a usar o antigo como desktop (usando um monitor, teclado e mouse externos), ou desmontá-lo e vender as peças.

Se, por outro lado, a imagem aparece corrompida tanto no LCD quanto no monitor externo, então o problema está em outro lugar e não adianta perder tempo trocando o LCD. Se o problema estiver na memória, você pode solucioná-lo usando as dicas anteriores. Se, por outro lado, o defeito for no controlador de vídeo, a situação fica mais complicada, já que ele é integrado à placa-mãe ou ao próprio chipset, lhe obrigando a trocar toda a placa.

Muitos notebooks utilizam placas de vídeo dedicadas. Além das placas MXM e AXION, existem ainda formatos proprietários, usados em modelos específicos, como nesta placa com chipset ATI usada em um Acer Aspire 1350



*Placa proprietária usada em um Acer Aspire 1350 (à esquerda) e uma placa alternativa para o mesmo modelo, baseada em um chipset diferente*

Nesses casos, não apenas o chipset de vídeo, mas também a memória fazem parte de um módulo separado, que pode ser substituído em caso de problemas. Os módulos MXM e AXION são relativamente bem padronizados, de forma que você pode encontrar placas de reposição em lojas online diversas e nos sites de leilão, mas os módulos proprietários são componentes incomuns, que geralmente são vendidos apenas pelo próprio fabricante.

Uma observação é que em alguns notebooks determinadas configurações de resolução e refresh podem fazer a tela ficar fora de sincronismo, assim como em um monitor para micros de mesa. Se o LCD exibe corretamente a tela do setup, mas a imagem fica embaralhada depois do carregamento do sistema, muito provavelmente é este o caso. Lembre-se também de que a regra do mal contato também vale para notebooks. Antes de descartar qualquer componente, experimente sempre limpar os contatos, desencaixar e re-encaixar os conectores e assim por diante.

## Modem e placa wireless

Assim como nos desktops, os modems discados incluídos nos notebooks podem se queimar ao receber uma descarga através da linha telefônica. O único motivo disso ser menos comum nos notebooks é que eles não costumam ficar o tempo todo ligados na linha telefônica como é o caso de muitos desktops. Na maioria dos notebooks, o modem acaba nem sendo usado.

De qualquer forma, em caso de queima do modem, quase sempre o dano se restringe à placa MDC, que contém os componentes analógicos do modem.

Você pode substituir a placa por outra retirada de um note similar.



*Placa MDC com os componentes analógicos do modem*

Ao contrário do modem, é muito raro que a placa wireless apresente qualquer defeito, já que ela não é vulnerável a descargas externas como ele. A possibilidade da placa wireless se queimar ou apresentar defeito não é maior do que a do chipset da placa-mãe, por exemplo.

A maioria dos "defeitos" relacionados à placa wireless são relacionados a problemas de configuração. Por exemplo, na maioria dos notebooks, o botão que ativa e desativa o transmissor da placa wireless é controlado através de funções do ACPI, o que faz com que ele só funcione depois de instalar o driver ou o utilitário correspondente do fabricante.

Na maioria dos notebooks da Acer, por exemplo, você precisa instalar (além do driver) o "Launch Manager", caso

contrário você não consegue ativar o transmissor da placa Wireless no Windows. No Linux essa função é desempenhada por um módulo de Kernel que, nas distribuições recentes, vem pré-instalado no sistema. O processo manual seria carregar o módulo "acer\_acpi", usando o comando "modprobe acer\_acpi" e em seguida ativar o transmissor usando o comando "echo "enabled : 1" > /proc/acpi/acer/wireless". Esse é o tipo de coisa que pode ser feita automaticamente pelo sistema durante a fase de detecção do hardware e muitas distribuições realmente o fazem de forma automática.



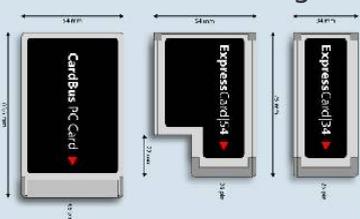
*Placa wireless ExpressCard (à direita), instalada em um notebook HP*

Além da questão dos drivers, temos os problemas normais relacionados à conexão com a rede wireless. O sinal pode estar sendo atenuado por paredes, lajes ou interferências presentes

no ambiente (como aparelhos de micro-ondas); ou o ponto de acesso pode ter sido configurado para não divulgar o ESSID, o que faz com que a rede não apareça no utilitário de conexão, até que você tente se conectar a ela manualmente, especificando o nome da rede.

Outra questão comum é que pontos de acesso 802.11g ou 802.11n podem ser configurados para não aceitar a conexão de placas de padrões anteriores, deixando de fora notebooks com placas 802.11b, por exemplo. Nesse caso não existe muito o que fazer além de mudar a configuração do ponto de acesso ou atualizar a placa do notebook.

Além da possibilidade de instalar uma placa PC-Card ou ExpressCard, é perfeitamente possível atualizar a placa Mini-PCI ou Express Mini do notebook caso desejado. Ambos são barramentos padronizados, de forma que do ponto de vista do hardware a alteração é perfeitamente normal. Apesar disso, existem casos de incompatibilidades entre novas placas e o BIOS. Nesse caso você recebe uma mensagem "Unsupported Card Detected" (ou similar) ao ligar o note e precisa ir atrás de uma atualização de BIOS ou de outra placa que seja compatível com o BIOS original.



## Comprando peças de reposição no exterior

Trocar o FL Inverter é uma tarefa simples, basta remover qualquer parafuso que o prenda à carcaça e soltar os dois conectores. O maior problema é como encontrar o substituto. O mesmo se aplica a telas, fontes de alimentação, cabos flat, teclados, baterias e outras peças de reposição.

O FL inverter, por exemplo, é uma peça relativamente barata, que custa de US\$ 60 a US\$ 120, dependendo do modelo. Aqui no Brasil, os preços variam muito, de acordo com onde pesquisar. É possível também encontrar alguns com bons preços nos sites de leilão.

Se você tem um cartão de crédito internacional, a melhor opção é comprar diretamente no exterior. Pesquisando pelo part number (que quase sempre vem decalcado no componente ou impresso em um adesivo) ou pelo modelo do note, você encontra diversas lojas que vendem peças de reposição. Alguns exemplos são:

<http://www.screenetekinc.com>

(página em português: <http://www.screenetekinc.com/telas-lcd.shtml>)

<http://www.sparepartwarehouse.com/>

<http://www.impactcomputers.com>

<http://www.laptoprepairco.com>

<http://www.crucial.com> (para módulos de memória difíceis de encontrar)

Outra opção é pesquisar no Ebay, onde você também encontra componentes usados a preços mais baixos: <http://ebay.com>. Ao pesquisar no Ebay, faça uma busca por "inverter" e a marca do notebook e procure por algum compatível com o modelo entre os anúncios. Não especifique diretamente o modelo, pois assim você reduz muito o número de resultados. Nem todo mundo especifica diretamente o modelo exato no título.

Na parte superior da lista, clique no link "Shipping to USA" e mude para "Brazil - BRA" no menu. Assim você pode ver diretamente os vendedores que oferecem a opção de envio para o Brasil, sem precisar ficar abrindo cada um dos anúncios. Em muitos casos, o vendedor pode não ter especificado o envio para o Brasil, mas aceitar enviar caso contactado diretamente. Outros especificam diretamente que enviam apenas para os EUA ou outros locais específicos.

List View | [Picture Gallery](#) Sort by: Time: ending soonest [Customize Display](#)

	Item Title	PayPal	Bids	Price*	Shipping to BRA	Time Left ▲
<input type="checkbox"/>	<a href="#">NEW HP LCD Inverter AS023172018 AS023172069 AS023172140</a>		- Buy It Now	\$34.99 \$34.99	\$14.99	3h 09m
<input type="checkbox"/>	<a href="#">NEW HP LCD Inverter AS023172018 AS023172069 AS023172140</a>		- Buy It Now	\$34.99 \$34.99	\$14.99	3h 15m
<input type="checkbox"/>	<a href="#">HP DV1000 DV1300 LCD INVERTER AS023172328 T18I077.03</a>		- Buy It Now	\$34.98	Not specified	3h 22m
<input type="checkbox"/>	<a href="#">NEW LCD Inverter HP ZT3000 ZX5000 ZV5000 &amp; COMPAQ X1000</a>		- Buy It Now	\$17.99 \$17.99	\$28.00	3h 52m
<input type="checkbox"/>	<a href="#">BRAND NEW LCD INVERTER F2330-60925 FOR HP XH N5000</a>		- Buy It Now	\$17.99 \$17.99	\$28.00	3h 52m
<input type="checkbox"/>	<a href="#">HP ZD7000 17" LCD INVERTER AS023172123 PWB-IV10150T/E3</a>		- Buy It Now	\$34.00 \$48.00	\$8.00	6h 32m
<input type="checkbox"/>	<a href="#">HP Pavilion ZV5000 ZT3000 Inverter Board PK070015900</a>		1	\$9.99	Not specified	6h 56m

Comprar peças no exterior usando o cartão de crédito é relativamente simples. Também é interessante ter uma conta no PayPal, o sistema de pagamento usado no Ebay, que é aceito por um grande número de lojas. Ele permite que você faça o pagamento sem precisar fornecer o número do cartão e possui alguns sistemas de proteção contra fraudes.

Com relação ao envio, você pode optar pelo envio via correios (USPS), Fedex ou outro sistema de envio expresso, como o UPS (note que UPS é diferente de USPS). Ao chegar no Brasil, o pacote passa pela receita, que vai decidir se os impostos se aplicam de acordo com o valor e o tipo de componente.

Por estranho que possa parecer, os impostos são calculados com base no valor total da compra, incluindo os produtos e o frete. Além dos 60% de impostos, você paga também uma taxa de ICMS (a alíquota varia de acordo com o estado; em São Paulo, por exemplo, é de 21%) e mais uma "taxa aduaneira", de pouco mais de 20 reais.

As opções de envio expresso internacional da UPS e da Fedex são muito caras, você acaba pagando 40 dólares ou mais mesmo para itens pequenos. Como além do custo do envio você vai pagar também impostos sobre ele, você acaba pagando, na prática, quase o dobro desse valor.

A melhor opção de envio para itens pequenos é o "USPS First Class Mail International", uma modalidade de envio prioritário oferecida pelos correios dos EUA. Nesta modalidade, um pacote vindo dos EUA demora de 7 a 10 dias corridos para chegar. Outra opção é o "USPS Priority Mail International", que é um pouco mais rápido, chegando, muitas vezes, em 5 dias. Pacotes vindos de outros países costumam demorar, em média, duas semanas, principalmente os vindos de países da ásia.

Normalmente os vendedores cobram uma taxa única de "Shipping and Handling" (envio e manuseio), que inclui o custo do envio e qualquer outra taxa que ele queira cobrar. É por isso que alguns podem cobrar 6 dólares e outros 20 pela mesma modalidade de envio. No mercado americano isso é considerado normal.

Existe a velha questão da isenção de compras de 50 dólares. Esta regra se aplica apenas para envios de pessoa física para pessoa física, para fins pessoais.

Se algum amigo dos EUA lhe mandar um mouse de presente, colocando dentro uma carta lhe dando os parabéns e o valor do mouse, somado com o custo do envio, não ultrapassar os 50 dólares, muito provavelmente o pacote será enquadrado na regra e você não pagará impostos. Mas, compras em geral não se enquadram nela.

A exceção fica por conta dos livros, que são isentos na maior parte dos países do mundo. Compras em que o envio é feito de forma eletrônica, como no caso de um software ou um e-book, também são isentas.

No caso dos serviços de entrega expressa (UPS, Fedex, etc.) é comum que a empresa pague as taxas de importação diretamente, para acelerar a liberação do pacote na alfândega e o entregador lhe apresente o comprovante e lhe cobre o valor ao entregar. Em alguns casos, o pagamento é feito através de um boleto entregue junto com o pacote.

Uma observação é que optando pelo envio via UPS ou Fedex você paga impostos praticamente sempre, mesmo no caso de presentes, pois o valor do envio sozinho já dá quase os 50 dólares permitidos. Você pode chegar então a casos extremos onde paga US\$ 45 de envio, mais R\$ 80 de impostos por um item de US\$ 6.

No caso dos pacotes enviados pelo correio (USPS ou similar), você recebe um aviso dos correios avisando da chegada do pacote, do endereço da agência (dos correios) onde ele está disponível e também dos valores a pagar. Você paga os impostos na própria agência, ao retirar o pacote. A principal vantagem é que o envio neste caso é muito mais barato, então você acaba pagando impostos praticamente apenas sobre o valor da mercadoria propriamente dita.

Pacotes que são considerados isentos são entregues diretamente, como qualquer outra correspondência. Em muitos casos você verá uma fita escrito "aberto pela aduana do Brasil", indicando justamente que o pacote foi aberto e o conteúdo conferido.

O valor da mercadoria é informado na "Customs Declaration", uma etiqueta padrão colada na parte externa do envelope. No caso de compras feitas em lojas, vale o valor da nota fiscal. Em casos onde o valor declarado é mais baixo que o real (o velho truque de declarar um valor menor para pagar menos imposto, ou para que o pacote se enquadre na regra dos 50 dólares), os impostos podem ser calculados usando uma tabela interna.

Lojas e empresas de informática pagam os mesmos impostos ao trazer produtos do exterior. Geralmente economizam no frete, por comprar vários

itens de cada vez, mas é só. Como qualquer empresa saudável precisa vender produtos com lucro, os preços acabam sendo quase sempre mais altos do que você pagaria ao comprar diretamente, incluindo todos os impostos. É muito comum que um FL inverter de US\$ 50 seja vendido por R\$ 300 ou uma bateria de US\$ 60 seja vendida por R\$ 400, por exemplo.

Comprar no exterior geralmente não é vantajoso ao comprar produtos comuns, como um processador ou um módulo de memória, por exemplo, ou no caso de itens pesados, como gabinetes e monitores, onde o custo do envio é muito alto.

Mas, no caso de itens raros, como peças de reposição para notebooks, baterias, cabos, adaptadores incomuns, etc. é muito mais prático e barato pesquisar diretamente e comprar no exterior. Recomendo fortemente que você pelo menos faça um teste, comprando algum item barato ou um livro.

#### 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.

# Configurando a rede no Windows

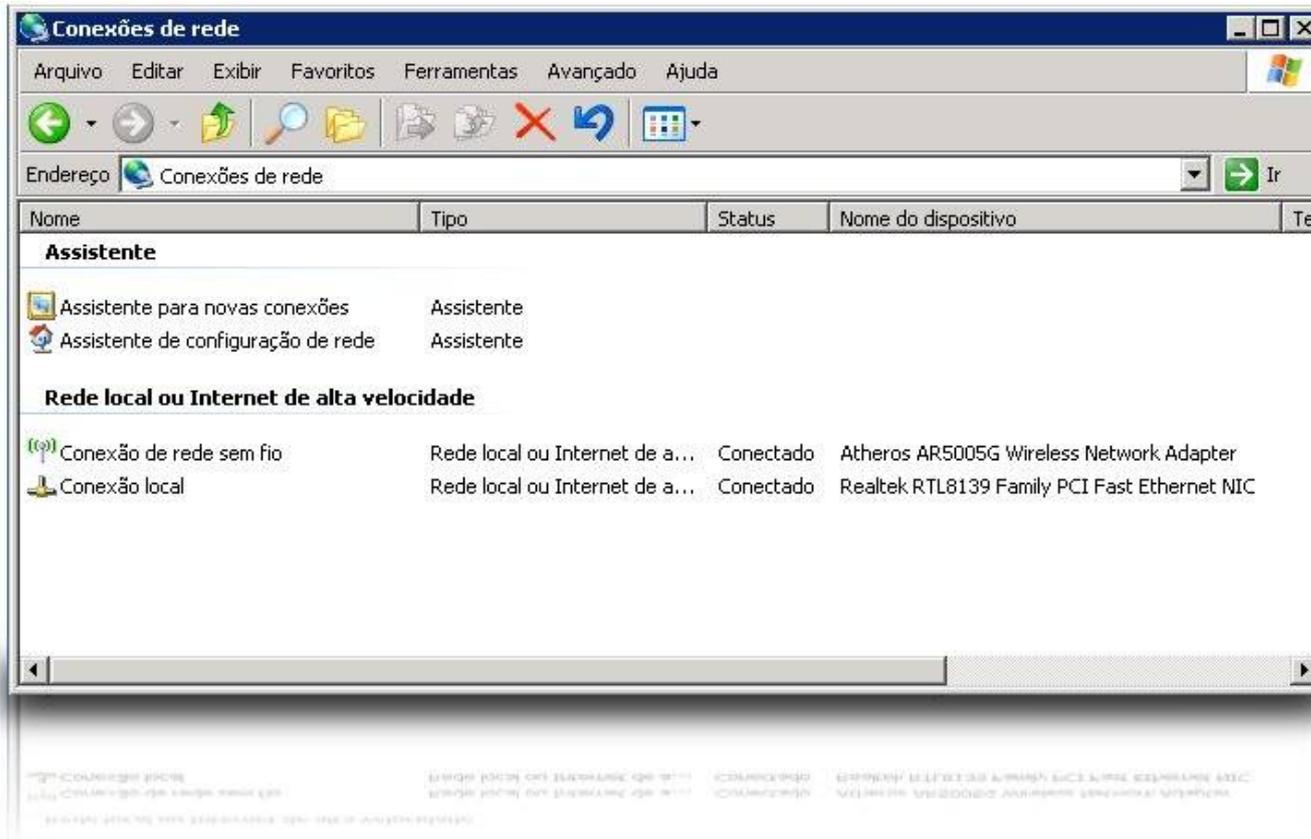
por **Carlos E. Morimoto**

*A configuração de rede no Windows é um assunto bastante conhecido, pois a configuração é bastante similar entre as diferentes versões do Windows e a configuração gráfica torna tudo mais simples. Entretanto, é justamente a aparente simplicidade que faz com que muitos recursos passem despercebidos. Vamos então a uma revisão de como configurar a rede nos clientes Windows, incluindo alguns tópicos avançados e o uso da linha de comando.*



No Windows, a configuração de rede vai dentro do "Painel de Controle > Conexões de rede", onde são listadas todas as interfaces de rede disponíveis. A configuração de rede no Windows é um assunto bastante conhecido, pois a configuração é bastante similar entre as diferentes versões do Windows e a configuração gráfica torna tudo mais simples. Entretanto, é justamente a aparente simplicidade que faz com que muitos recursos passem despercebidos. Vamos então a uma revisão de como configurar a rede nos clientes Windows.

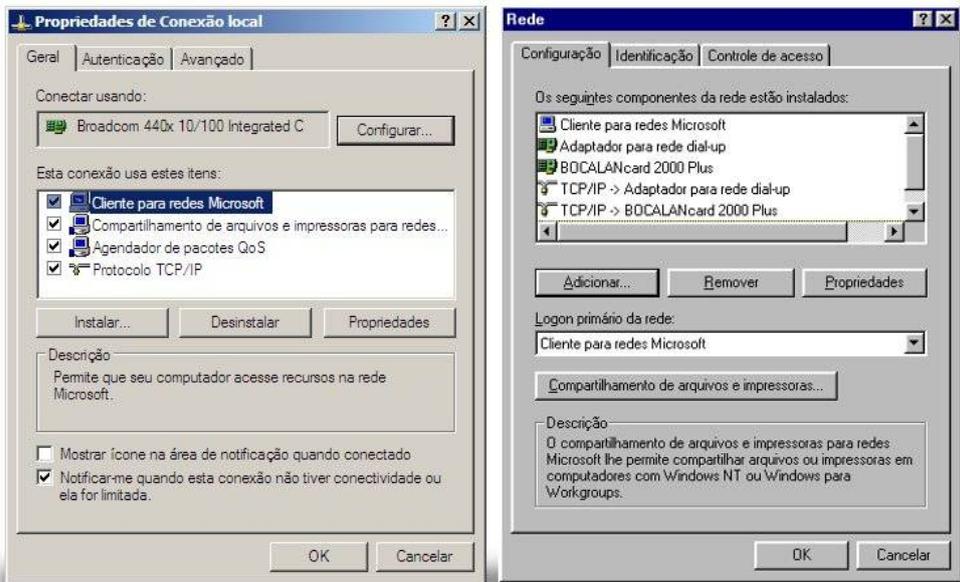
Começando do básico, a configuração de rede é dividida em duas partes. A principal vai no "Painel de Controle > Conexões de Rede" ("Conexões dial-up e de rede" no Windows 2000), onde ficam agrupadas as configurações de todas as interfaces de rede disponíveis, incluindo os modems e placas wireless:



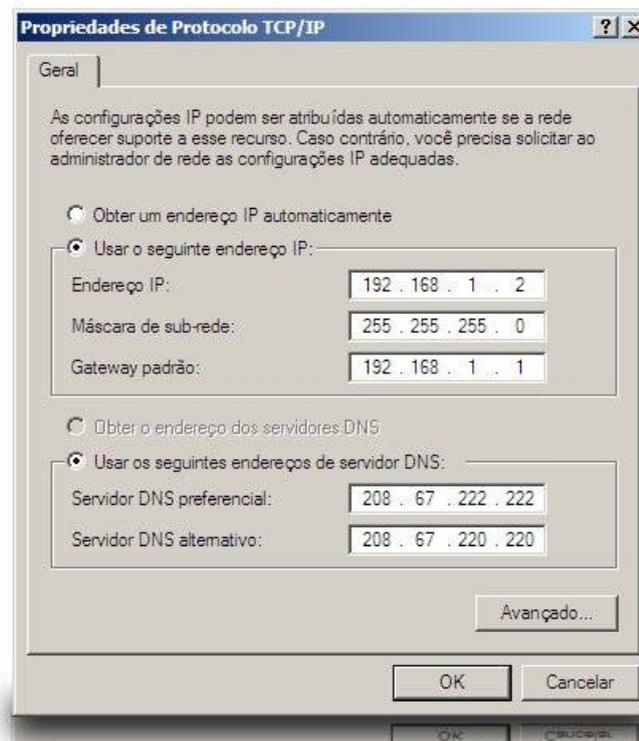
Dentro das propriedades de cada interface, vai uma lista dos protocolos e serviços disponíveis. O absoluto mínimo é o suporte a TCP/IP, que é instalado por padrão. Em seguida temos o "Cliente para redes Microsoft", que permite que você acesse compartimentos de rede em outras máquinas e o "Compartilhamento de arquivos e impressoras para redes Microsoft", que é o componente servidor, que permite que você compartilhe arquivos e impressoras com outras máquinas da rede.

O compartilhamento de arquivos e impressoras no Windows é baseado no protocolo SMB/CIFS, que é suportado no Linux através do Samba. Graças a ele, máquinas Linux podem participar da rede, tanto atuando como servidores, quanto como clientes.

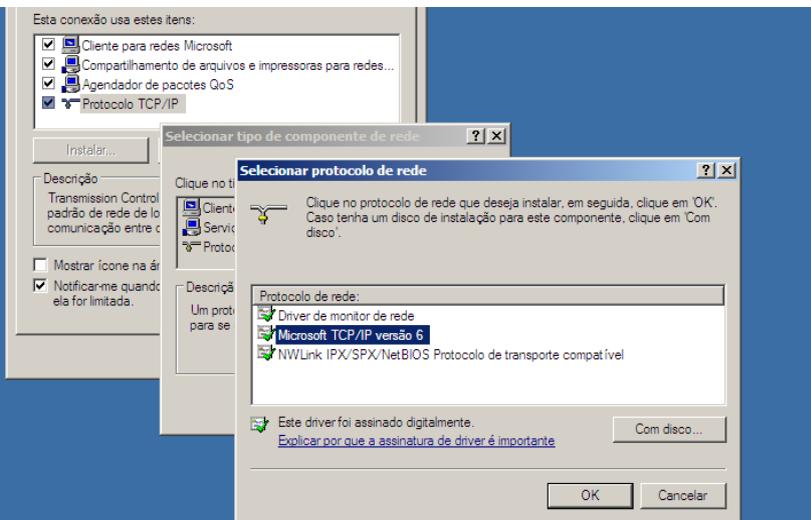
Até o Windows 98 a configuração da rede era baseado em um conceito mais confuso, onde os protocolos eram relacionados às interfaces e tudo era mostrado em uma única lista. Você podia então ter listadas três versões do TCP/IP, sendo uma para cada uma das placas de rede instaladas e outra para o modem, por exemplo. Aqui temos dois screenshots, com a configuração de rede no Windows 95/98 (à direita) em comparação com a interface atual, que foi adotada apenas a partir do Windows 2000:



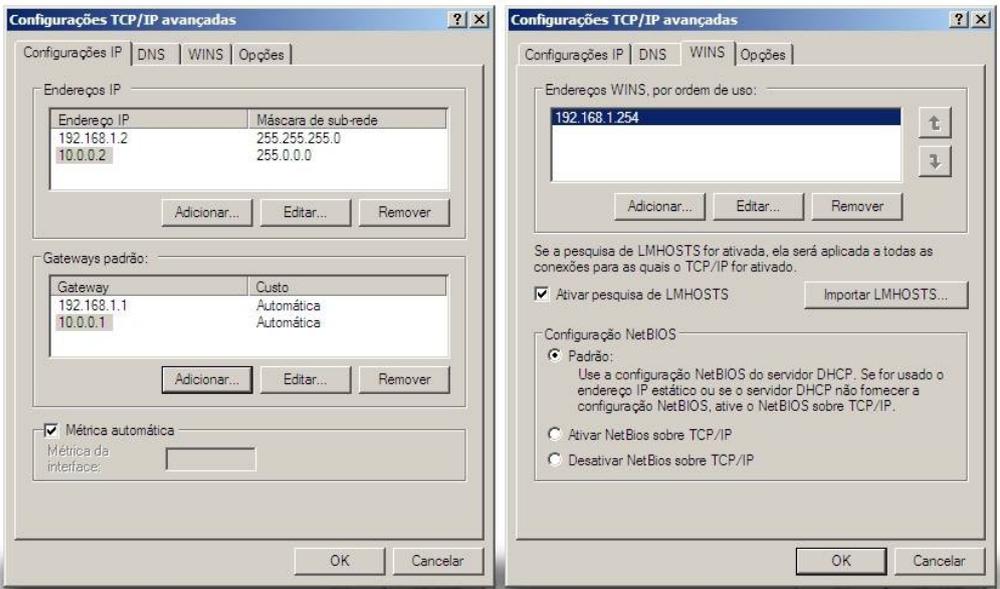
Voltando às propriedades da conexão, a configuração da rede vai dentro das propriedades do protocolo TCP/IP, onde você pode escolher entre ativar o cliente DHCP ou configurar manualmente os endereços. O segundo servidor DNS é desajável pela questão da redundância, mas não é obrigatório dentro da configuração. Além de usar os endereços de DNS do provedor, você pode usar um servidor de DNS instalado em uma máquina da rede local ou ainda um servidor de DNS público, como os do [opendns.com](http://opendns.com):



Você pode adicionar componentes de rede adicionais através do botão "Instalar", como em casos em que você precisar conectar a máquina Windows a uma rede Netware antiga, baseada no protocolo IPX/SPX, ou em que você precise adicionar o suporte ao protocolo IPV6 em uma máquina Windows XP:



Clicando no botão "Avançado" você tem acesso a algumas opções adicionais. À primeira vista, a aba "Configurações IP" parece inútil, já que ela simplesmente mostra o endereço IP e o endereço do gateway padrão definidos na tela principal, mas na verdade ela tem uma função importante, que é permitir que você configure endereços IP e gateway adicionais, como no meu exemplo, onde incluí o endereço IP "10.0.0.2":



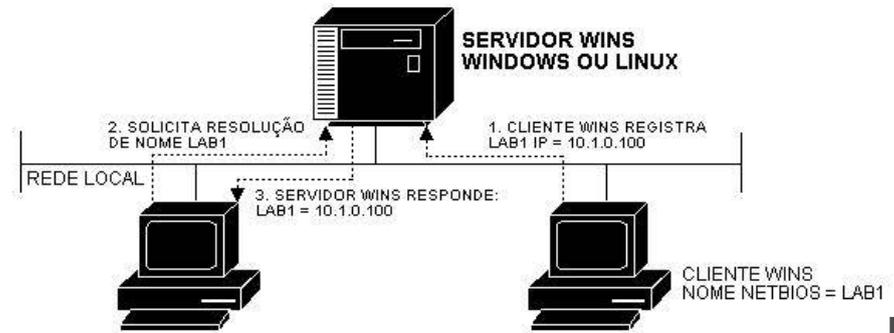
Isso faz com que o sistema crie um alias (um apelido) para a placa de rede e passe a escutar nos dois endereços. Se você configurar outro micro da rede para usar outro endereço dentro da mesma faixa, você conseguiria acessar o micro através do endereço "10.0.0.2" adicionado, da mesma forma que através do endereço principal.

Existem diversos possíveis usos para este recurso. Um exemplo comum é em casos em que você precisa acessar a interface de administração de um ponto de acesso ou um modem

ADSL que utiliza por padrão um endereço fora da faixa utilizada na sua rede, como por exemplo o "10.0.0.138". Uma forma simples de resolver o problema é adicionar um segundo endereço IP na mesma faixa de rede usada pelo modem, de forma que você possa fazer o acesso inicial à interface de configuração e alterar o endereço do dispositivo.

Continuando, temos a aba "WINS", que está relacionada à navegação na rede local. O WINS (Windows Internetworking Name Server) é um serviço auxiliar dentro das redes Microsoft, responsável pela navegação no ambiente de rede. A função de servidor WINS pode ser desempenhada tanto por um servidor Microsoft quanto por um servidor Linux rodando o Samba.

O uso do servidor WINS não é obrigatório, pois na ausência dele os clientes da rede simplesmente passam a utilizar pacotes de broadcast para localizarem os compartilhamentos, mas o uso de um servidor WINS torna a navegação no ambiente de redes mais estável. Usar um servidor WINS é também uma solução em casos em que você precisa combinar dois segmentos de rede ligados através de um roteador, já que o roteador descarta os pacotes de broadcast, fazendo com que os clientes de um segmento da rede não consigam enxergar os do outro, até que seja adicionado um servidor WINS central. Isso acontece por que as consultas ao servidor WINS são feitas diretamente ao endereço IP do servidor (e não através de pacotes de broadcast), o que permite que os pacotes passem pelo roteador.



Logo abaixo temos o campo "Configuração NetBIOS", que permite ativar ou desativar o suporte ao NetBIOS, que é o protocolo responsável pela navegação no ambiente de redes (o WINS é um serviço que roda sobre o NetBIOS).

A partir do Windows 2000 passou a ser usado o CIFS, uma versão atualizada do antigo protocolo SMB, usado nas versões anteriores do Windows, que funciona de forma independente do NetBIOS. Entretanto, o suporte ao NetBIOS ainda é necessário para diversas funções, entre elas a navegação no ambiente de redes (sem ele você consegue apenas mapear os compartilhamentos diretamente) e para que os clientes Windows sejam capazes de fazer login em um servidor Samba configurado como PDC. O NetBIOS também é necessário caso você tenha máquinas com versões antigas do Windows na rede, já que as versões anteriores ao Windows 2000 ainda não oferecem suporte ao CIFS.

Na última aba, "Opções", você tem acesso ao recurso de filtragem de TCP/IP. À primeira vista, pode parecer tratar-se de um firewall simples, onde você pode especificar manualmente quais portas devem ficar abertas e fechar as demais, mas não é bem assim que ele funciona.

Você pode fazer um teste permitindo apenas a porta 80 TCP (http) e 53 UDP (dns). Normalmente isso seja suficiente para que você conseguisse navegar, mas nesse caso você vai perceber que o acesso é bloqueado completamente. Isso acontece por que a porta 80 é usada apenas para iniciar a conexão, a resposta do servidor remoto chega em uma porta alta aleatória, que nesse caso é bloqueada pelo sistema, impedindo que você navegue.

Diferente de um firewall, que permite bloquear conexões de entrada, mas permite o retorno de pacotes de resposta, este recurso simplesmente bloqueia todas as conexões fora das portas especificadas, o que faz com que ele seja útil apenas em situações bem específicas.



## QoS

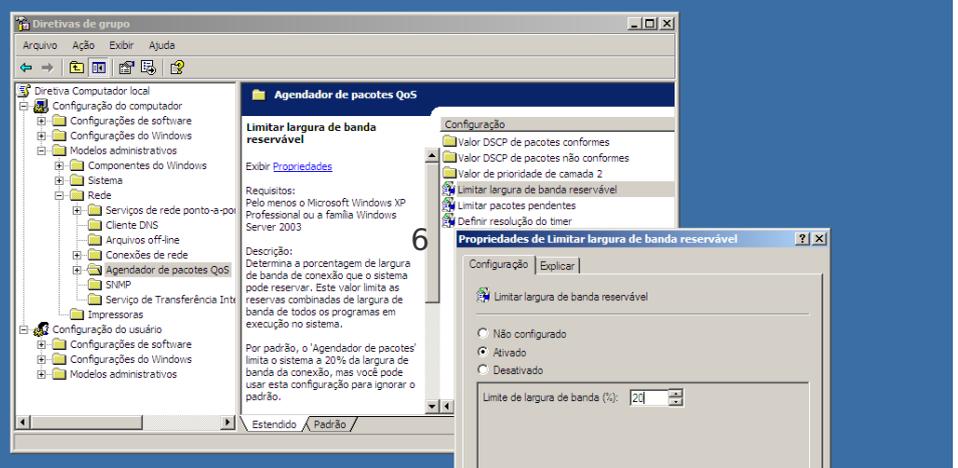
**Um recuso polêmico introduzido no Windows XP é o uso do "Agendador de pacotes QoS", destinado a otimizar a conexão. Ele aparece junto com os outros protocolos instalados nas propriedades da conexão e pode ser desativado caso desejado:**



A função do QoS é otimizar o tráfego da rede, priorizando o tráfego de aplicações que precisam de um tráfego constante, como aplicativos de VoIP, streaming de áudio e vídeo e jogos multiplayer. O uso do QoS oferece ganhos práticos em diversas áreas (reduz a necessidade de

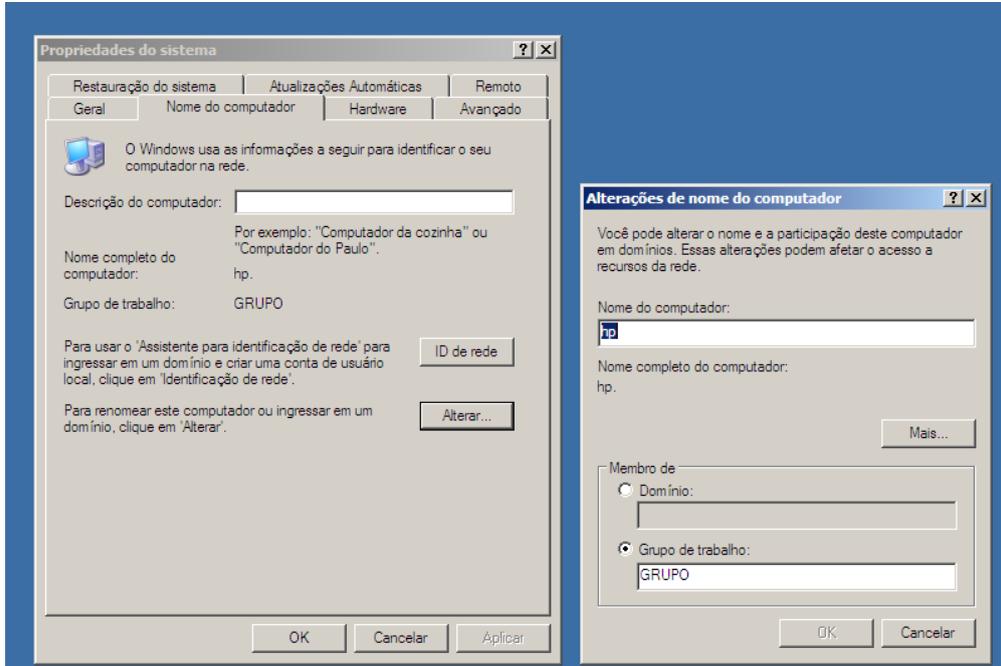
manter a conexão livre enquanto está falando no Skype, por exemplo), mas causa uma certa redução na banda total disponível, já que o sistema precisa manter uma determinada parcela da banda reservada.

Não existe configuração para o QoS disponível nas propriedades da conexão, mas ele pode ser configurado através do "gpedit.msc", que você chama usando o prompt do DOS ou o "Iniciar > Executar". Dentro do gpedit, a configuração do QoS vai no "Configuração do computador > Modelos administrativos > Rede > Agendador de pacotes QoS":



O default é reservar até 20% da banda. Este não é um valor fixo, mas apenas um limite; a percentagem realmente reservada varia de acordo com o uso. De qualquer forma, esta opção permite reduzir o valor, de forma a reduzir a perda, sem com isso precisar desativar o QoS completamente.

A pouco comentei que a configuração de rede é feita em duas partes. A segunda parte seria a configuração do nome da máquina e do grupo de trabalho ou domínio, que é feita através do Painel de Controle > Sistema > Nome do computador > Alterar:



O conceito de "Grupo de trabalho" existe desde o Windows 3.11, permitindo que um grupo de máquinas compartilhem arquivos e impressoras, sem necessidade de um servidor central. Configurando todos os micros da sua rede com o mesmo grupo de trabalho, todos aparecem no ambiente de redes e você não tem dificuldades em acessar os compartilhamentos.

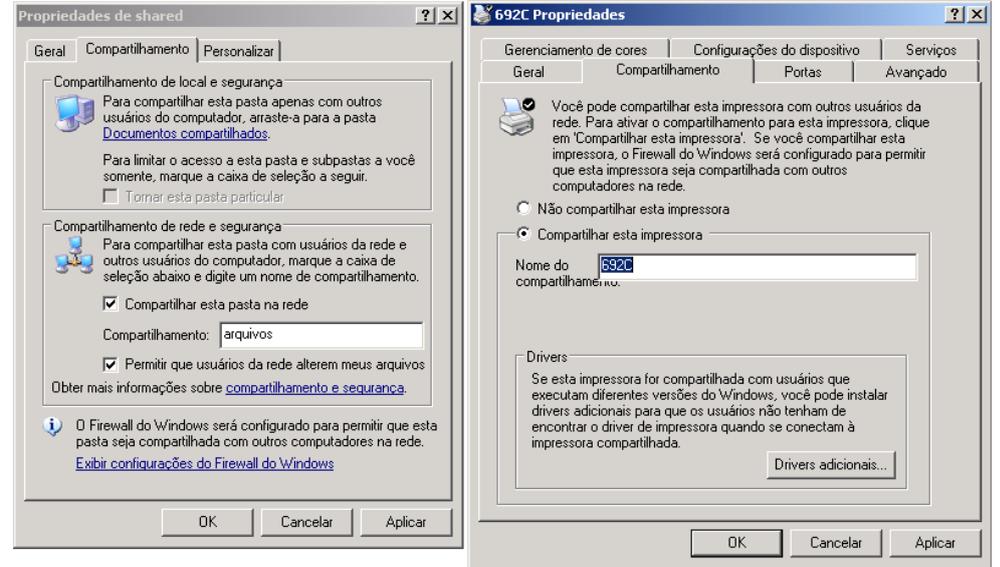
Em seguida temos o conceito de "Domínio", que se baseia no uso de um servidor central (chamado de Primary Domain Controller, ou PDC), que centraliza a autenticação dos usuários. A principal vantagem de utilizar um PDC é que você não precisa mais se preocupar em manter os logins e senhas das estações sincronizados em relação aos dos servidores. As senhas passam a ser armazenadas em um servidor central, de forma que qualquer alteração é automaticamente aplicada a toda a rede. O PDC permite inclusive o uso de perfis móveis, onde as configurações do usuário ficam armazenadas no servidor, permitindo que ele se logue em qualquer máquina.

## Compartilhamento de arquivos e impressoras

Mantendo o "Compartilhamento de arquivos e impressoras para redes Microsoft" ativo nas propriedades da rede, você pode compartilhar pastas clicando com o botão direito sobre elas e acessando a opção "Compartilhamento e segurança".

Marque a opção "Compartilhar esta pasta na rede" e, opcionalmente, a opção "Permitir que usuários da rede alterem meus arquivos" para tornar o compartilhamento leitura e escrita. Por padrão, o Windows XP utiliza uma pasta chamada "Arquivos compartilhados", que é a única compartilhada por padrão. Para compartilhar outras pastas, você precisa primeiro clicar sobre o link "Se você entende os riscos de segurança, mas deseja compartilhar arquivos sem executar o assistente, clique aqui", dentro da aba de compartilhamento.

O mesmo vale para as impressoras instaladas, que você pode compartilhar através do "Painel de Controle > Impressoras". Clique com o botão direito sobre ela e acesse a opção "Compartilhamento":



Você pode visualizar e acessar os compartilhamentos disponíveis nas outras máquinas da rede através do menu "Meus locais de rede", dentro do próprio Windows Explorer, o famoso "Ambiente de rede":



A navegação em redes Windows é um recurso que depende fortemente do envio de pacotes de broadcast (ou do uso de um servidor WINS) e da figura do "Master Browser", uma das máquinas da rede, eleita com a função de colocar "ordem na casa", localizando os compartilhamentos e entregando a lista para as demais. Em resumo, existem muitas coisas que podem dar errado, fazendo com que novos compartilhamentos demorem para aparecer, ou que micros configurados para usar diferentes grupos de trabalho (porém na mesma rede) não se enxerguem.

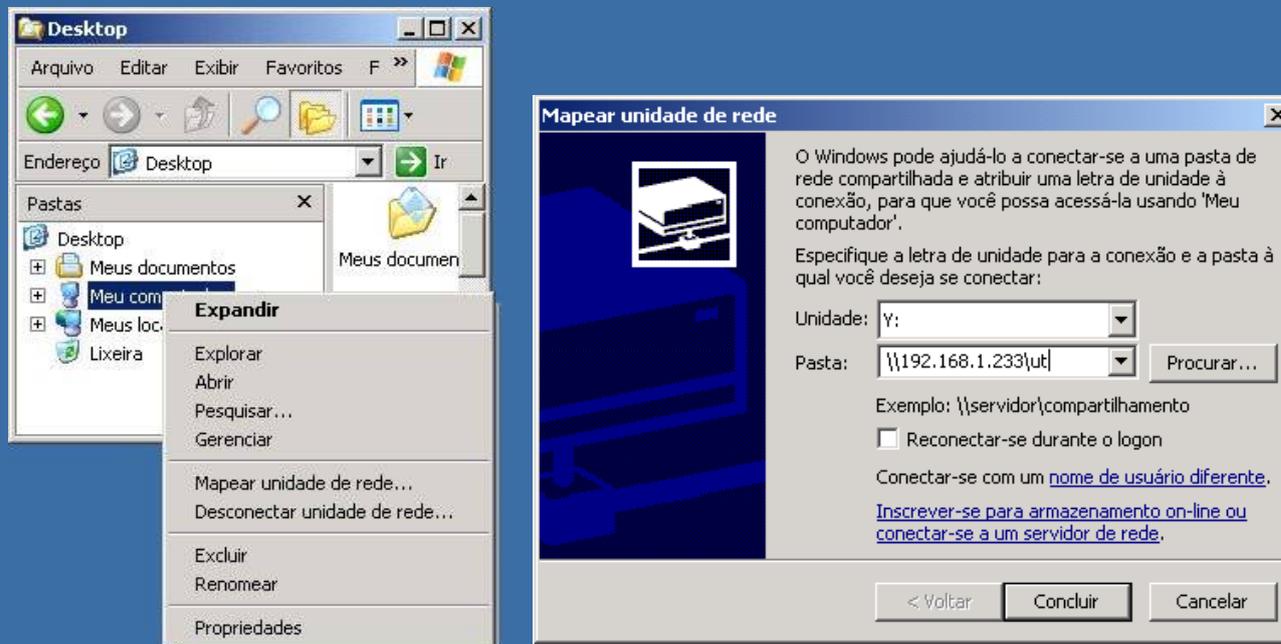
Nesses casos, você pode mapear o compartilhamento manualmente. Ainda dentro do Windows Explorer, clique com o botão direito sobre o "Meu Computador" e acesse a opção "Mapear unidade de rede". Na tela seguinte, escolha uma letra para a unidade e indique o endereço IP, ou nome do servidor, seguido pelo nome do compartilhamento, como em "\\servidor\ut" ou "\\192.168.1.233\ut".

Note que você usa duas barras invertidas antes do nome do servidor e mais uma barra antes do nome do compartilhamento. Ao acessar um servidor que fica ligado continuamente, você pode marcar a opção "Reconectar-se durante o logon", o que torna o mapeamento permanente:

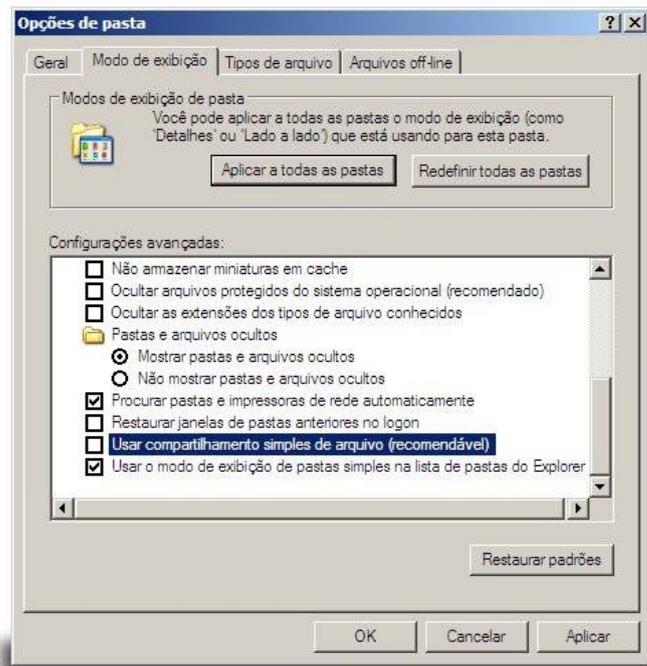
Versões antigas do Windows, incluindo o 3.11, 95, 98 e ME utilizam o modo de acesso baseado em compartilhamento por padrão. Neste modo de acesso, as pastas ficam disponíveis para acesso público dentro da rede e a única opção de segurança é proteger a pasta usando uma senha.

A partir do Windows 2000 passou a ser usado o modo de acesso com base no usuário, onde é necessário se autenticar (especificando um dos logins cadastrados na máquina) para ter acesso. Apesar disso, ao compartilhar uma pasta no Windows XP, todos os usuários da rede tem acesso o conteúdo, sem precisar especificar login e senha. Se você marcar a opção "Permitir que usuários da rede alterem meus arquivos", então além de qualquer um ter acesso, qualquer um poderá alterar o conteúdo da pasta.

Isso acontece devido ao sistema "simple sharing" (compartilhamento simples de arquivo), que é ativado por padrão no Windows XP. Ele flexibiliza as permissões de acesso, de forma a facilitar as coisas para usuários não técnicos. Por baixo dos panos, todos os acessos passam a ser mapeados para a conta "guest" (ativa por padrão), o que permite que usuários remotos sem login válido acessem os compartilhamentos diretamente. Este recurso é também chamado de "force guest".

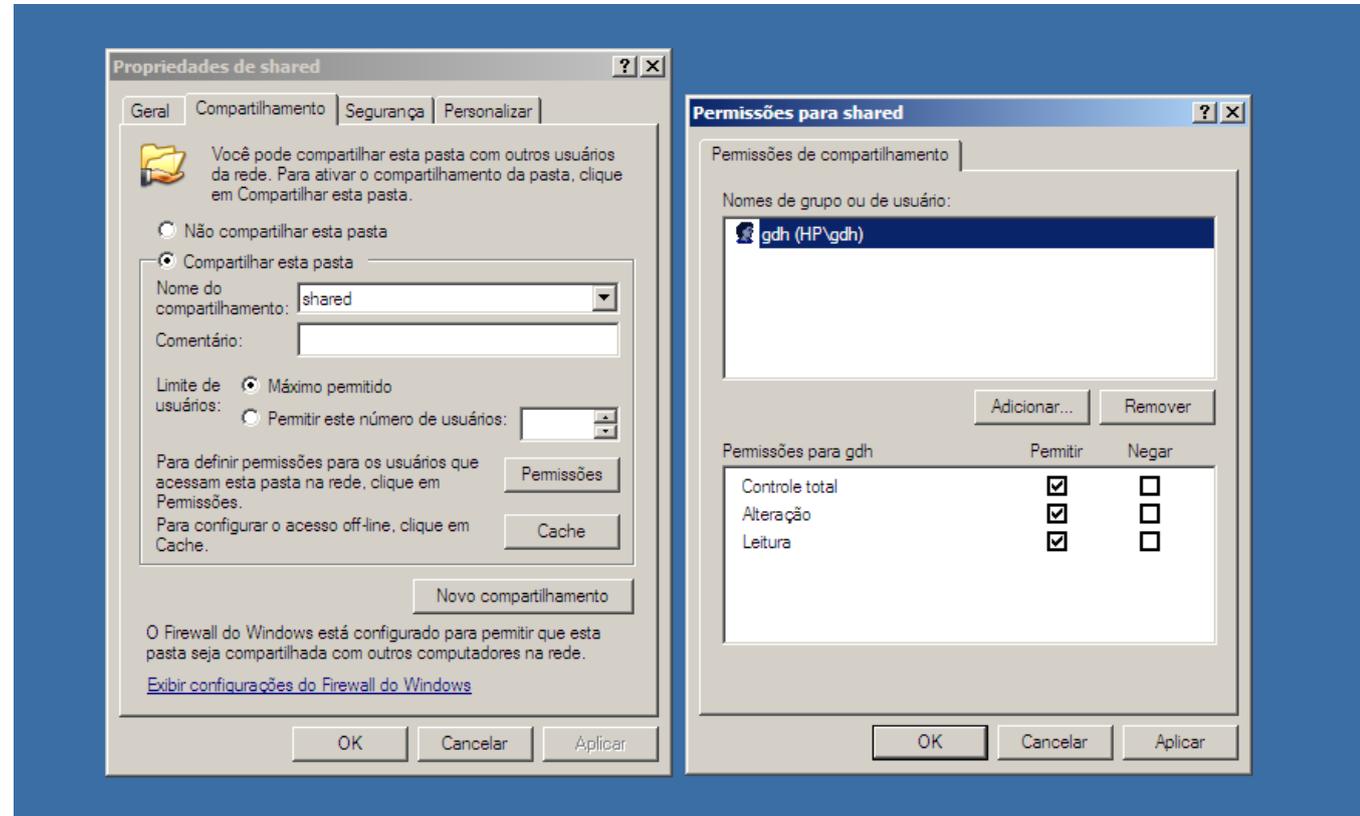


Para definir permissões de acesso, é necessário primeiro desativá-lo, o que é feito no menu de opções de pasta ("Ferramentas > Opções de pasta" no Explorer), desmarcando a opção "Modo de exibição > Arquivos e pastas > Usar compartilhamento simples de arquivo":



A partir daí, você pode ajustar as permissões de acesso através da opção "Compartilhamento > Permissões" nas propriedades da pasta (note que existe um menu separado para a configuração das permissões locais (Segurança > Avançado), que não é o que queremos nesse caso).

Por padrão, o grupo "Todos" tem acesso à pasta, o que efetivamente permite o acesso de qualquer um (apenas leitura no XP SP2 e acesso completo no XP original). Ao ajustar as permissões, remova o grupo "Todos" e adicione manualmente os usuários ou grupos que devem ter acesso à pasta, ajustando as permissões de acesso para cada um:



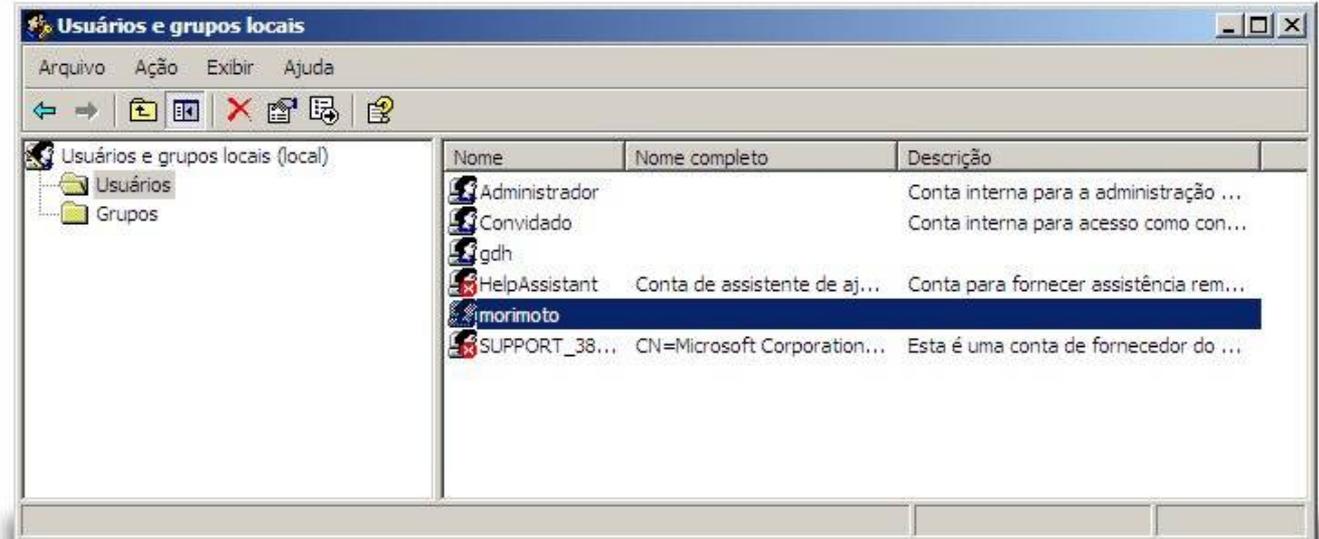
Quando um usuário remoto acessa o compartilhamento, o sistema verifica as credenciais de acesso e a partir daí autoriza ou não o acesso, baseado nas permissões definidas. O cliente envia por padrão o login e senha que foram usados para fazer logon. Se eles forem recusados pelo servidor, o usuário vê uma tela de autenticação como a abaixo, onde deve especificar um login e senha válidos no servidor. Este modo de acesso é similar ao que temos em um servidor Samba configurado com a opção "security = user" (usada por padrão) ou no Windows 2000:



A exceção fica por conta dos clientes que ainda utilizam o Windows 95/98/ME. Nesses, o cliente usa o login e senha usados para fazer logon na rede e simplesmente exibe uma mensagem de erro se o acesso for negado, sem exibir a janela de autenticação. Se você simplesmente pressionar ESC na tela de logon, você não consegue acessar compartilhamentos em máquinas configuradas para utilizar o modo de segurança com base no usuário.

Você pode gerenciar os logins de acesso na máquina com o Windows 2000/XP/Vista através do **lusrmgr.msc** ("local user manager", também conhecido como "looser manager" ;), chamado através do prompt do DOS (este utilitário está disponível apenas no XP Professional, não no Home ou Starter).

Ele oferece um volume muito maior de opções que o "Painel de controle > Contas de usuário", incluindo a exibição de contas ocultas, configuração de grupos, desativação temporária de contas, entre outros. Ele também permite desativar a conta guest, o que é desejável ao desativar o simple sharing e especificar manualmente quais usuários devem ter acesso. A conta guest faz parte do grupo "Todos" (que, como vimos, é incluído por padrão nas permissões de acesso de todos os novos compartilhamentos), o que permite o acesso de usuários não autenticados.

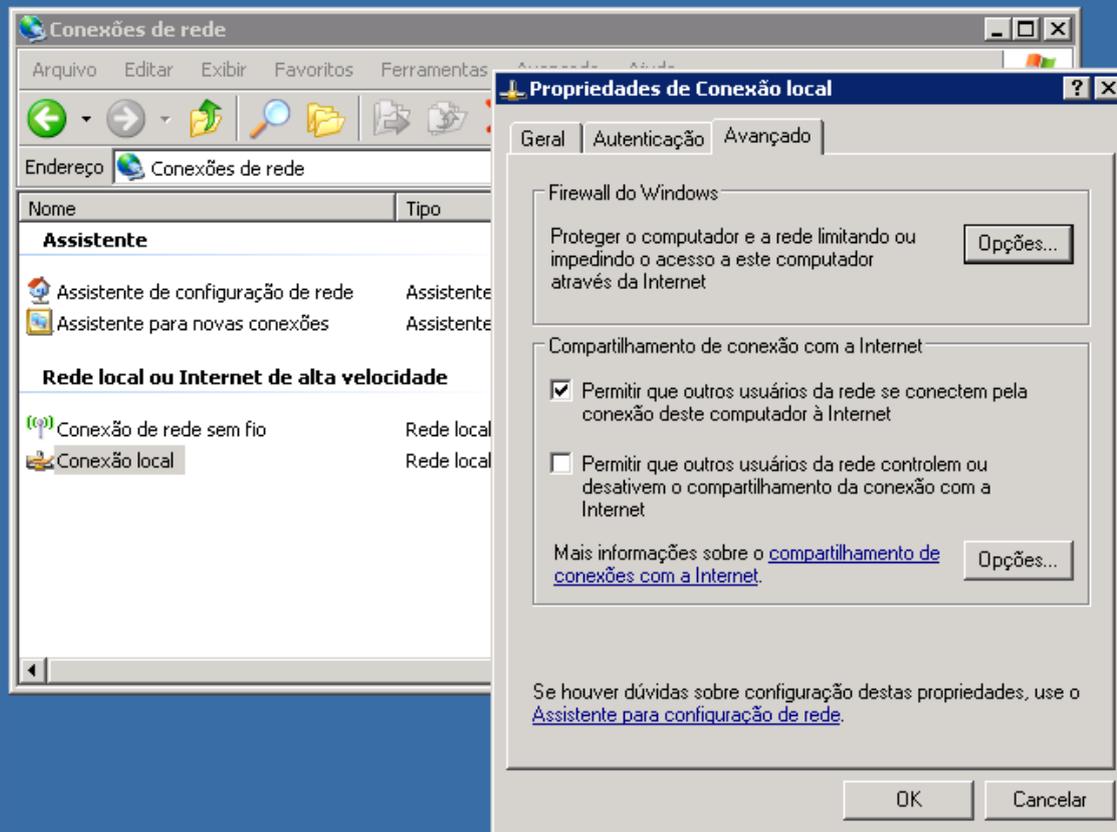


O maior problema em utilizar o controle de acesso com base no usuário é que você precisaria criar contas idênticas em todas as máquinas (para que elas acessem os compartilhamentos automaticamente, sem exibir o prompt de logon) ou pelo menos distribuir as senhas para que os outros usuários da rede possam especificá-las manualmente ao acessar os compartilhamentos. É justamente por isso que a Microsoft decidiu criar o simple sharing em primeiro lugar.

A solução para o problema seria utilizar um domínio, o que permite centralizar a autenticação em um servidor central (o PDC), resolvendo o problema das senhas. Você passa então a criar todas as contas (de todos os usuários da rede) no servidor e os usuários fazem logon nas estações utilizando qualquer uma das contas criadas. A função de PDC pode ser também assumida por um servidor Samba.

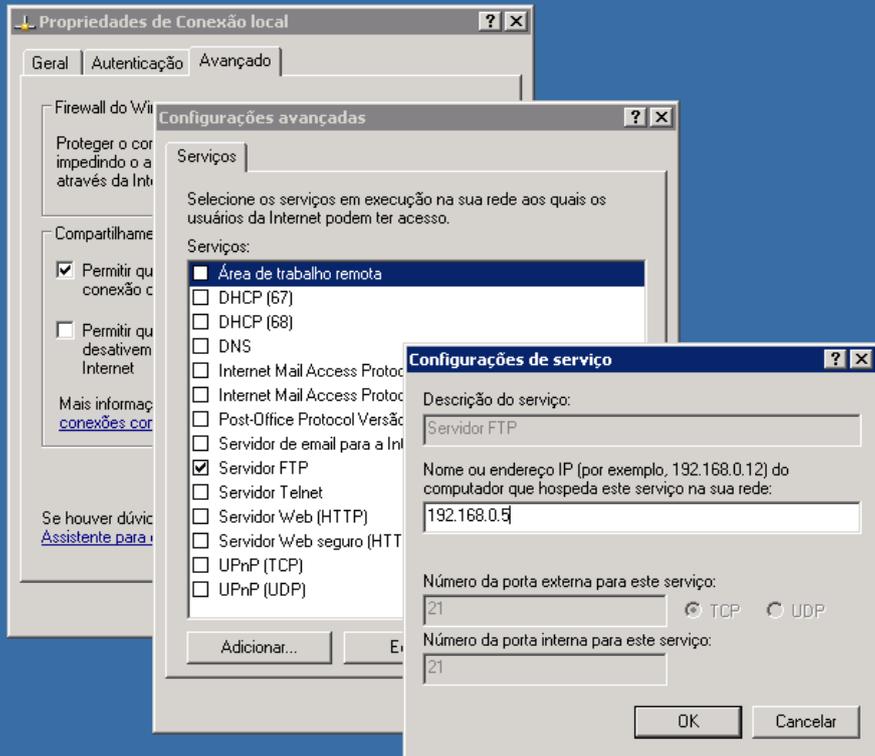
## Compartilhamento e conexões de ponte

O Windows inclui um sistema simples de compartilhamento de conexão, o famoso ICS (Internet Connection Sharing), disponível desde o Windows 98 SE. Para ativá-lo, acesse as propriedades da interface que recebe a conexão com a web (pode ser um modem discado inclusive) e, na aba Avançado, marque a opção "Permitir que outros usuários da rede se conectem pela conexão deste computador à Internet":



Ativar o compartilhamento faz com que a interface da rede local seja reconfigurada com o endereço "192.168.0.1" (não é uma opção, você é apenas notificado da mudança), o que vai quebrar a conectividade com outros micros da rede caso você utilize uma faixa de endereços diferente. O principal motivo é que o ICS inclui um servidor DHCP, que passa a fornecer endereços dentro da faixa "192.168.0.x" para os demais micros da rede. Como a faixa de endereços deste DHCP interno não é configurável, você acaba sendo obrigado a adotar o uso desta faixa de endereços no restante da rede. O principal motivo da faixa de endereços "192.168.0.x" ser a mais usada em redes locais é justamente por que ela é a utilizada pelo ICS. Ele inclui também um proxy DNS, que permite que o endereço IP da máquina que está compartilhando seja usado como DNS na configuração dos clientes. Na verdade ele apenas redireciona as requisições para o DNS do provedor.

Clicando no "Opções" você tem a opção de ativar o encaminhamento de portas (port forwarding) para micros da rede interna, de forma que eles possam ser acessados via internet. Isso é útil caso você deseje rodar servidores diversos nos micros da rede interna, ou precise rodar programas que precisem de portas de entrada:

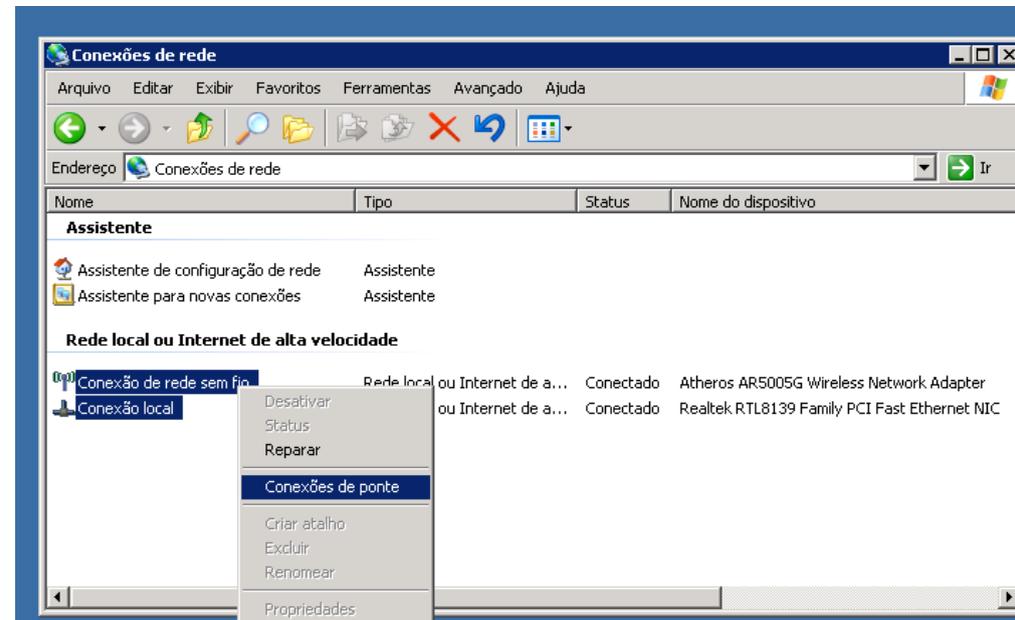


Ao usar uma máquina Windows para compartilhar a conexão, é indispensável ativar um firewall, seja o próprio firewall do Windows ou seja um software avulso, como o ZoneAlarm. Já é temerário conectar uma máquina Windows desprotegida diretamente, o que dizer de usá-la como gateway da rede, ligada 24 horas por dia a uma conexão de banda larga. O ICS pode ser usado como quebra-ganho em situações onde você precisa compartilhar a conexão rapidamente, mas do ponto de vista da segurança é mais recomendável compartilhar a conexão utilizando um servidor Linux, mantendo as estações Windows protegidas dentro da rede local.

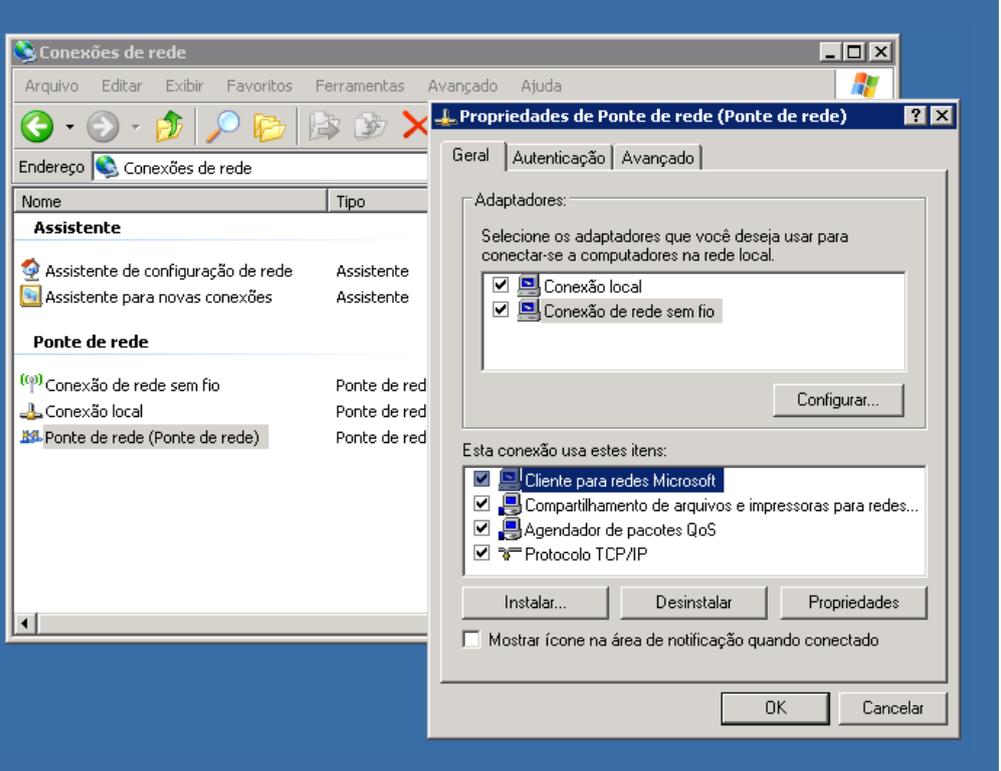
Outra dica é que, ao usar uma máquina XP com duas ou mais conexões de rede, é possível também criar uma ponte (bridge connection) dentre elas, permitindo que os micros conectados a cada uma das duas interfaces se enxerguem mutuamente.

Imagine uma situação onde você tenha três micros e precisa configurar rapidamente uma rede entre eles para jogar uma rodada de Doom 3, sem usar um switch. Se um dos micros tiver duas placas de rede (mesmo que seja uma placa cabeada e uma placa wireless), você pode usar cabos cross-over ou conexões wireless add-hoc para ligar os outros dois micros a ele. Inicialmente, o micro com as duas placas enxergaria os outros dois, mas os dois não se enxergariam mutuamente. A ponte resolve este problema, permitindo que os três se enxerguem e façam parte da mesma rede.

Para ativá-la, selecione as duas placas com o mouse, clique com o botão direito e acesse a opção "Conexões de ponte":



Isso faz com que seja criada uma nova interface, chamada de "ponte de rede". Imagine que esta interface é como um switch, ao qual as duas interfaces estão ligadas, ou que as duas interfaces foram unificadas em uma nova placa de rede. Nas propriedades da ponte de rede você pode ver os adaptadores que fazem parte e ajustar as propriedades de rede (IP, máscara, gateway, etc.) para ela:

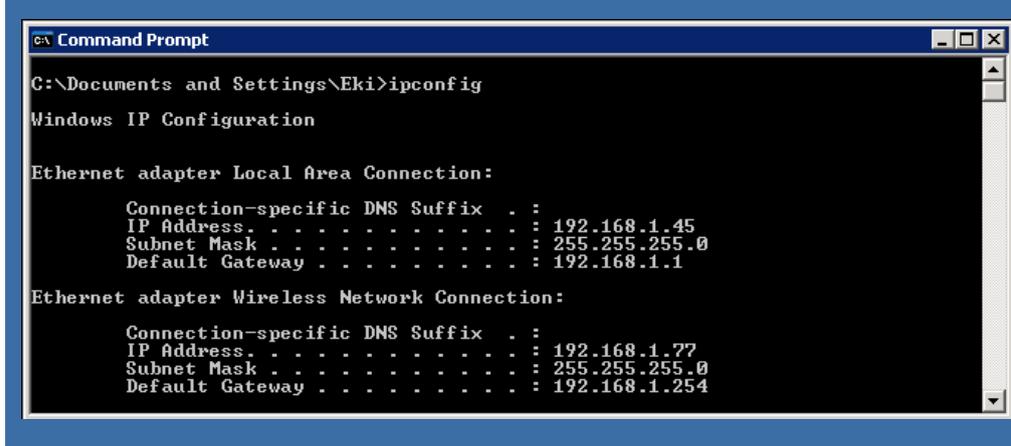


A ponte de rede recebe um IP próprio e é através dele que o PC passa a ser acessado pelos micros ligados às duas interfaces. Com a ponte, o sistema passa a encaminhar os frames que chegam a uma interface para a outra, de forma que os micros ligados aos dois segmentos passam a se enxergar manualmente.

## Comandos

Apesar do prompt de comando ser muitas vezes marginalizado e relegado a segundo plano, o Windows XP e o Vista oferecem um arsenal razoável de comandos de configuração da rede, que podem ajudá-lo a ganhar tempo em muitas situações.

Por exemplo, ao configurar a rede via DHCP, você pode checar rapidamente qual endereço IP está sendo usado por cada micro usando o comando "**ipconfig**" dentro do prompt do MS-DOS:



Para ver também o endereço MAC da placa de rede (necessário, por exemplo, para liberar o acesso à rede wireless na configuração do ponto de acesso, ao configurar restrição de acesso com base no endereço MAC das placas) e outras informações, adicione o parâmetro /all, como em:

```
C:\> ipconfig /all
```

Outro comando que pode ser usado para ver rapidamente o endereço MAC da placa de rede é o "**getmac**".

Ao configurar o sistema para obter a configuração da rede via DHCP, você pode usar o comando `ipconfig` para liberar o endereço obtido via DHCP (desconfigurando a rede) ou para renová-lo, o que pode ser útil em caso de problemas ou em situações onde você acabou de mudar a configuração do servidor DHCP e precisa agora fazer com que os clientes renovem os endereços para obterem a nova configuração. Outro exemplo são casos em que o micro falha em renovar o empréstimo do endereço obtido via DHCP (o que é relativamente comum ao acessar via cabo, por exemplo) fazendo com que seja desconectado da rede.

Para liberar o endereço obtido via DHCP, use:

```
C:\> ipconfig /release
```

Para renovar o endereço, use:

```
C:\> ipconfig /renew
```

Caso você possua mais de uma interface de rede instalada, você deve especificar a interface (com o mesmo nome que ela aparece no "Painel de Controle > Conexões de rede") no comando, entre aspas, como em:

```
C:\> ipconfig /release "Conexão local"
C:\> ipconfig /renew "Conexão local"
```

Se o comando falhar, muito provavelmente o seu servidor DHCP está fora do ar, ou existe algum problema no cabeamento da rede que esteja impedindo a comunicação, como um conector mal-crimpado ou uma porta queimada no switch.

Para ver outras máquinas que fazem parte do mesmo grupo de trabalho da rede Windows, incluindo máquinas Linux compartilhando arquivos através do Samba você pode usar o comando `"net view"`. Ele mostra uma lista das máquinas, similar ao que você teria ao abrir o ambiente de redes, mas tem a vantagem de ser mais rápido:

```
C:\> net view
```

Para visualizar quais pastas seu micro está compartilhando com a rede de forma rápida (para confirmar se um novo compartilhamento foi ativado, por exemplo), você pode usar o comando `"net share"` e, para ver quais máquinas estão acessando os compartilhamentos nesse exato momento, usar o `"net use"`.

```
C:\> net share
C:\> net use
```

O Windows cria dois compartilhamentos administrativos por padrão, o `C$` (que compartilha todo o conteúdo do drive `C:\`) e o `IPC$`, usado para trocar informações de autenticação e facilitar a transmissão de dados entre os micros. Estes compartilhamentos podem ser usados para acessar os arquivos da máquina remotamente, mas apenas caso você tenha a senha de administrador.

```

Microsoft Windows XP [versão 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\gdh>net share

Nome          Recurso          Observação
-----
C$            C:\              Recurso compartilhado padrão
ADMIN$       C:\WINDOWS      Administração remota
IPC$         C:\              IPC remoto
shared       C:\shared
SharedDocs   C:\DOCUMENTS AND SETTINGS\ALL USERS\DOCUMENTOS

Comando concluído com êxito.

C:\Documents and Settings\gdh>
    
```

É possível também usar o comando "net use" para mapear compartilhamentos de rede de outras máquinas via linha de comando. Nesse caso, você deve especificar a letra que será atribuída ao compartilhamento, seguida pelo "\\servidor\compartilhamento", como em:

```
C:\> net use G: \\servidor\arquivos
```

Ele vai solicitar o login e senha de acesso, caso exigido e a partir daí você pode acessar os arquivos através do drive G:\. O resultado é o mesmo de mapear o compartilhamento clicando sobre o "Meu Computador", apenas feito de forma diferente.

Para desconectar o compartilhamento, use o parâmetro "/delete", como em:

```
C:\> net use G: /delete
```

Ao contrário do que temos no Linux, o prompt do Windows não é case sensitive, de forma que tanto faz digitar "net use G: /delete", quanto "NET USE G: /DELETE" ou "NeT uSe G: /deleTE".

Você pode também fazer toda a configuração da rede via linha de comando usando o "netsh". Na prática, não existe nenhuma grande vantagem sobre configurar pelo Painel de controle, mas não deixa de ser um truque interessante.

Para configurar a rede, especificando manualmente os endereços, você usaria:

```
C:\> netsh int ip set address name="Conexão Local" source=static 192.168.0.22 255.255.255.0 192.168.0.1 1
```

... onde o "Conexão Local" é o nome da conexão de rede (da forma como aparece no painel de Conexões de rede do Painel de controle), seguido pelo endereço IP, máscara e gateway da rede. Não se esqueça do número "1" no final, que é um parâmetro para a configuração do gateway.

Para configurar o DNS, você usaria:

```
C:\> netsh int ip set dns "Conexão Local" static 200.204.0.10
```

Para configurar os endereços e DNS via DHCP, você pode usar os comandos:

```
C:\> netsh int ip set address name="Conexão Local" source=dhcp
C:\> netsh int ip set dns "Conexão Local" dhcp
```



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.

## ➤ Um trambolho na tampa do seu notebook

A Asus anunciou o primeiro notebook dual-screen do mercado, o W5Fe. A segunda tela é posicionada na tampa externa do notebook e funciona como uma espécie de palmtop, permitindo ter acesso aos contatos, e-mails e algumas outras funções sem precisar abrir o notebook. Não se trata de uma simples tela, mas sim de um "quase palmtop" com processador e memória próprios, que se comunica com o sistema operacional principal (por enquanto apenas o Vista suporta este recurso) a fim de obter os dados que são exibidos.

A idéia parece boa a princípio, mas na prática duvido muito que sobreviva, pois ela viola alguns princípios básicos.

O primeiro é a questão do custo. A segunda tela custa quase tanto quanto um palmtop completo. Estima-se que adicione algo entre US\$ 200 a 300 ao custo de produção do notebook. Considerando que os notebooks de baixo custo custam cerca de US\$ 600, nos EUA, a idéia encareceria demais os produtos para ser usada em qualquer lugar fora dos tops de linha. Você pode argumentar que a tecnologia tende a cair de custo caso seja usada em larga escala, mas se isso fosse verdade, os palmtops e smartphones também cairiam de preço na mesma proporção, pois os componentes usados são essencialmente os mesmos.

O segundo é a questão do tamanho e peso. Todos os fabricantes estão trabalhando no sentido de criar notebooks mais leves e esta idéia vai na contra mão, adicionando cerca de 200 gramas ao peso do equipamento, sem falar no aumento da espessura da tela. Você pode notar que no W5Fe temos um "calombo", justamente causado pela inclusão dos componentes adicionais.

O terceiro é que ele reduz a autonomia das baterias, outro ponto importante nos notebooks atuais.

O quarto é que quem carrega um notebook, normalmente também possui um palmtop ou celular, que já é usado para armazenar informações sobre contatos e compromissos, que seria uma das principais funções da segunda tela. Um software que sincronizasse o palmtop/celular com relação às informações com relação às informações armazenadas no notebook, via bluetooth, seria mais útil.

E, finalmente, temos o quinto, que é a questão da praticidade. Se você já teve o trabalho de tirar o notebook da mala, por que ficar perdendo tempo com uma tela minúscula, que mostra informações limitadas, ao invés de abrir de uma vez e usar o sistema principal? O sistema depende que o sistema operacional principal esteja em modo stand-by, onde você obtém seu ambiente de trabalho em poucos segundos.



Leia mais em:

<http://gizmodo.com/gadgets/laptops/asus-mutated-laptop-spotted-overseas-206329.php>

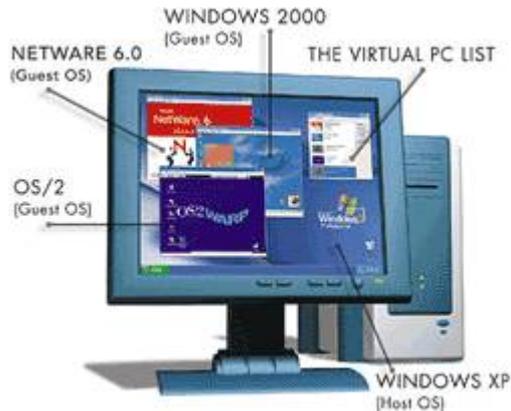
Postado por Carlos E. Morimoto

## Microsoft libera Virtual PC 2007

A Microsoft disponibilizou para download gratuito a tão aguardada versão de seu software de virtualização, lançado na segunda feira. O Virtual PC 2007 agora possui suporte para o Windows Vista, tanto no sistema virtualizado quanto no servidor, e possui a vantagem de utilizar as novas tecnologias providas da Intel e AMD.

Está disponível para download gratuito a versão 32 e 64 bits. A Microsoft prometeu a gratuidade de software assim que tornou freeware também o Virtual PC 2004 SP1, em julho do ano passado.

Neste contexto, o Microsoft Virtual PC 2007 agora passa também a concorrer com outros programas de virtualização freeware, como o famoso VMWare. Para desempate, só mesmo os usuários poderão dizer.



Leia mais em:

[http://www.betanews.com/article/Microsoft\\_Issues\\_Free\\_Virtual\\_PC\\_2007/1171916802](http://www.betanews.com/article/Microsoft_Issues_Free_Virtual_PC_2007/1171916802)

Download:

<http://www.microsoft.com/windows/products/winfamily/virtualpc/default.aspx>

Postado por **Júlio César Bessa Monqueiro**

## ➤ Celular: a nova forma de pagamento está em testes

*"Basta pegar o telefone, fazer o pedido normalmente e informar o número do celular, que foi previamente cadastrado. Do outro lado da linha, o dono do estabelecimento digita o número do celular e o valor da compra no terminal enquanto a VisaNet encaminha os dados da transação via mensagem de texto ao cliente." Segundo a [IDG Now!](#), é assim que está funcionando o novo método de pagamento - que está em fases de testes, com alguns táxis e disque-pizzas, em São Paulo, capital.*

A nova onda de pagamentos está sendo feita com a nova tecnologia da empresa Visa, que está levando o novo meio à vários países. A primeira empresa de celular a estreiar o serviço é a Oi, com o Oi Paggo, que pretende estar em toda a sua área de cobertura até o final do ano. Sistema semelhante está sendo utilizado pela Claro, em parceria com a Visa e Cinemark, para compras de ingressos pelo aparelho celular.

*"Mas quem mais vai ganhar com isso, não tenham dúvidas, são as administradoras dos cartões e os bancos. (...) Outro que ganha é o governo. (...) Enquanto quase todos nós procuramos formas de ganhar algum dinheiro, acabamos de ser contemplados com mais uma forma de gastá-lo. Aproveitemos, pois." [disse André Caramuru, colunista da IDG Now!](#).*

Para ver como funciona o sistema, acesse o especial que a IDG fez sobre o novo serviço, em:

[http://idgnow.uol.com.br/especiais/pagamento\\_celular](http://idgnow.uol.com.br/especiais/pagamento_celular)

Postado por **Júlio César Bessa Monqueiro**

## ➤ OpenSUSE abandona ZENWorks e abre YaST para comunidade

O OpenSUSE abandonará o sistema de atualizações ZENWorks - que inclusive não tem muitos fãs - para deixar a atualização mais agradável, e também está abrindo o YaST para a comunidade com fins de desenvolvimento. O gerente do projeto SUSE, Andreas Jaeger disse "O OpenSUSE está focando num software nativo de gerenciamento usando o YaST e o Libzypp, a biblioteca de gerenciamento de pacotes"

Com isso, o sistema ZENWorks passa apenas para a linha dos produtos corporativos da Novell, como o SUSE Enterprise Desktop e Server. Lembramos que, num passado recente, o OpenSUSE foi usado como uma distribuição de testes para os produtos oficiais, assim como o Fedora e o Red Hat.

Assim, a partir do OpenSUSE 10.3 Alpha 4, não terá mais o ZENWorks e sim ferramentas nativas para atualização, desenvolvendo o atual Opensuse-updater e integrando-o ao YAST e Zypper.

Além disso, a abertura do desenvolvimento para a comunidade do YAST e várias outras bibliotecas, como a Libzypp, segundo Andreas, aumentará a velocidade e qualidade do mesmo, que passaram para a GPLv2. Assim, o YaST pode ser inclusive usado por outras distribuições, expandindo a facilidade para o mundo Linux. Andreas também disse que em breve terão repositórios SVN de acesso público .



Leia mais em:

<http://www.desktoplinux.com/news/NS5210817105.html>

Postado por **Júlio César Bessa Monqueiro**

## ➤ Análise e testes com a nova plataforma Santa Rosa

O site AnandTech publicou uma análise sobre a nova plataforma para os processadores Centrino, chamada "Santa Rosa", que está sendo lançado com diversos notebooks desde o início deste mês.

Para quem não conhece o Santa Rosa, uma introdução extraída de outra notícia:

ova plataforma provavelmente virá com a tecnologia Robson que oferece um buffer em memória flash para o HD, usada conjuntamente com o ReadyDrive do Windows Vista, e segundo Carlos E. Morimoto "além dos ganhos de desempenho, sobretudo a potencial redução no tempo de boot, o buffer permite que o HD fique mais tempo em modo de economia de energia, já que não é preciso "acordar" o HD ao salvar arquivos ou quando o sistema precisa atualizar arquivos de log, por exemplo, operações que podem ser realizadas no buffer. Isto acaba tornando a tecnologia bastante interessante para os notebooks, onde o HD chega a representar um quarto do consumo elétrico total.(...) Temos também a tecnologia Robson, desenvolvida pela Intel, onde temos um buffer similar, instalado na placa mãe." (Veja mais sobre a tecnologia em: <http://www.guiadohardware.net/artigos/sdds-hhds-readyboost-readydrive/>).

No texto, o site fala sobre a história da evolução deste processador e suas arquiteturas, as novidades com relação à arquitetura, processamento, desempenho e chipsets, além de suportar a conectividade (como Wi-Fi), e testes com a nova tecnologia Robson.

Também é feito testes em diversos segmentos como renderização, codificação, etc, apresentando também gráficos sobre consumo de energia.



Leia mais em:

<http://anandtech.com/cpuchipsets/intel/showdoc.aspx?i=2985>

Postado por **Júlio César Bessa Monqueiro**

## ➤ Transmissão de energia elétrica sem fios já é realidade

O conceito de wireless, obviamente, não é recente na informática. A maioria dos periféricos, celulares, notebooks e outros aparelhos possuem alguma compatibilidade, e muitas pessoas já usam o mínimo de fios em suas casas. Porém, entre Internet, rede e conexão de componentes de forma sem fio, existe um em cada aparelho que com certeza você ainda não largou. É o fio de energia elétrica.

Pesquisadores do Massachusetts Institute of Technology, o MIT, criaram uma tecnologia experimental para transmitir eletricidade sem necessidade de fios.

Embora limitada, a tecnologia é similar ao uso de laser ou microondas, possuindo portanto um efeito negativo significativo que ainda precisa ser alterado, por causar ainda alguns malefícios ao meio, e muitas interferências em ambientes que possuem metais.

Chamado de "WiTricity" pelos pesquisadores, os mesmos conseguiram acender uma lâmpada de 60 Watts a sete pés da fonte energética, usando um sistema de indução eletromagnética, caindo radicalmente de eficiência quando é afastado mais ainda, como era de se esperar.

O atual experimento usa uma bobina com raio de 20 polegadas de diâmetro, e é capaz de operar em uma distância de aproximadamente 2 metros. Boa notícia principalmente para quem usa notebook, embora a tecnologia ainda esteja em modo "embrião", é o início para uma nova geração num futuro médio-distante.



Leia mais em:

<http://www.dailytech.com/MIT+Engineers+Unveil+Wireless+Power+System/article7632.htm>

Postado por **Júlio César Bessa Monqueiro**

## ➤ Boas novas para quem usa o MS Office e o Open Office

A Sun (mantenedora oficial do OpenOffice) lançou um plugin de importação/exportação do formato de arquivo OpenDocument (ODF - Open Document Format), utilizado pelo OpenOffice/BROffice, para a suíte de escritório da Microsoft.

É o "Sun ODF Plug-in for Microsoft Office", para o MS Office 2000, XP (2002) e 2003. Depois de instalado, ele ativa no MS Office o salvamento de arquivos no formato OpenDocument, adicionando o formato na janela "salvar como" do Word, permitindo ainda escolher o ODF para formato padrão de arquivo. No Excel e no PowerPoint, no entanto, os add-ons ainda ficam disponíveis como novos itens no menu Arquivo, e não diretamente na janela de salvamento (similar a um "Exportar").

O instalador tem cerca de 30 MB, baixe em:  
<http://javashoplmsun.com/ECom/docs/Welcome.jsp?...>

Mais informações sobre ele (e o OpenOffice em geral):  
<http://www.sun.com/software/star/openoffice/>

Paralelamente, surgem melhorias no projeto em desenvolvimento da Microsoft e Novell, o "OpenXML/ODF Translator Add-On for Office". Ele permite trabalhar com textos, planilhas e apresentações, mas as extensões para o Excel e PowerPoint ainda estão em versões bem iniciais de desenvolvimento, apresentando uma compatibilidade um tanto fraca. Além disso, estas extensões valem apenas para as versões 2003 e 2007 do MS Office.

Aproveitando o tema, com a chegada do Office 2007 e seu novo formato de arquivo (onde vem um "x" depois da extensão e os arquivos são gravados compactados, como ".docx", ".ppsx", ".xlsx"...), a Microsoft lançou pouco depois do Office 2007 o "Compatibility Pack for the 2007 Office System", que permite que as versões anteriores do Office (especialmente o 2003, ou as versões anteriores do Office 2000/2002 com as mais recentes atualizações) abram, editem e salvem arquivos no novo formato, auxiliando a compatibilidade para quem precisa ver arquivos desse tipo.



Download do Compatibility Pack for the 2007 Office System:  
<http://www.microsoft.com/downloads/details.aspx?...>

Postado por Marcos Elias Picão

## iPhone desbloqueado no Brasil, para operadoras nacionais

iPhone desbloqueado, agora no Brasil. "MacMasi" e "Paulo Stool", estudantes de São Paulo, estão oferecendo o serviço de desbloqueio do iPhone por cerca de R\$ 600.

Originalmente o iPhone só é compatível com a rede da operadora norte americana AT&T.

Após modificações, o aparelho aceita cartões GSM de praticamente qualquer operadora nacional. Além disso, eles trabalham de forma a permitir a troca do chip com o aparelho ligado.

Montaram um site, tentando vender o serviço de desbloqueio, o [www.desbloeio.br.com.br](http://www.desbloeio.br.com.br). No site há fotos e até um vídeo do iPhone desbloqueado. Veja esse, onde usam o iPhone para fazer e receber ligações de operadoras nacionais, com chips retirados de celulares.

**No vídeo anunciam** também outro projeto: a tradução dos menus do iPhone para o português. O trabalho não foi bem uma "nova descoberta", eles agem seguindo o "passo a passo" com 10 etapas, publicado no blog americano de George Hotz ( [www.iphonejtag.blogspot.com](http://www.iphonejtag.blogspot.com) ). Entre os ingredientes está um ferro de solda, certamente não é para qualquer um ;) Afinal o iPhone já custa caro, e facilmente desastrados poderiam danificá-lo.



Mais informações:

<http://info.abril.com.br/aberto/infonews/082007/30082007-14.shl>

e claro, o site deles:

<http://www.desbloeio.br.com.br/>

Postado por Marcos Elias Picão

## Finalmente acesse seu Speedy sem provedor

Depois de muita luta, a Telefonica finalmente dispensou, de acordo com regra estabelecida pela Anatel, o uso obrigatório de um provedor de conteúdo para acesso à Internet no seu serviço banda larga Speedy, que funciona em todo o Estado de São Paulo.. Agora, você não precisa mais pagar a velha taxa de 19,90 ou 24,90 por mês :-)

***O comunicado oficial saiu dia 26/09, veja:***

<http://www.speedyvantagens.com.br/cadmus/em/comunicado/index.html>

Confira também o anúncio pela Folha de São Paulo:

<http://www1.folha.uol.com.br/foha/dinheiro/ult91u323739.shtml>

### COMUNICADO IMPORTANTE

A Telefônica comunica a você cliente Speedy que, de acordo com decisão judicial não definitiva, a partir de hoje (quarta-feira, 26 de Setembro de 2007) oferece a conexão à internet através do login 'internet@speedy.com.br' e da senha 'internet'.

Você, que já possui um provedor, poderá conectar-se com o login e senha acima descritos. Entretanto, este acesso não inclui os serviços, tais como: e-mail, conteúdos de acesso restrito, entre outros.

Ainda, a decisão não cancela os serviços de provedores de internet já contratados por você, que poderá optar por continuar **fazendo o login** através do provedor que já contratou e manter os serviços que possui atualmente. Neste caso, você não precisa fazer nada – basta continuar utilizando o login e senha do seu provedor atual. Caso queira entender melhor estes serviços e/ou alterá-los entre em contato com o seu provedor.

Mais informações podem ser obtidas em nossa Central de Relacionamento.



[www.telefonica.com.br](http://www.telefonica.com.br)

Postado por **Júlio César Bessa Monqueiro**

## Asus anuncia oficialmente o Eee PC

A Asus finalmente lançou "formalmente" a sua linha Eee PC, de sub-notebooks. Ultra portáteis e aparentemente mais adequados para crianças, muito se tem ouvido falar ultimamente.

São 4 modelos baseados num padrão composto por:

- \* tela de 7 polegadas;
- \* processador Intel Celeron 900 MHz;
- \* porta de rede ethernet 10/100 Mbps e rede sem fio 802.11b/g;
- \* "HD" de memória flash;
- \* 3 portas USB;
- \* slot para cartão de memória SD;
- \* auto-falantes estéreo e microfone embutidos;
- \* e uma versão do Linux Xandros.

O modelo básico, 2G Surf, vem com 2 GB de memória flash e 256 MB de memória DDR2. O 4G tem o dobro de espaço e memória (4 GB e 512 MB, respectivamente).

As máquinas superiores incluem também uma webcam de 0.3 megapixel no topo da tela. Também superam em tempo de bateria - 2,8 horas com uma bateria de 4400mAh, e 3,5 horas com uma unidade de 5200mAh. Os modelos 2G Surf e 4G Surf são os mais básicos, e o 4G (sem o "Surf") e 8G são os mais avançados.

O topo de linha da série, 8G, vem com 1 GB de memória DDR2 e 8 GB de memória flash.

Apesar de terem sido projetados para rodar Linux, eles podem rodar o Windows XP também.

A Asus disponibiliza os drivers necessários para quem quiser usar o Windows. A versão on board do Linux vem com software de e-mail, navegação na internet e aplicações para internet via rádio, além de ferramentas parecidas com o Skype, Google Docs e Wikipedia.

Modelos diferentes deverão ser adequados a diferentes países. A Asus da Inglaterra, a princípio, oferecerá apenas o 4G, por £219. Há alguns adicionais ainda com preços não divulgados.

Apresentação (comercial) do Eee PC, e especificações:

<http://eeepc.asus.com/en/>

Fonte:

[http://www.reghardware.co.uk/2007/10/16/asus\\_launches\\_eee\\_pc/](http://www.reghardware.co.uk/2007/10/16/asus_launches_eee_pc/)

Aproveite para ler aqui no GdH a análise do Intel Classmate, do Carlos E. Morimoto, que vale também para o Eee PC:

"O Intel Classmate é um mini-notebook voltado para uso na educação, concorrente direto do OLPC. Ele utiliza um Celeron de 900 MHz, com 256 MB, 2 GB de memória Flash, usada como espaço de armazenamento, wireless, portas USB e uma tela de 800x480. Você pode notar que esta é a mesma configuração do Asus Eee e isso não é mera coincidência, já que o Eee é uma versão de consumo do Classmate, voltado para o público em geral. Se você está interessado no Asus Eee, esta análise também será do seu interesse, já que, excluindo a aparência externa, os dois são praticamente idênticos."

Leia mais em:

<http://www.guiadohardware.net/analises/intel-classmate-parte1/>

Postado por Marcos Elias Picão

## ➤ Compiz Fusion no Ubuntu 7.10

O usuário Edvan Macedo publicou no site Viva o Linux um artigo interessante chamado "Compiz Fusion no Gutsy", que mostra "de forma bem prática a instalação e configuração dos efeitos visuais em 3D na área de trabalho do Ubuntu 7.10. Veja a descrição:

"Essa vai pro pessoal que acabou de instalar seu Ubuntu Gutsy e não sabe muito bem como configurá-lo. Leia, faça e deixe seu computador impressionar qualquer um!"

Veja um trecho (introdução):

*"Por padrão o Ubuntu Gutsy já vem com o Compiz-Fusion incluso, no entanto, logo de inicio, percebe-se que este não está muito na sua frente, claro, nada mais que uma boa procura pelo menu Ubuntu, mas como irei mostrar tudo mastigado, aí vai o caminho (...).*

*os efeitos Extras que vem por default no Ubuntu Gutsy não demonstram todo o poder do Compiz-Fusion, fazendo poucos efeitos. Para que possamos acrescentar novos efeitos e configurarmos do jeito que quisermos (LIVRE), é bom fazermos os seguintes procedimentos (...).*"

O artigo está dividido entre as seguinte seções:

- 1. Adicionando novas configurações**
- 2. Configuração**



compiz fusion

Leia mais em:

<http://www.vivaolinux.com.br/artigos/verArtigo.php?codigo=7486>

Postado por **Júlio César Bessa Monqueiro**

## ➤ Liberado Windows Server 2008 RC1

A MS disponibilizou para download o Release Candidate 1 do Windows Server 2008.

O sistema de servidor deverá ficar pronto até o começo do ano que vem, com data prevista de lançamento no primeiro trimestre de 2008 (mais precisamente, 27 de fevereiro). Antes se esperava para o final de 2007, mas em agosto a MS decidiu deixar para o ano que vem mesmo.

O Release Candidate é uma versão quase final, onde se trabalham em correções e ajustes, sem ficar adicionando muita coisa. Pessoalmente testei algumas coisas com o Release Candidade 0 (a versão anterior ao RC1), e pareceu bastante estável.

O Windows Server 2008 vem para substituir o Windows Server 2003, com novos recursos de servidor e o kernel baseado no Windows Vista SP1 (além da interface e aplicações).

O Server 2008 se diferencia radicalmente do Windows NT, 2000 e 2003 para servidor. Ele possui um modo de operação onde apenas o servidor em si é instalado, sem interface gráfica (à la Linux :p). Isso é muito útil em servidores remotos, onde toda a operação é feita via linha de comando, além de deixar o sistema muito mais leve e próprio para a tarefa que vai fazer na maior parte do tempo. Durante a instalação, é possível optar por instalá-lo com ou sem suporte a interface gráfica.

Além disso, há significantes implementações em recursos como o Active Directory, serviços de terminal e acesso remoto, segurança, virtualização e ferramentas de administração do servidor. É planejado para ter várias edições, tais como:

**Standard Edition (x86 e x86-64)**

**Enterprise Edition (x86 e x86-64)**

**Datacenter Edition (x86 e x86-64)**

**Web Server (x86 e x86-64)**

**Storage Server (x86 e x86-64)**

**Small Business Server (x86-64)**

**Essential Business Server 2008 (x86-64)**

E também é esperada uma versão para processadores Itanium.

Interessados em baixar a versão RC0, que funciona até 30 de julho de 2008 (ou por 30 dias, se você não quiser fazer o registro gratuito), podem entrar em:

<http://www.microsoft.com/windowsserver2008/audsel.mspix>

Informações não oficiais sobre novos recursos dessa versão de servidor do Windows você pode ler na Wikipedia:

[http://en.wikipedia.org/wiki/Windows\\_Server\\_2008](http://en.wikipedia.org/wiki/Windows_Server_2008)

E também na página da Microsoft, é claro:

<http://www.microsoft.com/windowsserver2008/>

No evento de lançamento do Windows Server 2008, no ano que vem, também serão anunciados oficialmente o SQL Server 2008 e o Visual Studio 2008. Caso lhe interesse as versões trial desses outros produtos, baixe aqui:

SQL Server 2008:

<http://www.microsoft.com/sql/2008/prodinfo/download.mspix#EDB>

Visual Studio 2008:

<http://msdn2.microsoft.com/en-us/vstudio/aa700831.aspx>

Postado por Marcos Elias Picão

## ➤ O Asus Eee foi superado

Com o lançamento do Eee PC, a Asus tinha em mãos a possibilidade de abocanhar uma grande fatia do mercado de notebooks, atendendo a um novo e crescente mercado: o de pessoas que querem um notebook simples, compacto e barato para acessar a web, ler e-mails e fazer outras tarefas simples.

O projeto do Eee PC atende perfeitamente bem a estes objetivos, oferecendo um processador de baixo consumo, um desempenho suficiente para o que se propõe e uma boa autonomia de baterias, como você pode ler em mais detalhes no [meu artigo anterior sobre o assunto](#).

Se o Eee tivesse sido lançado dentro da faixa de preço original (US\$ 200 a US\$ 250) e fosse produzido em volume suficiente, sem dúvida seria um sucesso de vendas. A Asus estaria mais de um ano na frente dos concorrentes e teria tempo de sobra para se estabelecer no mercado e aperfeiçoar o projeto antes que o Eee tivesse um concorrente de peso.

Entretanto, não foi isso o que aconteceu. Aproveitando o hype e o grande interesse em torno do Eee, a Asus optou por aumentar o preço, vendendo a versão inicial do Eee por US\$ 400 e, apesar da generosa margem de lucro, não foi capaz de atender à demanda, perdendo sua janela de oportunidade. Com isso, outros fabricantes começaram a se mexer.

Surgiu então o CloudBook da Everex, que provavelmente será sucedido por outros mini-notebooks semelhantes. Ele é baseado no projeto NanoBook da Via, uma plataforma de mini-notebooks de baixo custo, com telas de 7", destinados a serem vendidos abaixo da marca dos US\$ 500.

O CloudBook é similar em tamanho e em conceito ao Asus Eee e é vendido por US\$ 399, exatamente o mesmo preço do Eee de 4 GB. Ele também tem uma tela de 7", com resolução de 800x480 usa uma bateria com 4 células, mas ele ganha do Eee em dois quesitos:

O primeiro é o armazenamento. O Eee usa um SSD de 4 GB, que apesar de oferecer bons tempos de acesso, oferece taxas de leitura e escritas muito inferiores às de um HD de 2.5", sem falar no pouco espaço de armazenamento, já que o sistema operacional deixa pouco mais de 2.5 GB disponíveis para o usuário. O CloudBook, por sua vez, oferece um HD de verdade, com 30 GB de espaço, suficiente para armazenar uma quantidade maior de arquivos e permitir a instalação de mais de um sistema operacional em dual-boot. A desvantagem de usar um HD tradicional seria o tamanho, o custo e o consumo elétrico, mas pelo visto a Everex conseguiu compensar estas desvantagens com ganhos em outros componentes.

O CloudBook ganha do Eee também no quesito processamento. O Eee utiliza um Celeron-M underclockado para 630 MHz, enquanto o CloudBook usa um Via C7-M de 1.2 GHz. Clock por clock, o C7 perde para o Celeron-M, mas quando comparamos um C7 de 1.2 GHz com um Celeron operando a quase metade disso, a vantagem do C7 é óbvia.

Se você está pensando em comprar um Eee, aconselho a esperar mais alguns meses. Além do CloudBook, temos outros concorrentes surgindo no horizonte.



Postado por Carlos E. Morimoto

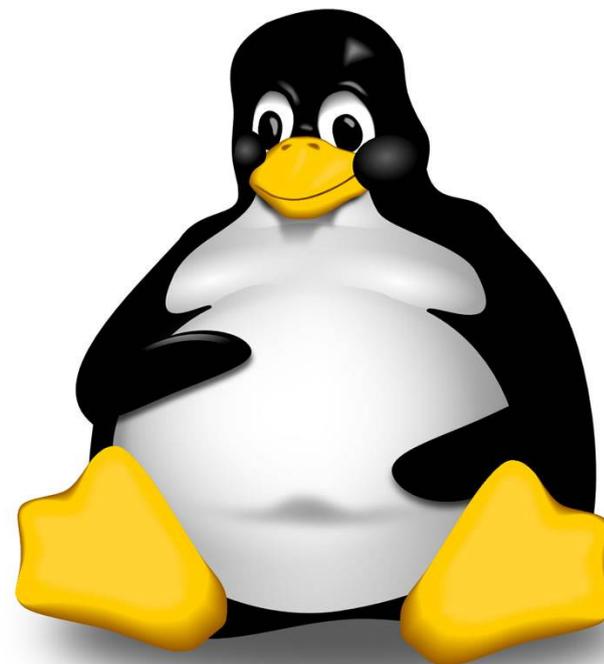
## Linux Next: o projeto de um kernel mais limpo e testado

O kernel 2.6.x é o que usamos hoje na maioria das distribuições - exceto algumas que, por motivo de leveza (para micros antigos) ou conceito (Slackware) usam ainda a antiga 2.4.x. Contudo, para ter cada vez mais recursos, o kernel foi inchando, inchando, e isso não é nada agradável aos desenvolvedores, sendo muito difícil de se manter, especialmente para garantir que não terá regressões e que todos os componentes sejam testados.

É justamente neste meio que entra a idéia do "Linux Next", que nasceu como um sonho de Andrew Morton, mantenedor do atual kernel, vendo que ele e outros estavam começando a ter problemas. Ele propôs a criação da árvore "linux-next", que receberia por dia várias árvores de subsistemas do Linux e testes de compilação depois de terem sido aplicadas. Parece simples, mas não é.

O desenvolvedor do kernel Stephen Rothwell anunciou que faria parte do grupo. Embora o processo seja mais complicado e trabalhoso, garantiria um código de maior qualidade e muito mais limpo que o atual, com muito mais testes.

Contudo, ainda há uma grande discussão, "saudável" segundo a fonte, no Linux Kernel Mailing List (LKML), sobre como seria o tal linux-next sem a participação de liderança de Linus Torvalds. O negócio é esperar para ver, afinal os participantes estão bem otimistas.



Leia mais em:

<http://osnews.com/story/19340/Linux-Next-Begins-to-Take-Shape>  
<http://blog.internetnews.com/skerner/2008/02/-linux-next-begins-to-take-sha.html>

Postado por **Júlio César Bessa Monqueiro**

## Google retira alguns mapas do ar a pedido do Pentágono

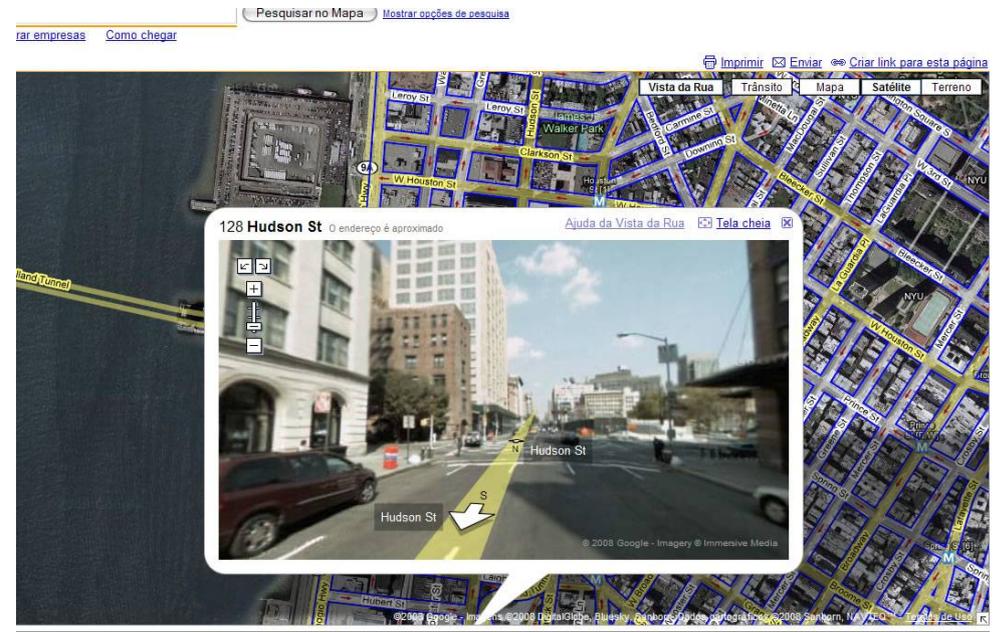
Até onde vai a liberdade de expressão em vias públicas? O Google Maps, popular serviço do Google que exibe mapas de vários países, retirou alguns mapas das proximidades do Pentágono, dos EUA.

A alegação do Pentágono é que algumas fotos presentes no Google Maps ofereciam ameaças à segurança das bases militares dos EUA.

O general Gene Renuart, chefe do comando militar responsável pela defesa nacional, falou que o Pentágono vinha conversando com o Google sobre os possíveis riscos, e esperavam a cooperação do Google para remoção das imagens selecionadas do serviço Street View (Vista da Rua).

A vista da rua do Google Maps fornece visão de 360º, ao nível do chão, em mais de 30 cidades dos EUA. É possível fazer um passeio virtual pelas ruas, dirigindo nas cidades já mapeadas. Qualquer um pode conferir no <http://maps.google.com>, acessando algum endereço dos EUA - clique em "Vista da rua".

É o tradicional caso onde todos são iguais perante a lei, mas alguns são "mais iguais" do que outros. Medidas como essa tiram a liberdade de expressão, a imparcialidade de um serviço fornecido publicamente. O Google Maps, assim como o Google Earth, já vêm sendo criticados por diversos governos há tempos. Convenhamos, é muito complicada a questão de proibir a exibição de vias públicas na web. Legalmente o Pentágono não teria esse poder (ou não deveria ter).



Leia mais em:

[http://www.techradar.com/news/internet/google/\[...\]images-from-google-maps-258983](http://www.techradar.com/news/internet/google/[...]images-from-google-maps-258983)

Postado por Marcos Elias Picão



O maior fórum de informática do país:  
**2.500.000 mensagens**  
**175.000 membros**

#### **Hardware:**

*Hardware Geral*  
*Overclock, Tweaks e Eletrônica*  
*Case Mod e Ferramentas*  
*Notebooks, Palms, Câmeras, Telefonia*  
*Sugestões de Compra*  
*Drivers, BIOS e Manuais*

#### **Linux:**

*Linux Geral*  
*Instalação e configuração*  
*Suporte a hardware e drivers*  
*Aplicativos, produtividade e multimídia*  
*Compatibilidade com aplicativos Windows*  
*Servidores Linux*

#### **Software e Redes:**

*Windows e Programas*  
*Redes, Servidores e acesso à web*  
*Mac e Apple*

#### **Multimídia:**

*Placas 3D*  
*Video, Codecs e DVD*  
*Gravação de CDs e DVDs*

**Participe você também:**

<http://guiadohardware.net/comunidade/>

Já visitou o  
**Guiadohardware.NET** hoje?

acesse:

<http://guiadohardware.net>

Seu verdadeiro guia de informação na internet

