

Aula 12

A célula de memória dinâmica

Uma célula de armazenamento DRAM muito utilizada esta mostrada na figura A.12.1. A célula consiste de um único transistor MOS canal n, conhecido como *transistor de acesso* (*access transistor*), e um capacitor de armazenamento. A célula é conhecida apropriadamente pelo nome de **célula de um transistor**. A porta do transistor é conectada à linha de palavra e sua fonte (dreno) é conectada à linha de bit. Observe que apenas uma linha de bit é usada em DRAM's, em contraste ao caso das SRAM's em que linhas de *bit* e *bit* são utilizadas.

A célula DRAM armazena seu bit de informação como carga no capacitor C_S . Quando a célula está armazenando um 1, o capacitor está carregado com $(V_{DD} - V_T)$; quando um 0 está armazenado, o capacitor é descarregado à tensão zero. Devido as correntes de fuga a célula precisa ser periodicamente regenerada.

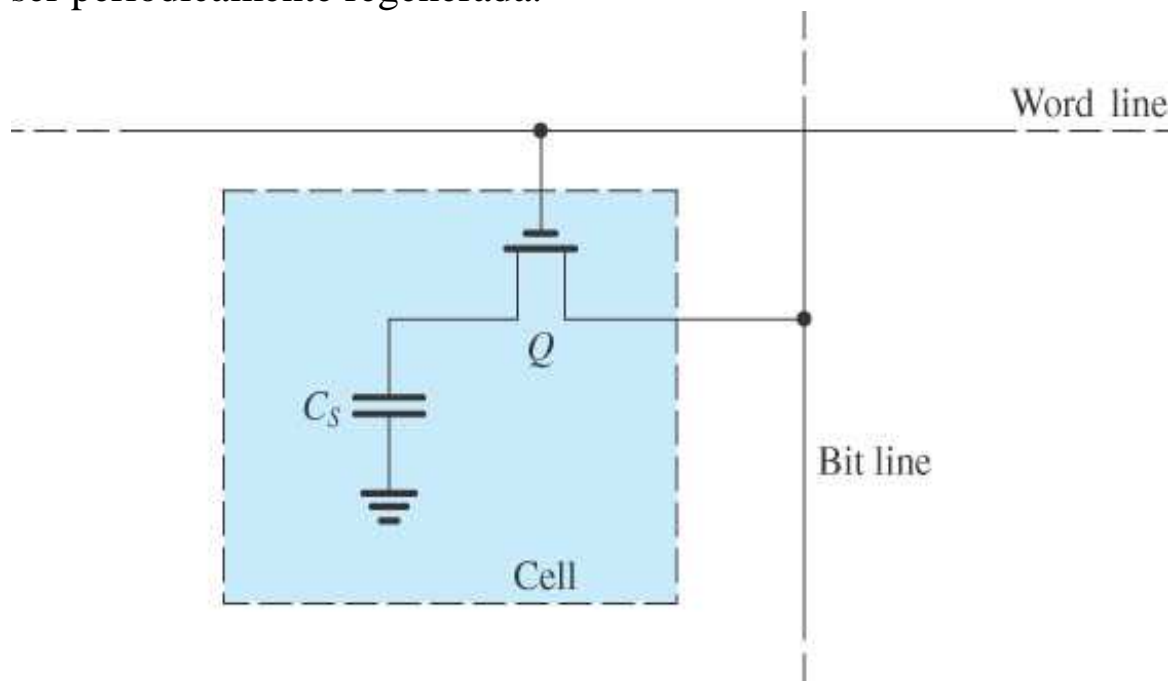


Figura A.12.1 A célula de memória RAM dinâmica de um transistor.

Funcionamento da célula DRAM

Como na RAM estática, o decodificador de linha seleciona uma linha particular elevando a tensão de sua linha de palavra. Isso faz com que todos os transistores de acesso da linha selecionada conduzam, conectando os capacitores de armazenamento de todas as células na linha selecionada com as suas respectivas linhas de *bit*. Portanto, cada capacitor de célula C_S é conectado em paralelo com a correspondente capacitância da linha de bit C_B , como indicado na figura A.12.2. Deve-se notar que C_S é da ordem de 30 a 50 fF, enquanto C_B é cerca de 30 a 50 vezes maior.



Figura A.12.2 Circuito equivalente quando a tensão da linha de palavra é selecionada.

Operação de leitura

Se a operação é de leitura, a linha de bit é pré-carregada com $V_{DD}/2$. Para determinar a variação na tensão da linha de bit que resulta da conexão de um capacitor de célula C_S a ela, considere a tensão inicial do capacitor da célula como V_{CS} ($V_{CS} = V_{DD} - V_T$ quando 1 está armazenado e $V_{CS} = 0$ V quando 0 está armazenado). Usando a lei de conservação de cargas, podemos escrever

$$C_S V_{CS} + C_B \frac{V_{DD}}{2} = (C_B + C_S) \left(\frac{V_{DD}}{2} + \Delta V \right)$$

da qual podemos obter para ΔV

$$\Delta V = \frac{C_S}{C_B + C_S} \left(V_{CS} - \frac{V_{DD}}{2} \right) \quad (\text{A.12.1})$$

e como $C_B \gg C_S$,

$$\Delta V \cong \frac{C_S}{C_B} \left(V_{CS} - \frac{V_{DD}}{2} \right) \quad (\text{A.12.2})$$

Agora, se a célula está armazenando um 1, $V_{CS} = V_{DD} - V_T$ e

$$\Delta V(1) \cong \frac{C_S}{C_B} \left(\frac{V_{DD}}{2} - V_T \right) \quad (\text{A.12.3})$$

e se a célula está armazenando um 0, $V_{CS} = 0$ e

$$\Delta V(0) \cong -\frac{C_S}{C_B} \left(\frac{V_{DD}}{2} \right) \quad (\text{A.12.4})$$

Como geralmente C_B é muito maior de que C_S , essas tensões de leituras são muito pequenas. Por exemplo, para $C_B = 30.C_S$, $V_{DD} = 5 \text{ V}$ e $V_T = 1,5 \text{ V}$, $\Delta V(0) = -83 \text{ mV}$ e $\Delta V(1) = 33 \text{ mV}$.

Observe também que um processo de leitura é destrutivo, já que a tensão sobre C_S não será mais $(V_{DD} - V_T)$.

A variação na tensão da linha de bit é detectada e amplificada pelo amplificador sensor da coluna. O sinal amplificado é então aplicado ao capacitor de armazenamento, restaurando assim seu sinal ao nível apropriado ($V_{DD} - V_T$ ou 0 V). Dessa forma, todas as células na linha selecionada são restauradas. Simultaneamente, o sinal na saída do amplificador sensor da coluna selecionada é levado à linha de dados de saída do chip pela ação do decodificador de coluna.

Operação de escrita

A operação de escrita acontece de forma similar à operação de leitura, exceto que o bit de dados que deve ser escrito e que está aplicado à linha de entrada de dados é também aplicado pelo decodificador de coluna à linha de bit selecionada. Assim, se um bit de dado a ser escrito for 1, a tensão da linha B será elevada para V_{DD} (isto é, C_B é carregado até V_{DD}). Quando o transistor de acesso de uma célula particular é ligado, seu capacitor C_S é carregado até $V_{DD} - V_T$; assim 1 é escrito na célula. De forma simultânea, todas as outras células da linha selecionada são simplesmente restauradas em uma operação de restauração.

Embora as operações de escrita e leitura resultem em uma restauração (*refresh*) automáticas de todas as células da linha selecionada, deve-se fazer uma restauração periódica da memória inteira a cada 5 a 10 ms, de acordo com a especificação do chip.

Amplificadores sensores e Decodificador de endereço

O projeto desses circuitos, comumente chamados *circuitos periféricos* da memória, apresenta desafios e desperta grande interesse para os projetistas de circuitos integrados: melhorar o desempenho dos circuitos periféricos pode resultar em chip de memórias mais rápidas e com maior capacidade e ainda dissipando menos potência.

O amplificador sensor

Em ordem de importância, depois das células de memórias, o amplificador sensor é o bloco mais crítico em um chip de memória.

Há várias configurações de amplificadores sensores em uso. Descrevemos aqui um amplificador sensor diferencial que emprega realimentação positiva. O amplificador deve receber da célula de memória um tensão diferencial da ordem de 30 mV a 500 mV, dependendo do tipo de memória e topologia da célula. O amplificador sensor, por sua vez, responde gerando um sinal com chaveamento lógico máximo (0 a V_{DD}) em seus terminais de saída. O circuito que apresentaremos possui uma propriedade pouco como: **seus terminais de entrada e saída são os mesmo.**

Antes de apresentarmos o circuito do amplificador sensor vamos repetir aqui uma célula de memória indicada na figura A.12.3. O amplificador sensor ‘sentir’ e definir as tensões nas linhas de bit. Assim, a propriedade acima é extremamente interessante na estrutura da memória como um todo.

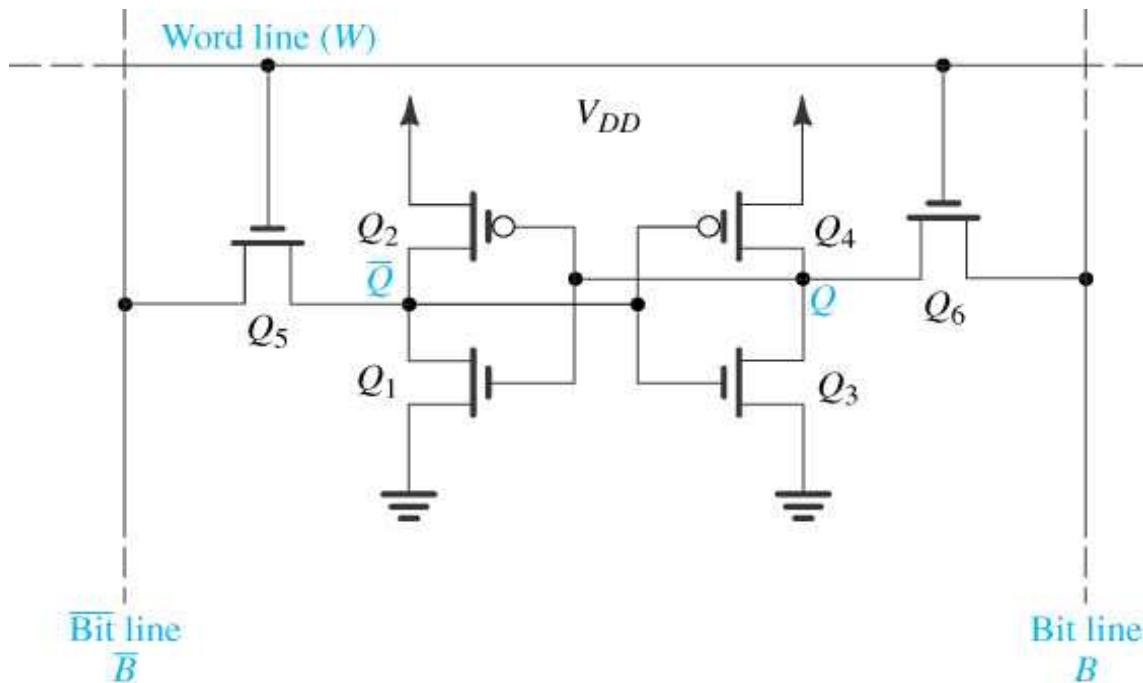


Figura A.12.3 Célula de memória

Um amplificador sensor com realimentação positiva

A figura A.12.4 mostra o amplificador sensor juntamente com outros circuitos periféricos de um chip de memória. Observe que o amplificador sensor nada mais é que um *latch* formado pelo acoplamento cruzado de dois inversores CMOS: um inversor é implementado pelos transistores Q_1 e Q_2 e o outro, pelos transistores Q_3 e Q_4 . Os transistores Q_5 e Q_6 atuam como chaves que conectam o amplificador sensor ao terra e ao VDD somente quando uma de ação de sensoriamento de dados é necessária. Se esse não for o caso, ϕ_S é baixo e o amplificador sensor é desligado. Isto conserva consumo de potência, uma característica importante, já que geralmente há um amplificador sensor por coluna (outra razão para formado retangular da memória), resultando em milhares de amplificador sensores por chip. Observe novamente que os terminais x e y são tanto de entrada e de saída do amplificador.

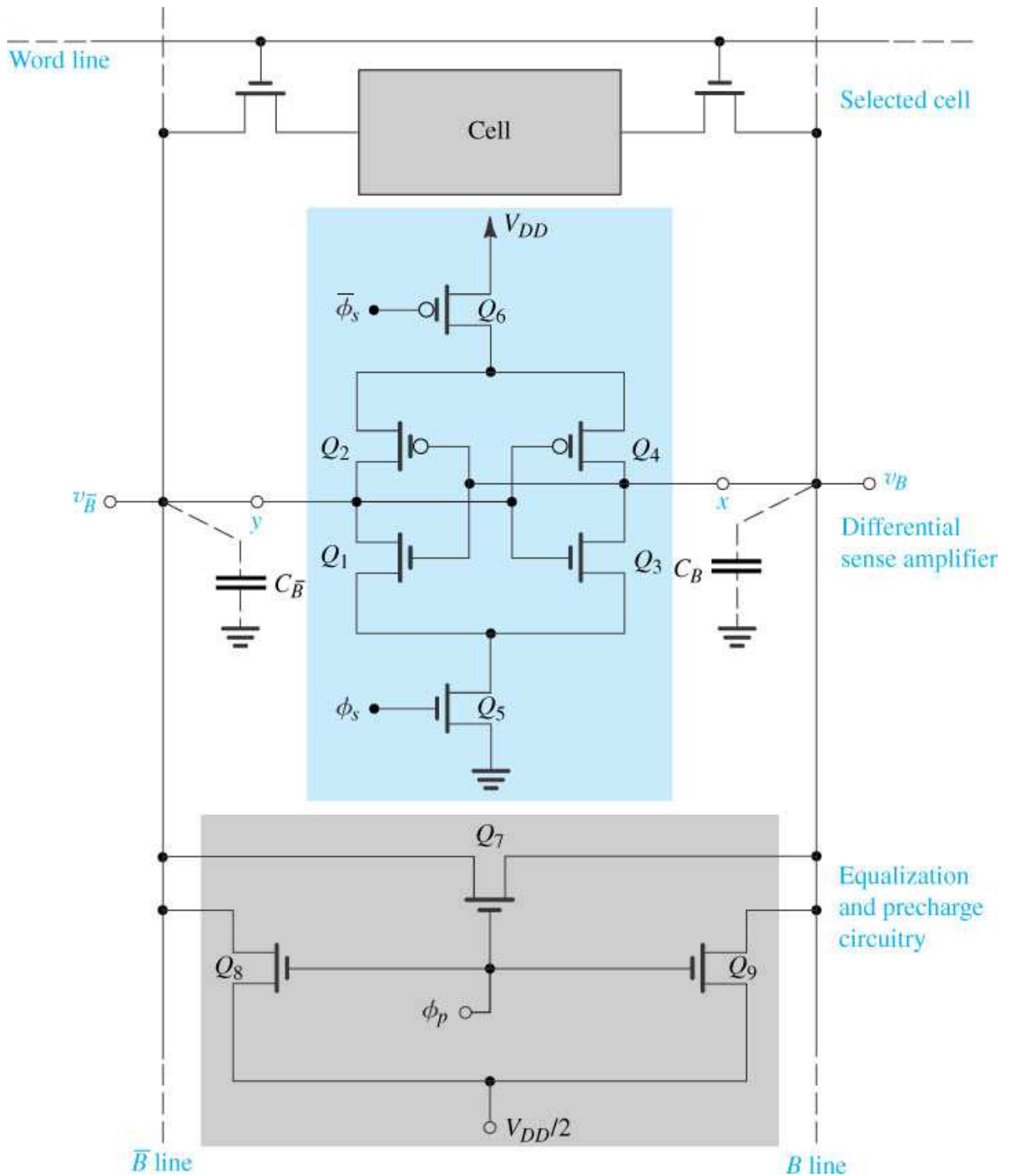


Figura A.12.4 Um amplificador sensor diferencial conectado às linhas de bit de uma coluna particular.

Funcionamento do Amplificador sensor

O amplificador deve ser capaz de detectar um pequeno que surge entre B e \overline{B} . O amplificador amplifica esse sinal e fornece de volta (saída) um sinal com uma excursão cheia (0 a V_{DD}). Por exemplo, se, durante uma operação de leitura, a célula tinha 1 armazenado, então uma pequena tensão positiva surgirá entre B e \overline{B} , com v_B maior que $v_{\overline{B}}$. O amplificador então fará com que v_B suba para V_{DD} e $v_{\overline{B}}$ desça para 0 V. Essa saída 1 é então direcionada para o pino de E/S do chip pelo decodificador de coluna (não mostrado) e ao mesmo tempo é utilizada para reescrever (no caso de uma célula DRAM), executando assim a operação de restauração.

A figura A.12.4 mostra também o circuito de pré-carga e equalização. A operação desse circuito é simples e imediata: quando ϕ_P vai para nível alto antes da operação de leitura, todos os três transistores conduzem. Enquanto Q8 e Q9 pré-carregam as linhas B e \overline{B} em $V_{DD}/2$, o transistor Q7 acelera esse processo equalizando as tensões em ambas as linhas. Essa equalização é crítica para a operação adequada do amplificador sensor: qualquer diferença de tensão presente entre B e \overline{B} antes do início da operação de leitura pode resultar em uma interpretação errônea do amplificador sensor em relação a seu sinal de entrada. Na mesma figura, mostramos apenas uma das células da coluna em questão, justamente a célula cuja linha de palavra está acionada. Essa célula pode ser tanto uma célula SRAM como uma célula DRAM. É claro que todas as outras células nessa coluna não estarão conectadas às linhas B e \overline{B} .

Para detalhar mais o funcionamento dês amplificador, vamos considerar uma seqüência de eventos durante a operação de leitura:

1. O circuito de pré-carga é ativado ($\phi_p = 1$). Isso define uma tensão equalizada de valor $V_{DD}/2$ nas linhas B e \bar{B} . Quando o circuito de pré-carga é desativado ($\phi_p = 0$) as linhas B e \bar{B} ficam flutuando (alta impedância).
2. A linha de palavra vai para nível alto, conectando a célula às linhas B e \bar{B} , com v_B maior que $v_{\bar{B}}$ se a célula acessada estiver armazenando 1 ou com v_B menor que $v_{\bar{B}}$ se a célula estiver armazenando 0.
3. Com um sinal de tensão diferencial adequado, presente entre B e \bar{B} , o amplificador sensor é ligado conectando-o ao terra e a V_{DD} através de Q_5 e Q_6 , ativado pelo sinal de controle ϕ_s . Como, inicialmente, os terminais de entrada dos inversores estão em $V_{DD}/2$, esses inversores estarão operando em sua região de transição, na qual o ganho é elevado. Logo, inicialmente o *latch* estará operando em seu ponto de equilíbrio instável. Assim, dependendo do sinal entre os terminais de entrada, o *latch* se deslocará para um de seus dois pontos de equilíbrio estável. Isso, como já vimos, ocorre pela ação regenerativa do *lacth*. A figura A.12.5 ilustra claramente esse ponto apresentando as formas de onda do sinal na linha de bit tanto para uma operação de leitura 1 quanto para leitura 0

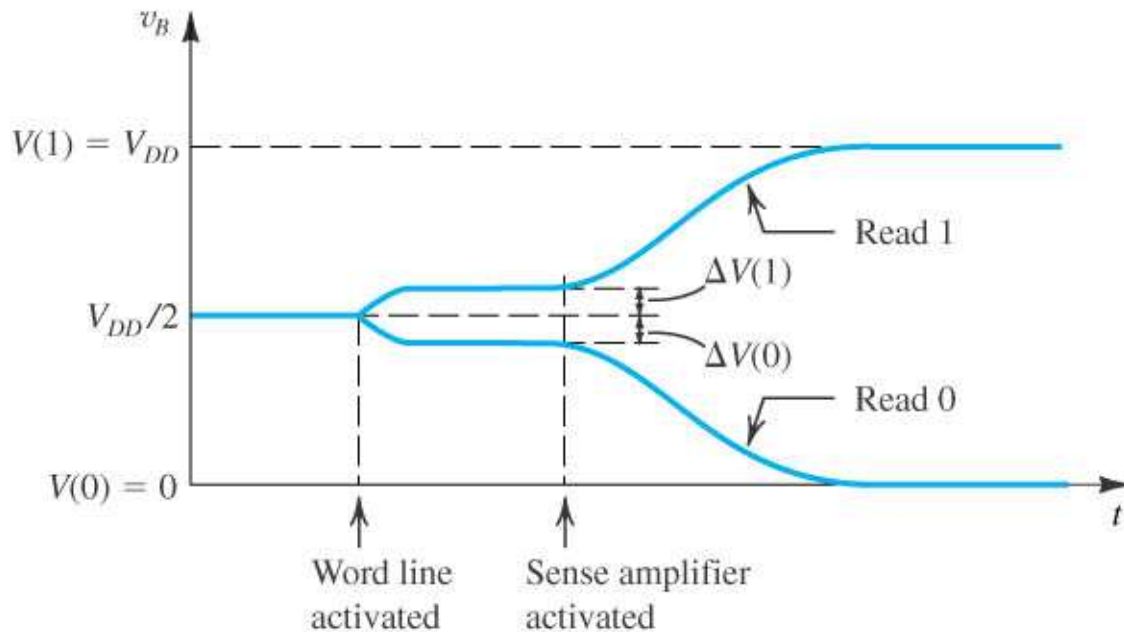


Figura A.12.5 Formas de onda em v_B antes e depois de ativar o amplificador sensor.

Operação em DRAM

O amplificador sensor descrito anteriormente é adequado para entradas diferenciais. Assim, dessa forma ele é rejeita sinais de modo comum (presente igualmente nas duas entradas). Como as células DRAM têm uma única saída os projetistas de CI inventaram um esquema engenhoso para resolver esse problema. O método é ilustrado na figura A.12.6

Basicamente, cada linha de bit é dividida em duas metades idênticas. Cada meia linha é conectada a metade das células na coluna e a uma célula adicional, conhecida como *célula fictícia* (*dummy cell*), que possui um capacitor de armazenamento $C_D = C_S$. Quando uma linha de palavra do lado esquerdo é selecionada para leitura, a célula

fictícia do lado direito (controlada por $\overline{\phi_D}$) também é selecionada. Dessa forma, a célula fictícia funciona como a outra metade da célula DRAM diferencial. Quando a meia linha de bit da esquerda está em operação, a meia linha de bit da direita atua como seu complemento (ou linha \overline{B} e vice-versa. A operação segue essa seqüência:

1. As duas metades da linha são pré-carregadas até $V_{DD}/2$ e uma linha de palavra é selecionada e a célula fictícia do outro lado é habilitada.
2. A meia linha conectada à célula selecionada gerará um pequeno incremento de tensão em torno de $V_{DD}/2$. Enquanto isso, a outra meia linha terá sua tensão mantida em $V_{DD}/2$.
3. O amplificador sensor é habilitado e detecta esta pequena diferença de tensão exatamente como numa memória SRAM.

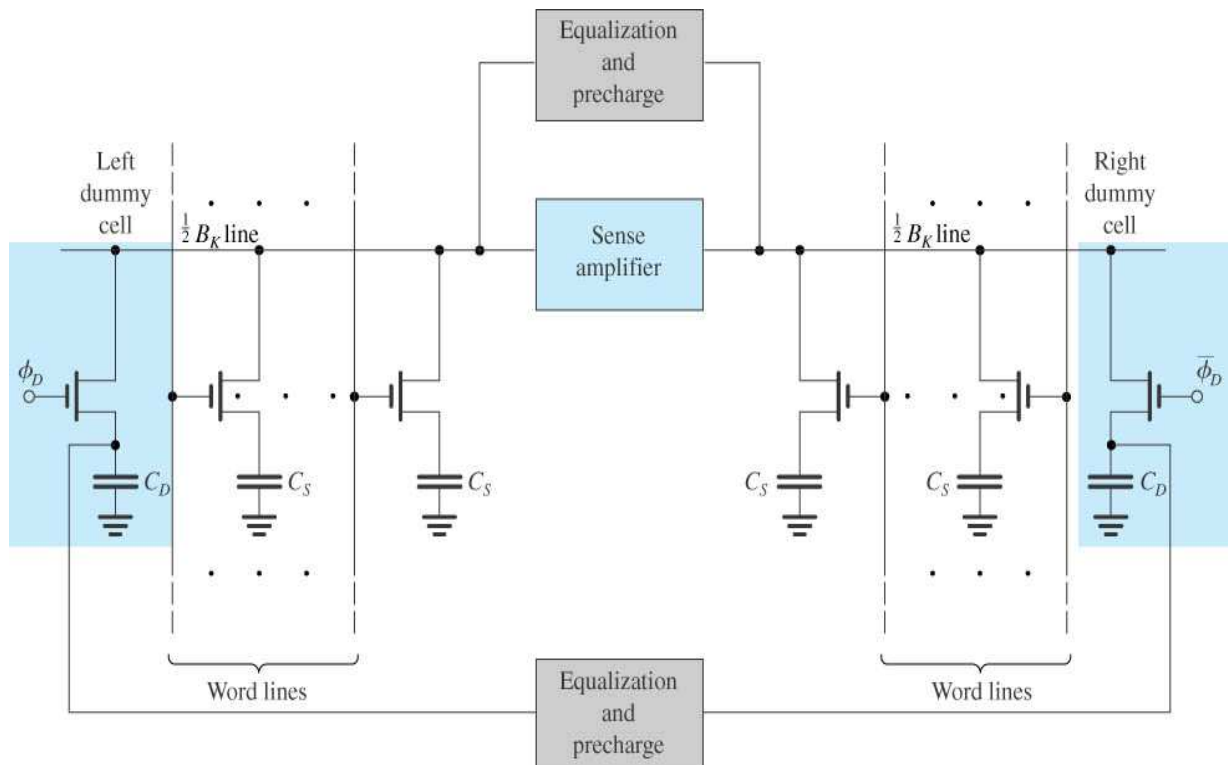


Figura A.12.6 Arranjo para obter a operação diferencial de uma célula DRAM.

O decodificador de endereço de linha

O decodificador de endereços de linha é necessário para selecionar uma dentre as 2^M linhas de palavras em resposta a um endereçamento de entrada de M bits. Como exemplo, considere o caso em que $M = 3$ e chame os três bits de A_0 , A_1 e A_2 e as oito linhas de palavras de W_0, \dots, W_7 . convencionalmente, a linha de palavra W_0 será ativada quando $A_0 = 0$, $A_1 = 0$ e $A_2 = 0$, de tal forma que podemos expressar W_0 como uma expressão booleana de A_0 , A_1 e A_2

$$W_0 = \overline{A_0} \overline{A_1} \overline{A_2} = \overline{A_0 + A_1 + A_2}$$

Assim, a seleção de W_0 pode ser obtida por uma porta NOU de três entrada, na qual estão ligados os bits A_0 , A_1 e A_2 e em cuja saída esta ligada a linha de palavra 0. A linha de palavra W_3 , por exemplo, será acionada quando $A_0 = 1$, $A_1 = 1$ e $A_2 = 0$, ou seja,

$$W_3 = A_0 A_1 \overline{A_2} = \overline{\overline{A_0} + \overline{A_1} + A_2}$$

Uma abordagem simples para realizar essas funções NOU é obtida pela estrutura matricial apresentada na figura A.12.7. O circuito mostrado é dinâmico. Anexado a cada linha possui um transistor de canal p associado que é ativado antes do processo de decodificação usando o sinal de pré-carga ϕ_p . Durante a pré-carga ($\phi_p = 0$), todas as linhas de palavras são puxada para V_{DD} . Supõe-se que nesse momento os bits de entrada de endereçamento ainda não foram aplicados e todas as entradas estão em nível baixo; portanto, não há necessidade de incluir no circuito o transistor de avaliação utilizado em portas lógicas dinâmicas. Então, a operação de decodificação se inicia quando os bits de endereçamento e seus complementos são aplicados. Note que os transistores NMOS são posicionados de tal forma que as linhas de

palavras não selecionadas sejam descarregadas. Para qualquer combinação de entrada, apenas uma linha de palavra não será descarregada e, portanto, sua tensão permanece alta, em V_{DD} . Por exemplo, a linha 0 estará em nível alto quando $A_0 = 0$, $A_1 = 0$ e $A_2 = 0$; essa é a única combinação que fará com que todos os três transistores conectados à linha 0 estejam cortados. Da mesma forma, a linha 3 possui transistores conectados a $\overline{A_0}$, $\overline{A_1}$ e A_2 e assim ela estará em nível alto apenas quando $A_0 = 1$, $A_1 = 1$ e $A_2 = 0$ e assim por diante. Depois que a saída do decodificador estabiliza, as linhas de saída são conectadas às linhas de palavra da matriz, geralmente via portas de transmissão controladas por relógios. Esse decodificador é conhecido como decodificador NOU. Observe que, em virtude da operação de pré-carga, esse circuito decodificador não dissipa potência estática.

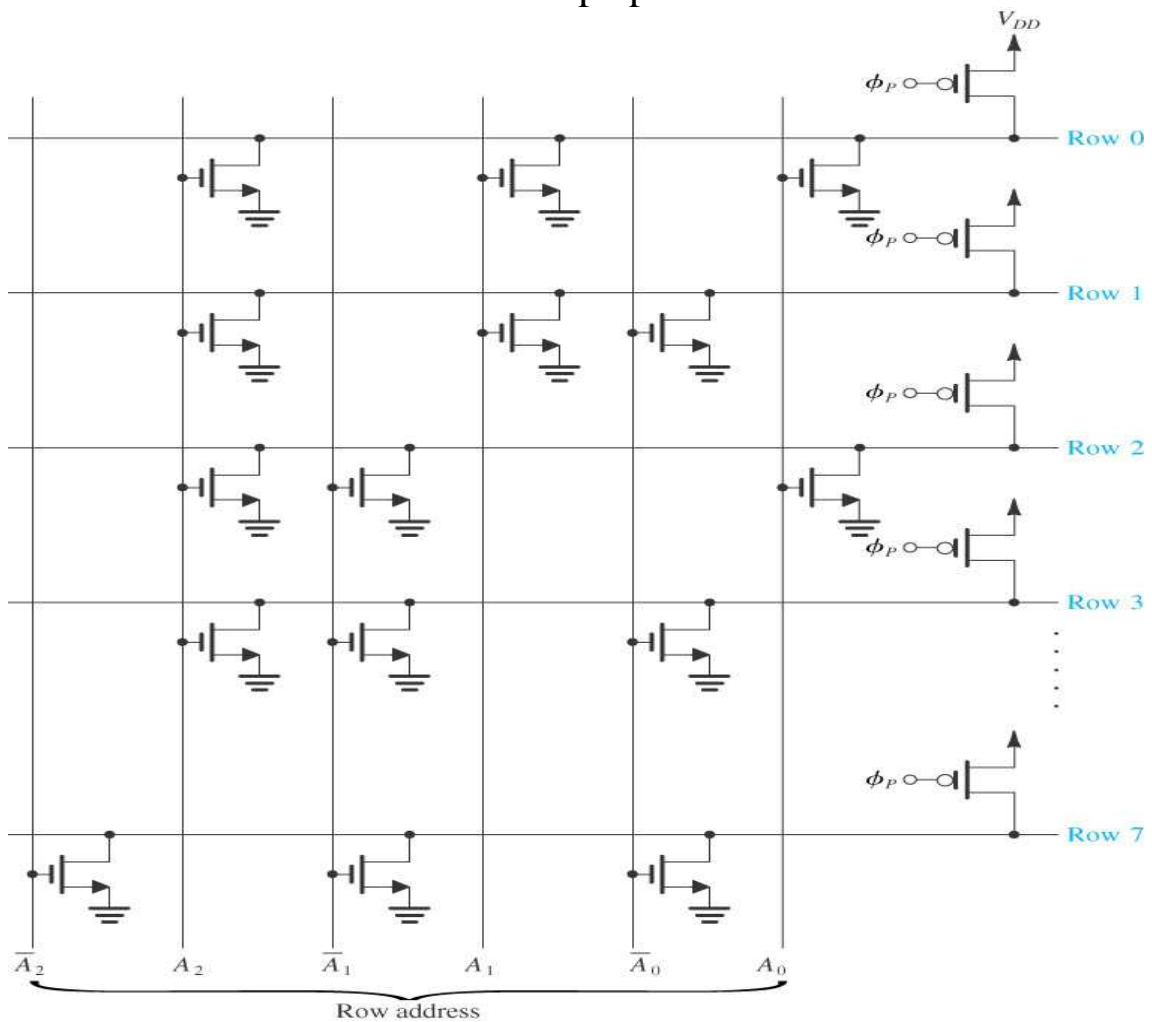


Figura A.12.7 Um decodificador de endereços NOU na forma de matriz.

O decodificador de endereço de coluna

A função do decodificador de endereços de coluna é conectar uma das 2^N linhas de bits à linha de dados *E/S* do chip. Como tal, ele é um multiplexador e pode ser implementado usando lógica de transistor de passagem, como mostrado na figura A.12.8. Nesse caso, cada linha de bit é conectada à linha de dados por meio de um transistor de passagem NMOS. As portas dos transistores de passagem são controladas por 2^N linhas, uma das quais é selecionada por um decodificador NOU similar àquele utilizado para decodificar os endereços de linha.

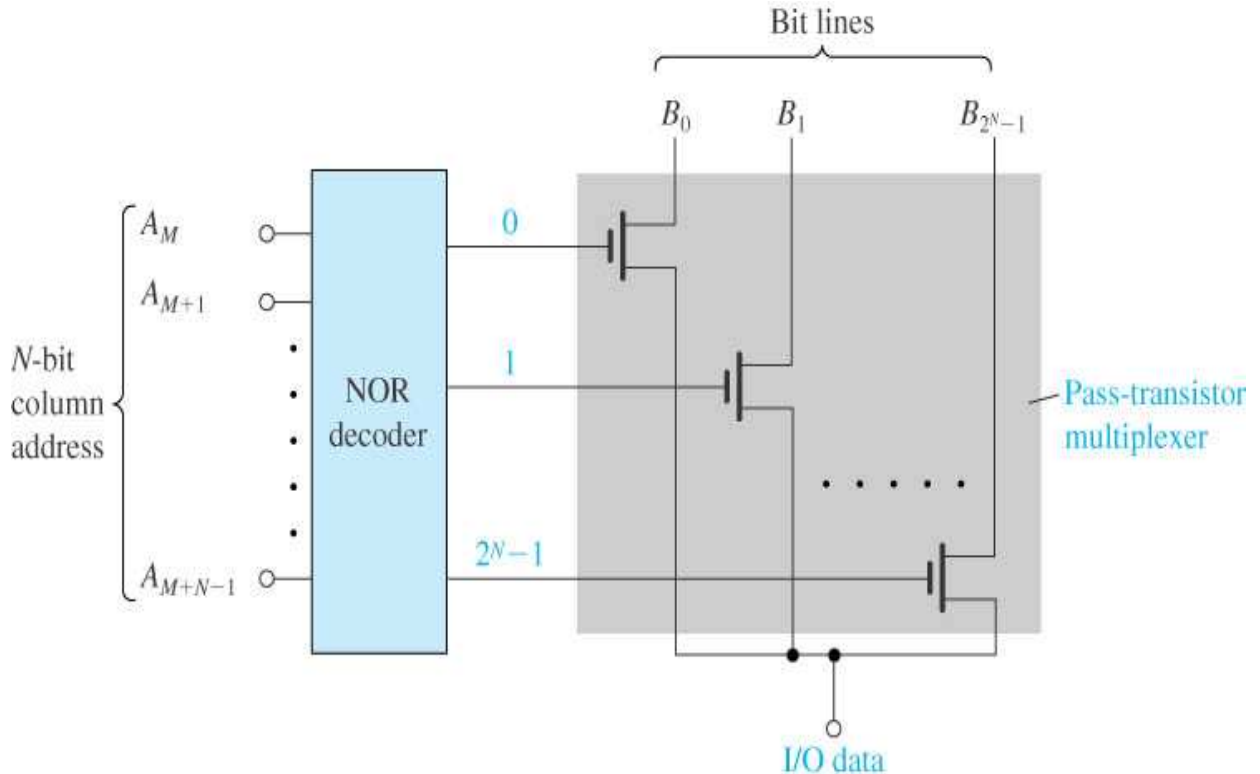


Figura A.12.8 Um decodificador de coluna implementado pela combinação de um decodificador NOU e um multiplexador com transistor de passagem.

Uma outra implementação do decodificador de coluna que usa um número menor de transistores (em detrimento da velocidade) é mostrada na figura A.12.9. Esse circuito, conhecido como *decodificador em árvore*, tem uma estrutura simples de transistores de passagem. Infelizmente, como um número significativo de transistores pode estar no caminho do sinal, a resistência das linhas de bit aumenta e a correspondente velocidade diminui.

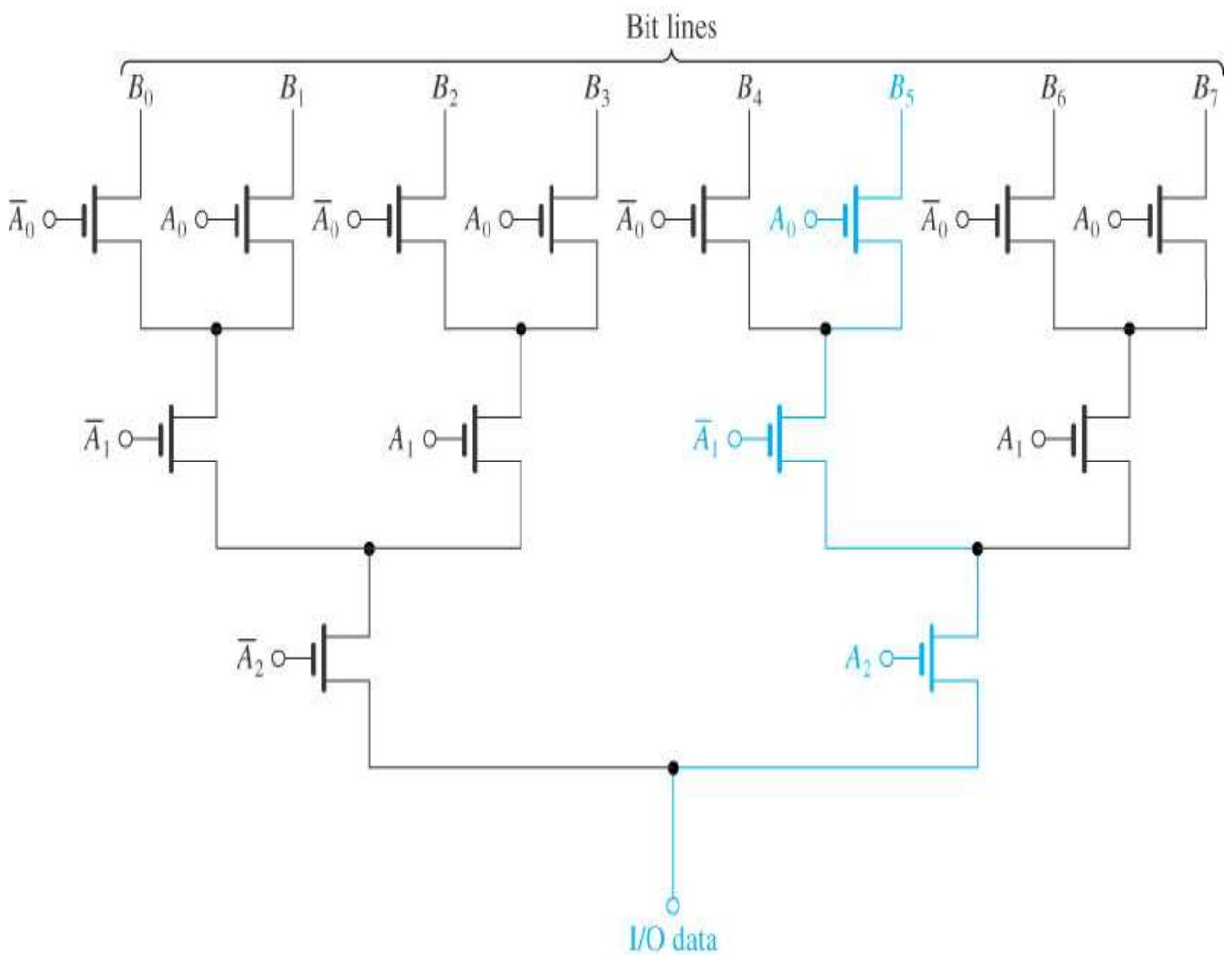


Figura A.12.9 Um decodificador de coluna em árvore.

Memória apenas de leitura (ROM)

A memória apenas de leitura (*read-only memory* – ROM) é uma memória que contém apenas dados fixos. Existem vários tipos de memórias apenas de leitura. Como a ROM fixa, à qual nos referimos apenas de ROM; a ROM programável (*programmable ROM* – PROM); e a ROM programável e apagável (*erasable programmable ROM* – EPROM).

Uma ROM pode ser vista como um circuito lógico combinatório em que a entrada é a coleção de bits de endereçamento da ROM e a saída é o conjunto de bits de dados recuperados a partir da localização endereçada.

A ROM MOS

A figura A.12.10 mostra uma ROM MOS simplificada de 32 bits (8 palavras x 4 bits). Conforme indicado, a memória consiste em uma matriz de MOSFET canal tipo n cujas portas são conectadas às linhas de palavras cujas fontes estão aterradas e cujos drenos estão conectados às linhas de bits. Cada linha de bit está conectada à fonte de alimentação por um transistor de carga PMOS, da mesma forma que uma lógica pseudo-NMOS já estudada. Há um transistor NMOS em uma célula particular se está estiver armazenando um zero; se a célula estiver armazenando 1, não há necessidade desse transistor. Esta ROM pode ser considerada como contendo oito palavras de 4 bits cada. O decodificador de linhas seleciona uma das oito palavras colocando 1 lógico nas portas dos transistores das linhas correspondentes que puxam a linha de bits para baixo. As linhas de bits que estão conectadas às células sem transistor permanecerão com um nível alto.

A desvantagem do circuito da ROM mostrada é que a dissipação de potência inerente a uma lógica pseudo-NMOS. Esta dissipação de

potência pode ser eliminada com uma mudança. Em vez de aterrar os terminais de porta dos transistores PMOS, podemos conectar esses transistores à linha de pré-carga como numa lógica dinâmica.

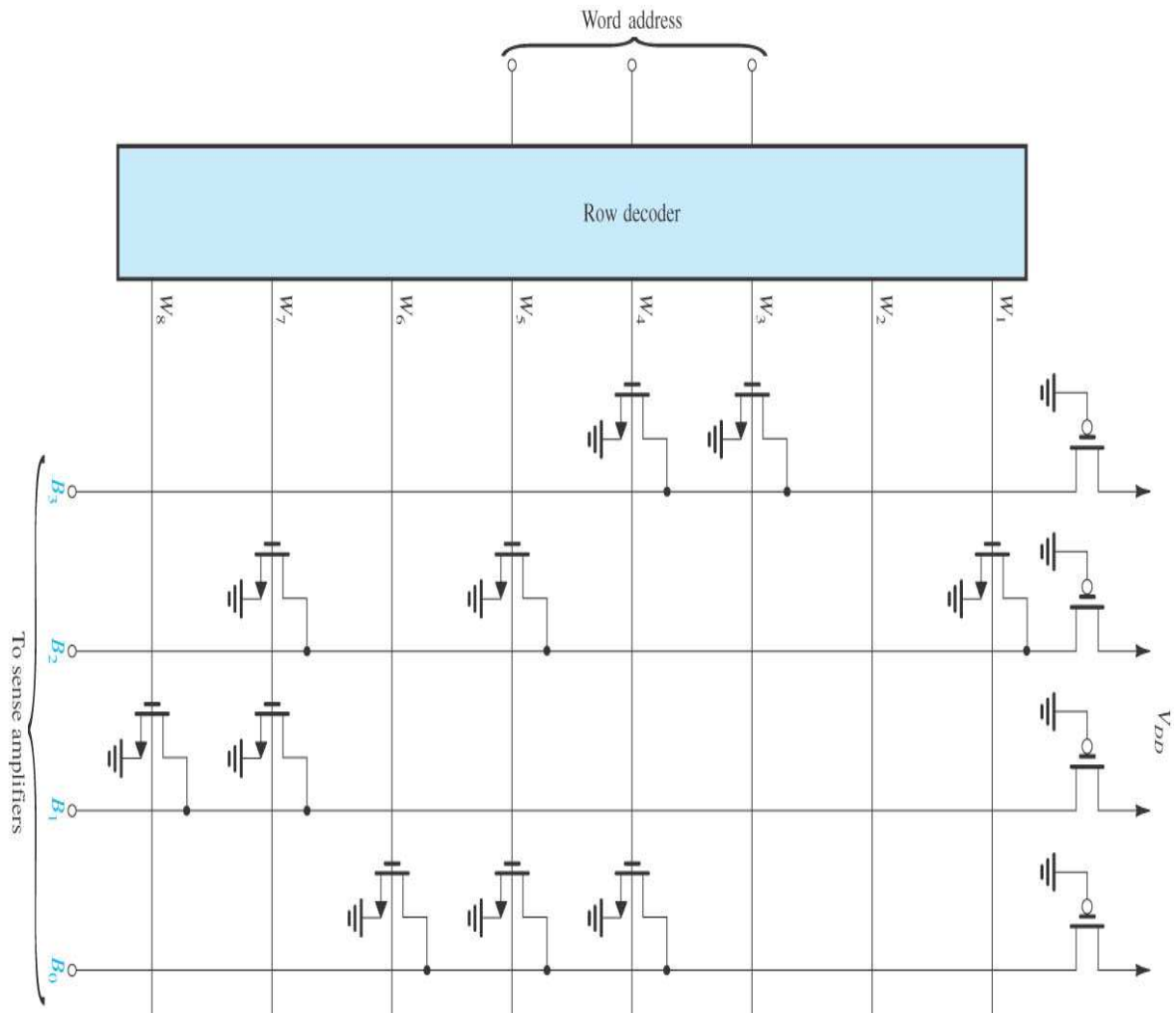


Figura A.12.10 Uma ROM MOS simples organizada como 8 palavras x 4 bits.

As ROM Programáveis por máscara

Os dados armazenados nas ROMs discutidas anteriormente são determinadas no momento da fabricação do chip, de acordo com as especificações do usuário. Porém, para evitar que se faça um projeto dedicado para cada ROM encomendada (processo caro), as ROM são fabricadas usando um processo conhecido como **programação por máscara**. Que de forma resumida, consiste em apenas definir a última máscara de metalização que define as interconexões como uma etapa onde o traçado desejado que depende do conteúdo da ROM, ou seja, as portas dos transistores nos quais serão armazenados 1 não serão conectadas.

As ROM Programáveis (PROM e EPROM)

As PROMs são ROM que podem ser programadas pelo usuário, mas apenas uma vez (também chamadas de OTP). Um arranjo típico empregado nas PROM com TBJ envolve o uso de fusíveis de silício policristalino para conectar o emissor de cada TBJ à coluna de dígitos correspondentes. Dependendo do conteúdo desejado de uma célula, o fusível pode ser deixado intacto ou queimado utilizando-se uma corrente alta.

Uma ROM programável e apagável, ou EPROM, é uma ROM que pode ser apagada e reprogramada inúmeras vezes. É, portanto, o tipo mais versátil (Hoje, memória flash predomina) apenas de leitura. Apesar do processo de reprogramação ser bastante lento.

As EPROM atuais usam variantes da célula de memória cuja secção transversal é mostrada na figura A.12.11. A célula é basicamente um único transistor canal n tipo enriquecimento com duas

portas feitas de silício policristalino. Uma das portas está flutuando e é apropriadamente chamada **porta flutuante**. A outra porta, denominada de porta de seleção, funciona do mesmo modo que a porta de um MOSFET tipo enriquecimento normal.

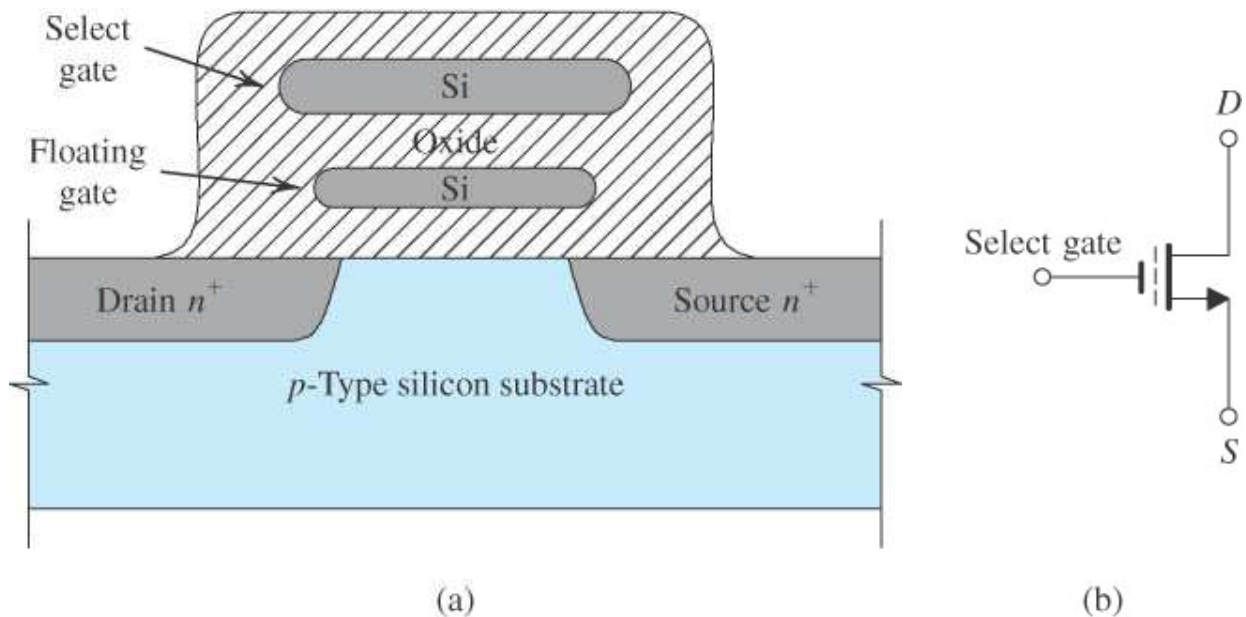


Figura A.12.11 a) Secção de corte e b) símbolo do transistor de porta flutuante usado como célula EPROM.

Antes de a célula ser programada, não há cargas na porta flutuante e o dispositivo funciona como um MOSFET canal n tipo enriquecimento normal. Portanto, exibe a característica $i_D - v_{GS}$, mostrada como curva a) na figura A.12.12. Observe que, nesse caso, a tensão de limiar (V_t) é muito baixa. Esse estado do transistor é conhecido como **estado não programado**. Esse é um dos dois estados em que o transistor de porta flutuante pode se encontrar. Tomaremos arbitrariamente o estado não programado para representar 1 armazenado. Isto é, diz-se que o transistor de porta flutuante cuja

característica $i_D - v_{GS}$ é mostrada como a curva a) na Figura A.12.12 está armazenando 1.

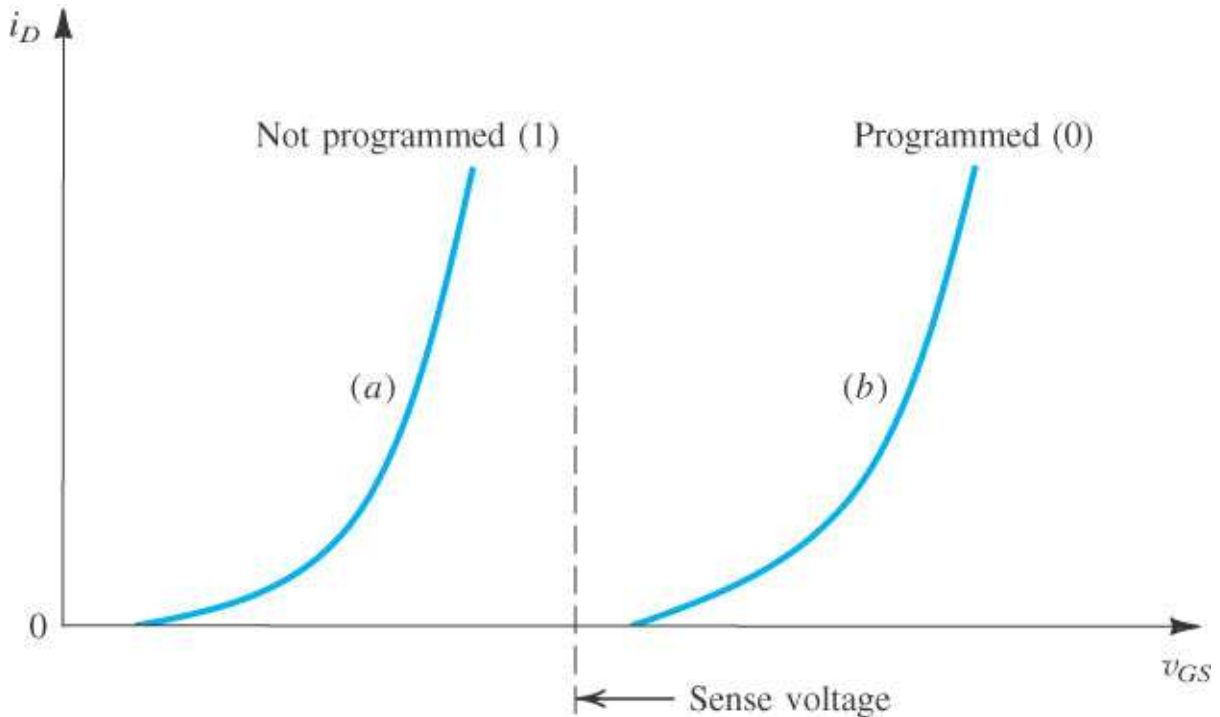


Figura A.12.12 Ilustração do deslocamento na característica $i_D - v_{GS}$ de um transistor de porta flutuante como resultado da programação.

Para programar o transistor de porta flutuante, é aplicada uma tensão elevada (cerca de 166 a 20 V) entre o dreno e a fonte. Ao mesmo tempo, é aplicada uma tensão elevada (cerca de 25 V) em sua porta de seleção. A figura A.12.13 mostra o MOSFET de porta flutuante durante a programação. Na ausência de cargas na porta flutuante, o dispositivo se comporta como um MOSFET de canal n tipo enriquecimento normal. Uma camada de inversão tipo n é criada na superfície da lamina como resultado da tensão positiva elevada na porta de seleção. Devido a elevada tensão aplicada entre o dreno e a fonte acelera os elétrons através do canal. Quando estes elétrons alcançam o dreno, eles

adquirem energia suficientemente alta são chamados de *elétrons quentes*. A tensão elevada de porta estabelece um campo elétrico no óxido isolante. Esse campo elétrico atrai os elétrons quentes e os acelera (através do óxido) em direção à porta flutuante. Desse modo, a porta flutuante fica carregada e a carga nela acumulada fica aprisionada.

Felizmente, esse processo de carregamento da porta é autolimitante. A carga negativa que se acumula na porta flutuante reduz a intensidade do campo elétrico no óxido até o ponto em que, finalmente, ele se torna incapaz de continuar acelerando os elétrons quentes.

O efeito da carga acumulada na porta flutuante é de deslocar a curva característica $i_D - v_{GS}$ do transistor. A leitura do conteúdo da célula é fácil: uma tensão V_{GS} , entre os valores de tensão de limiar alto e baixo é aplicado à porta de seleção. Se houver condução tem 1 armazenado, caso contrário terá zero.

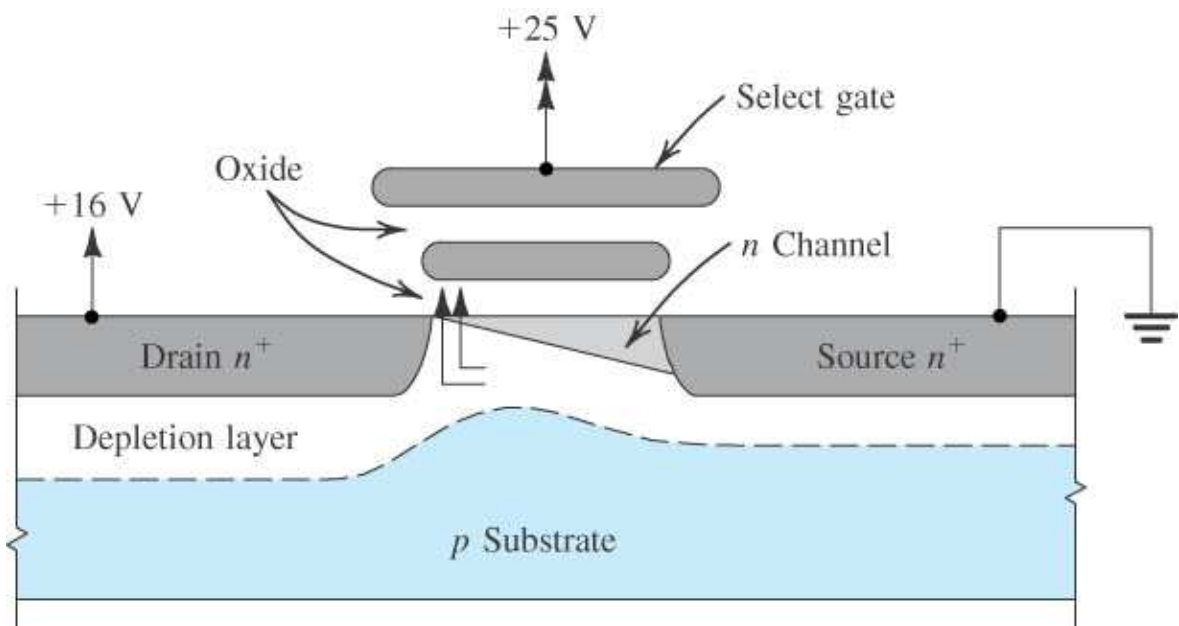


Figura A.12.13 O transistor de porta flutuante durante a programação.