

Lista 1 de Arquitetura de Computadores (BCC263) - Parte B

Matheus Peixoto Ribeiro Vieira - 22.1.4104

O vídeo “How does Computer Memory Work?” do canal “Branch Education” explica, de uma forma muito completa e bem detalhada, o funcionamento da memória de acesso aleatório, mais conhecida como memória RAM.

Em sua introdução, é explicado a importância da memória RAM para armazenar informações que são usadas pelo processador, mas que estão na memória secundária, muito mais lenta do que na principal. Assim, é comparada a transferência do SSD com a RAM diretamente para o processador com uma tartaruga e um avião a jato, respectivamente.

Em seguida é explicado que o processador pode mapear e transferir dados entre a memória principal e secundária quando for necessário, e que a conexão entre todos os componentes é feito pela placa-mãe, onde que, para uma memória DDR5, a memória se divide em canais A e B, onde cada um pode enviar, independentemente, 32 bits de dados (Imagem 1).

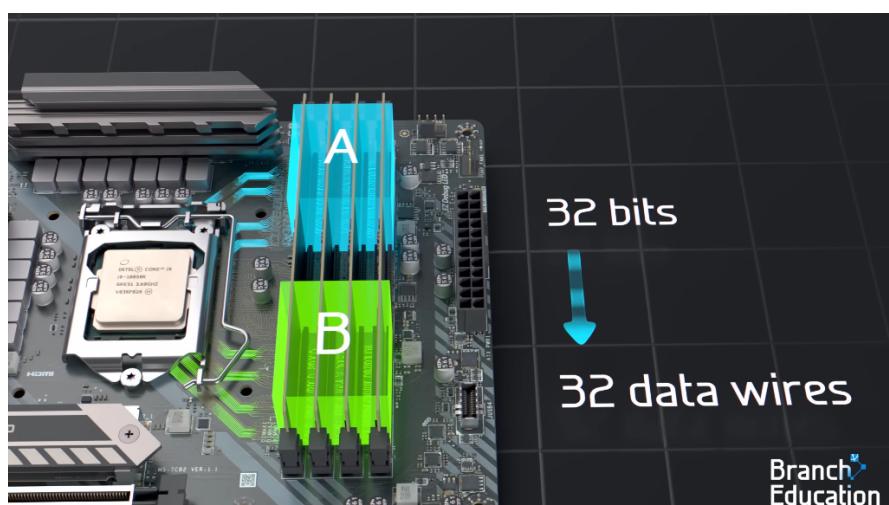


Imagen 1 - Dual channel

Na próxima etapa, é mostrado algo mais oculto, o funcionamento de um dos chips da memória (Imagen 2) - o funcionamento se replica para os demais -, que é onde os dados realmente ficam salvos (Imagen 3). Dessa forma, dentro deles, há duas matrizes de interconexão, que conectam a matriz de bola com a matriz onde os dados são, de fato, gravados e lidos em 8 grupos de bancos, divididos em 4 partes cada, totalizando 32 bancos, onde cada um deles possui milhões de células.

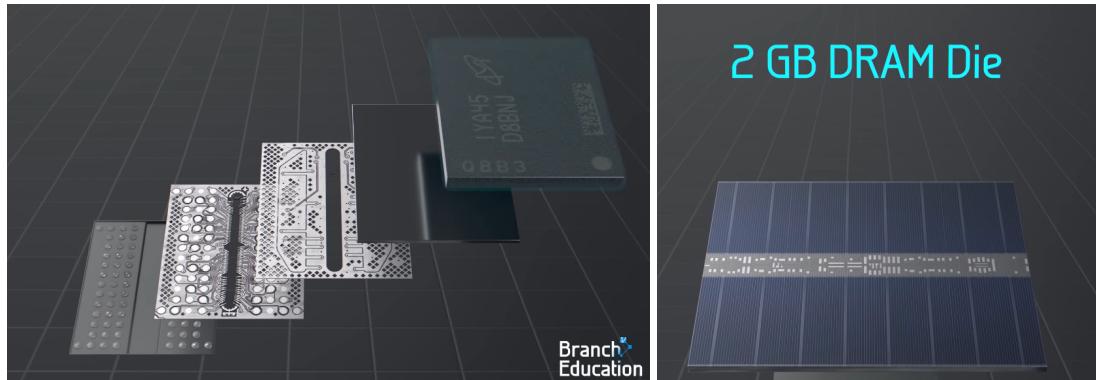


Imagen 2 - Chip de memória

Imagen 3 - Matriz

Assim, é mostrado, de forma bem ilustrativa e explicativa, o processo onde as bilhões de células são selecionadas para que um dado seja armazenado. Sendo que, para isso, é necessário receber 31 bits de informação, com 3 deles selecionando o grupo de bancos, 2 para selecionar o banco, 16 para a linha e 10 para a coluna. Pode-se ver, na imagem 4, uma representação de um banco.

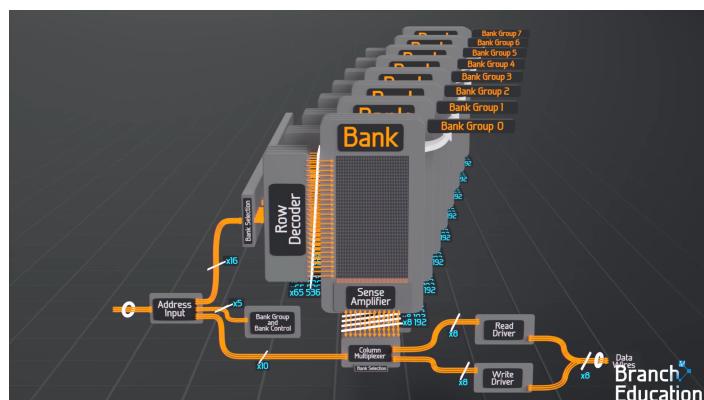


Imagen 4 - Representação de um banco

Sequentemente, o espectador é apresentado à célula de memória, conhecida como 1T1C (Imagen 5), formada por um capacitor que irá representar a carga e um transistor, que irá permitir a leitura e a escrita dos dados.

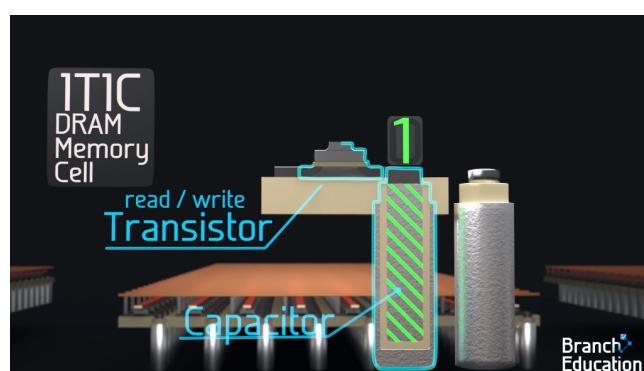


Imagen 5 - Célula de memória

Assim, para acessar as informações corretamente, ainda é necessário um fio de linha chamado de wordline, que ativa ou desativa o transistor para a passagem de energia, e um fio de coluna conhecido como bitline, que transfere energia para o capacitor através de um canal controlado pelo transistor (Imagen 6).

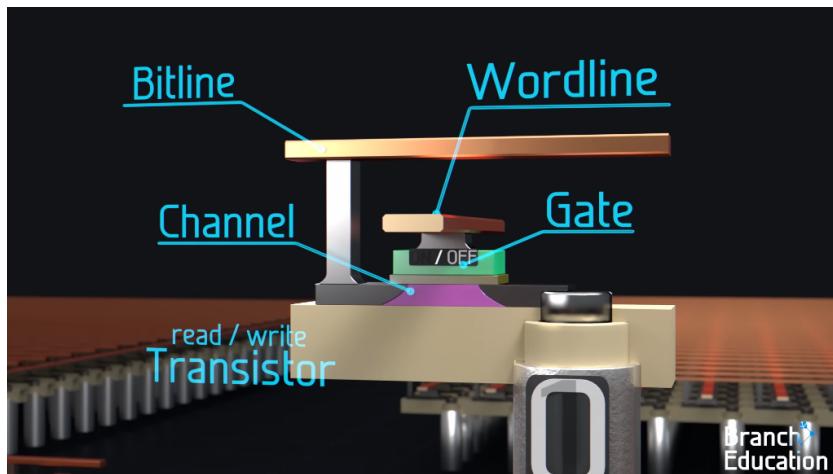


Imagen 6 - Bitline, Wordline, Canal e Transistor

Aplicando tensão ao wordline, o transistor permite que elétrons passem pelo canal e vá até o bitline, permitindo o acesso ao capacitor, e deixando-o com carga (1) ou descarregado (0).

Em seguida, após a apresentação de todos os conceitos e como é a estrutura da memória RAM, é, finalmente, apresentada a leitura e escrita de dados.

Dessa forma, quando os 31 bits são recebidos e tratados, o banco é selecionado, em seguida, toda a linha desejada recebe uma carga, causando uma oscilação no transistor, permitindo um pouco de passagem de carga.

Logo em sequência, todos os bitlines são ativados e os dados começam a percorrer até encontrarem um amplificador que verifica a carga da linha que começa com 0.5V. Se a carga aumentar, então o sinal é 1, o capacitor recarrega e a informação é transmitida. Quando o valor da é 0, o capacitor recebe uma pequena carga, que logo será perdida, e o amplificador para de fornecer energia.

No final, os dados chegam a um multiplexador de coluna, que recebe todos os 0 e 1 e aceita os da coluna solicitada, enviando a informação para o processador.

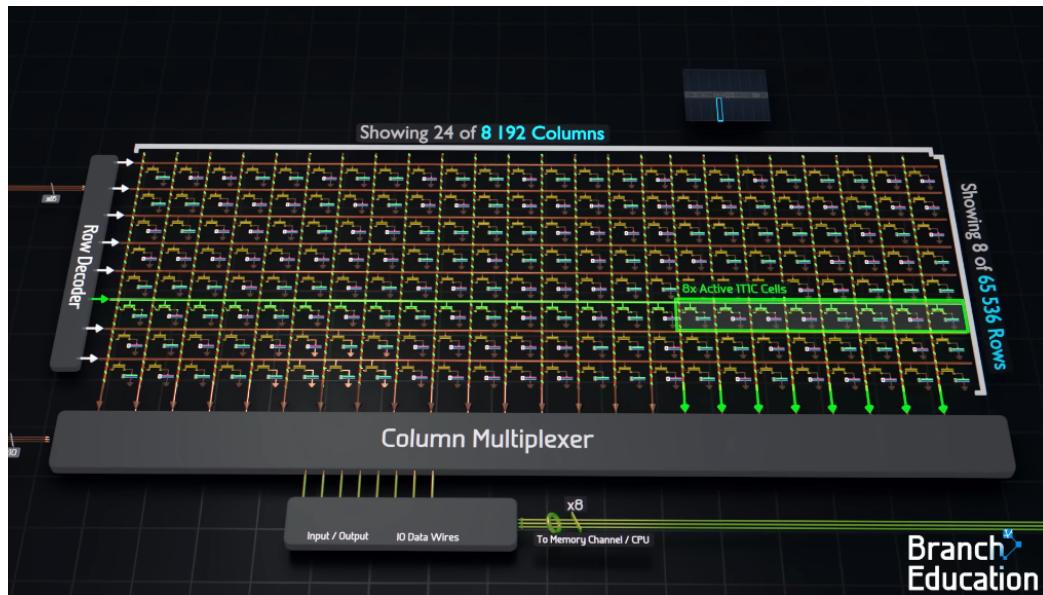


Imagen 7 - Matriz de células da memória

Após a explicação da mágica de seleção de informações, é mostrado como funciona o processo inverso, a escrita, que possui um multiplexador de colunas mais poderoso que o de leitura, podendo, assim, sobrescrever os dados.

Todavia, os capacitores e transistores são muito pequenos e, após um pequeno tempo, a sua carga baixa. Assim, foi mostrado que, em uma situação real, a perda é compensada por um valor de energia um pouco maior.

Assim, sabendo que os capacitores precisam de energia para manterem os seus dados, é explicado o processo de carregamento dos mesmos, algo que ocorre 16 vezes por segundo.

Também é explicado que quando um dado é acessado a partir de uma linha já aberta para a leitura, isto é chamado de row hit, o que ajuda a economizar tempo, já que não é feito todo o processo de procura e abertura de transistores. Mas quando um dado não é encontrado na mesma linha, isso é chamado de row miss.

Sobre as medidas de tempo (Imagen 8), é explicado rapidamente sobre a latência de recebimento de dados da coluna (CAS Latency); o tempo para isolar uma linha (RAS to CAS); o tempo para carregar uma célula antes de abrir (Precharge tRP) e o tempo de ativamento de linha antes da pré carga (Activate to Precharge Delay).

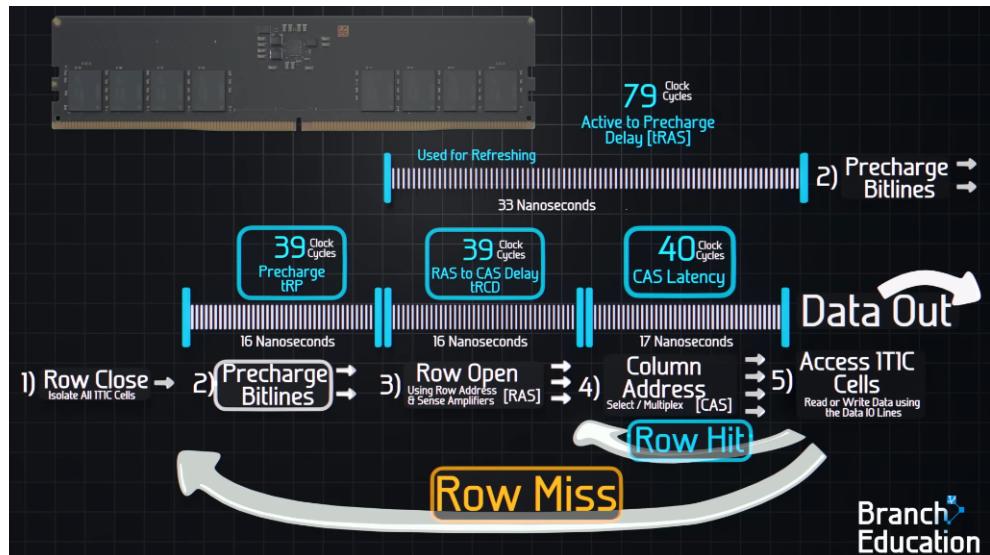


Imagen 8 - Medidas de tempo e siglas

Também é mostrado algumas técnicas de otimização, e entre elas, pode-se citar o Burst Buffer (Imagen 9), que lê 128 colunas de uma vez e manda as informações necessárias para a CPU, sendo que ele é muito válido para maiores quantidades de row hit, mas funciona caso encontre somente uma palavra. Dessa forma, a escrita funciona da mesma forma, só que ao contrário.

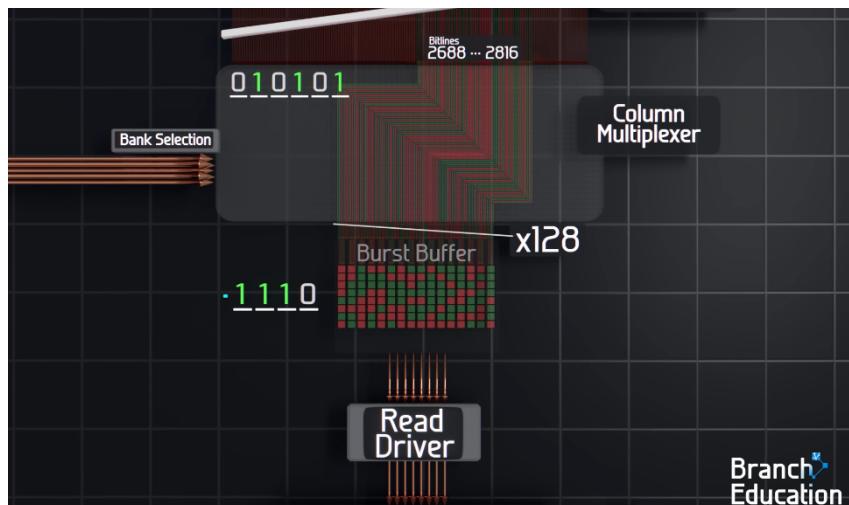


Imagen 9 - Burst buffer

Em suma, pode-se concluir que o vídeo apresenta de uma forma bem clara e demonstrativa os principais conceitos e funcionamento de uma memória RAM, um item extremamente importante para todos os tipos de computadores.