

O Transistor de Efeito de Campo Metal-Óxido-Semicondutor MOSFET

Jacobus W Swart

Vimos em capítulo anterior o efeito do campo elétrico ou da tensão aplicada, a uma porta de um capacitor MOS, ou seja, como, variando a tensão V_G , variam os diagramas de bandas, de densidades de cargas, de campo elétrico e de potencial elétrico num eixo perpendicular à superfície. Estes resultados resultam da análise eletrostática unidimensional, pelo uso das equações básicas de semicondutores e da equação de Gauss e/ou de Poisson.

Esta análise resultou entre outros, na determinação da tensão de limiar clássica da estrutura MOS, definida como a condição onde o potencial de superfície torna-se de valor igual a duas vezes o potencial de Fermi, ou seja, considerando substrato tipo p, temos:

$$\phi_S = 2\phi_F \quad (1)$$

$$\text{onde: } \phi_F = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right) \quad (2)$$

$$V_T = V_{FB} + 2\phi_F + \frac{1}{C_O} \cdot \sqrt{2 \cdot q \cdot \epsilon_{Si} \cdot N_A \cdot 2\phi_F} \quad (3)$$

onde:

$$V_{FB} = -\frac{Q_O}{C_O} + \Phi_{MS} \quad (4)$$

Q_O = carga efetiva de interface SiO₂-Si, por unidade de área.

Φ_{MS} = diferença de função trabalho entre o metal e o semicondutor.

$C_O = \frac{\epsilon_{ox}}{t_{ox}} =$ capacitância de placas paralelas do dielétrico de porta
por unidade de área.

Estes conceitos citados acima constituem os fundamentos para o entendimento do transistor MOSFET ou simplesmente MOS. O princípio básico do transistor MOS é na verdade bem simples e foi proposto e patenteado já em 1928, por Lilienfeld, um homem muito à frente do seu tempo. Dizemos à frente do seu tempo, pois a realização física do transistor MOS não foi possível na época, pela não maturidade tecnológica. A Fig. 1 ilustra um desenho esquemático do transistor, como apresentado na patente. A limitação tecnológica da época refere-se ao não controle e alta densidade de estados e cargas de superfície do semicondutor. Esta alta densidade de estados de superfície produzia uma blindagem do semicondutor, impedindo assim uma modulação da densidade de portadores, portanto, da condutância entre os contatos de fonte e dreno, pela tensão de porta. Finalmente, apenas em 1960, obteve-se sucesso na fabricação do transistor MOS, na Bell Labs, por D. Kahng e M. Atalla. A Fig. 2 mostra um desenho esquemático do transistor MOS tipo nMOS (substrato p). O transistor MOS é um dispositivo de 4 terminais, sendo estes: fonte, dreno, porta e substrato. O transistor pMOS é complementar ao nMOS, ou seja, é formado por substrato tipo n e regiões de fonte e dreno tipo p.

Neste capítulo apresentaremos os princípios físicos do transistor MOS e os modelos básicos de operação.

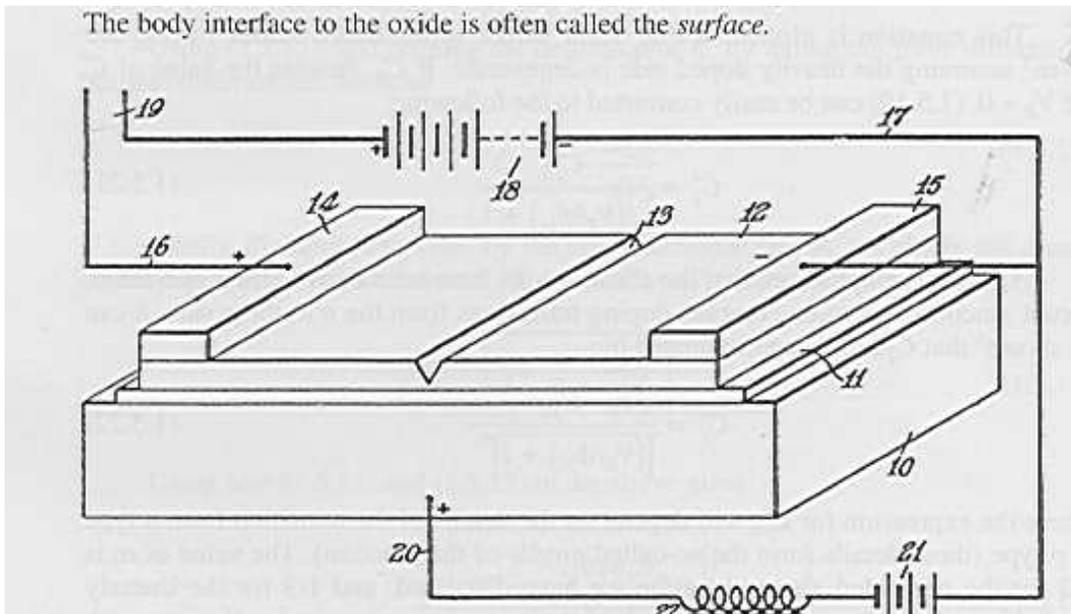


Fig. 1 Desenho esquemático do transistor MOSFET como apresentado por Lilienfeld, em 1928.

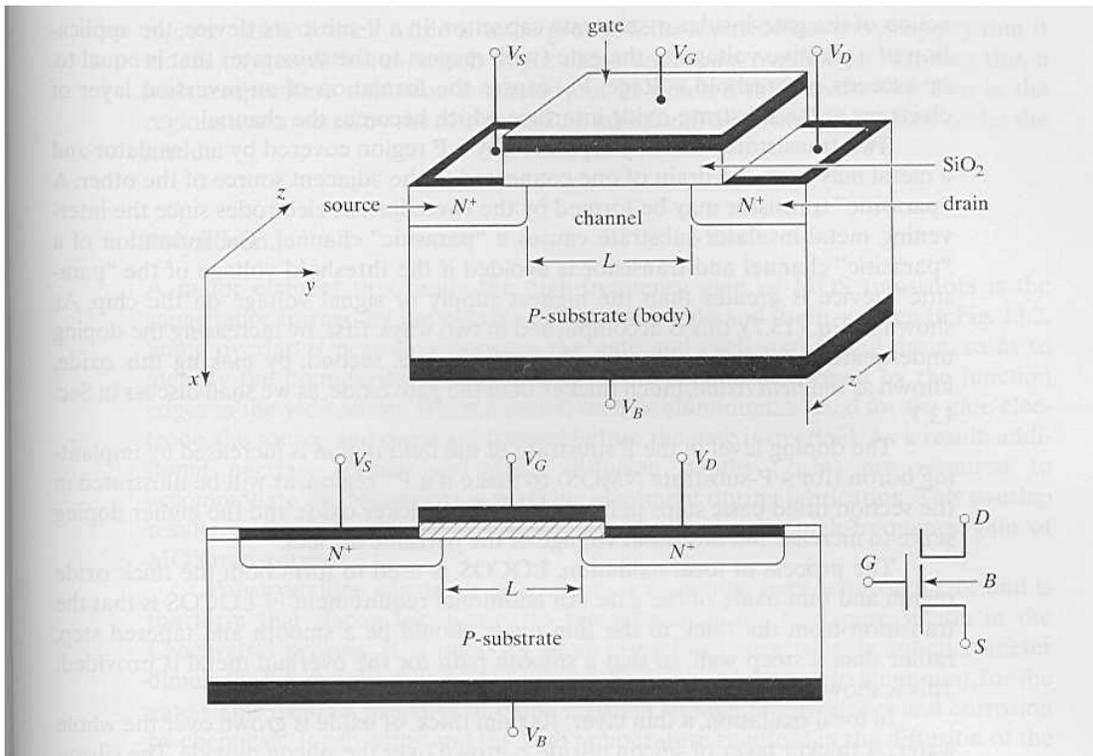


Fig 2 Desenho esquemático da estrutura moderna do transistor MOSFET em perspectiva, corte em secção transversal e o símbolo do transistor nMOS

1. MOS de três terminais ou diodo controlado por porta

A Fig. 3 ilustra a estrutura de um MOS de 3 terminais ou diodo controlado por porta. Esta estrutura não tem aplicação prática como dispositivo, mas é de extrema relevância para o entendimento do funcionamento do transistor MOS, ou MOS de 4 terminais. O MOS de 3 terminais corresponde a um meio transistor, omitindo-se o seu dreno.

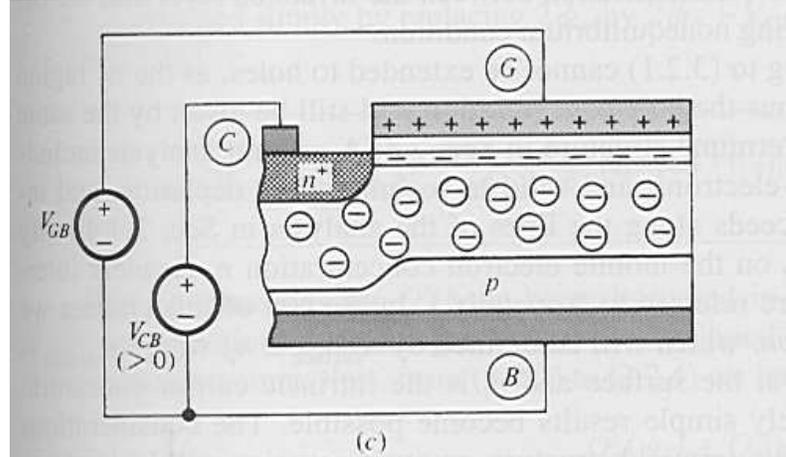
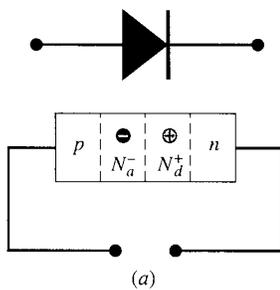


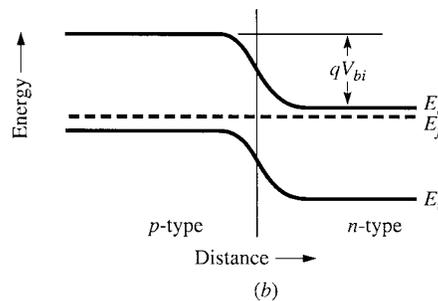
Fig. 3 Estrutura de diodo controlado por porta o MOS de 3 terminais.

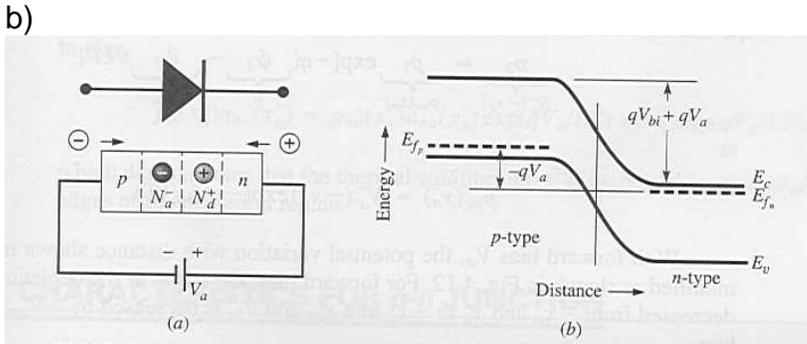
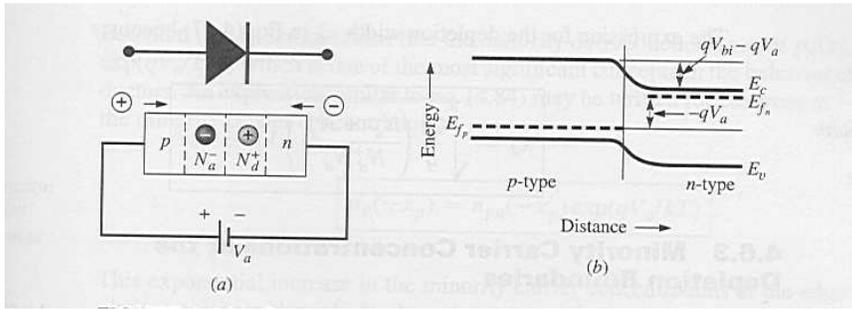
Para entender o efeito do diodo sobre a análise do capacitor MOS, devemos juntar os conceitos do diodo pn e do capacitor MOS. Na Fig. 4 repetimos as estruturas de bandas de diodos pn, sem e com polarização, sendo esta direta e reversa. Observa-se que com polarização direta ocorre uma redução da barreira da junção enquanto com polarização reversa esta barreira aumenta. Para os casos de polarização da junção, ou seja, com a estrutura fora do equilíbrio térmico, define-se níveis de quase-Fermi, que são assumidos constantes dentro das regiões de depleção (existem argumentos convincentes que justificam esta aproximação). Dentro da região de depleção da junção, os dois níveis de quase-Fermi são separados em energia, com valor dado por $q \cdot V_a$. Conhecidos os valores dos níveis de quase-Fermi podemos determinar as concentrações dos portadores pelas relações de Boltzmann.

A Fig. 5 mostra os diagramas de bandas bi-dimensionais, de MOS de 3 terminais, com diodo sem polarização e com polarização reversa. O caso do diodo com polarização direta não tem interesse para estudo de transistores MOS, dado que nunca devemos polarizar as junções de fonte e dreno diretamente em relação ao substrato.



a)





c) Fig. 4 Diagramas de bandas de diodo pn, a) em equilíbrio, b) com polarização direta e c) com polarização reversa.

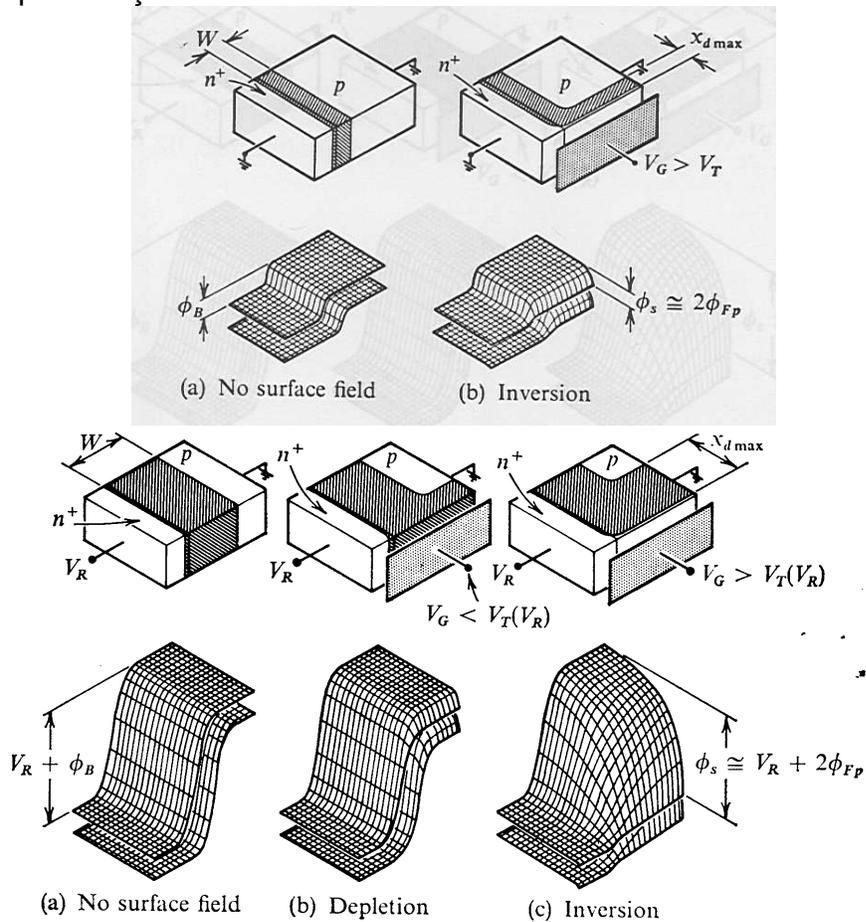


Fig. 5 Diagramas de bandas bi-dimensionais de estrutura MOS de 3 terminais, a) em equilíbrio e b) com polarização reversa do diodo.

No caso do diodo com polarização nula em relação ao substrato, a estrutura permanece em equilíbrio. Desta forma, o comportamento do capacitor MOS não sofrerá alteração em relação ao caso sem diodo, ou seja, valem os mesmos diagramas de bandas, de densidade de cargas, de campo elétrico e de potencial elétrico do caso do capacitor MOS convencional. Ocorrerá apenas uma única alteração, com relação ao comportamento da curva C-V de alta freqüência da porta em relação ao substrato. Esta curva será o mesmo ao da curva C-V de baixa freqüência do capacitor. A explicação deste resultado é fácil de obter ao re-lembrar o motivo da diferença do comportamento das curvas C-V de baixa e alta freqüência do capacitor MOS convencional. No caso do MOS de 3 terminais, assim como no caso do capacitor MOS em baixa freqüência, a capacitância volta ao valor de C_{MAX} para V_G na região de inversão forte. O motivo disto é atribuído à habilidade dos portadores do canal responderem à variação de V_G , curto-circuitando assim a capacitância da região de depleção abaixo da porta. No caso do capacitor MOS medido em baixa freqüência, esta resposta dá-se pela geração e recombinação de portadores (tempo de vida bem menor que o tempo de variação da tensão), enquanto que no caso do MOS de 3 terminais, a resposta dos portadores de canal dá-se pelo suprimento ou drenagem de portadores pelo diodo justaposta ao canal (no capacitor MOS convencional não existe esta fonte).

A grande diferença, no entanto, entre o MOS de 2 terminais e o MOS de 3 terminais ocorre com a polarização reversa do diodo, como ilustra o diagrama de bandas da Fig. 5b. Mesmo com V_G com valor na região de inversão do MOS de 2 terminais, pode não ocorrer inversão no MOS de 3 terminais. Isto se deve ao fato do diodo drenar todos os portadores do canal enquanto o seu potencial de superfície for inferior ao potencial do diodo, $V_D + V_{bi}$. Apenas quando o potencial de superfície do canal tornar-se de valor semelhante ao do diodo poderemos manter uma camada de inversão na superfície do semiconductor.

Enquanto o diodo polarizado impedir a inversão da superfície ($V_G < V_{TB}$, onde V_{TB} é o novo valor de tensão de limiar, dependente de V_D), as relações de densidade de portadores (Boltzmann com quase-Fermi) e a equação de Gauss devem ser respeitadas. A ausência da carga de inversão deve ser compensada por um incremento da carga de depleção para neutralizar a carga na porta, que aumenta com a tensão V_G . Esta carga de depleção está relacionada com o potencial de superfície, pela relação normal de uma camada de depleção obtida pela aproximação de depleção:

$$Q_D = -\sqrt{2q \cdot \epsilon_{Si} \cdot N_A \cdot \phi_S} \quad (5)$$

A inversão irá ocorrer quando o potencial de superfície tornar-se igual a $V_D + V_{bi}$ ou $V_D + 2\phi_F$. Isto altera o valor da tensão de limiar para V_{TB} dado por:

$$V_{TB} = V_{FB} + V_D + 2\phi_F + \frac{1}{C_O} \cdot \sqrt{2 \cdot q \cdot \epsilon_{Si} \cdot N_A \cdot (V_D + 2\phi_F)} \quad (6)$$

Em muitos casos, desejamos o valor V_G de limiar não em relação ao substrato, mas sim em relação à tensão do diodo (normalmente a fonte do transistor MOS). Neste caso devemos apenas aplicar a relação de mudança de referencial:

$$V_{TD} = V_{FB} + 2\phi_F + \frac{1}{C_o} \cdot \sqrt{2 \cdot q \cdot \epsilon_{Si} \cdot N_A \cdot (V_D + 2\phi_F)} \quad (7)$$

No caso do transistor ou MOS de 4 terminais, esta expressão é apresentada normalmente como tensão de limiar de porta em relação ao terminal de fonte do transistor. Portanto devemos substituir a tensão V_D por V_{SB} ou ainda por $-V_{BS}$. Obtém-se desta forma o V_T do transistor dado por:

$$V_T = V_{FB} + 2\phi_F + \frac{1}{C_o} \cdot \sqrt{2 \cdot q \cdot \epsilon_{Si} \cdot N_A \cdot (2\phi_F + V_{SB})} \quad (8)$$

A Fig. 6 mostra curvas C-V de dispositivo MOS de 3 terminais, tendo a tensão do diodo como parâmetro. Observa-se que quanto maior a tensão reversa do diodo, maior a tensão de limiar de inversão e menor a capacitância mínima obtida antes da inversão, o que corresponde a uma região de depleção maior e uma tensão de superfície também maior. O entendimento desta série de curvas ajuda em muito o entendimento do transistor MOS, incluindo o porque do aumento da tensão de limiar do transistor quando aumentarmos a polarização reversa da fonte do transistor em relação ao substrato.

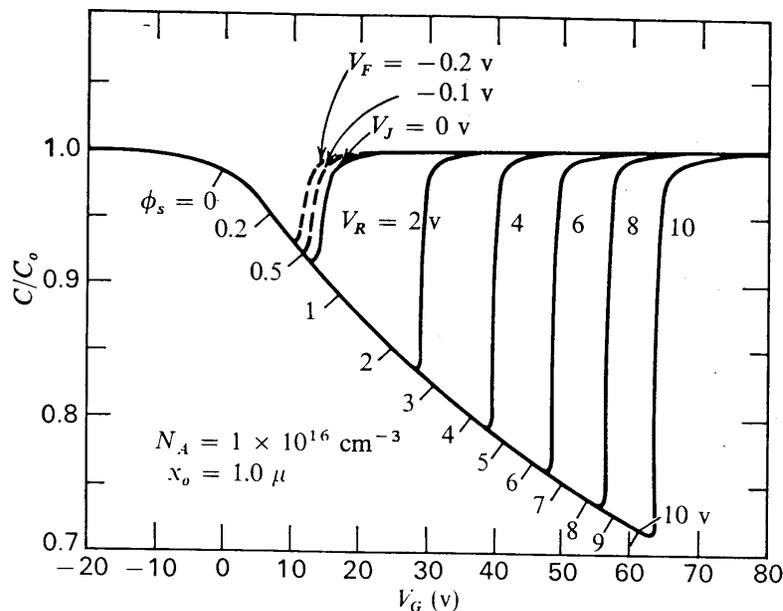


Fig. 6 Curvas C-V de dispositivo MOS de 3 terminais, com tensão do diodo como parâmetro.

2. Regiões de operação

Dependendo da polarização dos 4 terminais do transistor, definem-se basicamente 3 regiões de operação do mesmo: corte, linear e saturação. A Fig. 7 mostra uma série de curvas características de um transistor nMOS. O transistor nMOS funciona com tensões de porta e dreno positivas em relação à fonte, passando corrente positiva do dreno para a fonte. O transistor pMOS por outro lado, funciona com tensões de porta e de dreno negativas em relação à fonte, passando corrente negativa do dreno para a fonte. Apresentaremos nossa análise, considerando transistores tipo nMOS.

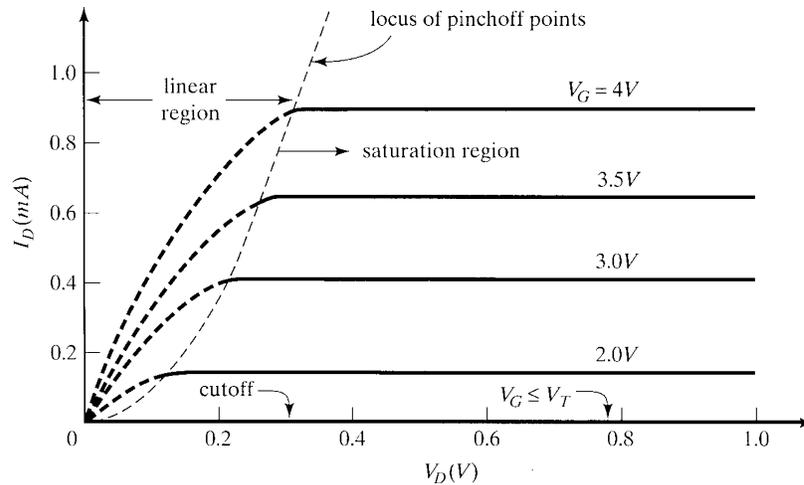


Fig. 7 Curvas características de transistor nMOS, com indicação das regiões de operação: corte, linear e saturação.

Para tensão de porta menor que a tensão de limiar, V_T , do transistor, a densidade de cargas no canal é nula ou muito pequena. Desta forma, a corrente que fluirá entre dreno e fonte também será nula ou muito pequena (desprezível em escala linear). Nestas condições, o transistor está em região de corte, ou ainda, em região sub-limiar.

Para tensão de porta acima do valor da tensão de limiar e tensão de dreno com valor pequeno, o transistor está na região linear ou também chamado de região triodo. Nesta região, a corrente é diretamente proporcional às tensões de porta e de dreno. Como a densidade de cargas no canal é diretamente proporcional à tensão V_{GS} , a condutância, ou a corrente I_{DS} , entre dreno e fonte, também segue esta mesma relação com V_{GS} . No entanto, como mostraremos no próximo item, a condutância de canal apresenta uma dependência com a tensão de dreno. Mais especificamente, a condutância diminui com a tensão V_{DS} , resultando num aumento não linear da corrente I_{DS} com V_{DS} .

Agora, quando a tensão de dreno, V_{DS} , passa de um certo valor, a corrente I_{DS} , fica aproximadamente constante. Esta região é a chamada de saturação. A tensão V_{DS} a partir da qual a corrente satura, é chamada de tensão de saturação, V_{Dssat} . Esta tensão apresenta uma dependência com a tensão de porta aplicada, como veremos também no próximo item.

3. Relações de Corrente-Tensão do Transistor MOS

No desenvolvimento de modelos de corrente do transistor faremos uso das seguintes hipóteses ou aproximações:

- A mobilidade de portadores de canal é constante.
- O gradiente de potencial na direção y , de fonte para dreno, é muito menor que o gradiente de potencial na direção x , de porta para substrato. Esta aproximação é conhecida como aproximação de canal gradual e permite simplificar a equação de Poisson tridimensional para unidimensional. Desta forma, todos os resultados da análise de densidades de carga de inversão e de depleção do capacitor MOS, bem como do diodo controlado por porta, podem ser aplicados diretamente ao transistor.

- Temos o canal formado em toda a superfície, desde a fonte até o dreno (região linear ou triodo).

A Fig. 8 mostra um desenho esquemático do transistor nMOS polarizado em região linear, onde temos um canal formado (inversão forte) desde a fonte até o dreno. A fonte é considerada aterrada, ou seja, o terminal de referência para as tensões de porta, dreno e substrato. Mostraremos que a densidade carga de inversão não é constante ao longo do canal, mas ao contrário, reduz-se da fonte até o dreno ($V_{DS} > 0$).

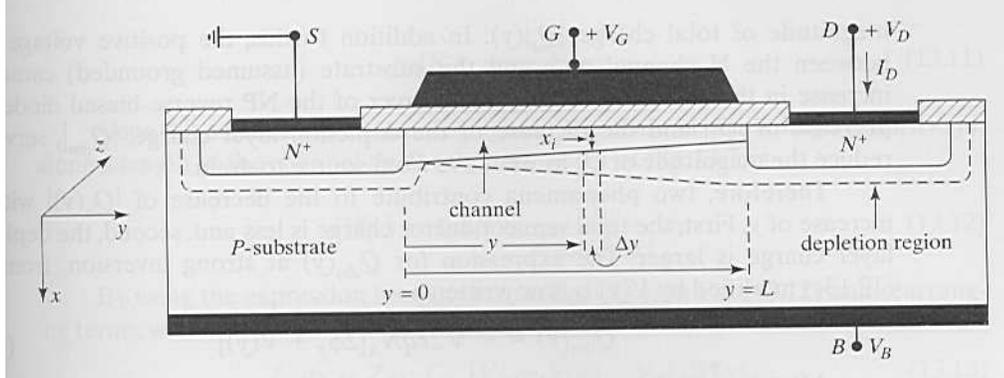


Fig. 8 Desenho esquemático de transistor nMOS polarizado na região linear de operação.

De acordo com a teoria desenvolvida no capacitor MOS e do diodo controlado por porta podemos escrever:

$$V_{GB} = V_{FB} + V_{OX}(y) + \phi_S(y) \quad (9)$$

O potencial de superfície no ponto y pode ser desmembrado como (superfície com canal formado):

$$\phi_S(y) = 2\phi_F + V(y) \quad (10)$$

onde $V(y)$ é tensão reversa aplicada entre o canal e o substrato, variando desde V_{SB} (tensão entre fonte e substrato) em $y=0$, a V_{DB} (tensão entre dreno e substrato) em $y=L$.

V_{OX} é a queda de tensão sobre óxido e pode ser expressa por (pela lei de Gauss):

$$V_{OX}(y) = -\frac{Q_S}{C_o} = -\frac{Q_I + Q_D}{C_o} \quad (11)$$

onde, Q_S é a densidade de carga no semiconductor, Q_I é a densidade de carga de inversão no canal e Q_D é a densidade de carga de depleção. Estas cargas variam com y .

Das relações acima podemos escrever:

$$Q_I(y) = -C_o[V_{GB} - V_{FB} - 2\phi_F - V(y)] - Q_D(y) \quad (12)$$

Na condição de canal formado temos que:

$$Q_D(y) = -\sqrt{2q \cdot N_A \cdot \epsilon_{Si} [2\phi_F + V(y)]} \quad (13)$$

$$Q_I(y) = -C_o [V_{GB} - V_{FB} - 2\phi_F - V(y)] + \sqrt{2q \cdot \epsilon_{Si} \cdot N_A [2\phi_F + V(y)]} \quad (14)$$

De posse da densidade de carga de inversão do canal podemos determinar a sua condutância e a relação I-V, como veremos nos próximos subitens.

3.1 Modelo de Lei Quadrática, Shichman e Hodges ou SPICE Nível 1

Neste modelo, assume-se a seguinte aproximação: a largura, portanto também a carga, da região de depleção é constante, desde a fonte até o dreno e dado pelo seu valor próximo à fonte. Assim temos:

$$Q_D(y) = -\sqrt{2q \cdot N_A \cdot \epsilon_{Si} (2\phi_F + V_{SB})} \quad (15)$$

$$Q_I(y) = -C_o [V_{GB} - V_{FB} - 2\phi_F - V(y)] + \sqrt{2q \cdot \epsilon_{Si} \cdot N_A (2\phi_F + V_{SB})} \quad (16)$$

Combinando as expressões (8) e (16) obtemos:

$$Q_I(y) = -C_o [V_{GB} - V_T - V(y)] \quad (17)$$

Sobre um intervalo dy do canal teremos uma queda de tensão dada por:

$$dV(y) = I_D \cdot dR(y) \quad (18)$$

O corrente I_D é a corrente que passa pelo canal e deve ser o mesmo valor para qualquer ponto y (continuidade de corrente). A resistência incremental do canal é dada por:

$$dR(y) = \frac{1}{W \int_0^{x_i} \frac{1}{\rho} dx} dy \quad (19)$$

onde ρ é a resistividade, W a largura do canal, x_i é a profundidade do canal. Por comodidade, trabalharemos com o inverso da resistência incremental. Lembramos ainda a expressão da resistividade de elétrons no canal n :

$$\rho = \frac{1}{q\mu_n \cdot n} \quad (20)$$

onde μ_n é a mobilidade de elétrons (assumida constante em x e em y). Assim podemos escrever o inverso da resistência incremental como:

$$\frac{1}{dR(y)} = \frac{W}{dy} \mu_n \int_0^{x_i} qn(x) dx = \frac{W}{dy} \mu_n |Q_I| \quad (21)$$

Substituindo (21) em (18) temos:

$$dV(y) = I_D \frac{dy}{W\mu_n |Q_I|} \quad (22)$$

$$I_D dy = W\mu_n |Q_I| dV(y) \quad (23)$$

Integrando esta expressão em y, de 0 a L, ou em V, de 0 a V_{DS} , resulta:

$$I_D \int_0^L dy = W\mu_n \int_0^{V_{DS}} |Q_I| dV(y) \quad (24)$$

Integrando o lado esquerdo desta equação obtemos:

$$I_D = \frac{W}{L} \mu_n \int_0^{V_{DS}} |Q_I| dV(y) \quad (25)$$

A equação (25) é a equação básica para o desenvolvimento de vários modelos de relação I-V de MOSFET.

Substituindo a relação (17) em (25), considerando ainda $V_{GS}=V_{GB}-V_{SB}$ e $V_{YS}(y)=V(y)-V_{SB}$, temos:

$$I_D = \frac{W}{L} \mu_n C_o \int_0^{V_{DS}} [(V_{GS} - V_T) - V_{YS}(y)] dV(y) \quad (26)$$

Efetuada a integral acima, resulta:

$$I_D = \frac{W}{L} \mu_n C_o [(V_{GS} - V_T) - \frac{V_{DS}}{2}] V_{DS} \quad (27)$$

A relação (27) é válida até o valor de V_{DS} correspondente ao ponto de máximo da curva da Fig. 9 (linha sólida). Tomando o derivada nula da curva obtém-se o valor de $V_{DS}=V_{DSsat}$:

$$V_{DSsat} = (V_{GS} - V_T) \quad (28)$$

Observa-se da curva da Fig.9, bem como da equação (29), que o seu comportamento é bem linear para V_{DS} pequeno ($V_{DS} \ll V_{GS}-V_T$). Para V_{DS} já não tão próximo da origem, a curva torna-se cada vez menos linear, correspondendo a um aumento da resistência de saída do transistor. Este efeito deve-se à redução da carga no canal, à medida que aumentamos a tensão de dreno (ver relação (17)).

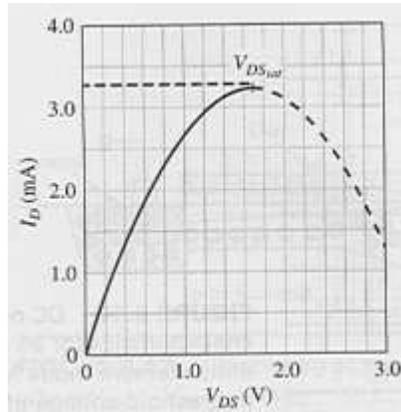


Fig. 9 Curva de corrente de dreno versus tensão V_{DS} , para um valor fixo de V_{GS} , de acordo com a relação (27).

A relação (17) mostra que a carga de canal, Q_1 , reduz-se da fonte ao dreno, tendo em vista que $V(y)$ aumenta em direção ao dreno. No entanto, não necessitamos determinar a função $V(y)$ para a obtenção da corrente I_D . Mesmo assim, poderemos obtê-lo por cálculo numérico. A Fig. 10 mostra a variação qualitativa de $V(y)$ da fonte até o dreno (considerou-se $V_{SB}=0$ neste caso). Observa-se que $V(y)$ não varia linearmente, mas sim com uma variação mais forte. Fisicamente isto se deve à continuidade da corrente I_D ao longo do canal. Se $Q_1(y)$ diminui, o campo elétrico de arraste dos portadores deve aumentar para manter a corrente constante. Como o campo elétrico no canal é dado pela derivada de $V(y)$ em relação a y , esta derivada deve ser uma função crescente, justificando assim este aumento mais forte que linear de $V(y)$.

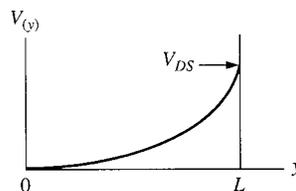


Fig. 10 Variação da tensão de canal, $V(y)$, em relação à tensão de fonte, ao longo do canal.

O modelo da lei quadrática é bastante simples e imprecisa. A sua utilidade serve, sobretudo para análise preliminar de circuitos ou cálculos manuais. Em análise de circuitos por programa de simulação tipo SPICE recomenda-se o uso de modelos mais precisos.

3.2 Modelo e Corrente de Saturação

Substituindo o valor de V_{Dssat} , dado por (28) em (27) obtém-se o valor da corrente de saturação:

$$I_{Dsat} = \frac{W}{2L} \mu_n C_O (V_{GS} - V_T)^2 \quad (29)$$

Para $V_{DS} > V_{Dssat}$, a corrente I_D não varia mais, em primeira aproximação.

Fisicamente, o ponto de saturação corresponde à situação em que Q_1 torna-se “nula”, ou seja, ocorre a “constricção do canal” (*pinch-off*). No ponto de constricção do

canal ocorre uma inversão do campo elétrico no óxido, como ilustram os resultados numéricos apresentados nas figuras 11 e 12. Entre a fonte, $y=0$, e o ponto de constricção, $y=y'$, valem as hipóteses adotadas para o desenvolvimento do modelo de corrente do transistor, porém para $y>y'$, a hipótese de canal gradual deixa de valer, ou seja, a variação do campo elétrico longitudinal deixa de ser desprezível em relação à variação do campo elétrico transversal. Neste momento necessitaríamos de cálculos bidimensionais para calcular as densidades de cargas, campos elétricos, potenciais, valor de y' e a corrente I_D .

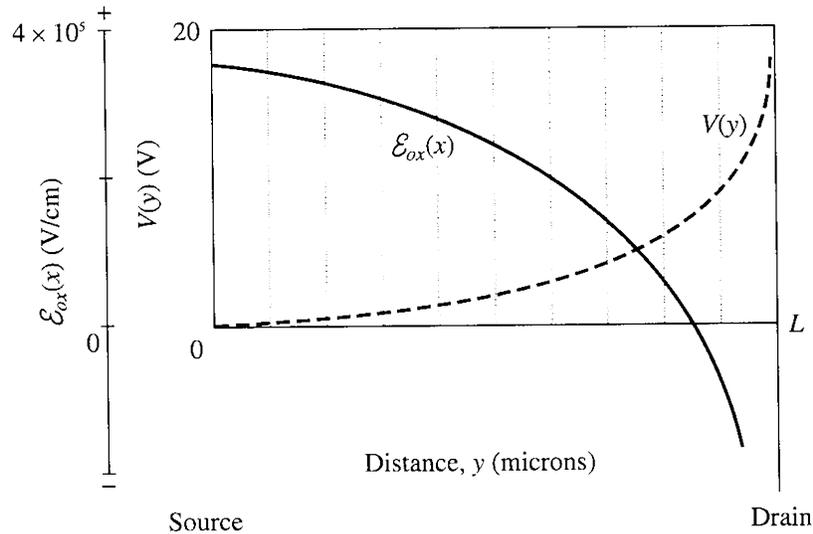


Fig. 11 Variação do potencial $V(y)$ e do campo elétrico transversal no óxido, ao longo do canal para transistor MOS polarizado na região de saturação.

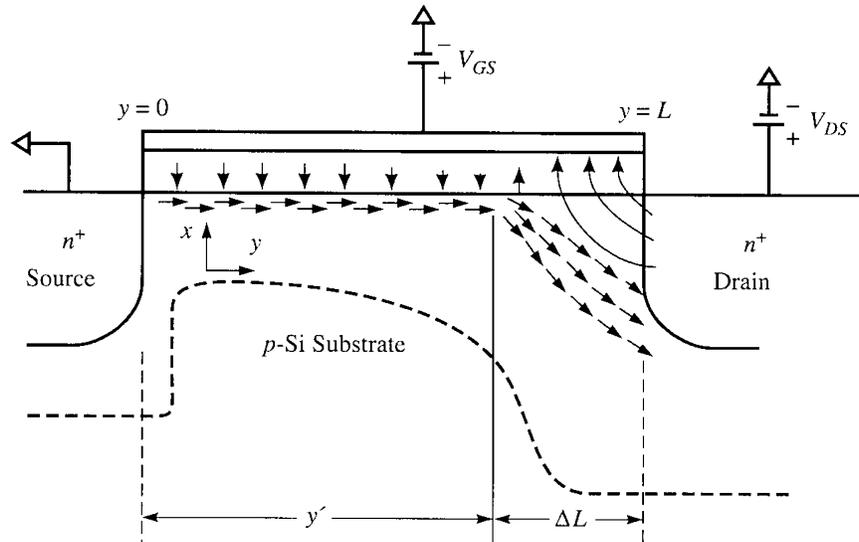


Fig. 12 Secção em corte de MOSFET em saturação, com indicação de linhas de corrente (flechas horizontais) e direção do campo elétrico (flechas verticais). Linha tracejada representa a borda da região de depleção; y' é o ponto de constricção do canal e ΔL é a distância entre o dreno e o ponto y' .

Na obtenção do modelo de corrente na região linear foi considerada apenas a corrente de deriva longitudinal dos portadores (é a única relevante nestas condições).

No entanto, na região física do transistor, com $y > y'$, a corrente de deriva longitudinal torna-se totalmente desprezível frente às correntes de difusão e deriva transversal.

A solução para este problema complexo é calcular a corrente, considerando apenas a região do canal com forte inversão, ou seja, a região com $y < y'$. Lembramos que no ponto y' sempre teremos $V(y') = V_{DBsat}$, independente do valor ou posição y' . Assim podemos calcular a corrente usando o modelo linear, tomando o cuidado de substituir V_{DS} por V_{DSSat} , mesmo para $V_{DS} > V_{DSSat}$. É isto que fizemos na obtenção da relação (29).

Para $V_{DS} = V_{DSSat}$, o ponto y' ocorre em $y = L$. Para $V_{DS} > V_{DSSat}$, no entanto, o ponto y' distancia-se cada vez mais de $y = L$, aumentando o valor de ΔL , definido na Fig.12. Desta maneira, para continuar calculando corretamente a corrente I_D , devemos substituir o valor de L por $(L - \Delta L)$ na expressão (29). Isto faz com que a corrente de saturação aumente levemente com o valor de V_{DS} , para $V_{DS} > V_{DSSat}$.

Uma boa analogia do comportamento aproximadamente constante da corrente de saturação é o caso do fluxo de água de um rio com uma queda d'água. O fluxo é determinado pelo suprimento de água do rio e independente da altura da cascata. A cascata pode ser comparada ao trecho de constrição, ΔL , do transistor, sendo altura da cascata correspondente à tensão ($V_{DS} - V_{DSSat}$) que cai sobre o trecho ΔL .

3.3 Modelo de Carga de Corpo, Ihanola e Moll ou SPICE nível 2.

O modelo da lei quadrática considera o efeito da tensão $V(y)$ ou V_{DS} sobre a variação da carga de inversão do canal, porém despreza o seu efeito sobre a variação da carga de depleção. O modelo de carga de corpo acrescenta este efeito ao modelo anterior. Fora este efeito, o modelo assume as mesmas hipóteses adotadas pelo modelo da lei quadrática.

Podemos usar a mesma equação (25), porém com a carga Q_i dada pela relação (14), sem aproximação, mantendo a variável $V(y)$ no termo da raiz quadrada, correspondente à carga de depleção. Efetuando a operação da integração da equação (25) obtemos o seguinte:

$$I_D = \frac{W}{L} \mu_n C_o \left\{ (V_{GS} - V_{FB} - 2\phi_F - \frac{V_{DS}}{2}) V_{DS} - \frac{2}{3} \frac{\sqrt{2q\epsilon_{Si} N_A}}{C_o} \left[(V_{DS} + 2\phi_F + V_{SB})^{3/2} - (2\phi_F + V_{SB})^{3/2} \right] \right\} \quad (30)$$

A tensão de dreno de início da saturação pode ser obtida pela equação (14), como a condição de constrição do canal ($Q_i = 0$). A solução de $V(y)$ desta operação nos fornece o valor da tensão dreno de saturação.

$$V_{DSSat} = (V_{GS} - V_{FB} - 2\phi_F) + \frac{q\epsilon_{Si} N_A}{C_o^2} \left[1 - \sqrt{1 + \frac{2C_o^2 (V_{GS} - V_{FB})}{q\epsilon_{Si} N_A}} \right] \quad (31)$$

Nota-se que este modelo já é consideravelmente mais complexo que o da equação (27). A Fig. 13 compara as curvas características calculadas pelos dois modelos apresentados. Observa-se uma grande diferença nas curvas, sobretudo quando o valor da tensão de porta for maior. Isto mostra a importância de considerar o

efeito da variação da carga de depleção ao longo do canal no desenvolvimento do modelo, como feito neste item.

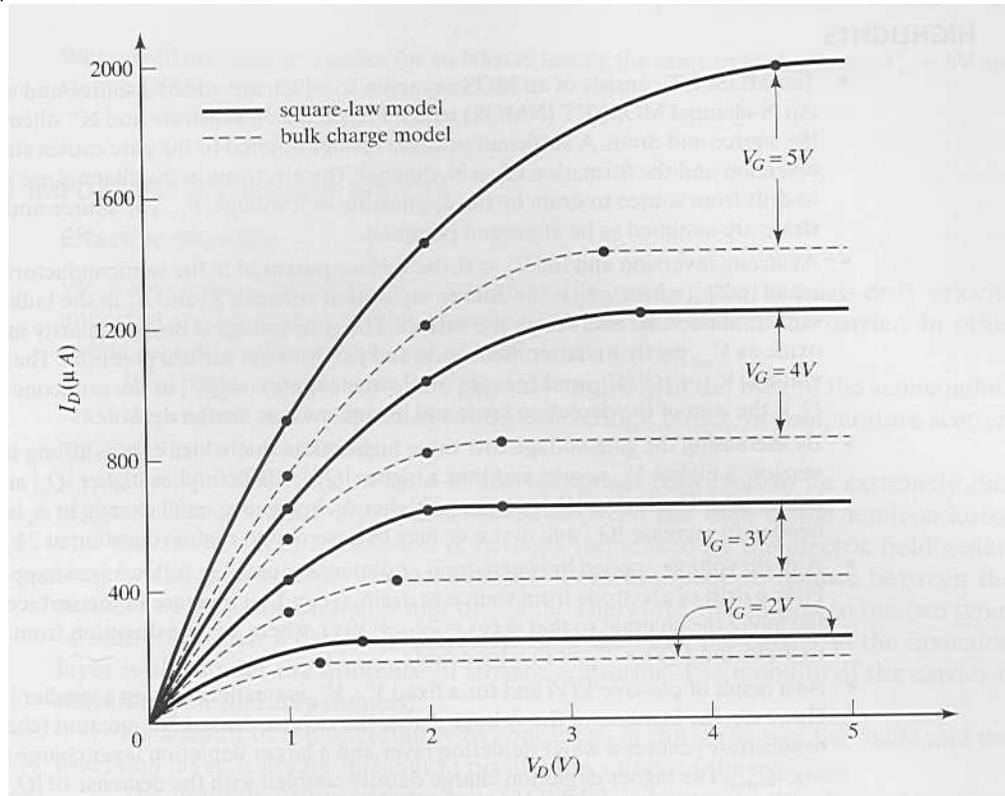


Fig. 13 Curvas características calculados pelos modelos da lei quadrática (equações 27 a 29) e da carga de corpo (equação 30 combinada com 31).

3.5 Corrente Sub-Limiar

Os modelos SPICE nível 1 e nível 2 apresentados acima assumem I_D nulo para tensão de porta V_G menor que V_T . Na verdade a corrente não se anula instantaneamente em $V_G=V_T$, como mostram as curvas de I_D versus V_G , na Fig. 14. Na escala linear (Fig. 14a), a corrente aparentemente se anula em tensão próxima a V_T , porém, em escala logarítmica nota-se que a corrente cai exponencialmente para $V_G < V_T$ (Fig. 14b). Ou seja, na região sub-limiar, a corrente tem um comportamento exponencial com a tensão de porta. Nesta região, o transistor apresenta uma semelhança com o transistor bipolar de junção em termos de comportamento. Inclusive, o mecanismo de transporte predominante da corrente é o de difusão, como no BJT, ao contrário da operação do MOSFET na região linear, onde o mecanismo predominante é o de deriva, como apontamos anteriormente.

A curva da Fig. 14a representa ainda uma das maneiras de extrair o valor de V_T do transistor, pela extrapolação da reta ao eixo da abscissa, de acordo com a equação (29). Esta curva normalmente é obtida pela medida da corrente I_D versus $V_{GS}=V_{DS}$ (isto garante que o transistor nMOS esteja sempre em saturação, desde que o valor de V_T seja positivo).

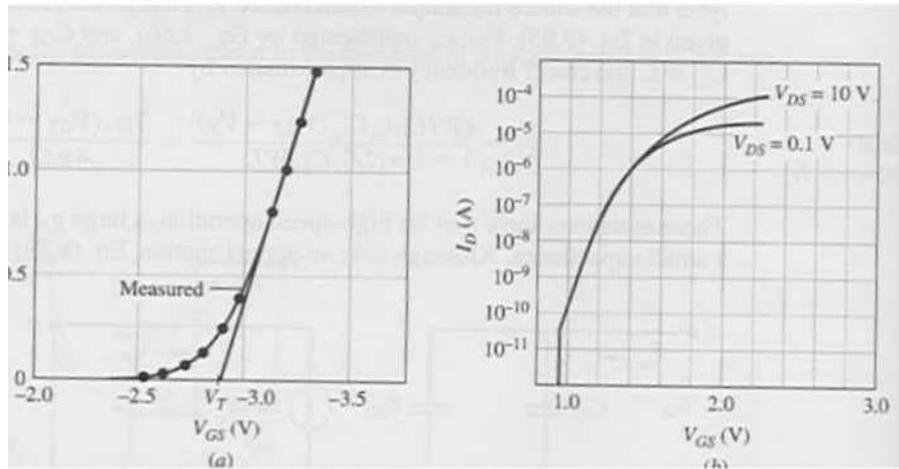


Fig. 14 a) Curva da raiz quadrada de I_D e b) $\log(I_D)$ versus V_G em torno e abaixo de V_T .

Uma expressão aproximada da corrente nesta região, válida para $V_{SB}=\text{fixo}$, é:

$$I_D = \frac{W}{L} I'_M e^{q(V_{GS}-V_T)/nkT} (1 - e^{-qV_{DS}/kT}) \quad (32)$$

onde:

$$I'_M = \mu \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{2\phi_F + V_{SB}}} \left(\frac{kT}{q} \right)^2$$

$$n = 1 + \frac{\gamma}{2\sqrt{2\phi_F + V_{SB}}}$$

γ é o parâmetro de efeito de corpo, a ser definido no item 4.2.

Um parâmetro usualmente empregado como característica da corrente sub-limiar de um MOSFET é o parâmetro S , de inclinação da corrente, definido como a variação de V_G correspondente à variação de uma década na corrente I_D , dado por:

$$S \equiv V_{GS}(I'_D) - V_{GS}(0.1I'_D) \quad (33)$$

Demonstra-se que em muitos casos vale a seguinte aproximação, comumente empregada:

$$S \cong \frac{kT}{q} \ln \left[1 + \frac{\epsilon_{Si} t_{ox}}{\epsilon_{ox} x_{d \max}} \right] \quad (34)$$

No passado, a corrente sub-limiar não representava muita relevância, tendo em vista que se usavam tensões relativamente altas no circuito. Atualmente, pelo alto nível de integração, bem como pelas várias aplicações que requerem baixo consumo de potência (sistemas móveis e outros), há uma demanda enorme por reduzir cada vez mais a tensão de alimentação. Isto faz com que os transistores operem freqüentemente

próximo à região de limiar e/ou sub-limiar. Assim, a apropriada modelagem da corrente nesta região torna-se cada vez mais importante. Uma questão mais complexa ainda refere-se à transição entre as regiões claramente sub-limiar e claramente linear. A corrente nesta região normalmente é modelada por equações empíricas de ajuste gradual entre as duas.

4. Efeitos Secundários Importantes

4.1 Variação da Mobilidade

Teoria de transporte de portadores por deriva ensina que a mobilidade, em material de corpo, definida como a razão entre a velocidade do portador e o campo elétrico aplicado, varia com a frequência de espalhamentos dos portadores com a rede e com as impurezas. Desta forma, ela normalmente é reduzida pelo aumento da concentração de impurezas (dopantes), bem como da temperatura. Agora, nos transistores MOS, o transporte dos portadores se dá na superfície do semiconductor. Isto dá origem a mais uma forma de espelhamento dos portadores, pelo “choque” do portador com a barreira de potencial da interface e/ou com a rugosidade desta mesma interface. Como consequência, a mobilidade de superfície é tipicamente bem mais baixa que a de corpo. Quanto maior o campo elétrico transversal ao canal, maior a frequência de colisões com a interface e menor será a mobilidade. Uma fórmula empírica, muitas vezes usada para modelar a mobilidade efetiva dos portadores no canal do MOSFET, é dado em (35). Outras fórmulas são também encontradas na literatura e usados nos diversos modelos.

$$\mu = \frac{\mu_0}{1 + \Theta(V_{GS} - V_T) + \Theta_B V_{SB}} \quad (35)$$

onde μ_0 é a mobilidade efetiva para V_{GS} próximo a V_T e Θ e Θ_B são parâmetros empíricos.

A Fig. 15 apresenta uma curva de I_D versus V_{GS} para $V_{DS}=0,1$ V. Para esta polarização de dreno estaremos em região linear na maior parte da excursão V_{GS} . Assim, de acordo com a equação (27), a tangente da curva fornece o valor da mobilidade (assumido conhecido os valores de W , L e C_O). Observa-se que a tangente é cada vez menor à medida que V_{GS} aumenta, o que está de acordo com a equação (35). A análise da variação da mobilidade com V_{GS} permite a determinação do parâmetro Θ . A extrapolação da maior tangente e sua intersecção com a abscissa representam uma outra forma de obtenção do V_T .

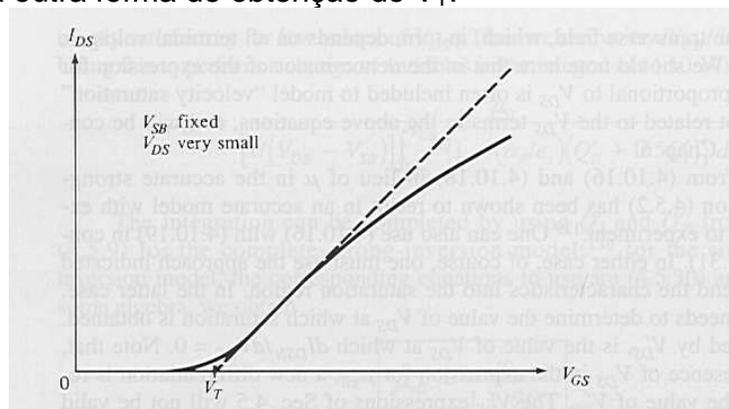


Fig. 15 Curva I_D versus V_{GS} para V_{DS} fixo e pequeno.

4.2 Efeito da Polarização de Substrato

Como nos vimos no item 1, a polarização reversa de um diodo, no nosso caso, a fonte do transistor MOS, altera o valor da tensão de limiar, como expressa pela relação (8). Esta alteração do V_T por sua vez, tem forte efeito sobre a corrente I_D (expressões 27 e 29).

Em muitos casos a fonte do transistor MOS está ligada à linha de terra do CI, porém isto nem sempre é verdade. Não é incomum ter dois transistores nMOS, por exemplo, colocados em série (uma porta NAND de duas entradas ou mais, por exemplo). Neste caso, a fonte do segundo transistor não está aterrada mas sim ligado ao dreno do primeiro. Ou seja, o potencial de fonte do segundo transistor é variável, e como conseqüência, o seu V_T será variável, afetando diretamente sua corrente. Isto faz com que o circuito torne-se mais lento. Este efeito será menor, quanto menor a variação do V_T com a polarização da fonte em relação ao substrato, V_{SB} . Podemos reescrever a equação (8) na forma (36) abaixo, onde o parâmetro γ é chamado de fator de corpo, dado pela expressão (37). Portanto, o ideal é ter o fator de corpo o menor possível. No entanto, por outras considerações não podemos reduzir a dopagem do substrato a nosso bel prazer, de forma que devemos conviver com um dado fator de corpo, caracterizando um dos parâmetros importantes do transistor, sendo um dos parâmetros de modelos SPICE.

$$V_T = V_{FB} + 2\phi_F + \gamma\sqrt{(2\phi_F + V_{SB})} \quad (36)$$

$$\gamma = \frac{1}{C_O} \cdot \sqrt{2 \cdot q \cdot \epsilon_{Si} \cdot N_A} \quad (37)$$

O parâmetro γ pode ser obtido pela medida de V_T para vários valores de V_{SB} , seguido da análise da curva de V_T versus raiz quadrada de $(2\phi_F + V_{SB})$.

4.3 Modulação do Comprimento Efetivo de Canal

A Fig. 16 ilustra a variação da corrente I_D na região de saturação do transistor. Este efeito já foi descrito no item 3.2 e explicado pelo efeito da redução do comprimento da região de inversão do canal, como ilustrado na Fig. 12. Este efeito é também chamado de modulação do comprimento efetivo de canal. A variação da corrente I_D com V_{DS} , na região de saturação, pode ser modelada pelo termo $(1 + \lambda V_{DS})$, como mostrado na expressão seguinte, do modelo da lei quadrática. Forma semelhante pode ser empregada com outros modelos de corrente. A Fig. 16 também indica como, pela inclinação da curva de corrente na região de saturação podemos determinar o valor de λ .

$$I_{Dsat} = \frac{W}{2L} \mu_n C_O (V_{GS} - V_T)^2 (1 + \lambda V_{DS})$$

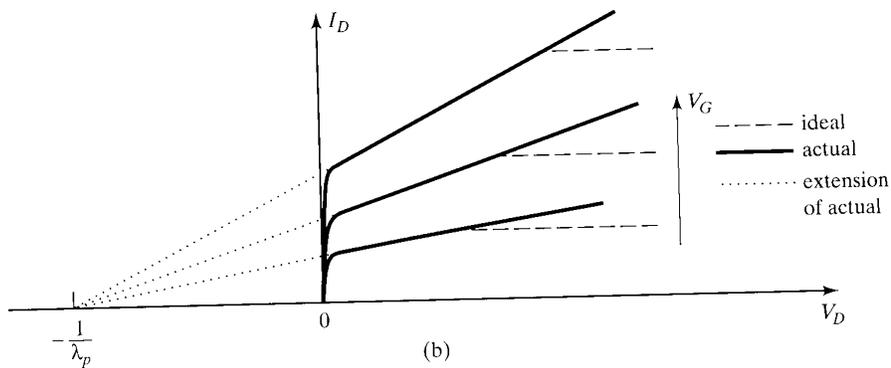


Fig. 16 Ilustração (exagerada) do efeito de modulação de comprimento efetivo de canal e a determinação do parâmetro λ .

4.4 Efeito de Temperatura

As características I-V de MOSFET são afetados por temperatura por duas maneiras:

a) O aumento da temperatura aumenta a frequência de colisões de portadores com a rede, reduzindo assim a mobilidades e como conseqüência a corrente de dreno. Este efeito é dominante para altas correntes, ou seja, para $V_{GS} \gg V_T$.

b) O aumento da temperatura reduz o potencial de Fermi. Isto pode ser visto pela relação (2), onde a variação da concentração intrínseca de portadores, n_i , é o fator predominante (este aumenta exponencialmente com a temperatura). A redução de ϕ_F reduz o valor de V_T , aumentando como conseqüência a corrente I_D . Este efeito é predominante para baixas correntes, ou seja, para V_G próximo a V_T .

A Fig. 17 mostra o efeito global da temperatura sobre a corrente de dreno. Observa-se que o transistor MOS possui um ponto de operação onde não ocorre variação com temperatura, onde os 2 efeitos descritos acima se cancelam mutuamente. Isto pode ser usado no projeto de circuitos em aplicações especiais. A variação da corrente de dreno é apropriadamente incluída nos modelos SPICE. Em geral, o efeito da temperatura sobre as características de MOSFET é muito menor que no caso de BJT.

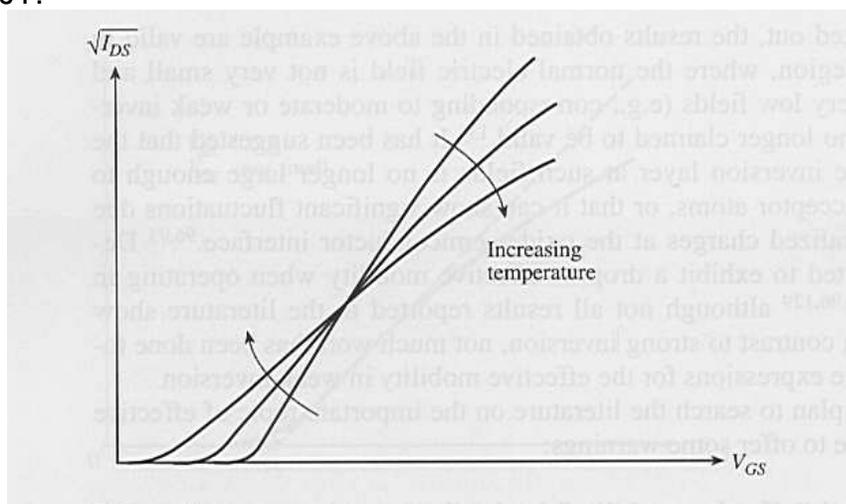


Fig. 17. Variação da corrente I_D versus V_{GS} , com aumento da temperatura.

4.5 Isolação entre MOSFETs em Circuitos Integrados

Num CI com transistores MOSFET, o processo de isolação entre os transistores é muito mais simples que no caso de CI's com transistores BJT, onde necessita-se realizar difusões para formar junções de isolação. No CI MOS, para garantir isolação, basta evitar a formação de um canal de inversão na superfície, nas regiões passivas, chamadas de regiões de campo. Isto deve estar garantido mesmo que passe por cima do óxido de campo uma linha metálica com a máxima tensão empregada no circuito. Para tanto, basta que esta região de campo tenha um V_T maior (bem maior de preferência) que V_{DD} , de alimentação.

Temos duas maneiras para controlar o V_T da região de campo (no transistor ativo vale o mesmo procedimento):

a) Aumentar a espessura do óxido da região de campo. Como pode ser visto pela relação (3), quanto maior a espessura do óxido, menor a capacitância C_O e maior o valor de V_T . A Fig. 18 mostra a estrutura esquemática de um transistor MOS com inclusão das regiões periféricas de óxido de campo espesso.

b) Aumentar a dopagem de substrato próximo à superfície do semiconductor. O aumento de N_A aumenta diretamente o valor de V_T como se observa da relação (3). Este aumento de dopagem pode-ser obtida por uma implantação iônica específica, com dopante do mesmo tipo ao do substrato, normalmente antes de realizar a etapa de oxidação de campo.

Vale ressaltar o efeito da carga efetiva de interface, Q_O , sobre o V_T , como podemos também observar da relação (3). Como a carga Q_O é sempre positiva, seu efeito é sempre desviar o V_T para a esquerda, no eixo de V_G . Isto ajuda a isolação entre transistores tipo pMOS, porém age em sentido contrário para a isolação entre transistores tipo nMOS. Atualmente isto não representa um grave problema, tendo em vista que temos um bom controle sobre a redução do valor de Q_O e adicionalmente, a técnica de implantação de íons nos permite corrigir adequadamente a dopagem de superfície, compensando o desvio negativo provocado pelo efeito da carga de interface. No passado, no entanto, este representou uma grande dificuldade, a tal ponto de se optar inicialmente (anos 60 e 70) pela tecnologia pMOS de CI's e não pela tecnologia nMOS, mesmo sendo o nMOS um dispositivo mais rápido que o pMOS ($\mu_n \sim 3\mu_p$).

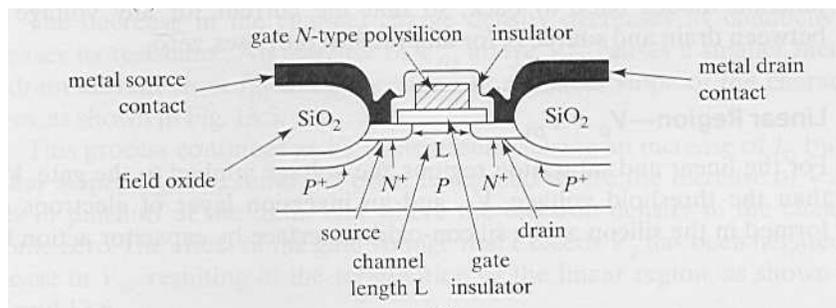


Fig. 18 Desenho esquemático da estrutura completa de um transistor nMOS, com inclusão da região periférica de óxido de campo. Note um óxido de campo espesso e um aumento da dopagem de substrato abaixo deste óxido.

4.6 Manipulação Destrutiva de MOSFETs

O uso de dielétrico de porta, normalmente dióxido de silício, apresenta uma das grandes vantagens do MOSFET, conferindo-lhe altíssima impedância de entrada,

comparada aos transistores BJTs. No entanto, este fato também traz uma grande desvantagem. O dióxido de silício apresenta ruptura dielétrica para campos elétricos da ordem de 2×10^7 V/cm. Assim, um transistor com espessura de óxido de porta de 10 nm, rompe com tensão da ordem de 20 V.

Uma pessoa pode carregar-se eletrostaticamente pela fricção com o ar, e/ou andando sobre um carpete. Ao tocar em seguida, um terminal ligado à porta de um transistor MOS, transfere esta carga, podendo dar origem a uma enorme tensão, da ordem de mil V por exemplo, dado a capacitância de porta ser muito pequena e a tensão gerado estar relacionada por Q/C . Esta tensão facilmente está em excesso à tensão de ruptura do dielétrico.

Para reduzir ou evitar esta ruptura destrutiva do MOSFET, devemos tomar o cuidado de aterrar o corpo, conectando o braço à terra. Adicionalmente é importante colocar dentro do CI, em paralelo ao terminal de porta do transistor de entrada do circuito, um dispositivo de proteção, que permita o escoamento de carga de manipulação, evitando assim o acúmulo de carga na porta do transistor. É comum incluir um diodo reversamente polarizado neste dispositivo de proteção. O diodo possui uma corrente reversa que permite descarregar a carga, e/ou o mesmo pode romper, descarregando a mais rapidamente, lembrando que o rompimento de um diodo não é destrutivo, como o é o dielétrico de porta.

5. Tipos de MOSFETs

No desenvolvimento deste capítulo consideramos sempre transistores tipo nMOS e sempre consideramos seu V_T como sendo um valor positivo. Este transistor é chamado de transistor nMOS de enriquecimento. Existe, no entanto, o transistor complementar ao primeiro, o chamado transistor pMOS de enriquecimento. A modelagem deste transistor é similar, bastando a troca de nomes de alguns parâmetros e alguns sinais nas fórmulas. O transistor pMOS de enriquecimento apresenta um V_T negativo. Por definição, transistor de enriquecimento é um transistor normalmente cortado, ou seja, se a tensão de porta for nula, a corrente será nula. Esta definição se aplica tanto ao nMOS como ao pMOS.

Tanto o nMOS como o pMOS têm também o transistor do tipo depleção, que significa um transistor normalmente conduzindo, ou seja, mesmo com a tensão de porta nula, pode passar corrente pelo transistor. Assim, o nMOS de depleção apresenta V_T negativo enquanto o pMOS de depleção apresenta V_T positivo.

Similar ao explicado sobre o controle de V_T das regiões de campo, no item 4.5, podemos controlar o V_T alterando, por exemplo, a dopagem na região do canal. Para tornar o transistor do tipo depleção podemos por exemplo reduzir a dopagem, ou ainda mais drasticamente, formar um canal por uma junção metalúrgica na superfície, por uma dopagem do tipo oposto ao do substrato.

A Fig. 19 ilustra os 4 tipos de transistores, incluindo um esquemático de sua estrutura, suas curvas características e seu símbolo. Existem, no entanto, diferentes símbolos empregados para transistores MOS, como podemos ver, para o caso do transistor nMOS e pMOS tipo enriquecimento, na Fig. 20.

| Type | Circuit Symbol | Cross Section | Output Characteristics | Transfer Characteristics |
|--|----------------|---------------|------------------------|--------------------------|
| <i>n</i> -Channel Enhancement (Normally Off) | | | | |
| <i>n</i> -Channel Depletion (Normally On) | | | | |
| <i>p</i> -Channel Enhancement (Normally Off) | | | | |
| <i>p</i> -Channel Depletion (Normally On) | | | | |

Fig. 19 Estrutura, curvas características e símbolo dos 4 tipos de MOSFETs: a) nMOS de enriquecimento, b) nMOS de depleção, c) pMOS de enriquecimento, d) pMOS de depleção.

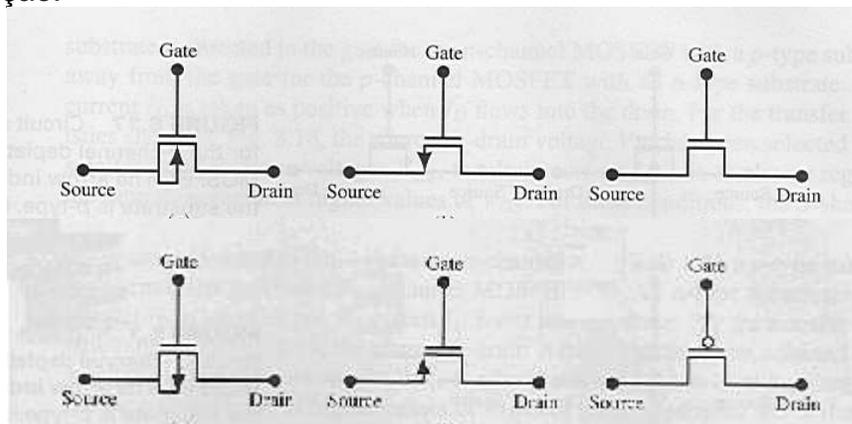


Fig. 20 Diferentes tipos de símbolos usados para representar transistores MOS, a) nMOS tipo enriquecimento e b) pMOS tipo enriquecimento. No símbolo da esquerda, a flecha indica o sentido do diodo entre substrato e canal. No símbolo do centro, a flecha indica o sentido da corrente pelo fonte do transistor (sai no nMOS e entra no pMOS). O símbolo da esquerda é o símbolo simplificado (no pMOS há um círculo de negação na entrada; pMOS funciona como lógica negativa)

6. Referências:

- 1 – H. Craig Casey, Jr, “Devices for Integrated Circuits”, John Wiley & Sons, 1999.
- 2 –Kanaan Kano, “Semiconductor Devices”, Prentice Hall, 1998.
- 3- Paolo Antognetti & Giuseppe Massobrio, “Semiconductor Device Modeling with Spice”,
- 4- R. Jacob Baker, Harry W. Li, David Boyce, “CMOS - circuit design, layout and simulation”,
- 5 - Yuhua Cheng, Chenming Hu, “Mosfet Modeling & Bsim user’s guide”,
- 6- Yannis Tsvividis, “Operation and Modeling of the Mos Transistor”, 2nd edition, WCB/McGrow Hill, 1999.
- 7- Orcad’s manual
- 8- IEEE - Transactions on Electron Devices n9 - September 1983 (1219-1228)
9. http://sc.tamu.edu/help/hspice/html_doc/manual/hspice-128.html