

EL652: Seminario de Diseño de Circuitos Integrados Análogos

Apunte de Clase, Parte 1

Introducción

La tecnología CMOS (*complementary metal oxide semiconductor*) ha dominado en el ámbito de la fabricación de circuitos integrados por los últimos 25 años, y probablemente lo siga haciendo por otros 25 años. ¿Por qué? La tecnología CMOS es robusta, fácil de fabricar, tiene bajo consumo de potencia, es de bajo costo, y, lo más importante, es escalable. Este último punto fue observado por Gordon Moore (fundador de Intel) el año 1965, en un artículo donde proyectaba que la cantidad de transistores dentro de un circuito integrado se duplicaría cada 18 años. Esta tendencia, denominada Ley de Moore, se ha mantenido hasta el día de hoy, llegando a dispositivos fabricables con largo de canal del orden de 32 nanómetros.

Los sistemas electrónicos actuales pueden contener desde algunos miles a algunos millones de transistores (caso en el cual se denominan circuitos VLSI, *very large scale integration*), y en muchos casos algunos billones de transistores o más (circuitos ULSI, *ultra large scale integration*). Estos sistemas pueden contener secciones que se encargan de procesar señales digitales o análogas, o que constituyen una interfaz entre ambos tipos de señales (también denominados circuitos de señal mezclada, fundamentalmente los conversores digital análogo y análogo digital). La complejidad de estos circuitos, sumado al hecho que las pequeñas dimensiones de fabricación introducen infinidad de problemas que pueden afectar la etapa de diseño, han motivado un renacimiento del diseño de circuitos a nivel de transistor. En esta clase, se introducirá el tema del diseño de circuitos integrados a nivel de transistor, que para el caso de circuitos digitales se denomina diseño "a medida" (*custom design*), en oposición a la síntesis automática de circuitos lógicos con librerías de celdas estándar o en plataformas FPGA. Estos métodos son utilizados para partidas de fabricación de baja escala, que necesitan un tiempo de diseño reducido. Sin embargo, la mayoría de los chips de producción en masa son ejemplos de diseño "a la medida". En el caso de los circuitos análogos, este es el método tradicional de diseño, y requiere profesionales con muy alto nivel de especialización y conocimientos. Se describirá de manera general el flujo de diseño en cada una de sus etapas, y las metodologías utilizadas en cada una de ellas.

Por otra parte, se hará una introducción a la estructura física y al modelo de canal largo del transistor MOS. Se ilustrarán las curvas características y las regiones de operación, en base a los supuestos considerados válidos para cada caso.

Diseño de Circuitos Integrados CMOS

El flujo de diseño de circuitos integrados a nivel de transistor (Figura 1-1) consiste en una definición de las características generales y especificaciones del circuito, y en particular la determinación de los terminales de entrada y salida, seguido por un proceso de cálculo a mano y elaboración del esquemático circuital. Posteriormente, se hace una simulación del circuito mediante herramientas computacionales (por ejemplo, SPICE), para verificar el funcionamiento y ajustar algunos parámetros de manera empírica, para así cumplir con los requerimientos del circuito. El proceso de implementación física o *layout*, que se lleva a cabo a continuación, en general la ejecutan profesionales especialistas, pero un diseñador de circuitos debe conocer esta etapa y los efectos parásitos de las diversas estructuras implicadas en el diseño (por ejemplo, capacitancias, inductancias, junturas PN, transistores parásitos, entre otros), para proveer de directrices apropiadas al diseñador de layout. Posteriormente a la verificación y re-simulación del circuito considerando el layout, se pasa a una etapa de prototipaje, que permite una producción final del circuito.

La fabricación de circuitos integrados se lleva a cabo en fábricas especializadas (por ejemplo, TSMC, ONSemiconductors, IBM, AMS, entre muchas otras), a través de un proceso que consta de varias operaciones que buscan construir las diversas estructuras de los dispositivos que forman parte del circuito, y que serán revisadas más adelante. Estos pasos requieren la producción de máscaras de fabricación, que permiten manejar el efecto de la luz sobre materiales fotosensibles, que luego servirán para definir patrones de dopado y deposiciones de diversos materiales sobre el silicio. La fabricación se lleva a cabo sobre una placa circular de silicio denominada oblea (*wafer*), donde se fabrican múltiples chips, que son cortados y que conforman la unidad denominada comúnmente *die*. Esto se muestra en la Figura 1-2.

El costo fijo de producir una partida de circuitos integrados es muy grande debido al costo altísimo de las máscaras de fabricación; por ello, se ha tendido a producir obleas cada vez más grandes, en la medida que su característica mecánica lo permite, lo que posibilita fabricar más circuitos con una misma máscara. En la Figura 1-3 se muestran obleas con diámetros de 150, 200 y 300 milímetros, siendo la de fabricación más actual la última.

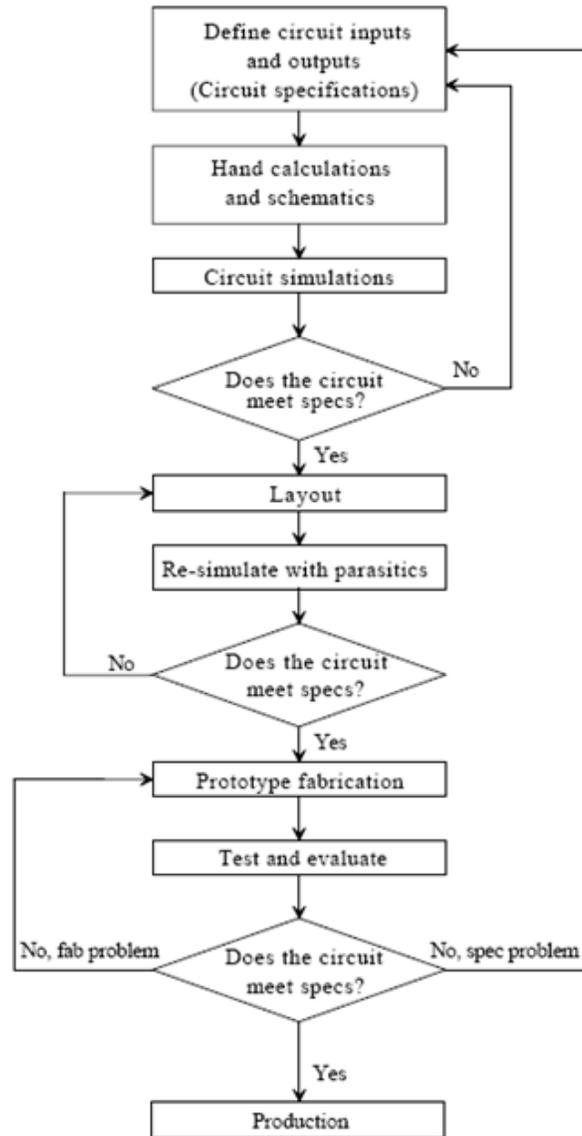


Figura 1-1: Flujo de diseño de circuitos integrados a nivel de transistor

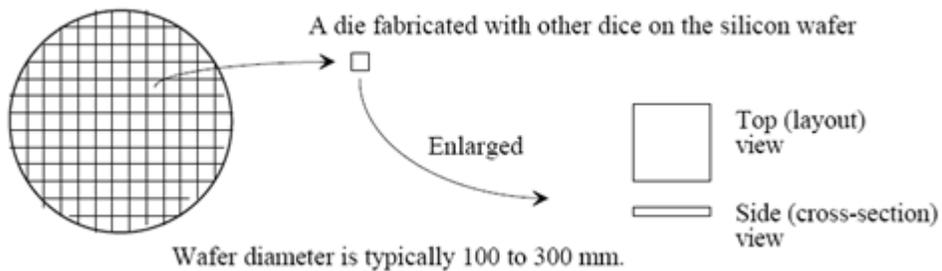


Figura 1-2: Esquema de las obleas de silicio y los *die* que contienen a cada chip.

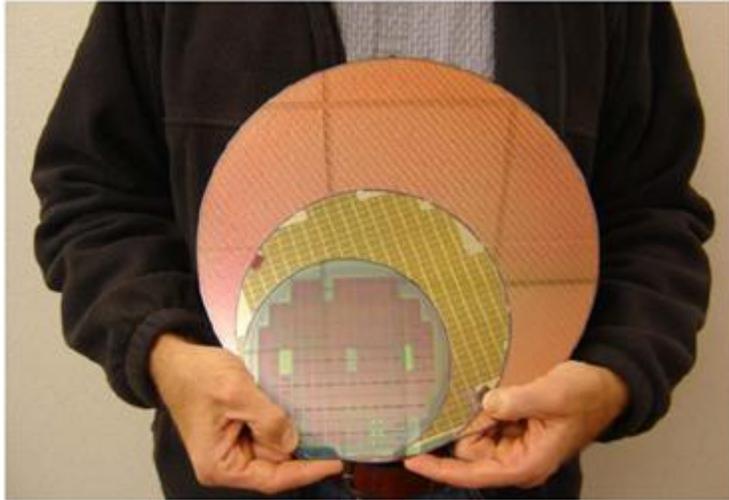


Figura 1-3: Obleas de silicio de 150, 200 y 300 milímetros

Proceso de fabricación estándar de circuitos CMOS

En esta sección, se presenta un flujo típico de proceso de fabricación de tecnologías CMOS. La Figura 1-4 muestra los pasos de fabricación tanto para un dispositivo NMOS como para uno de tipo PMOS.

En general, se asume que la oblea, una vez fabricada de silicio puro, es dopada con contaminantes de tipo P (átomos de Boro). En el primer paso mostrado (Figura 1-4 (a)) se asume que la etapa de fabricación de la fosa tipo N (*N-well*) ya ha sido llevada a cabo. Se genera una capa de óxido que sirve para separar al silicio de la capa de nítrido que se depositará sobre la oblea. Sobre el nítrido, se generan los patrones de foto-resistencia que permitirán formar el óxido de aislamiento de los dispositivos (*field oxide*, FOX). Luego de pulir la superficie, las cavidades donde se formará el óxido de aislamiento son generadas, como se muestra en la Figura 1-4 (b). La exposición a oxígeno y el posterior pulido de la oblea permiten obtener la estructura de la Figura 1-4 (c). Este conjunto de operaciones se denominan STI (*shallow-trench isolation*) y le dan el nombre al tipo de proceso, en contraste con la tecnología anterior llamada LOCOS (*local oxidation of silicon*). El paso siguiente consiste en el implante de impurezas que permitan dar valores apropiados al voltaje de umbral de cada dispositivo, localizados mediante patrones de foto-resistencia.

A continuación, se procede a formar la puerta del dispositivo. Para ello, en primer lugar se deposita un óxido de muy buena calidad (i.e. que no contenga impurezas que puedan afectar el funcionamiento del dispositivo), sobre el cual se deposita y se define el patrón del poli-silicio; el corte transversal de la estructura resultante se muestra en la Figura 1-4 (e).

Una serie de implantes se aplican a continuación. En primer lugar, se hace un implante superficial que se difunde un cierto tramo por debajo del óxido (ver Figura 1-4 (f)), y que impide la formación de campos de muy alto valor que generan efectos negativos sobre el funcionamiento del dispositivo (fundamentalmente, una corriente no despreciable por el terminal de cuerpo debido al

efecto *hot carrier*). A continuación, se hace crecer óxido a los costados del poli-silicio; esta estructura se llama *spacer*, y tiene como función actuar como máscara para el implante fuerte de fuente y drenaje. Éste último se muestra en la Figura 1-4 (g); se debe notar que la puerta actúa como máscara, impidiendo el paso del implante hacia el canal. Finalmente, se aplica una capa de metal (titanio o cobalto) templado, que forma una capa denominada *salicide* (*self-aligned silicide*), y cuya función es disminuir la resistividad del poli-silicio y de las regiones de dopado fuerte.

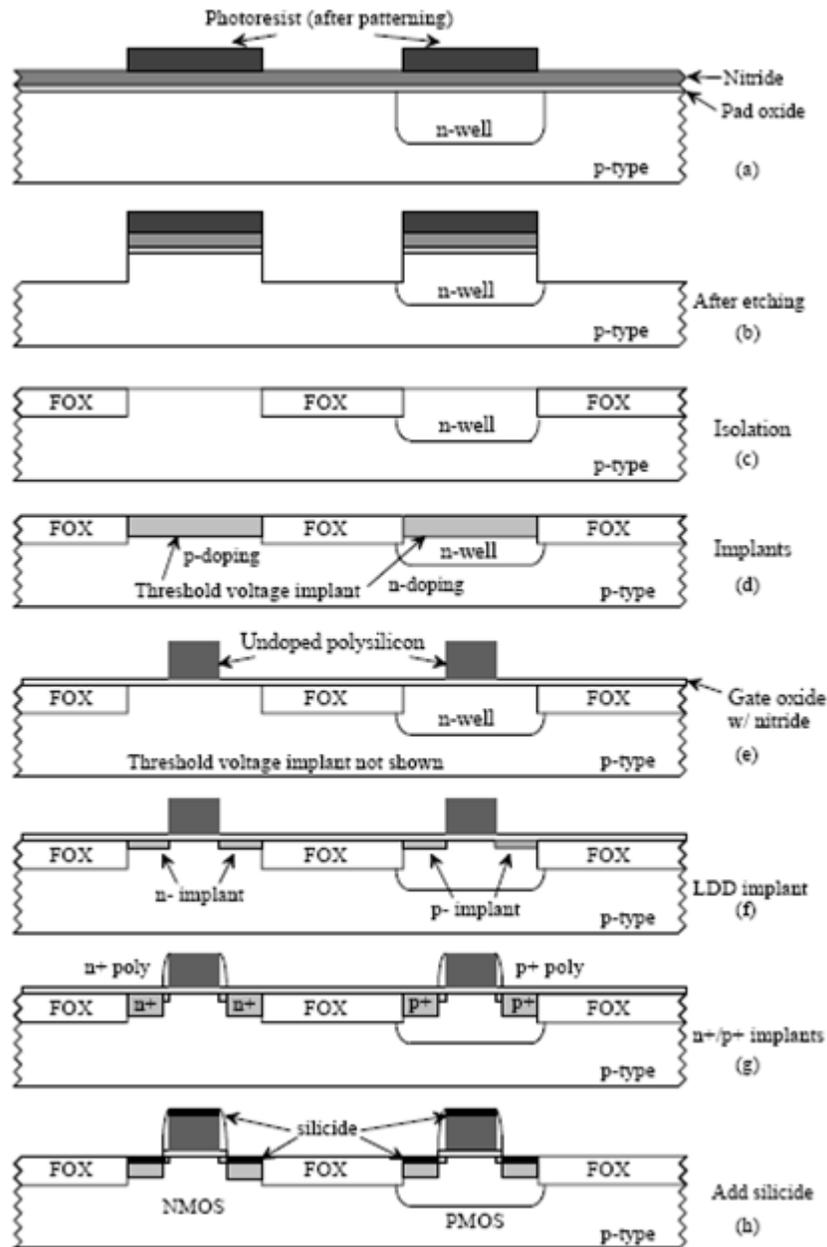


Figura 1-4: Flujo de proceso de fabricación estándar CMOS

El Transistor MOS

Estructura y funcionamiento físico

El transistor MOS es un dispositivo semiconductor consistente en una estructura de cuatro terminales compuesta por distintos materiales dispuestos según una configuración geométrica determinada. La Figura 1-1 muestra una representación esquemática del MOSFET.

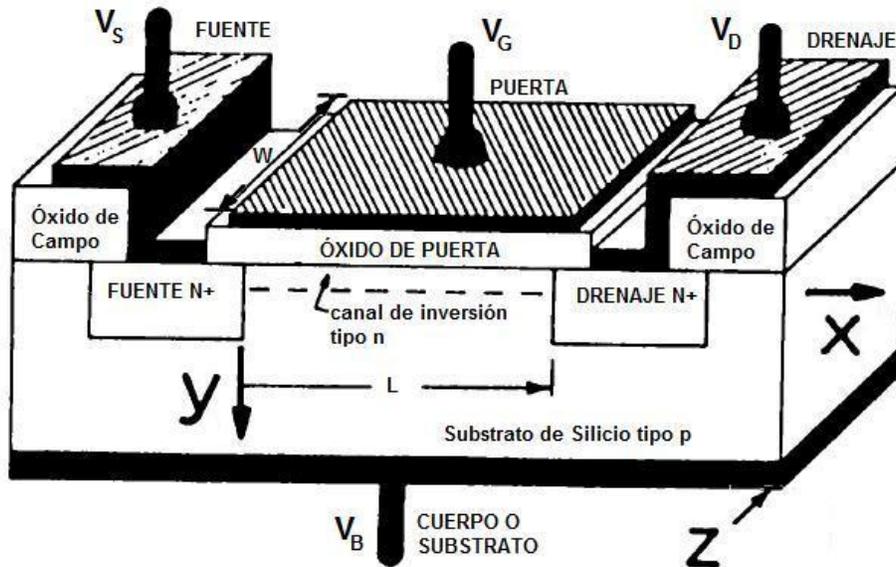


Figura 1-5: Esquema del transistor MOS tipo N (NMOS)

El MOSFET opera a través del principio de modulación de conductividad sobre una capa semiconductor fina, denominada canal, usando un campo eléctrico controlador para cumplir funciones de amplificación o switching entre dos de sus cuatro terminales (entrada, salida y comunes) adheridos a la capa de semiconductor mencionada.

El dispositivo se construye sobre un sustrato (también denominado cuerpo) de material semiconductor, el cual va conectado a un terminal metálico, y sobre cuya superficie se generará una región de conductividad variable con la polarización denominado canal. El tamaño del canal define la característica de conducción del dispositivo. Como se muestra en la Figura 1-2, las dimensiones del canal se denotan con las letras L (*length* o largo) y W (*width* o ancho).

La modulación de la conductividad del canal se logra mediante la acumulación de portadores de signo positivo o huecos, caso en el cual se denomina canal P, o de signo negativo o electrones, caso en el cual se denomina canal N. En la mayoría de los casos, el sustrato es construido de silicio dopado con átomos donores (en el caso de dispositivos de canal n ó NMOS) o aceptores (en el caso de dispositivos de canal p ó PMOS). Una capa de material dieléctrico (en general, dióxido de silicio) se dispone sobre la estructura. La interfaz entre el cuerpo y el óxido se denomina superficie. Se genera un electrodo de baja resistividad sobre el óxido, que se conecta al terminal

de puerta, y que se construye comúnmente de silicio poli-cristalino altamente dopado. Dos regiones fuertemente dopadas con átomos de igual signo que el canal se implantan a los costados del dispositivo, y corresponden a las regiones de fuente y drenaje (fuente y drenaje en Figura 1-2, respectivamente), que van conectadas a terminales metálicos. El dopado fuerte produce baja resistividad y alta disponibilidad de portadores para la conducción.

El voltaje de encendido del MOSFET es aplicado al electrodo conductor de puerta para crear un campo eléctrico en la dirección perpendicular a la capa semiconductor del canal, de manera de modular la conductancia laminar de la capa en el plano horizontal. Este voltaje de puerta aplicado modula la corriente que pasa por el canal entre los contactos de drenaje y fuente.

Los experimentos de caracterización electrónica y la teoría matemática del transistor MOS comenzaron en 1959, cuando fue posible fabricar óxidos de silicio sobre semiconductores de silicio por Atalla, Tannenbaum y Scheiber en los Laboratorios de Bell Telephone. Desde los primeros intentos por modelar el transistor MOS, se ha supuesto que su funcionamiento se puede describir en dos dimensiones, correspondientes a las de la cara del corte que se muestra en la Figura 1-2. Para hacer el problema en dos dimensiones manejable, se han empleado derivaciones basadas en la física de los dispositivos, que permiten separarlo en dos problemas acoplados de una dimensión. El mayor esfuerzo del desarrollo avanzado de modelos compactos para MOSFET ha sido encontrar algoritmos numéricos y formulaciones analíticas linealizadas de este acoplamiento de manera de proveer una extracción rápida y precisa de los parámetros de los modelos. Estos parámetros son posteriormente usados en para computar las características de un amplio rango de diseños de transistores para los miles a millones de transistores que contienen hoy en día los circuitos integrados, usando herramientas de simulación de circuitos como SPICE para predecir el desempeño de éstos.

Funcionamiento electrónico: modos de operación

La aplicación de supuestos que simplifican el entendimiento y la formulación matemática de la operación física del MOSFET permite derivar un modelo de canal largo (es decir, que no incluye el efecto de los fenómenos que se hacen presentes en geometrías bajo 100 nanómetros). Este modelo está definido por regiones de operación según los diversos modos en que puede operar el dispositivo.

La operación básica del MOSFET descrita por el modelo de canal largo asume que el fenómeno de conducción está confinado en el canal bajo el óxido (ver Figura 1-6); este supuesto se basa en el hecho que la dimensión del canal es lo suficientemente extensa como para desprestigiar los efectos de borde. Además, para el modelo básico se asume que toda la corriente que entra por el drenaje sale por la fuente, con corrientes nulas por los terminales de puerta y cuerpo. En el caso de la corriente de puerta, el óxido de puerta impide el paso de corriente; por otra parte, la región de empobrecimiento bajo la capa de inversión aísla de manera análoga el terminal de cuerpo. Estas corrientes pueden llegar a tener una magnitud considerable para dispositivos de canal pequeño.

El modelo de canal largo se basa en el concepto de voltaje de umbral (V_t). En la versión más simple de esta formulación, utilizada para cálculo manual, la corriente de canal debido al mecanismo de desplazamiento producido por el voltaje no nulo entre drenaje y fuente (V_{DS}) es no nula sólo cuando el voltaje de puerta referenciado al voltaje de fuente (i.e. V_{GS}) es mayor que V_t . En caso contrario, la corriente es nula. En modelos más avanzados, la corriente sub-umbral (cuando $V_{GS} < V_t$) se modela con un valor no nulo considerando la corriente generada por mecanismo de difusión, de manera semejante a como ocurre en los dispositivos BJT.

El supuesto que el voltaje a lo largo del canal varía de forma gradual en comparación con el voltaje vertical permite separar el problema de modelación en dos dimensiones en dos problemas unidimensionales. En primer lugar, se tiene el problema en dirección vertical, que relaciona la carga en el canal con el voltaje de puerta (este problema se denomina “capacitor MOS”). En segundo lugar, se tiene el problema en dirección horizontal, que relaciona la corriente en el canal y el voltaje que cae a través de éste.

La corriente de desplazamiento en cualquier punto del canal generada por el efecto del voltaje V_{DS} sobre los portadores (electrones en el caso de la Figura 1-6) presentes en la superficie debido al voltaje $V_{GS} > V_t$ se puede expresar como se muestra en la ecuación (1.1).

$$i = \frac{dq}{dt} = \frac{dq}{dx} \cdot \frac{dx}{dt} \quad (1.1)$$

En un diferencial dx del canal, el diferencial de carga dq existente toma la forma de la ecuación (1.2).

$$dq = -C_{ox} (W \cdot dx) [V_{GS} - V(x) - V_{th}] \quad (1.2)$$

En la expresión anterior, C_{ox} corresponde a la capacitancia por unidad de área de la puerta; W corresponde al ancho del canal; V_{GS} corresponde al voltaje entre puerta y fuente; V_{th} es el voltaje de umbral; finalmente, $V(x)$ es el voltaje en un punto X del canal.

Por otra parte, la velocidad de los portadores se relaciona con la magnitud del campo a través de la expresión (1.3).

$$\bar{v} = \frac{dx}{dt} = -\mu_n \cdot E(x) \quad (1.3)$$

La constante de movilidad μ_n establece una relación de proporcionalidad entre la velocidad y el campo eléctrico en un punto X del canal. Para dispositivos de canal corto, la relación de proporcionalidad ya no se cumple, debido al aumento excesivo del campo eléctrico; este fenómeno se conoce como saturación de velocidad.

Es posible relacionar el campo eléctrico con el voltaje en el canal a través de la ecuación de Poisson, que aplicada al caso de estudio toma la forma mostrada en la expresión (1.4).

$$E(x) = -\frac{dv(x)}{dx} \quad (1.4)$$

Por la definición del problema, la corriente de drenaje a fuente tiene sentido contrario al considerado. Con esto, y según las simplificaciones mencionadas, la corriente buscada se puede expresar como se muestra en la ecuación (1.5).

$$i_D = -i = \mu_n C_{OX} W \left[v_{GS} - v(x) - V_{th} \right] \frac{dv(x)}{dx} \quad (1.5)$$

Integrando la corriente entre $X = 0$ y $X = L$, y el lado derecho de la expresión entre $v(0) = 0$ y $v(L) = v_{DS}$, se obtiene la expresión (1.6). Se debe notar que, dados los supuestos, la corriente de canal es constante a lo largo de éste, y por lo tanto la variable puede salir de la integral.

$$i_D = \mu_n C_{OX} \frac{W}{L} \left[(v_{GS} - V_{th}) v_{DS} - \frac{1}{2} v_{DS}^2 \right] \quad (1.6)$$

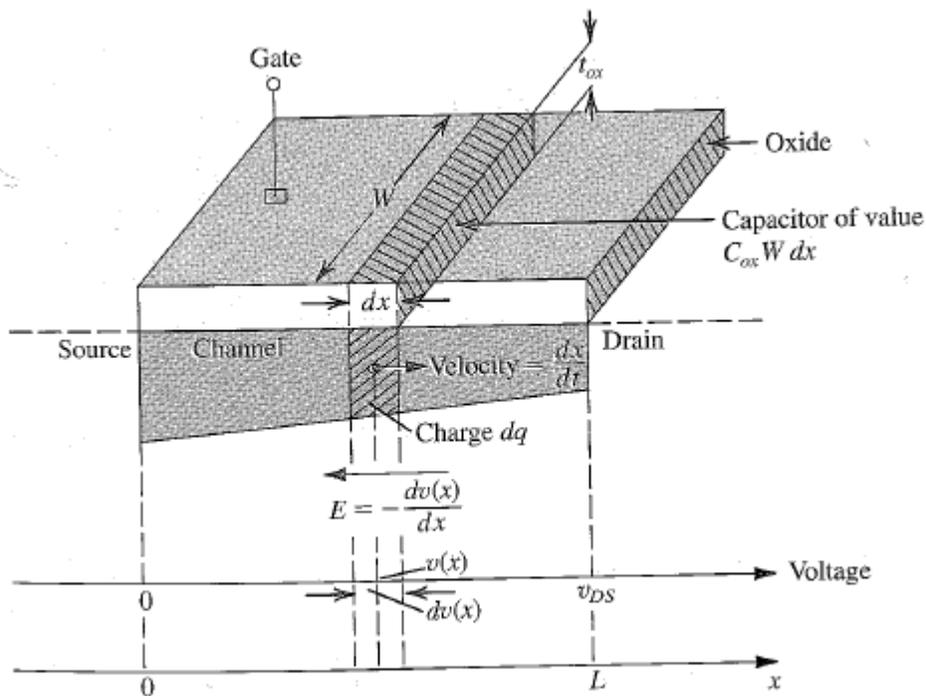


Figura 1-6: Esquema del funcionamiento del canal del transistor NMOS

Desde el punto de vista circuital, el dispositivo se considera un elemento de cuatro terminales: Puerta (Gate, anotado con la letra G), Drenaje (Drain, anotado con la letra D), Fuente (Source, anotado con la letra S) y Cuerpo (Body, anotado con la letra B). Sin embargo, en una gran cantidad de aplicaciones la fuente está conectada al cuerpo, y por lo tanto el dispositivo se trata como un elemento de tres terminales. Los símbolos circuital y las convenciones de signo de corrientes y

voltajes en los transistores NMOS y PMOS se muestran en la Figura 1-2. Debe notarse que las variables eléctricas son siempre positivas definidas de esta forma.

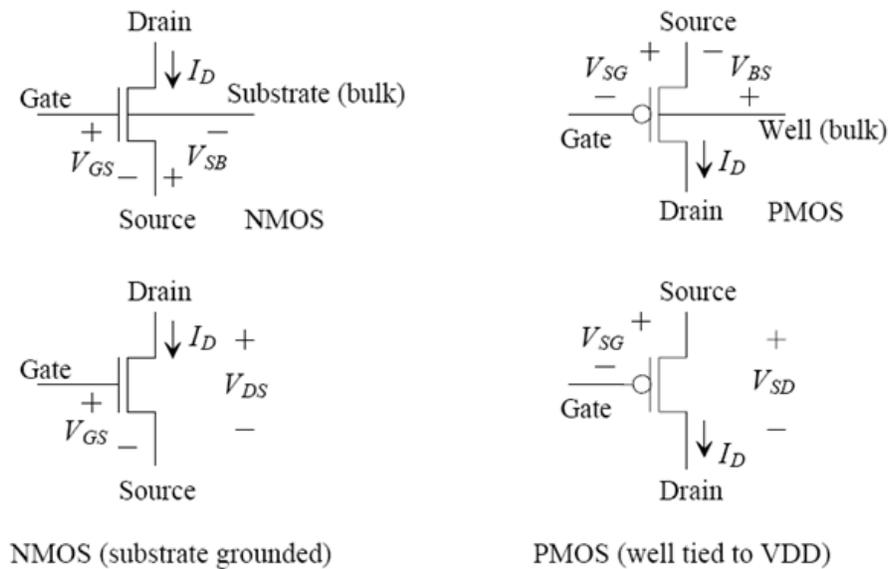


Figura 1-7: Esquema circuital de los transistores NMOS y PMOS

Para estudiar las curvas características que describen gráficamente el funcionamiento de los circuitos en los transistores MOS, se ilustrará el caso del dispositivo de canal N (el PMOS actúa de manera complementaria).

En primer lugar, si se considera un voltaje constante entre drenaje y fuente y un voltaje nulo entre cuerpo y fuente, y se mide la corriente ante un aumento progresivo del voltaje entre puerta y fuente, se obtiene la relación $I_D V_{GS}$, también denominada curva de encendido. La Figura 1-3 muestra el esquema circuital de la prueba y el resultado obtenido (curva del lado izquierdo).

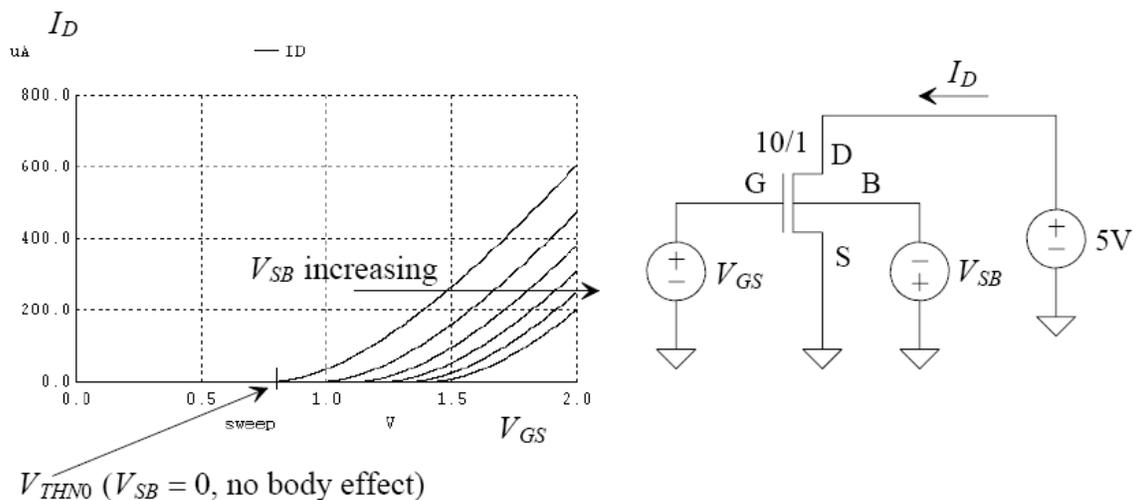


Figura 1-8: Medición de curva característica $I_D V_{GS}$

Una analogía que permite entender de manera intuitiva lo que ocurre es la siguiente: el MOSFET corresponde a una llave de agua en una casa, cuyo nivel de encendido está dado por el voltaje V_{GS} . El voltaje V_{DS} corresponde a la presión constante proporcionada por la empresa proveedora de agua. Como se muestra en la Figura 1-3, existe un valor de V_{GS} a partir del cual la corriente comienza a circular desde el drenaje hacia la fuente, denominado voltaje de umbral (V_{THN0}); en la analogía, corresponde al punto de activación de la llave, que permite la salida de agua.

La ecuación que describe esta característica (para $V_{GS} > V_{th}$) se muestra en (1.7).

$$I_D = \frac{K_P}{2} \cdot \frac{W}{L} \cdot (V_{DS} - V_{th})^2 \quad (1.7)$$

Ahora, si se fija un voltaje $V_{GS} > V_{THN0}$, al medir la corriente de canal I_D para una variación gradual del voltaje entre drenaje y fuente, se obtiene la característica $I_D V_{DS}$. En la Figura 1-4 se ilustra el esquema circuital y la forma típica de la característica.

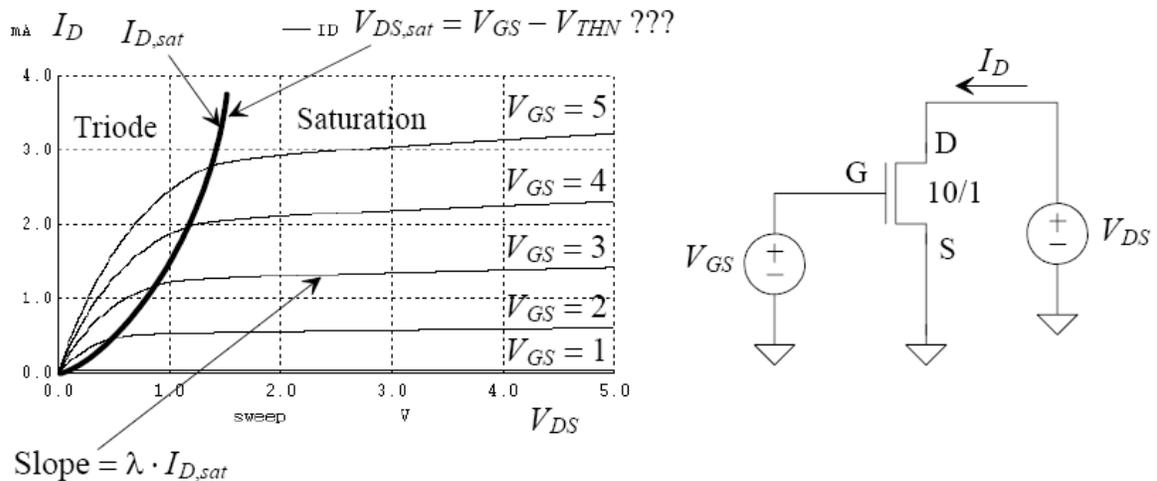


Figura 1-9: Medición de curva característica IDVDS

Siguiendo con la analogía, en este caso se tiene la llave de agua encendida ($V_{GS} > V_{THN0}$), pero la presión entregada por la empresa proveedora va de cero a un cierto valor máximo. A medida que la presión (V_{DS}) aumenta, también lo hace la cantidad de agua que fluye (I_D); sin embargo, hay un punto en el cual la compañía proporciona toda la presión necesaria para la cantidad que estoy solicitando (dada por el diámetro de la cañería), lo cual produce un enclavamiento del flujo de agua, como se muestra en el gráfico de la Figura 1-4 para I_D . Esta zona se denomina región de saturación (debido a que la corriente I_D se satura a partir de cierto punto); antes del enclavamiento, la región se denomina lineal o de triodo, pues el aumento de I_D se relaciona linealmente con el incremento de V_{DS} .

Un caso trivial ocurre cuando se varía V_{DS} cuando $V_{GS} < V_{THN0}$; en tal caso, I_D es cero en todo el intervalo de variación.

Lo anterior es válido también para PMOS, considerando que los signos se invierten (e.g. utilizamos $V_{SG} = -V_{GS} > V_{THPO}$ para activar el dispositivo). Como ejemplo, la figura 1-5 muestra la característica $I_D V_{DS}$ para el transistor PMOS.

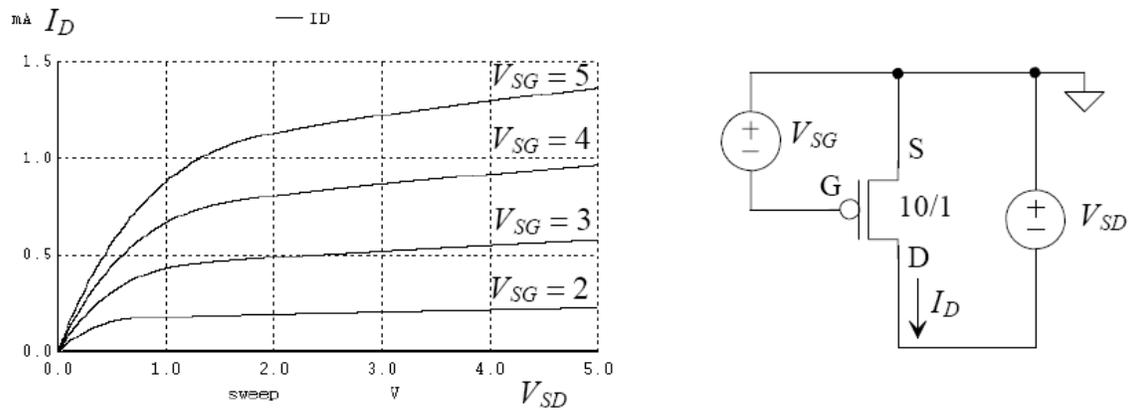


Figura 1-10: Medición de curva característica IDVDS para PMOS

Se debe notar que: $V_G = V_{DD} - V_{SG}$. Otro detalle importante es que el cuerpo está unido a la fuente, que está conectada a V_{DD} (para asegurarse que el diodo desde N-well de cuerpo a P+ de fuente esté en inversa).

Referencias

1. Baker, R. J. 2004 CMOS Circuit Design, Layout, and Simulation, Segunda Edición. Wiley-IEEE Press.
 - a. Capítulo 1 para Introducción
 - b. Capítulo 4 para Proceso de Fabricación Estándar
 - c. Capítulo 6 para Curvas Características
2. A History of MOS Compact Modeling, Sah.
3. The MOS Transistor, Tsividis.
4. Microelectronic Circuits, Sedra & Smith.