

MICROELECTRÓNICA

TEMA 3: EL TRANSISTOR MOS DE VOLUMEN.

ANÁLISIS AMPLIADO.

3.1	Velocidad de respuesta
3.2	Resistencias en serie
3.3	Corriente subumbral
3.4	Escalamiento
3.5	Efectos de canal corto
3.5.1	Velocidad límite
3.5.2	Variación del V_T
3.5.3	Variación del V_T por canal estrecho
3.5.4	Portadores calientes. Ruptura por avalancha.
3.5.5	Ruptura por “punch-through”

El transistor MOS de volumen, o sea, el que está fabricado sobre una oblea semiconductor cristalino que le sirve de substrato. Se ha ido desarrollando desde su invención en 1960 hasta la actualidad siguiendo la llamada “Ley de Moore”. Según esta ley la densidad de portadores en una pastilla de silicio, “chip”, deberá duplicarse cada 18 meses. Para que esto se cumpla se tienen que reducir las dimensiones del transistor, incrementar sus prestaciones y reducir su consumo de energía, todo esto incrementando el rendimiento en la fabricación (“yield”). Todo se ha cumplido por lo cual los transistores han ido modificándose de año en año para pasar de longitudes de canal de 10 μm a las dimensiones nanométricas que tienen en la actualidad. El funcionamiento físico de los transistores ha ido modificándose y el conocimiento de sus propiedades y limitaciones se hace de la mayor importancia. En el presente capítulo se analizarán algunos de los aspectos que afectan al transistor a medida que se van reduciendo sus dimensiones.

3.1 Velocidad de respuesta

La velocidad de respuesta de un transistor MOS está dada por el tiempo requerido para transferir correctamente una señal a la entrada del transistor, por ejemplo, el V_{GS} , a la salida de este, por ejemplo, I_{DS} . Los elementos que influyen en este tiempo de retardo, o velocidad de respuesta, son:

- a) El tiempo de tránsito a lo largo del canal de un portador desde la fuente hasta el drenaje;
- b) La carga y descarga de los capacitores inherentes al dispositivo, capacitores intrínsecos;
- c) La carga y descarga de los capacitores extrínsecos o parásitos.

Tiempo de tránsito.

El portador es acelerado por el efecto del campo eléctrico longitudinal E_y . Su velocidad es igual a $v = \mu E_y$. Consideramos que la caída de potencial a lo largo del canal es uniforme, por lo cual el potencial en cada punto del canal será igual a:

$$V(y) = \frac{V_D}{L} y \quad (3.1)$$

De la expresión de la velocidad podemos despejar el tiempo de transito t_{tr} según:

$$t_{tr} = \int_0^L \frac{dy}{\mu E_y(y)} \quad (3.2)$$

A partir de la expresión de corriente (2.5) sabemos que:

$$j = qn\mu E_y = \frac{I}{x_c W}, \quad (3.3)$$

Despejando E_y , considerando (2.57) y que se alcanzo la saturación:

$$E_y = \frac{I}{qn x_c W \mu} = \frac{\frac{W}{L} C_0 \mu \frac{V_{GT}^2}{2}}{Q_n W \mu} = \frac{\frac{W}{L} C_0 \mu \frac{V_{GT}^2}{2}}{-C_0 (V_{GT} - V(y)) W \mu} = -\frac{V_{GT}^2}{2L(V_{GT} - V(y))}, \quad (3.4)$$

Sustituyendo en (3.2) e integrando se obtiene la expresión para el tiempo de tránsito:

$$t_{tr} \approx \frac{L^2}{\mu V_{GT}}. \quad (3.5)$$

Como un ejemplo podemos calcular en forma aproximada este tiempo para el caso de $L=1 \mu\text{m}$, $\mu=500 \text{ cm}^2/\text{Vs}$ y $V_{GT}=2 \text{ V}$, obteniendo que $t_{tr}=10 \text{ ps}$, o sea que la frecuencia esta en el orden de 100 GHz. Este tiempo es mucho menor de lo que puede obtenerse con este transistor, por lo cual se puede concluir que el tiempo de tránsito no es la principal limitante de la velocidad de operación de un TMOS. El tiempo de transito depende de la longitud del canal al cuadrado, por lo que la reducción de la L reduce en gran medida este tiempo.

Capacitores intrínsecos.

En la figura 3.1 se muestran los capacitores intrínsecos de un TMOS. Son aquellos propios de la estructura como la capacitancia de compuerta C_0 y las capacitancias de las uniones P-N entre fuente y substrato C_{sb} y entre drenaje y substrato C_{db} .

Capacitores extrínsecos.

Estos son los capacitores que no son parte del principio de operación del transistor y deben ser reducidos lo más posible. Los capacitores de traslape (*overlap*) entre la compuerta y la fuente, C_{GS} , y la compuerta y el drenaje, C_{GD} , son los típicos capacitores parásitos que deben ser reducidos lo más posible.

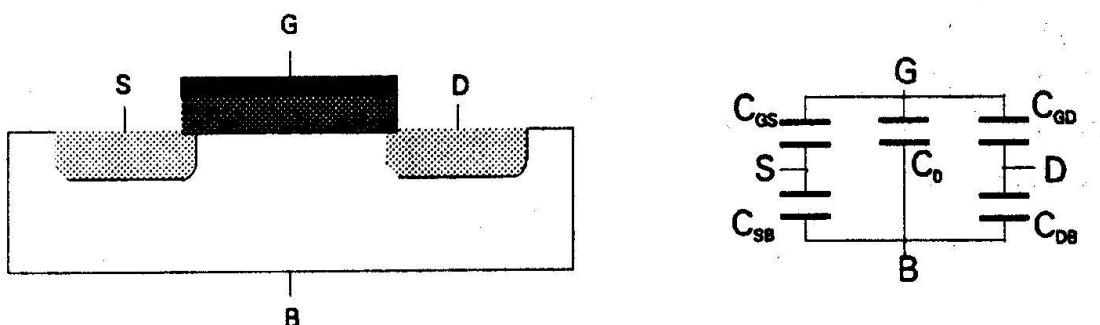


Fig. 3.1 Capacitancias en un transistor MOS, intrínsecas y parásitas.

3.2 Resistencias en serie.

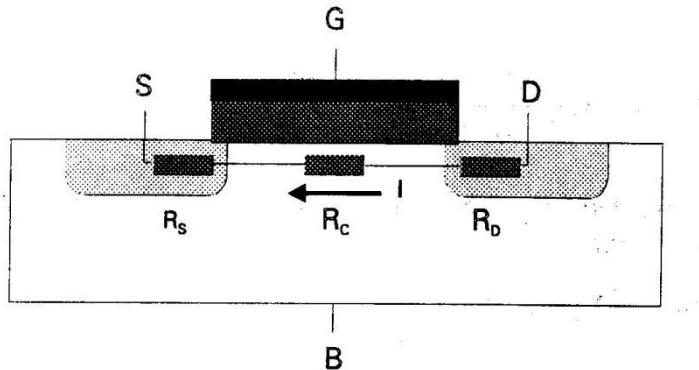


Fig. 3.2 Resistencias en un transistor MOS. R_s - resistencia de contacto y de la región del surtidor; R_d - resistencia del contacto y de la región del drenador.

La corriente en un TMOS fluye entre los electrodos externos de fuente y drenaje, S y D. El análisis realizado hasta el momento corresponde a la corriente a través del canal. Como se puede ver de la figura 3.2 además del canal existe en S y D un contacto metálico al semiconductor, que aunque es bajo, es diferente de cero. Además, entre el contacto metálico y el comienzo del canal en S o D hay una zona de semiconductor altamente dopado que tiene una resistencia dada, también diferente de cero. Ambas resistencias conforman una resistencia en serie en la fuente, R_s , y en el drenaje, R_d , y sumadas las dos se tiene la resistencia total en serie, $R_T = R_s + R_d$.

El efecto físico que introducen estas resistencias es el siguiente. Para un voltaje externo aplicado entre S y D, la caída de potencial en las resistencias hace que el voltaje real aplicado al canal se vea reducido, reduciendo la corriente.

Un cálculo de primera aproximación de este efecto es el siguiente.

La corriente en la región lineal se expresa como:

$$I_{ds} = \mu C_0 \frac{W}{L} \left[(V_{gs} - V_T) V_{ds} - \frac{V_{ds}^2}{2} \right] \quad (3.6)$$

Donde los voltajes V_{ds} y V_{gs} están considerados con respecto a los bordes del canal. Si los potenciales externos son V_{DS} y V_{GS} , entonces:

$$V_{DS} = V_{ds} + I_{DS}(R_T), \quad (3.7)$$

y

$$V_{GS} = V_{gs} + I_{DS}R_s. \quad (3.8)$$

Sustituyendo (3.7) y (3.8) en (3.6) se obtiene una ecuación de segundo orden para I_{DS} . Despreciando el término cuadrático y despejando se llega a la siguiente expresión para la corriente en función de las resistencias en serie:

$$I_{DS} = \frac{W}{L} \mu C_0 \frac{(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}}{1 + \frac{W}{L} \mu C_0 [(R_T)(V_{GS} - V_T) - R_D V_{DS}]} = \frac{\mu \left[\frac{W}{L} C_0 (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right]}{1 + \frac{W}{L} \mu C_0 [(R_T)(V_{GS} - V_T) - R_D V_{DS}]} \quad (3.9)$$

Comparando (3.6) y (3.9) se ve que el efecto de reducción de la corriente por las resistencias en serie es similar a que la movilidad se reduzca en ese mismo grado. En muchas ocasiones es difícil separar cuando la caída de corriente se produce por la movilidad y cuando por la resistencia en serie. Es importante señalar que la resistencia en serie no afecta la movilidad directamente.

Desde el punto de vista de diseño es necesario reducir la R_T lo más posible.

3.3 Corriente subumbral

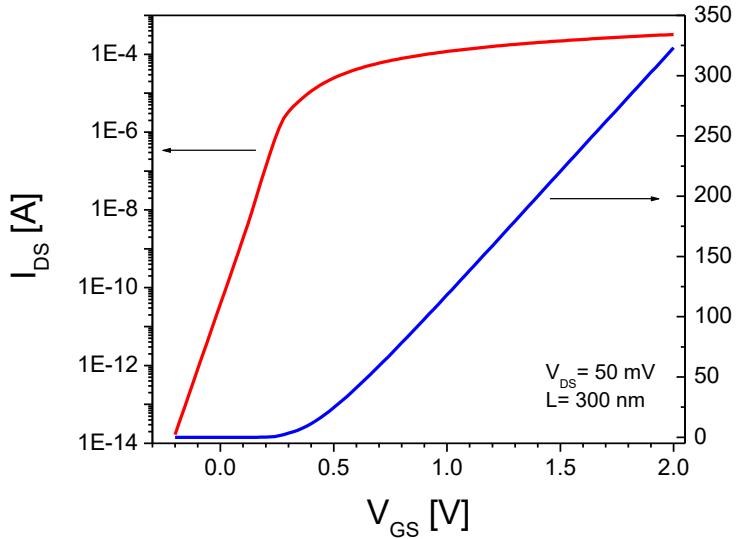


Figura 3.4 Característica transferencial en la región lineal. Curva lineal y curva semilogarítmica.

El análisis realizado hasta el momento es válido cuando el voltaje aplicado V_{GS} es mayor que el voltaje de umbral V_T . Como se ve en la figura 3.4 cuando la corriente se va reduciendo, cerca de V_T , la curva pasa de ser lineal a tener una curvatura dada, de forma tal que cuando se aplica un $V_{GS}=V_T$ la corriente es diferente de cero. Esta región se conoce como *región subumbral* y la corriente en la misma como corriente subumbral.

Las condiciones para calcular esta corriente tienen en cuenta que el potencial superficial está por debajo de $2\phi_F$, por lo tanto, está en inversión débil, y consideramos:

- 1) La concentración superficial está invertida, pero con poca concentración de carga móvil;
- 2) Como hay poca carga a lo largo del canal, el campo eléctrico longitudinal es mucho menor que el transversal;
- 3) Como el canal está casi cerrado junto al drenaje, se considera que la concentración de portadores junto a S es de n_s y junto a D de 0, desplazándose los portadores por difusión de S a D.

La densidad de carga será igual a:

$$\rho(u, v) = -qN_a [1 - e^{-u} + e^{u-2u_f-v} - e^{-2u_f}] \approx -qN_a [1 - e^{-u} + e^{u-2u_f-v}] \quad (3.10)$$

$$Es(u_s, v) = \sqrt{\frac{2qNa\phi}{\epsilon_s}} \sqrt{u_s + e^{u_s-2u_f-v}} \quad (3.11)$$

$$Q_n = \varepsilon_s E_s - Q_B = \sqrt{2qN_a \varepsilon_s \phi t} u_s \left[\sqrt{1 + \frac{e^{u_s - 2u_f - v}}{u_s}} - 1 \right] \approx \sqrt{\frac{qN_a \varepsilon_s \phi t}{2u_s}} \cdot e^{u_s - 2u_f} \cdot e^{-v} \quad (3.12)$$

$$I_{DSsub} = -AD_n \frac{d(qn)}{dy} = -WD_n \frac{d(qnx_c)}{dy} = -WD_n \frac{Q_n^D - Q_n^S}{L - 0} = \frac{W}{L} \phi t \mu_n (Q_n^S - Q_n^D) \quad (3.13)$$

si $D_n = \phi t \cdot \mu_n$; $V(0) = 0$ y $V(L) = V_D$

$$I_{DSsub} = \frac{W}{L} \mu_n \sqrt{\frac{qN_a \varepsilon_s}{2\varphi_s}} \phi t^2 \cdot e^{\frac{\varphi_s - 2\varphi_f}{\phi t}} \left[1 - e^{-\frac{V_D}{\phi t}} \right]. \quad (3.14)$$

Se ve en (3.14) que la corriente varía con el potencial superficial con un carácter exponencial, que en una curva semilog corresponde a una línea recta, como se ve en la figura 3.4. Para caracterizar esta región se utiliza un parámetro que se llama barrido subumbral o pendiente subumbral S. Se define como el incremento del voltaje que se debe hacer en la compuerta para que la corriente varíe una década (10 veces). Matemáticamente se define como:

$$S = \frac{1}{\frac{d(\log(I))}{dV_G}} = 2.3 \frac{1}{\frac{d(\ln(I))}{dV_G}} \quad (3.15)$$

$$V_G = V_{FB} + \varphi_s + \gamma \sqrt{\varphi_s}$$

$$\frac{dV_G}{d\varphi_s} = 1 + \frac{\gamma}{2\sqrt{\varphi_s}} = 1 + \frac{\sqrt{2qN_a \varepsilon_s}}{2C_o \sqrt{\varphi_s}} = 1 + \frac{\frac{\varepsilon_s}{2\varepsilon_s \varphi_s} \sqrt{\frac{qN_a}{2\varepsilon_s \varphi_s}}}{C_o} = 1 + \frac{x_d}{C_o} = 1 + \frac{C_s}{C_o} \quad (3.16)$$

Calculando la derivada en (3.15)

$$\frac{d(\ln(I))}{dV_G} = \frac{d(\ln(I))}{d\varphi_s} \frac{d\varphi_s}{dV_G} = \frac{1}{\phi t} \frac{1}{1 + \frac{C_s}{C_o}} \quad (3.17)$$

$$S = 2.3 \phi t \left(1 + \frac{C_s}{C_o} \right) \approx 60 \left(1 + \frac{C_s}{C_o} \right) \approx 60 \left(1 + \frac{\varepsilon_s}{\varepsilon_o} \frac{x_o}{x_d} \right) \quad \left[\frac{mV}{decada} \right] \quad (3.18)$$

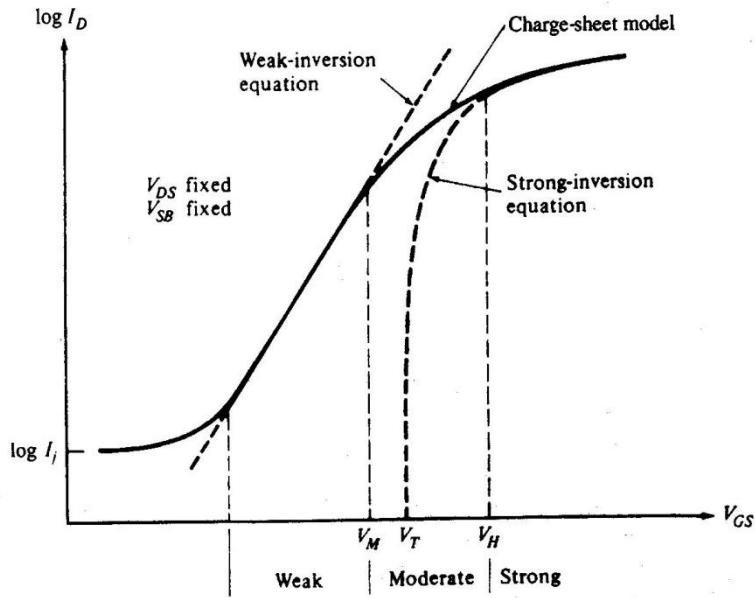


Figura 3.5 Diferentes partes de la corriente subumbral.

3.4 Escalado (“Scaling”)

La reducción de las dimensiones de los transistores recibe el nombre de *escalado*. Como en un proceso industrial el escalado conlleva el incremento o la reducción de los parámetros que definen un proceso, o dispositivo, para que con las nuevas dimensiones funcione como era antes del escalado.

Generalmente en la industria el escalado se hace en forma incremental, pasando de una maqueta experimental, a un prototipo, a una línea industrial. En la microelectrónica es al revés, se requiere reducir las diferentes dimensiones y variar los parámetros del transistor de forma de obtener un transistor más pequeño con un comportamiento similar al de partida. O sea, un transistor de canal corto que se comporte como uno de canal largo.

La reducción de estas dimensiones, para seguir la Ley de Moore en los circuitos integrados, se realizó durante muchos años utilizando unas reglas de escalado. Estas reglas consistían en definir un factor de variación α por el cual se multiplicaban o dividían las magnitudes.

Para realizar se suponían algunas condiciones previas como son: escalado a voltaje constante, cuando el voltaje de la fuente se mantiene fijo; escalado a campo constante, cuando el campo interno en el canal se mantiene fijo y un escalado mixto que tomaba varios criterios.

Un ejemplo de usar estos criterios se muestra en la tabla 3.1. Como se puede ver en la tabla, cada parámetro tiene una limitación para seguir reduciéndose, por eso es necesario variar cada parámetro en forma independiente. Cuando la reducción se realizó hasta un micrómetro, se utilizaba casi en todo un mismo parámetro. Para bajar de la micra se requiere un ajuste parámetro por parámetro, haciendo el trabajo de escalado una tarea ardua y costosa.

TABLA 3.1 Ejemplo de escalado donde el parámetro α es diferente para cada uno.

Parámetro	Regla utilizada	Factor limitante	1972	1992	Relación
Longitud del canal [μm]	$1/\alpha$	Fotolitografía	6	0.1	1/60
Anchura del canal [μm]	$1/\alpha$	Fotolitografía			1/60
Espesor del dieléctrico [nm]	$1/\alpha$	Corriente de tuneleo	100	3	1/30
Profundidad de la unión P-N [μm]	$1/\alpha$	Resistencia de capa	0.8	0.04	1/20
Concentración de dopaje [cm^{-3}]	α	Corriente de fuga de la unión	10^{16}	10^{18}	100
Voltaje de alimentación [V]	$1/\alpha$	Límite de V_T Velocidad	5	1.5	1/3
Voltaje umbral [V]	$1/\alpha$	Corriente de fuga	0.8	0.4	1/2
Corriente de saturación /W [$\mu\text{A/mm}$]	1		50	500	10
Retardo de propagación	$1/\alpha$		4000	15	1/270

3.5 Efectos de canal corto

Cuando se comienzan a reducir las dimensiones de la longitud de los canales de los TMOS comienzan a aparecer diferentes fenómenos físicos que modifican el comportamiento del transistor. Estos efectos se conocen como *efectos de canal corto* (“short channel effects, SCE”). Entre los habituales podemos señalar:

Saturación de la velocidad de los portadores.

Variación del voltaje umbral.

DIBL – reducción del potencial en la fuente por el voltaje en el drenaje.

Incremento de la pendiente subumbral S.

Corriente de compuerta: tuneleo directo, GIDL.

Rupturas: avalancha, “punch through”.

3.5.1 Velocidad de saturación.

Los portadores de carga con una masa efectiva m_n se desplazan en el semiconductor con una velocidad térmica v_t y sabiendo que la energía térmica por cada grado de libertad, que son tres, es igual $kT/2$, tendrán una energía cinética igual a:

$$\frac{1}{2}m_n v_t^2 = \frac{3}{2}kT \quad (3.19)$$

Cuando hay un campo eléctrico se aceleran los portadores, con una velocidad de desplazamiento para un campo longitudinal E_y igual a:

$$v_d = -\mu E_y = \frac{|\mu E_y|}{\left[1 + \left(\frac{E_y}{E_c}\right)^\alpha\right]^\frac{1}{\alpha}} \quad (3.20)$$

Donde α es igual a 2 para los electrones y 1 para los huecos. El valor de la velocidad máxima se conoce como velocidad de saturación, v_{sat} y se define como:

$$v_{dmax} = v_{sat} = \mu E_c = \mu \frac{V_{dsat}}{L}. \quad (3.21)$$

¿Qué ocurre con la corriente cuando se llega a la velocidad de saturación?

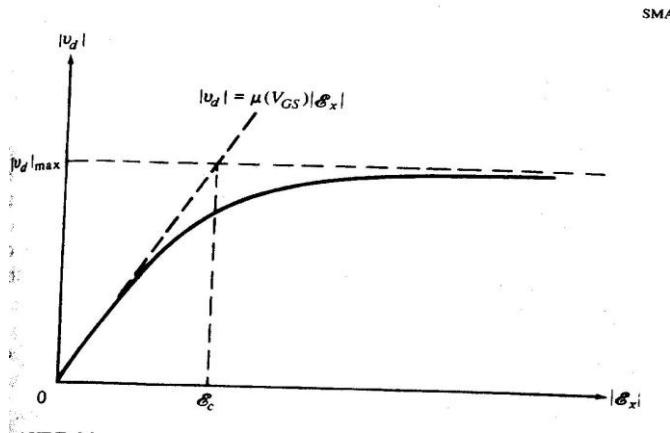


Figura 3.5a Velocidad de los portadores en función del campo eléctrico

Como conocemos,

$$I_D dy = -Z \mu Q_n dV_D; \quad E_y = -\frac{dV}{dy} \quad \text{and} \quad 1 + \left| \frac{V_D}{LE_c} \right| \Rightarrow 1 + \left| \frac{\mu V_D}{Lv_{sat} L} \right| \quad (3.22)$$

Y sustituyendo en la corriente resulta:

$$I_D = \mu W Q_n E_y = W Q_n \frac{\mu E_y}{1 + \frac{E_y}{E_c}} = -\frac{W \mu}{1 + \left| \frac{\mu V_D}{v_{dsat} L} \right|} \left(-\frac{dV}{dy} \right) Q_n$$

$$I_D \int_0^L dy = -\frac{W \mu}{1 + \left| \frac{\mu V_D}{v_{dsat} L} \right|} \int_0^{V_D} Q_n dV$$

$$I_D = -\frac{W}{L} \frac{\mu}{1 + \left| \frac{\mu V_D}{v_{dsat} L} \right|} \int_0^{V_D} Q_n dV$$

Obteniéndose la expresión de corriente de primera aproximación igual a:

$$I_D = \frac{W}{L} C_o \frac{\mu}{1 + \left| \frac{\mu V_D}{L v_{sat}} \right|} \left[(V_G - V_T) V_D - \frac{V_D^2}{2} \right] \quad (3.23)$$

El voltaje de drenaje, V_D , para el cual se alcanza la condición de la velocidad de saturación será el nuevo voltaje de saturación V_{Dsat} . Este voltaje es menor que el calculado anteriormente por el efecto de pellizco (*pinch-off*) y dependerá de la longitud del canal del transistor.

El cálculo analítico de este voltaje se hace igual que antes, o sea, considerando que la derivada de la corriente se hace cero:

$$\frac{dI_D}{dV} = \frac{\mu Z Q_n}{L \left(1 + \left| \frac{\mu V_D}{L v_{sat}} \right| \right)} + \frac{Z \mu \int_0^{V_D} Q_n dV}{L^2 E_c \left(1 + \left| \frac{\mu V_D}{L v_{sat}} \right| \right)^2} = 0 \quad (3.24)$$

Resultando:

$$V_{Dsat} = |LE_c| \left[\sqrt{1 + \frac{2(V_G - V_T)}{|LE_c|}} - 1 \right] = \left| \frac{Lv_{sat}}{\mu} \right| \left[\sqrt{1 + \frac{2(V_G - V_T)}{\left| \frac{Lv_{sat}}{\mu} \right|}} - 1 \right]. \quad (3.25)$$

Un momento importante es el hecho que cuando se alcanza el V_{Dsat} por la saturación de velocidad la carga en el lado del drenaje es diferente de cero, a diferencia del caso anterior donde se consideraba igual a cero.

3.5.2 Variación del voltaje umbral por canal corto.

Hasta el momento hemos estudiado TMOS con un canal lo suficientemente largo como para que lo que ocurra junto a las regiones de S y D no sea tomado en cuenta. En la figura 3.7 se muestran dos transistores con diferentes L.

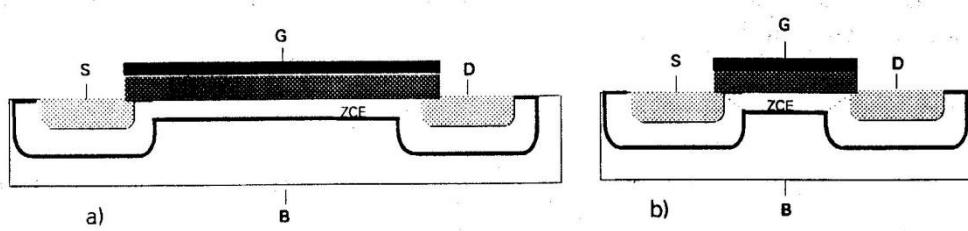


Fig. 3.6 Geometría trapezoidal del canal bajo la compuerta: a) canal largo; b) canal corto.

Lo característico de esas estructuras es que se unen las ZCE de las uniones y del canal. En el caso del canal largo la ZCE correspondiente al canal, bajo la compuerta, es mucho mayor que la corresponde a las uniones P-N. Al acortarse el canal estas regiones tienen dimensiones similares, lo cual cambia la distribución de carga, debido a que ahora la carga que controla la compuerta es menor. Realizaremos un cálculo aproximado para demostrar esto.

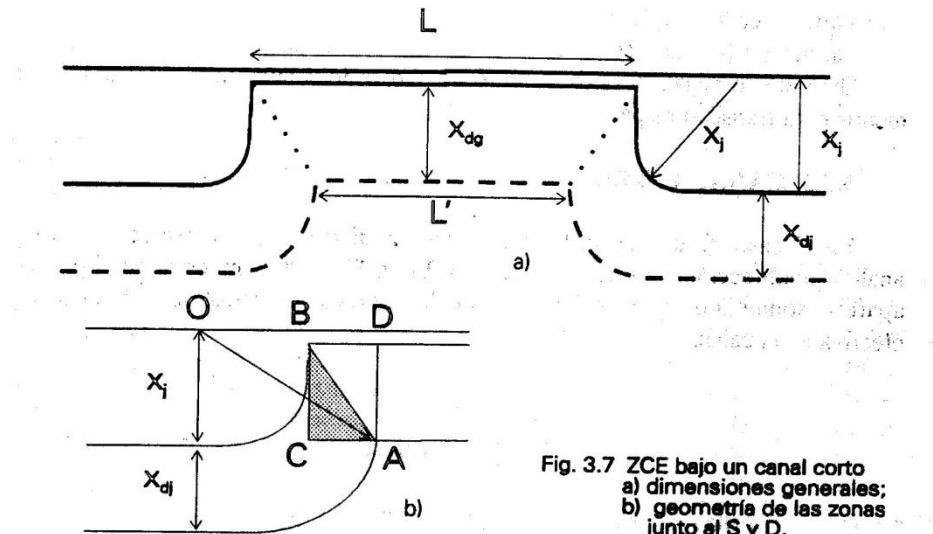


Fig. 3.7 ZCE bajo un canal corto
a) dimensiones generales;
b) geometría de las zonas junto al S y D.

El método geométrico que presentaremos a continuación se le conoce como método del trapecio o de Poon y Yau.

Si dividimos el área común bajo la compuerta en triángulos, y definimos x_{dg} la profundidad de la ZCE bajo la compuerta; x_{dj} la profundidad de la ZCE bajo la unión P-N y x_j la profundidad de la unión, tendremos:

$$BC = x_{dg} \quad AO = x_j + x_{dj} \quad (3.26)$$

$$DB = AC \quad x_{dj} \approx x_{dg}$$

Considerando $x_{dg}=x_{dj}$ por la regla de los lados de un triángulo

$$(x_j + x_{dj})^2 = x_{dj}^2 + (AC + x_j)^2 \quad (3.27)$$

Despejando AC:

$$AC = x_j \left[\sqrt{1 + \frac{2x_{dj}}{x_j}} - 1 \right]$$

(3.28)

podemos calcular el área de la región rayada que estará a ambos lados del canal, que es igual a:

$$x_{dj} \cdot x_j \left[\sqrt{1 + \frac{2x_{dj}}{x_j}} - 1 \right] \quad (3.29)$$

Ahora podemos calcular la carga bajo la compuerta en la zona del canal que puede controlar, la región trapezoidal. La carga total bajo la compuerta sería igual a Q_{Btotal} y la carga resultante en el trapecio es igual a Q_{Btrap} , son iguales a:

$$Q_{Btotal} = qN_a WLx_{dg} \quad (2.30)$$

$$Q_{Btrap} = Q_{Btotal} - qN_a W \left[x_{dg} x_j \left(\sqrt{1 + \frac{2x_{dj}}{x_j}} - 1 \right) \right] = Q_{Btotal} \left[1 - \frac{x_j}{L} \left(\sqrt{1 + \frac{2x_{dj}}{x_j}} - 1 \right) \right] \quad (2.31)$$

Como el voltaje umbral depende de la carga fija Q_B , cuando esta varía el V_T variará también, por eso:

$$\frac{\Delta V_T}{V_T} = \frac{V_T - V_T}{V_T} = 1 - \frac{Q_{Btrap}}{Q_{Btotal}} = \frac{x_j}{L} \left(\sqrt{1 + \frac{2x_{dj}}{x_j}} - 1 \right). \quad (2.32)$$

Como se ve de (3.32) al reducir la L se produce una reducción del V_T , efecto que se conoce como “roll-off”. Para evitar esto si se tiene una L dada se requiere hacer las uniones más superficiales y aumentar la concentración del dopaje en el substrato y las uniones para reducir la profundidad de las ZCE.

Otro efecto indeseable es el incremento del barrido subumbral S .

3.5.3 Variación del voltaje umbral por canal estrecho

Cuando la anchura del canal W se reduce, la distribución de la carga fija también varía. En la figura 3.8 se puede ver el área sombreada que incrementa la ZCE que puede controlar la compuerta.

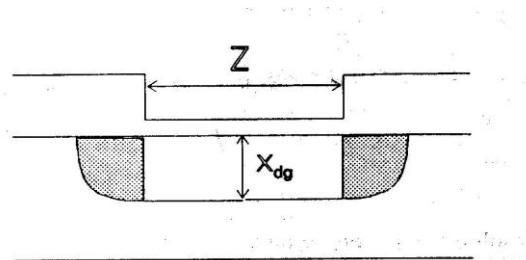


Fig. 3.8 Anchura del canal de un TMOS mostrando la ZCE que se prolonga más allá de los bordes.

Utilizaremos un cálculo similar al anterior basándonos en la geometría de la estructura de la figura 3.8. La carga inicial bajo la compuerta la llamaremos Q_{Bini} y la carga total considerando las zonas sombreadas como Q_{Btot} , ambas serán igual a:

$$Q_{Bini} = WLqN_a x_{dg} \quad (3.33)$$

$$Q_{Btot} = Q_{Bini} + \left(2 \frac{\pi x_{dg}^2}{4} \right) q N_a L = Q_{Bini} \left(1 + \frac{\pi x_{dg}}{2W} \right) \quad (3.34)$$

Por lo tanto, la variación de V_T será igual a:

$$\frac{\Delta V_T}{V_T} = + \frac{\pi}{2} \frac{x_{dg}}{W}. \quad (3.35)$$

En este caso el voltaje umbral se incrementa al reducirse la anchura del canal con respecto a la profundidad de la ZCE de la compuerta.

3.5.4 Portadores calientes. Ruptura por avalancha.

Los portadores de carga, por ejemplo, electrones tienen una energía cinética que está en equilibrio con la energía de la red cristalina. Esta energía corresponde a $kT/2$ por cada grado de libertad del átomo de la red. Para los tres grados de libertad de una estructura volumétrica la energía media será de $3kT/2$. Cuando el electrón es acelerado por el campo eléctrico adquiere una energía cinética superior a la de equilibrio, lo que se puede considerar como si el electrón tuviera una temperatura mayor, por lo cual se le denominan electrones calientes.

El efecto que pueden producir estos electrones energéticos es que choquen con un átomo de la red, extrayendo un electrón y dejando un hueco, o sea, creando un par electrón-hueco. Si la energía del electrón liberado es suficiente para poder liberar otro electrón y así sucesivamente, se puede producir una cadena de choques que generen pares, lo que llama avalancha. Este efecto puede llevar a un crecimiento abrupto de la corriente de drenaje produciendo una ruptura del transistor. Esta ruptura es reversible si el calentamiento del TMOS no sobrepasa su límite de ruptura física.

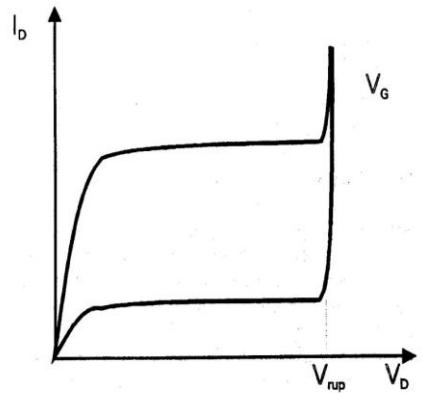


Fig. 3.9 Zona de ruptura en la característica I-V del TMOS

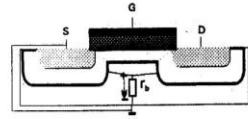


Fig. 3.10 Polarización de la juntura de la fuente por efecto de la corriente de substrato.

Es importante señalar que cuando se producen pares junto al D, los electrones se atraen hacia el D, y los huecos lo hacen por la región homogénea del substrato hacia el S y hacia el contacto del substrato. Como el substrato tiene una resistencia se produce una caída de potencial que por su polaridad tiende a polarizar en directa la unión S-B, incrementando la corriente más aún, o sea, se produce una retroalimentación positiva que lleva más rápidamente el transistor a la ruptura.

Si el campo transversal junto al drenaje es suficientemente alto, está avalancha puede producir otro fenómeno conocido como inyección en el dieléctrico. Estas cargas, huecos o electrones, se quedan atrapadas en el óxido produciendo un corrimiento del V_T . Este fenómeno es utilizado en las memorias no-volátiles EEPROM.

3.5.5 Ruptura por “punch-through”.

Si se incrementa el voltaje de drenaje por encima del de saturación, la ZCE de la unión drenaje-substrato comienza a avanzar recortando la longitud del canal. Si el voltaje se incrementa mucho, puede llegar el momento en que la ZCE de la unión fuente-substrato se toque con la de drenaje-substrato. Este fenómeno se conoce como “punch-through”, que es similar al que se puede producir en la base de un transistor bipolar, conlleva a un incremento brusco de corriente sin control, o sea, a una ruptura.

Si la potencia disipada durante este incremento de corriente se puede mantener debajo de un valor dado, el transistor no se rompe físicamente. Esto se logra si hay una resistencia limitadora en serie. Este tipo de rupturas se conoce como reversible.