

## Tipos e Formatos de módulos de memória

Embora seja brutalmente mais rápida que o HD e outros periféricos, a memória RAM continua sendo muito mais lenta que o 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, evoluindo das memórias FPM usadas nos micros 486, até as DDR2 usadas atualmente. O aspecto mais visível são os diferentes formatos de módulos de memória, sobre os quais aprenderemos mais neste tutorial.

*Texto escrito por Carlos E. Morimoto em 09/03/2007.*

*Adaptação: Amarildo Carneiro – 13/04/2008.*

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.

Ao contrário do processador, que é extremamente complexo, os chips de memória são formados pela repetição de uma estrutura bem simples, formada por um par de um transistor e um capacitor. Um transistor solitário é capaz de processar um único bit de cada vez, e o capacitor permite armazenar a informação por um certo tempo. Esta simplicidade faz com que os pentes de memória sejam muito mais baratos que os processadores, principalmente se levarmos em conta o número de transistores.

Um pente de 1 GB é geralmente composto com 8 chips, cada um deles com um total de 1024 megabits, o que equivale a 1024 milhões de transistores. Um Athlon 64 X2 tem "apenas" 233 milhões e custa bem mais caro que um pente de memória.

Existem basicamente dois tipos de memória em uso: SDR e DDR. As SDR são o tipo tradicional, onde o controlador de memória faz apenas uma leitura por ciclo, enquanto as DDR são mais rápidas, pois fazem duas leituras por ciclo. O desempenho não chega a dobrar, pois o acesso inicial continua demorando o mesmo tempo, mas melhora bastante.

Os pentes de memória SDR são usados em micros antigos: Pentium II e Pentium III e os primeiros Athlons e Durons. Por não serem mais fabricados, eles são atualmente muito mais raros e caros que os DDR, algo semelhante ao que aconteceu com os antigos pentes de 72 vias, usados na época do Pentium 1.

É fácil diferenciar os pentes SDR e DDR, pois os SDR possuem dois chanfros e os DDR apenas um. Essa diferença faz com que também não seja possível trocar as bolas, encaixando por engano um pente DDR numa placa-mãe que use SDR e vice-versa (a menos que você use um alicate e um martelo, mas a placa provavelmente não vai funcionar mais depois!).

Algumas poucas placas possuem os dois tipos de soquete, permitindo usar os dois tipos de acordo com a conveniência, mas sem misturar os dois.

Mais recentemente temos assistido a uma nova migração, com a introdução dos pentes de memória DDR2. Nelas, o barramento de acesso à memória trabalha ao dobro da frequência dos chips de memória propriamente ditos. Isso permite que sejam realizadas duas operações de leitura por ciclo, acessando dois endereços diferentes.

Como a capacidade de realizar duas transferências por ciclo introduzida nas memórias DDR foi preservada, as memórias DDR2 são capazes de realizar um total de 4 operações de leitura por ciclo,

uma marca impressionante :). Existem ainda alguns ganhos secundários, como o menor consumo elétrico, útil em notebooks.

Os pentes de memória DDR2 são incompatíveis com as placas-mãe antigas. Eles possuem um número maior de contatos (um total de 240, contra 184 dos pentes DDR) e o chanfro central é posicionado de forma diferente, de forma que não seja possível instalá-los nas placas antigas por engano. Muitos pentes são vendidos com um dissipador metálico, que ajuda na dissipação do calor e permite que os módulos operem a frequências 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.

Mas, 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:

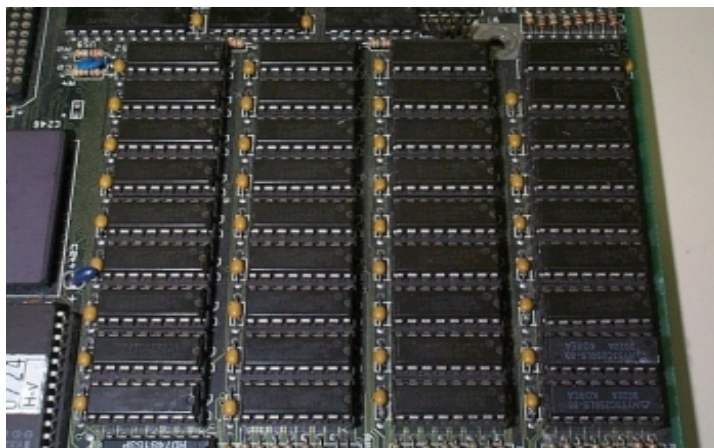
- **Quanto à tecnologia usada (FPM, EDO, SDRAM, DDR, DDR2, RDRAM)**
- **Quanto ao formato usado (DIP, SIMM, DIMM)**

Neste tutorial vamos aprender sobre os formatos de memórias usados nos micros PC:

## **Módulos DIP**

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:

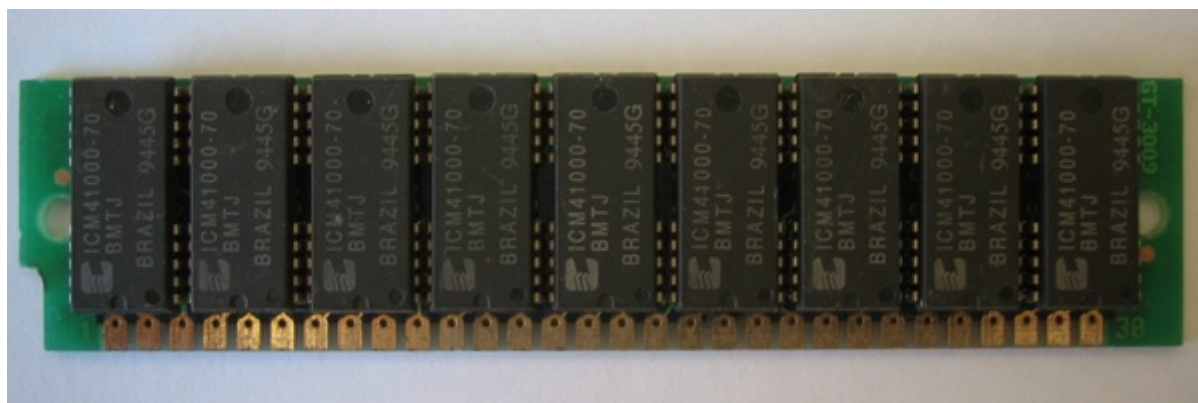


## Módulos de 30 vias

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. 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.



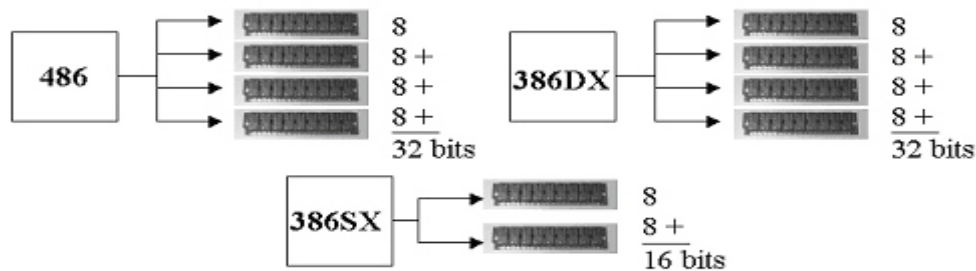
Estes módulos de 30 vias possuíam sempre 8 ou 9 chips de memória. Cada chip fornecia um único bit de dados em cada transferência, de forma que 8 deles formavam um módulo capaz de transferir 8 bits por ciclo. No caso dos módulos com 9 chips, o último era destinado a armazenar os bits de paridade, que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

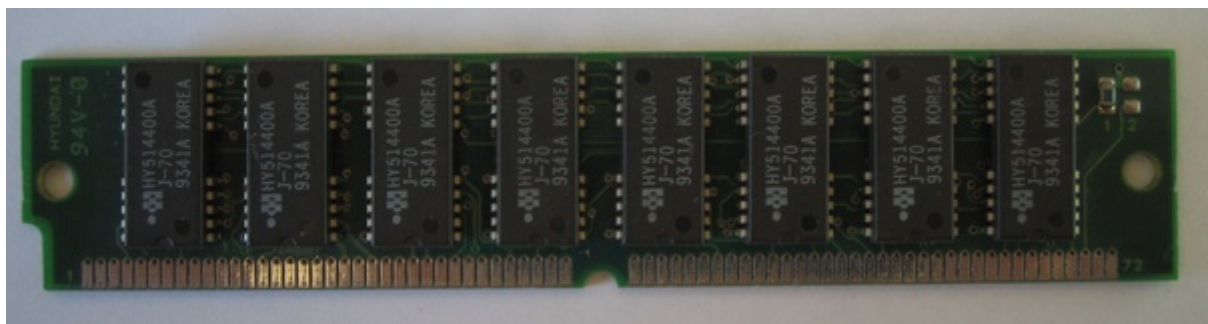
A exceção ficava por conta dos micros equipados com processadores 386SX, onde são necessários apenas 2 módulos, já que o 386SX acessa a memória usando palavras de 16 bits.



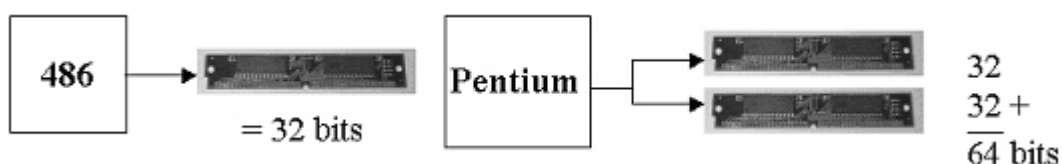
Apesar de serem muito mais práticos do que os chips DIP, os módulos SIMM de 30 vias ainda eram bastante inconvenientes, já que era preciso usar 4 módulos idênticos para formar cada banco de memória. Eles foram desenvolvidos pensando mais na questão da simplicidade e economia de custos do que na praticidade.

## Módulos de 72 vias

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.



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.

## Módulos DIMM

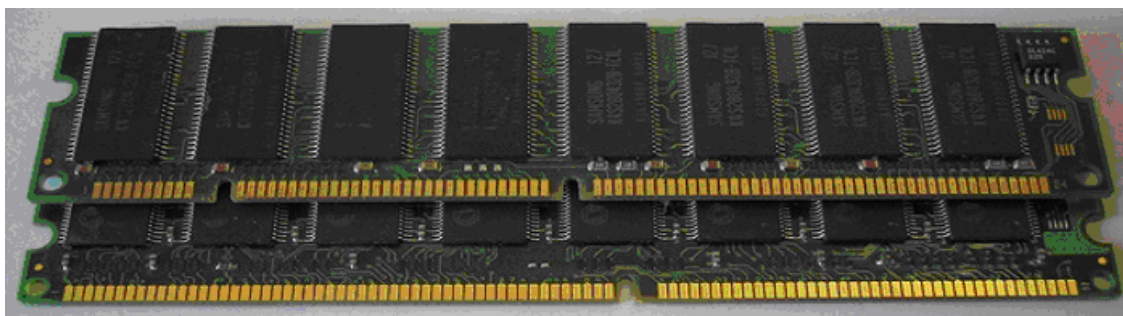
Finalmente, temos os módulos **DIMM**, usados atualmente. Ao contrário dos módulos SIMM de 30 e 72 vias, os módulos DIMM possuem contatos em ambos os lados do módulo, o que justifica seu nome, “Double In Line Memory Module” ou “módulo de memória com duas linhas de contato”.

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 onboard, 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 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:

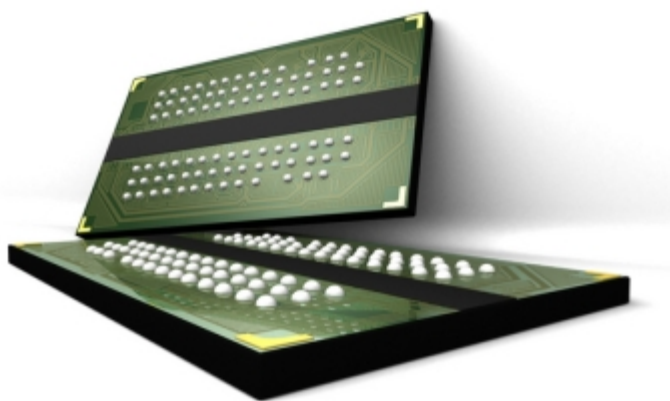


Os módulos DDR2 também utilizam um único chanfro, mas ele está posicionado mais à esquerda que o usado nos módulos DDR, de forma que é novamente impossível encaixar um módulo DDR2 numa placa antiga:



Isto é necessário, pois além das mudanças na forma de acesso, os pentes DDR2 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 surgiram no mercado alguns pentes de memória DDR que também utilizam chips BGA, mas eles são menos comuns.

Como os módulos DDR2 trabalham a frequências mais altas, o uso de dissipadores se tornou mais comum. Eles não são realmente necessários, mas a melhor dissipação do calor permite que o módulo trabalhe a frequências mais altas, por isso eles se tornaram norma nos módulos DDR2 de alto desempenho e, principalmente nos módulos "premium", destinados a overclock. Alguns fabricantes chegam a utilizar heat-pipes ou a oferecer coolers ativos, que podem ser instalados sobre os módulos.



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.

## Módulos RDRAM – Rambus DRAM

Rambus Dynamic Random Access Memory - Memória de Acesso Aleatório Dinâmica da Rambus  
Tecnologia de memória criada pela empresa Rambus.

Existem três tecnologias de memória RDRAM: Base RDRAM, Concurrent RDRAM e Direct RDRAM. As duas primeiras tecnologias foram usadas somente para a fabricação de memórias de vídeo, enquanto que a terceira pode ser usada como memória RAM do PC.

A tecnologia Rambus consiste em transmissões de poucos bits por vez (ex: 16 bits) porém com um clock muito elevado.

As memórias Direct RDRAM são soldadas em módulos chamados RIMM, que RIMM são classificados da seguinte forma:

- 1ª geração (16 bits): PC600 ou RIMM 1200 (1.200 MB/s), PC700 ou RIMM 1400 (1.400 MB/s) e PC800 ou RIMM 1600 (1.600 MB/s).
- 2ª geração (32 bits): PC1066 ou RIMM 4200 (4.200 MB/s).
- 3ª geração (64 bits): PC1333 ou RIMM 11G (10,7 GB/s).

Só é possível instalar módulos RIMM em placas-mãe que aceitam este tipo de memória. São poucas as placas-mãe com este recurso.

Normalmente a configuração usada é de dois canais, fazendo com que a taxa de transferência seja o dobro da nominal caso sejam usados dois módulos. Por exemplo, em uma placa-mãe usando o chipset Intel 850 e com dois módulos PC800 instalados, a taxa de transferência da memória será de 3.200 MB/s (1.600 MB/s x 2), já que este chipset usa a configuração de dois canais.

Esta tecnologia necessita de terminação resistiva. Por este motivo, todos os soquetes RIMM da

placa-mãe tem de obrigatoriamente estar preenchidos. No caso de não haver módulos de memória suficientes para preencher todos os soquetes, deve-se instalar um módulo chamado C-RIMM (Continuity RIMM), que é um módulo "vazio" responsável por fechar o circuito resistivo.

Atualmente as memórias RDRAM não são mais usadas, pois os fabricantes tinham que pagar royalties para a Rambus. Com o lançamento das memórias DDR-SDRAM o uso das memórias RDRAM perdeu o sentido.

## Módulos para notebooks

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:



## 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.

Esta simplicidade tem um custo, que é uma limitação no número de chips por módulos também no número de módulos que podem ser instalados na mesma placa mãe. Salvo raras exceções, os módulos unbuffered possuem no máximo 16 chips de memória e é possível projetar placas mãe com suporte para até 4 módulos.

Isto não é um problema nos desktops, onde normalmente não precisamos de mais do que 2 ou 4 GB de RAM, mas é uma grave limitação nos servidores, onde é comum o uso de mais memória.

Os módulos registered incluem chips adicionais (registradores) que funcionam como uma interface adicional entre o controlador e os chips. Eles permitem que o controlador suporte um número maior de módulos de memória e também que sejam usados módulos com mais chips, permitindo a instalação de quantidades muito maiores de memória. Muitas placas para servidores incluem 8 slots de memória, e existem módulos registered com 32 ou até mesmo 48 chips (sem contar os chips adicionais no caso dos módulos com ECC). É fácil reconhecer os módulos registered, devido à presença dos chips adicionais:



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.

## 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 absolvidos 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:

