

T1 SEMICONDUCTORES, UNIÓN PN Y TRANSISTORES MOS

T1.1. Semiconductores

T1.2. Unión PN

T1.3. Transistor MOS

La electrónica se dedica a manejar información y para ello utiliza dispositivos que aprovechan las propiedades de los electrones. Los primeros dispositivos electrónicos fueron las válvulas, derivadas del triodo de Lee de Forest; pero la electrónica actual no es una electrónica de válvulas sino una electrónica de transistores.

El transistor es el punto de partida y, a la vez, el impulsor de un continuado proceso de miniaturización de la electrónica que, a través de los circuitos integrados y de su complejidad cada vez mayor, llega hasta nuestros días.

Los transistores provienen de la utilización eficiente de los semiconductores y, en particular, de estructuras conformadas por zonas semiconductoras de distinto signo: P y N.

Este capítulo presenta en forma conceptual y muy simplificada los fundamentos del comportamiento de los semiconductores, sus dos tipos P y N y la unión entre ambos, hasta llegar a los transistores MOS como interruptores digitales.

Un semiconductor conduce (más bien poco) gracias a los pares electrón-hueco que se forman cuando, por efecto de la temperatura, un electrón deja libre un enlace de valencia («hueco»); pero el número de pares electrón-hueco a temperatura ambiente es insignificante y la resistencia de un semiconductor puro es muy alta.

Ahora bien, la adición de impurezas formadas por átomos de fósforo (con electrones de más, tipo N) o de boro (deficitario de electrones y, por ello, generador de huecos, tipo P) permite aumentar drásticamente la disponibilidad de portadores y lleva al semiconductor a conducir en forma «intermedia» (semi-conduce).

Por su parte, la unión de un semiconductor P y de otro N presenta propiedades no lineales, configurando un discriminador de polaridad: si le llega una señal con parte positiva y parte negativa «deja pasar» solamente una de las polaridades, según el sentido en que se encuentre conectada. Además, la unión PN en polaridad inversa constituye un excelente aislante (para separar zonas de difusión del sustrato que las contiene).

El transistor NMOS aprovecha el efecto capacitivo de una placa (Puerta) para generar un canal conductor entre dos zonas de difusión de tipo N (Fuente y Drenaje, que ofician simplemente como contactos). De esta forma configura un interruptor que conduce cuando la tensión de puerta es positiva y no lo hace si dicha tensión es nula; las características de este interruptor (cuasi-ideal) lo hacen un magnífico candidato para el álgebra de conmutadores que da origen a la electrónica digital.

T.1. Semiconductores

Los metales son materiales conductores de la electricidad debido a que en su estructura cristalina disponen de electrones libres que se mueven por efecto del campo eléctrico cuando se aplica una diferencia de potencial (una tensión eléctrica). La diferencia de potencial eléctrico acelera a los electrones en la dirección del campo eléctrico (aunque en sentido contrario al mismo, ya que su carga eléctrica es negativa); pero en este movimiento de avance los electrones chocan con los átomos de la red cristalina que conforma el metal y la pérdida de energía en tales choques determina un movimiento no acelerado con una velocidad promedio proporcional a la intensidad del campo.

Como materiales conductores que son, la resistencia eléctrica de los metales es muy baja; un hilo de cobre de 1 metro de longitud y 1 mm² de sección, a temperatura ambiente, presenta una resistencia de unos 20 mΩ.

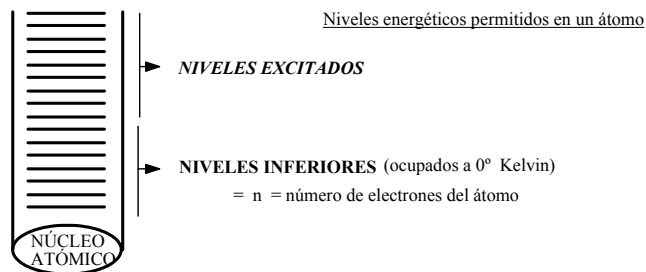
Al aumentar la temperatura, aumenta también el movimiento desordenado de los electrones debido a su agitación térmica y, con ello, aumentan los choques con los núcleos de la red cristalina y la mayor pérdida de energía determina una disminución de la velocidad promedio. Por ello, la resistencia de los metales aumenta al aumentar la temperatura; lo cual puede comprobarse fácilmente midiendo directamente con el polímetro la resistencia de una lámpara apagada y comparándola con el valor que se obtiene por aplicación de la ley de Ohm cuando la lámpara se encuentra encendida.

Existen en la naturaleza unos materiales cuya conductividad es intermedia entre la de los conductores metálicos y la de los aislantes y, además, su resistencia no aumenta con la temperatura sino que disminuye. Su comportamiento no puede ser explicado con el modelo de electrones libres con el que se describe el comportamiento de los conductores, sino que es necesario acudir a los principios cuánticos que rigen la configuración interna de los átomos.

Veamos algunas ideas elementales de la mecánica cuántica, es decir, de la organización de los electrones en los átomos:

Bohr postuló que los electrones de un átomo no pueden situarse en cualquier órbita, o sea, no pueden encontrarse en cualquier nivel energético respecto al núcleo, sino que existe un conjunto discreto de niveles permitidos. Pauli añadió que en un mismo nivel energético no puede haber más que un electrón.

A 0° de temperatura absoluta los electrones se sitúan en los niveles energéticos más bajos, pero a temperaturas superiores, $T > 0^\circ \text{K}$, la energía térmica correspondiente a la propia temperatura les permite ocupar niveles superiores: niveles excitados.

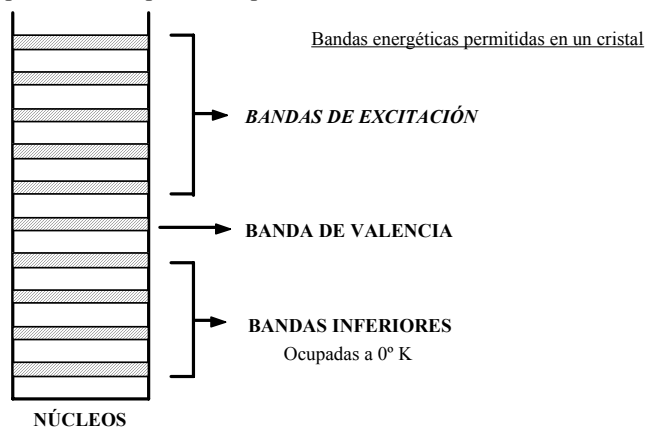


La probabilidad de que un electrón adquiriera una energía de excitación E_i viene dada por $p(E_i) = \alpha \cdot e^{-E_i/KT}$, donde K es la constante de Boltzmann y T la temperatura absoluta: para $T = 0^\circ K$ $p(E_i) = 0$; para $T > 0^\circ K$ $p(E_i) > 0$.

La probabilidad de ocupar un nivel excitado aumenta con la temperatura y es tanto mayor cuanto menor es la energía de excitación E_i .

El principio de exclusión de Pauli es válido también para un sistema molecular y para un sistema cristalino: al unirse dos átomos idénticos, desplazan sus niveles energéticos permitidos para que no coincidan dos electrones con la misma energía; este desplazamiento es, claro está, muy pequeño.

En un cristal de un elemento se encuentran multitud de átomos idénticos ligados por la red cristalina, los cuales dan lugar a un conjunto de bandas energéticas formadas por los niveles energéticos de los múltiples átomos; tales niveles eran coincidentes en los átomos aislados pero, al unirse, se han desplazado mínimamente y dan lugar a una banda de energía. Cada banda contiene multitud de niveles energéticos muy próximos, cada uno de los cuales puede ser ocupado sólo por un electrón.



En un sistema cristalino existen bandas energéticas permitidas que pueden ser ocupadas por un número discreto de electrones y que se encuentran separadas por intervalos de energía prohibidos.

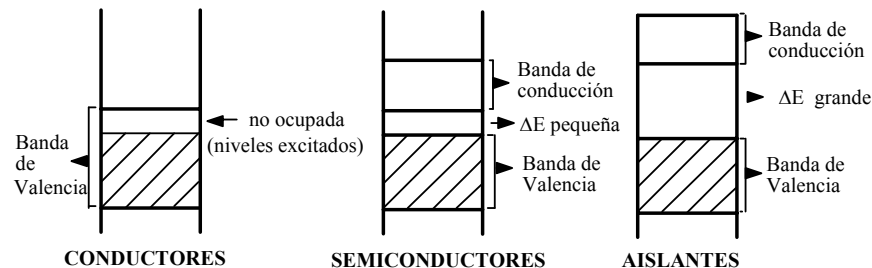
A 0° Kelvin se completan las bandas de energía inferior. La última banda energética que resulta ocupada se denomina banda de valencia; los electrones de esta banda, electrones de valencia, son los que pasarán a niveles excitados ya que los tienen más próximos que el resto de los electrones.

Para contribuir a la conducción eléctrica los electrones han de pasar a niveles excitados, pues han de adquirir la energía cinética correspondiente a su movimiento; de modo que, al aplicar un campo eléctrico, solamente contribuirán a la conducción aquellos electrones que puedan dar el salto a niveles energéticos excitados.

Aquellos materiales que tienen la banda de valencia incompleta, o sea no toda ella ocupada por electrones, son conductores; presentan niveles energéticos excitados lo suficientemente próximos, dentro de la misma banda de valencia, para que la energía correspondiente al campo eléctrico haga pasar electrones a ellos, incluso a temperatura de 0° Kelvin: en los conductores la banda de valencia coincide con la de conducción.

Los materiales cuya banda de valencia está completa no pueden conducir a $0^\circ K$, pues los campos eléctricos normales no son capaces de comunicar la energía necesaria para que los electrones de valencia salten a la banda superior, que será en este caso la banda de conducción. Dentro de estos materiales existe una diferencia importante:

- aquellos en los que el intervalo energético entre las bandas de valencia y de conducción es grande y que, por tanto, a temperatura ambiente apenas tendrán electrones en la banda de conducción y, en consecuencia, no conducirán
- y aquellos otros en que este intervalo es pequeño, de forma que a temperatura ambiente tienen un número apreciable de electrones excitados en la banda de conducción y, por tanto, conducen: éstos son los semiconductores.



Ésta es la diferencia entre conductores, aislantes y semiconductores. Los conductores tienen niveles excitados libres en la propia banda de valencia y los electrones pueden pasar a ellos por acción de un campo eléctrico, incluso a 0° absolutos; los semiconductores tienen dichos niveles de excitación en otra banda que se encuentra energéticamente próxima y pueden ocuparlos cuando la temperatura es mayor de $0^\circ K$; mientras que los aislantes tienen la banda de niveles excitados muy distante, de forma que el número de electrones que pasan a ella es despreciable.

En los semiconductores, al aumentar la temperatura, aumenta el número de electrones que pasan de la banda de valencia a la de conducción y contribuyen a la corriente eléctrica; por ello, su resistencia disminuye al aumentar la temperatura.

En el contexto del modelo cuántico de los semiconductores resulta que al pasar un electrón a la banda de conducción deja un hueco en la de valencia, es decir, un nivel energético que puede ser ocupado por otro electrón inferior, adquiriendo una energía de excitación. Este segundo electrón también puede contribuir a la conducción; de forma que contribuyen a la conducción tanto los electrones que han pasado a la banda de conducción como los niveles energéticos «huecos» que han quedado libres en la banda de valencia.

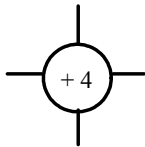
Los pares electrón-hueco se forman por efecto de la energía térmica: a temperatura superior a 0°K un electrón adquiere energía suficiente para pasar a un nivel excitado de la banda de conducción y deja un hueco en la banda de valencia. Un «hueco» contribuye a la corriente eléctrica al ser ocupados por un electrón de un nivel energético inferior y, como el hueco supone «ausencia de electrón», actúa como si fuera una carga positiva.

A la vez sucede que, en ocasiones, un hueco es cancelado por un electrón libre que pasa a formar parte del enlace vacío propio del hueco, cediendo la correspondiente energía; este fenómeno se denomina recombinación y supone pérdida de portadores.

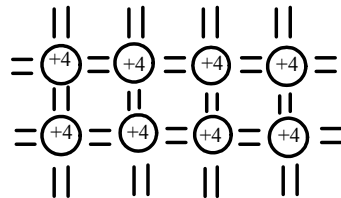
El número de portadores eléctricos disponibles se debe al balance entre la formación de pares electrón-hueco y la recombinación de los mismos; es, pues, una situación de equilibrio dinámico.

Consideremos una imagen más gráfica de los semiconductores: por razones tecnológicas, principalmente por razones de consistencia mecánica y de disponibilidad de ellos, los semiconductores utilizados son cristales constituidos por átomos de valencia 4, cuatro electrones en la última capa, principalmente el germano y el silicio.

Representación esquemática de un átomo de valencia 4

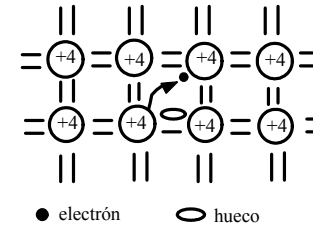


Representación esquemática y plana de un cristal de dichos átomos de valencia 4



La carga positiva indicada dentro del núcleo atómico (+4) queda compensada por los 4 electrones de valencia representados por las líneas periféricas; de forma que, como es obvio, el átomo con sus 4 electrones de valencia es eléctricamente neutro.

Cuando por efecto térmico, a $T > 0^\circ \text{K}$, un electrón adquiere energía suficiente para «saltar» de su enlace queda «libre», es decir, pasa a la banda de conducción, pero, asimismo, al ausentarse del enlace correspondiente deja un hueco (que representa una carga positiva).



Como resultado del paso de un electrón de la banda de valencia a la de conducción queda un hueco, un enlace vacante por ausencia de electrón, en dicha banda de valencia

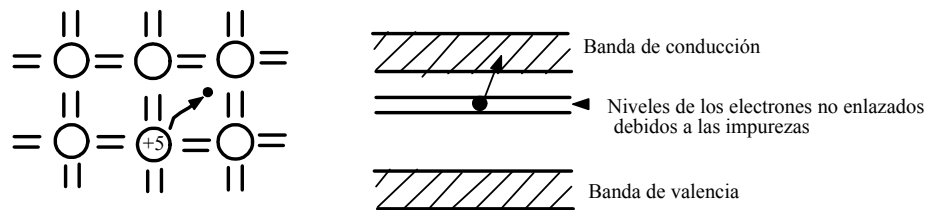
Otro electrón de un enlace próximo puede ocupar este hueco, dejando otro en su enlace anterior, y así sucesivamente, de forma que un hueco se mueve en sentido contrario a los electrones que pasan a ocuparlo.

Si este movimiento es debido a un campo eléctrico, el hueco se mueve como si tuviera carga positiva y el semiconductor contribuye a la conducción con dos tipos de portadores: los electrones en la banda de conducción que se mueven en sentido opuesto al del campo (cargas negativas) y los huecos en la banda de valencia que se mueven en el mismo sentido del campo (cargas positivas). Está claro que los huecos no tienen existencia propia, pero constituyen un modelo válido y simple para representar el movimiento de los electrones de valencia que pasan a ocuparlos.

En el caso del semiconductor más utilizado, el silicio, la concentración de átomos en la red cristalina es del orden de 10^{22} átomos/cm³ y, a temperatura ambiente (25° C), se produce una generación de pares electrón-hueco del orden de 10^{10} pares/cm³, es decir, solamente 1 átomo de cada billón (10^{12}) de átomos de silicio suelta un electrón y se queda con un hueco. Esta baja disponibilidad de portadores eléctricos hace que la conductividad de los semiconductores a temperatura ambiente sea muy baja; la resistencia de una línea de silicio de 1 metro de longitud y 1 mm² de sección, a temperatura ambiente, es del orden de 1000 MΩ.

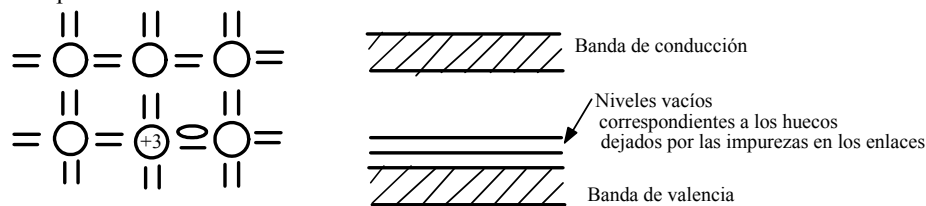
Pero la conductividad de los semiconductores puede ser considerablemente mayor si se aumenta la concentración de portadores eléctricos, añadiendo dentro de la red cristalina pequeñas cantidades de ciertas impurezas. La existencia de dos tipos de portadores (electrones y huecos) da lugar a dos tipos de semiconductores enriquecidos o «dopados»: semiconductores N, con predominio de electrones, y semiconductores P, con predominio de huecos; según que la impureza tengan exceso (fósforo) o defecto (boro) de electrones de valencia resulta un semiconductor tipo N o tipo P.

Si en una red de átomos de valencia 4 se incluyen, en forma muy diseminada, átomos de valencia 5, un electrón de cada uno de estos átomos quedará sin formar parte de los enlaces, por lo cual tendrá mucha mayor facilidad para «saltar» a la banda de conducción. Estos electrones no dejan huecos en la banda de valencia (no dejan enlaces libres), de forma que predominarán los electrones en la banda de conducción y éstos serán los portadores de corriente mayoritarios: semiconductor tipo N.



Si, por el contrario, los átomos de las impurezas son de valencia 3, quedará un hueco en los enlaces de cada átomo de impureza; estos huecos serán ocupados por electrones de la banda de valencia, dejando los correspondientes huecos en ésta pero sin generar electrones libres en la banda de conducción.

En este caso predominan los huecos como portadores de corriente: semiconductor tipo P.



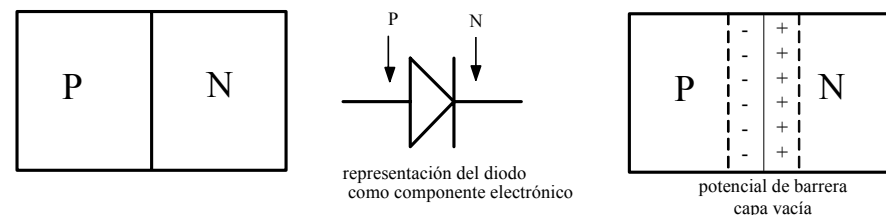
Un dopado normal suele ser del orden de 10^{20} átomos/cm², o sea, 1 átomo de impureza (fósforo o boro) por cada 100 átomos de silicio; ello supone una disponibilidad de portadores eléctricos de 10^{20} /cm² pues todos los átomos de impureza contribuyen con su portador (electrón / hueco) dado que el salto energético para ello es pequeño.

Esta concentración de portadores resulta ser 10.000 millones de veces superior a los portadores disponibles en el silicio puro (10^{10}) y la conductividad del silicio P o N resulta 100 millones de veces mayor que la del silicio sin impurezas. Téngase en cuenta que la conductividad no depende solamente de la concentración de portadores disponibles sino también de la movilidad de tales portadores, que es más baja en los semiconductores fuertemente dopados porque inserta un amplio número de núcleos con carga en la red cristalina; éstos atraen a los portadores mayoritarios y los frenan y, además, el número de choques contra la red cristalina aumenta con el dopado.

La resistencia de una línea de silicio dopado de 1 metro de longitud y 1 mm² de sección, a temperatura ambiente, es del orden de 10 Ω (sólo unas 500 veces superior a la del cobre).

T.2. Unión PN

Un diodo está constituido por una unión P-N, es decir, por la unión de dos semiconductores, uno de tipo P y otro de tipo N, con una superficie común.



Del lado P del diodo predomina la concentración de huecos y, en cambio, en el lado N predomina la concentración de electrones. La diferencia de concentración de portadores entre ambos lados origina sendas corrientes de difusión, de huecos hacia el lado N y de electrones hacia el lado P, que tienden a homogeneizar la concentración de portadores en todo el material.

Los electrones del lado N cercanos a la unión pasan al lado P, lo cual da origen a una carga eléctrica positiva en la zona N próxima a la unión (carga positiva que se produce por la pérdida de electrones, más el aporte de huecos desde P). Asimismo, la zona P próxima a la unión se carga en negativo (por la pérdida de huecos enviados a la otra zona, más el aporte de electrones que le llegan desde N).

Entre ambos lados de la unión se crea una especie de condensador cargado y, consecuentemente, un campo eléctrico que va de N a P y se opone a la difusión de los portadores. Enseguida se llega a un equilibrio en el que la tensión en la unión, asociada al campo eléctrico producido y denominada potencial de barrera (ϕ), adquiere un cierto valor que anula la corriente de difusión.

El potencial de barrera es negativo en el lado P y positivo en el N; de forma que si se conecta una tensión exterior con el terminal negativo hacia P y el positivo hacia N, polarización inversa, dicha tensión se suma al potencial de barrera y ambas juntas impiden la conducción a través del diodo.

En cambio, si la tensión es positiva hacia P y negativa hacia N, polarización directa, dicha tensión exterior se opone al potencial de barrera, disminuyéndolo; en este caso, existirán portadores (huecos del lado P y electrones del lado N) capaces de saltar energéticamente la barrera y contribuir a la conducción en el sentido del campo exterior aplicado, de P a N.

En consecuencia, un diodo sólo puede conducir en el sentido de P a N, que es en el indicado por la flecha en su representación esquemática, y no puede conducir en sentido contrario.



La intensidad que atraviesa el diodo será tanto mayor cuanto menor sea la barrera de potencial que se opone a ella en la unión, es decir, cuanto mayor sea la tensión aplicada al diodo. Esta tensión exterior ha de ser siempre inferior al valor del potencial de barrera propio del diodo, pues de otro modo la barrera se anularía totalmente, la intensidad aumentaría indefinidamente y produciría la destrucción de la unión.

La probabilidad de que un electrón tenga energía suficiente para atravesar la barrera viene dada por:

$$p = \alpha e^{-E/KT}, \text{ siendo } E \text{ la energía de la barrera de potencial.}$$

En una unión PN aislada dicha energía es $E = e \cdot \phi$, ($e =$ carga del electrón) y la probabilidad de atravesar la barrera resulta despreciable, no existiendo corriente a través del diodo. Menos aún si aplicamos una polarización inversa, $E = e \cdot (\phi + V)$, en cuyo caso la probabilidad es aún menor, pues la energía de la barrera de potencial es mayor.

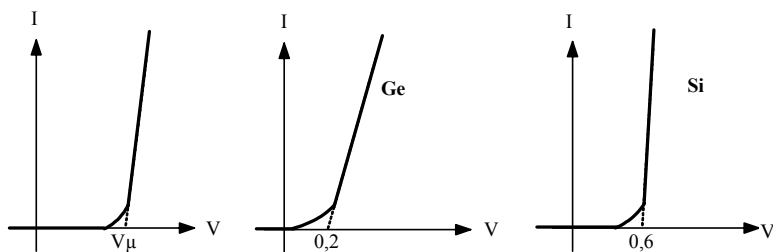
En cambio, en polarización directa la energía de la barrera es disminuida por la tensión exterior aplicada, $E = e \cdot (\phi - V)$, de forma que la probabilidad de atravesarla aumenta exponencialmente con dicha tensión V :

$$p = \alpha e^{-e \cdot (\phi - V)/KT} = \alpha e^{-e\phi/KT} \cdot e^{+eV/KT} = \alpha' e^{+eV/KT}$$

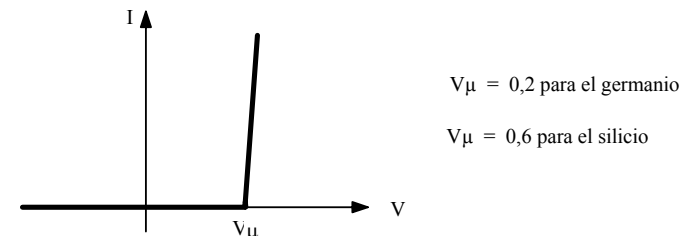
Éste será el tipo de dependencia de la corriente que atraviesa la unión PN con la tensión aplicada

$$I = \alpha'' e^{+eV/KT} = I_s \cdot e^{+eV/KT}$$

cuya representación da lugar a las gráficas de la figura siguiente, que coinciden con las medidas experimentales para diodos de germanio y de silicio.

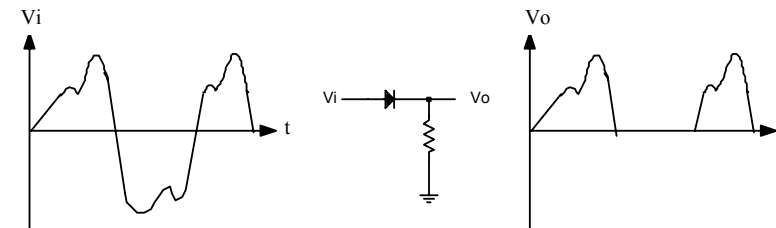


Estas gráficas se pueden linealizar con gran aproximación mediante dos rectas, una horizontal hasta V_μ y la otra prácticamente vertical.



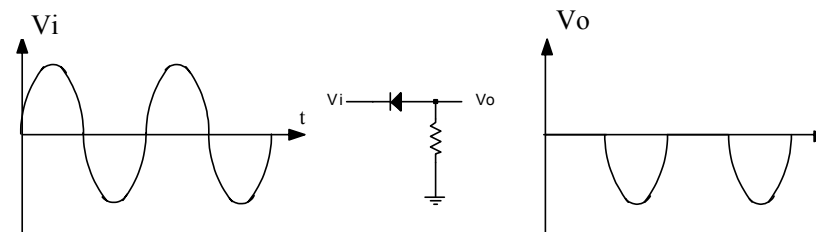
Es decir, un diodo no conduce mientras la tensión aplicada entre sus extremos sea inferior a V_μ y a partir de esta tensión conduce como si fuera prácticamente un cortocircuito; en realidad presenta una pequeña resistencia, una ligera inclinación de la recta, del orden de 10Ω para el germanio y de 2Ω para el silicio (dicha resistencia resulta francamente despreciable en la mayoría de los casos).

Configurando un circuito con un diodo de silicio y una resistencia de carga y aplicando a su entrada una señal con ambas polaridades, positiva y negativa:



pasa a la carga la parte de la onda que se encuentra por encima de los 0,6 voltios necesarios para la conducción del diodo.

Supongamos el diodo al revés y una onda senoidal en la entrada:



Por encontrarse el diodo al revés pasa a la carga la parte negativa de la onda, que es la que polariza al diodo en positivo; en concreto, el diodo deja pasar la parte de onda que se encuentra por debajo de -0,6 voltios.

Tal es la aplicación de los diodos en los circuitos electrónicos: dejar pasar las tensiones de una polaridad (bien positiva, bien negativa, según el sentido del diodo) y no las de la polaridad opuesta. Esta función se denomina «rectificar» una señal eléctrica; el diodo es un componente «rectificador».

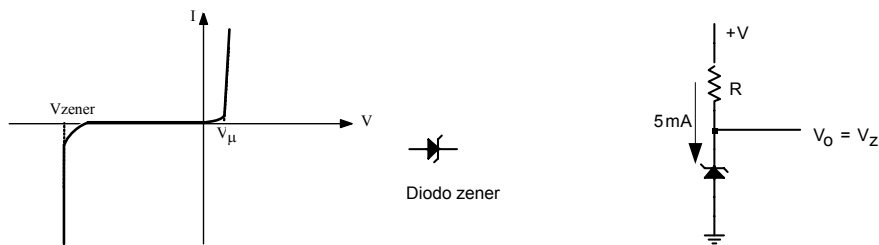
Realmente la operación de rectificar aparece cuando el diodo va seguido de un condensador que filtra la señal de salida y proporciona la envolvente de la onda, dando lugar a una «recta» (tensión continua) con un pequeño rizado; por ello, un nombre más apropiado para el diodo es el de «discriminador de polaridad».

Los diodos se fabrican difundiendo consecutivamente sobre un semiconductor impurezas tipo P por un extremo e impurezas tipo N por el opuesto; así se forman las dos zonas P y N sobre el mismo sustrato semiconductor. Según la intensidad máxima de corriente que vayan a conducir, se fabrican diodos de 0,1 A, 0,5 A, 2 A, 10 A, etc.

Una unión PN en polarización inversa (incluyendo el caso de $V_D = 0$) presenta una zona vacía de portadores, justo a ambos lados de la unión (en la propia barrera de potencial que la unión PN genera): capa de vaciamiento o capa vacía. La ausencia de portadores se debe a la difusión de los portadores propios hacia el otro lado de la unión y a la recombinación de los portadores que le llegan desde dicho otro lado (que se encuentran con una concentración amplia de portadores de signo opuesto que propicia la cancelación electrón-hueco).

La capa vacía de la unión constituye un excelente aislante que separa eléctricamente la zona semiconductor P de la zona semiconductor N. Este aislamiento propio de la unión PN es aprovechado en la configuración de circuitos integrados para «aislar» las regiones N que se fabrican dentro de un sustrato P y, viceversa, las regiones P fabricadas sobre un sustrato N.

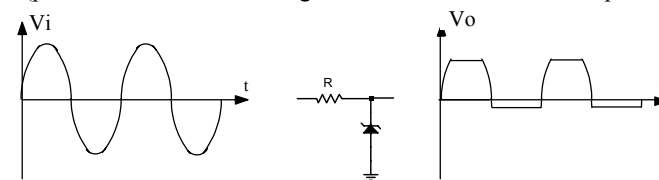
Ahora bien, cuando la tensión aplicada en polarización inversa aumenta, llega un momento en que el potencial creado en la unión, $\phi + V$, origina un campo eléctrico lo suficientemente fuerte como para romper enlaces atómicos, dejando libres huecos y electrones que contribuyen a la conducción y que no han de atravesar la barrera de potencial pues se generan en la propia barrera; este fenómeno se denomina efecto zener y los portadores así generados contribuyen a la conducción de forma que a dicha tensión, tensión zener, el diodo conduce en polarización inversa.



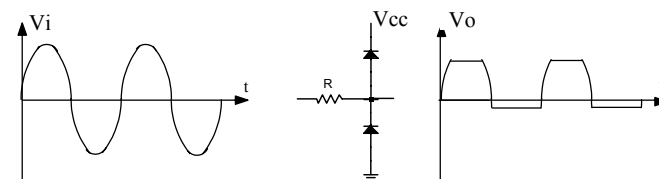
La tensión zener, en los diodos normales, es de varias decenas o de varios cientos de voltios y es otra característica a tener en cuenta junto con la intensidad máxima: la tensión zener de un diodo rectificador ha de ser superior al valor de la máxima tensión inversa que vaya a soportar, a fin de que no conduzca en polarización inversa. En tal sentido, un diodo rectificador viene caracterizado por dos parámetros: la intensidad máxima a conducir y la tensión máxima inversa a soportar (y así, hablaremos de un diodo de 1 A y 300 V y de otro de 10 A y 100 V, etc.)

Pero el efecto zener también es aprovechado en forma contraria en diodos que precisamente se polarizan en zona zener para utilizar dicha tensión como referencia (ver en la figura anterior el circuito de la derecha). Para ello se fabrican diodos con tensión zener de unos pocos voltios, denominados diodos zener: existe toda una gama de zener con valores de tensión que van desde los 2,7 V a los 24 V, muy útiles para configurar tensiones fijas de referencia, para lo cual han de polarizarse inversamente de manera que los atraviese una intensidad inversa (contraria a la flecha del diodo) entre 5 y 20 mA.

Un circuito resistencia-zener, tal como el representado en la figura siguiente, limita la tensión que recibe al intervalo $[-0,6 \text{ V} ; V_Z]$ ya que recorta las tensiones superiores a su tensión zener V_Z (para tensión de entrada positiva el diodo se encuentra en polarización inversa) y también recorta las tensiones negativas superiores a su tensión umbral como diodo $-0,6 \text{ V}$ (para tensión de entrada negativa conduce como diodo a partir de $0,6 \text{ V}$).

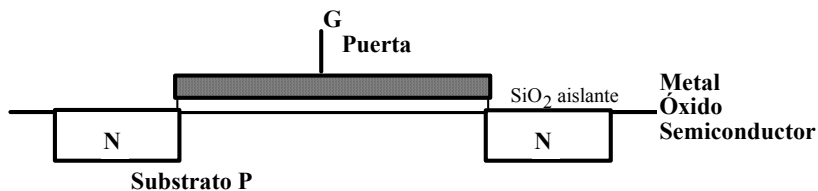


Un circuito limitador análogo puede configurarse con dos diodos en polarización inversa según la figura siguiente; en este caso los diodos recortarán las tensiones positivas superiores a $V_{CC}+0,6$ y las negativas inferiores a $-0,6 \text{ V}$, manteniendo la tensión dentro del intervalo $[-0,6 \text{ V} ; V_{CC}+0,6]$. Ambos circuitos son útiles para reducir al rango de tensiones digitales $[0 ; V_{CC}]$ pulsos de amplitud superior al mismo.



T.3. Transistor MOS

Un transistor NMOS es un interruptor controlado por un terminal de puerta **G**: la conducción se establece sobre un sustrato P entre dos difusiones N (fuente **S** y drenaje **D**) que actúan como contactos y la puerta configura un «condensador» intermedio cuya carga permite la conducción.



Si la tensión de entrada $V_i = V_G$ es suficientemente positiva, la puerta atrae a los electrones libres en el sustrato y forma un canal conductor entre fuente y drenaje: el transistor conduce. Esto sucede cuando la tensión $V_G > V_{TO}$, la tensión de puerta mayor que la tensión umbral de conducción.

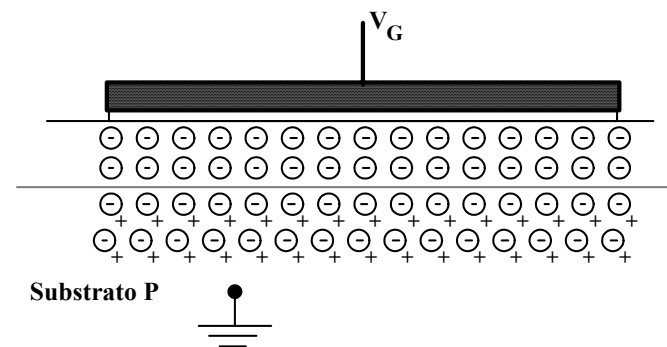
Si la tensión de puerta es claramente inferior al valor de la tensión umbral V_{TO} el canal no se forma y no hay paso de corriente entre las zonas de difusión (fuente y drenaje): el transistor no conduce.

En la zona de puerta (zona activa) la estructura que conforma el transistor está formada por tres capas: conductor – óxido – sustrato; en un principio tales capas fueron Metal – Óxido – Semiconductor (dando nombre al dispositivo: M - O - S), pero muy pronto el metal fue substituido por silicio policristalino fuertemente dopado: polisilicio – óxido – semiconductor. Se utiliza polisilicio en lugar de metal para reducir el potencial electroquímico que se genera en la interfase metal-óxido de silicio y que afecta fuertemente a la tensión umbral.

Analicemos en detalle la formación del canal y puesta en conducción de un transistor NMOS:

a) El sustrato de los transistores NMOS es de tipo P, dopado con átomos de Boro con sólo 3 electrones de valencia, los cuales al formar parte de la estructura cristalina del silicio presentan un hueco (un enlace no cubierto por electrón).

Cuando se aplica una tensión positiva a la puerta $V_G > 0$ respecto al sustrato (conectado a tensión de referencia 0 V) se forma una capa de vaciamiento, desprovista de portadores, debida a la repulsión electrostática que la tensión de puerta (+) ejerce sobre los huecos del sustrato. Al emigrar los huecos más allá de la capa de vaciamiento, los átomos de boro quedan cargados negativamente, pues la ausencia de hueco se debe a que un electrón lo ha cancelado (ha cubierto el enlace que se encontraba libre).



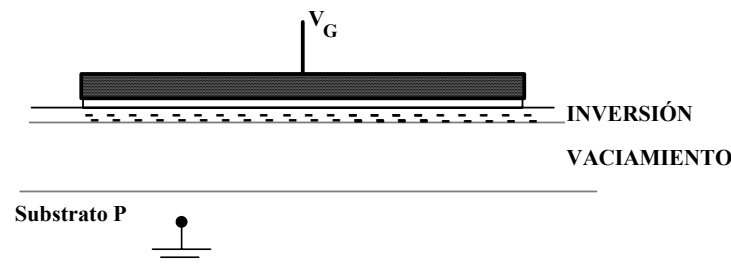
En la figura anterior, los círculos negativos representan a los átomos de impureza P (boro); todos ellos han perdido el hueco resultante de su déficit de electrones (han captado un electrón). En cambio, los signos + representan a los huecos: en la capa de vaciamiento no hay huecos porque han sido expulsados por repulsión electrostática provocada por la tensión positiva de la puerta; los átomos de boro inferiores, situados fuera de la capa vacía, mantienen su hueco y por ello su carga eléctrica es nula.

La densidad superficial de carga, es decir, carga por unidad de superficie Q será:

$$Q_v = q \cdot N_A \cdot ev$$

q = carga del electrón,
 N_A = dopado del sustrato = nº de huecos por cm^3 ,
 ev = profundidad de la capa de vaciamiento.

b) Cuando la tensión de puerta es suficientemente positiva (tensión umbral del transistor) se forma una capa superficial de electrones libres, arrastrados por el gradiente de tensión existente en la capa de vaciamiento; ésta es la capa de inversión: el semiconductor P cambia de signo por acumulación de electrones que dan lugar a una zona de tipo N. Dichos electrones libres proceden de los pares electrón hueco del propio silicio semiconductor: son, pues, portadores intrínsecos y su número es reducido.



Para formar una capa de inversión o canal conductor se necesita una tensión en la superficie del semiconductor Φ_B capaz de mantener una concentración de electrones en la banda de conducción equivalente al dopado del sustrato, es decir, el sustrato P se vuelve N en la zona superficial.

La carga total almacenada en el «condensador» puerta-substrato tiene dos componentes: carga correspondiente a la capa de vaciamiento + carga debida al canal o capa de inversión: $Q = Q_V + Q_I$, siendo Q carga por unidad de superficie. Por otra parte, dicha carga será proporcional a la diferencia de potencial existente entre puerta y substrato con un factor de proporcionalidad que expresa la capacidad de dicho condensador c_{ox} .

$$\left. \begin{aligned} Q &= Q_I + Q_V \\ Q &= c_{ox}(V_G - \Phi_B) \end{aligned} \right\} \begin{aligned} Q &= \text{carga por unidad de superficie} \\ c_{ox} &= \text{capacidad por unidad de superficie} \end{aligned}$$

$$Q_I = Q - Q_V = c_{ox}(V_G - \Phi_B) - Q_V = c_{ox}[V_G - (\Phi_B + Q_V / c_{ox})]$$

$$Q_I = C_{ox}(V_G - V_{TO}) \quad \text{donde} \quad V_{TO} = \Phi_B + Q_V / c_{ox}$$

V_{TO} es la tensión umbral de conducción del MOS: tensión de puerta necesaria para que se forme la capa de inversión.

Comentario: aunque este análisis es conceptualmente correcto, el valor concreto de V_{TO} no puede deducirse directamente del mismo ya que sobre V_{TO} inciden otros efectos de segundo orden relativamente complejos y, en la práctica, dicho valor lo fija el fabricante (mediante un proceso de implantación iónica en la superficie del substrato).

c) Si sobre el substrato en su zona superficial se induce, por algún medio, una tensión V' , dicha tensión actúa como referencia o nivel 0 de tensiones, de forma que la tensión efectiva sobre el condensador MOS será $V_G - V'$:

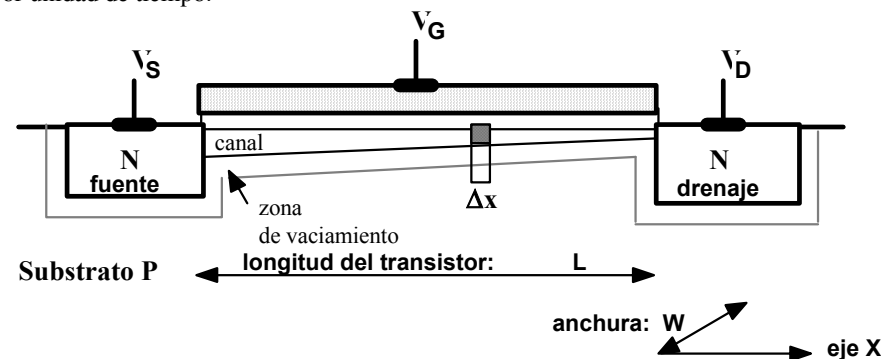
$$Q_I = c_{ox}(V_G - V' - V_{TO}).$$

Precisamente esto es lo que ocurre al polarizar la fuente y el drenaje del transistor MOS para que circule una corriente eléctrica a su través: las tensiones aplicadas a drenaje y fuente $V_D - V_S$ establecen una distribución de tensiones en la zona superficial del substrato, variando desde V_S (en el extremo de la fuente) hasta V_D (en el del drenaje).

Orientando el eje X en la dirección del canal (a lo largo del mismo, de fuente a drenaje), la concentración de portadores de carga en el canal será:

$$Q_I = c_{ox}(V_G - V_{TO} - V(x)) \quad \text{donde } V(x) \text{ varía entre } V_S \text{ y } V_D.$$

d) La diferencia de tensión entre fuente y drenaje $V_D - V_S$ da lugar a una intensidad de corriente I que corresponde a la carga que atraviesa una superficie perpendicular al eje X por unidad de tiempo.



La distribución de portadores libres en la capa de inversión variará a lo largo del canal (eje X): será mayor al lado de la fuente que en el del drenaje ya que $V_{DS} > 0$ y, por ello, $V_G - V_S > V_D - V_S$. En un elemento diferencial Δx la cantidad de carga que corresponde a los portadores libres, a todo lo ancho W del transistor será:

$$\Delta Q_I = [\text{densidad de carga}] \times \text{volumen} = Q_I \cdot W \cdot \Delta x = c_{ox} \cdot (V_G - V_{TO} - V(x)) \cdot W \cdot \Delta x$$

$$I = \frac{dQ_I}{dt} = \frac{\Delta Q_I}{\Delta x} \cdot v \quad \frac{\Delta Q_I}{\Delta x} \text{ es la carga de portadores libres por unidad de longitud}$$

v es la velocidad de los portadores que es proporcional al campo eléctrico: la constante de proporcionalidad es la movilidad de los portadores de carga μ

$$\left. \begin{aligned} \frac{\Delta Q_I}{\Delta x} &= c_{ox} \cdot W \cdot (V_G - V_{TO} - V(x)) \\ v &= \mu \cdot E = \mu \cdot \frac{dV}{dx} \end{aligned} \right\} I = \mu \cdot c_{ox} \cdot W \cdot (V_G - V_{TO} - V(x)) \cdot \frac{dV(x)}{dx}$$

Integrando la anterior ecuación diferencial a lo largo del canal resulta

$$\int_0^L I dx = \int_{V_S}^{V_D} \mu \cdot c_{ox} \cdot W \cdot (V_G - V_{TO} - V(x)) dv$$

$$I = \mu \cdot c_{ox} \cdot \frac{W}{L} \cdot \int_{V_S}^{V_D} (V_G - V_{TO} - V(x)) dv = \alpha \cdot A$$

donde $A = \int_{V_S}^{V_D} (V_G - V_{TO} - V(x)) dv$ expresa el efecto de las tensiones aplicadas sobre el transistor y $\alpha = \mu \cdot c_{ox} \cdot W/L$ incluye parámetros tecnológicos y geométricos:

K_p = coeficiente de transconductancia = $\mu \cdot c_{ox}$ (parámetros tecnológicos)

ff = factor de forma = W/L , cociente entre anchura W y longitud L del transistor.

Ambos influyen en proporcionalidad directa sobre la intensidad que conduce el transistor.

e) Para V_{DS} muy pequeño: $V_{DS} \ll V_{GS}$ la expresión $(V_G - V_{TO} - V(x))$ es prácticamente constante a lo largo de todo el canal: $(V_G - V_{TO} - V_S) = V_{GS} - V_{TO}$; la tensión aplicada es la misma a lo largo de todo el canal, de forma que el canal resulta plano, con una distribución de carga uniforme.

En tal caso, la integral es inmediata y la intensidad resultante será:

$$I_D = K_p \cdot \frac{W}{L} \cdot (V_{GS} - V_{TO}) \cdot V_{DS}$$

expresión que muestra una proporcionalidad directa entre la intensidad que pasa por el transistor y la tensión aplicada sobre el mismo, lo cual indica que el transistor se comporta como una simple resistencia cuyo valor será:

$$R_{eq} = \frac{I_D}{V_{DS}} = \frac{1}{K_p \frac{W}{L} (V_{GS} - V_{TO})}$$

valor que disminuye al aumentar la tensión de puerta V_{GS} . Esta región de funcionamiento del transistor se denomina zona lineal o zona óhmica.

Ésta es la situación booleana que corresponde a un transistor MOS en conducción: equivale a una resistencia cuyo valor puede hacerse adecuadamente pequeño a través del factor de forma, W/L ; la resistencia es inversamente proporcional a su anchura W .

f) Al aumentar V_{DS} el canal se hace más estrecho (presenta menor número de portadores) en el lado del drenaje (pues $V_{GD} < V_{GS}$) y, al aumentar la tensión del drenaje, llega un momento en que el canal se satura, es decir, la diferencia de tensiones $V_G - V(x)$ no supera la tensión umbral, $V_G - V(x) < V_{TO}$ y, por ello, en tal zona no hay capa de inversión que contribuya a la conducción: para $V(x) > V_G - V_{TO}$ no existe canal.

Esta situación de canal saturado se da si $V_D > V_G - V_{TO}$, en cuyo caso la integral **A** se extiende solamente a la zona en que $V_G - V_{TO} - V(x)$ es positiva y la intensidad resultante es:

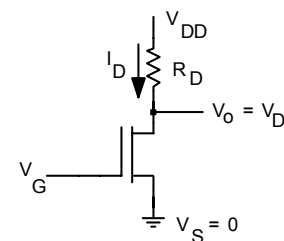
$$I_{D,sat} = \frac{K_p}{2} \frac{W}{L} (V_{GS} - V_{TO})^2$$

expresión que no depende de la tensión de drenaje y que representa la intensidad máxima que el transistor puede conducir para una tensión de puerta V_{GS} determinada; esta intensidad máxima (la de canal saturado) depende de la tensión de puerta, con la cual aumenta cuadráticamente.

g) Entre ambas zonas: zona lineal y canal saturado, la intensidad (resultante de integrar **A**) es:

$$I_D = K_p \cdot \frac{W}{L} \cdot \left(V_{GS} - V_{TO} - \frac{V_{DS}}{2} \right) \cdot V_{DS}$$

Etapa en fuente común:



$$V_S = 0$$

zona lineal

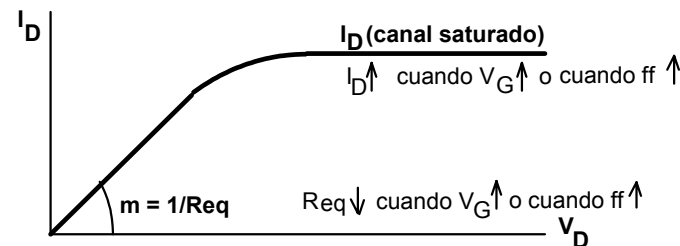
$$R_{eq} = K_p \frac{W}{L} \left(V_G - V_{TO} - \frac{V_D}{2} \right) V_D$$

canal saturado

$$I_{D,sat} = \frac{1}{K_p \frac{W}{L} (V_G - V_{TO})}$$

Para canal no saturado: $I_D = K_p \cdot \frac{W}{L} \cdot \left(V_G - V_{TO} - \frac{V_D}{2} \right) \cdot V_D$

Representando la intensidad que circula por el transistor (drenaje-fuente) I_D en función de la tensión entre sus terminales V_D se obtiene la curva característica de la conducción del transistor:



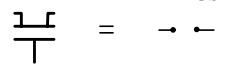
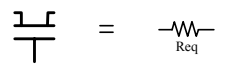
Actuando sobre las dimensiones geométricas del transistor (sobre su factor de forma $ff = W/L$), se modifica su resistencia en zona lineal y, en sentido contrario, la intensidad que conduce con canal saturado.

$$W \uparrow \text{ ó } L \downarrow \text{ (en ambos casos } ff \uparrow \text{) } \Rightarrow R_{eq} \downarrow ; I_D \text{ (canal saturado)} \uparrow ;$$

la magnitud de ambas variaciones es proporcional, con el mismo factor de proporcionalidad, a la modificación del factor de forma.

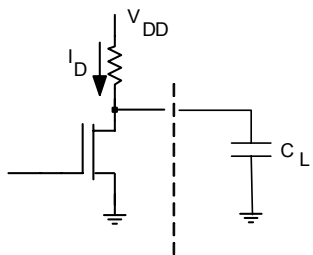
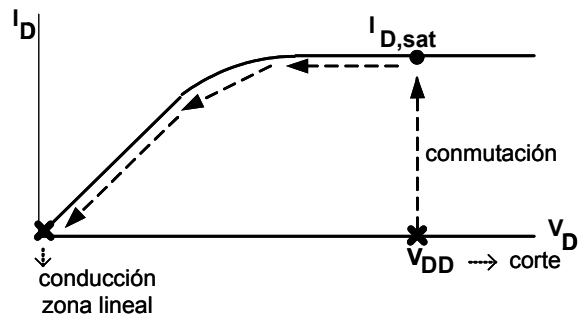
Comportamiento booleano:

Las situaciones booleanas de un transistor MOS corresponden a:

- transistor en corte $I = 0$ $V_{GS} < V_{TO}$

- transistor en zona lineal $V_{DS} = 0$ $V_{GS} > V_{TO}$


$$R_{eq} = \frac{1}{K_p \frac{W}{L} (V_G - V_{TO})}$$

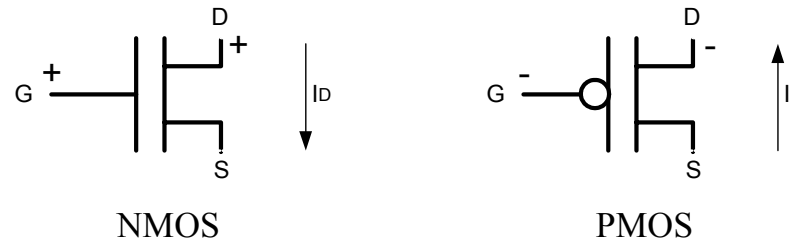
En la conmutación, en la puesta en conducción del transistor, éste recorre la curva característica correspondiente a $V_{GS} = V(1)$ desde la situación de canal saturado hasta alcanzar la zona lineal con $V_{DS} = 0$ V.



La $I_{D,sat}$ señala la máxima intensidad disponible por el transistor, con ella se inicia el proceso de conmutación, para descargar las capacidades equivalentes conectadas a su salida (debidas a otras puertas o dispositivos a los que el transistor comunica el valor booleano 0).

Transistores PMOS

En el transistor MOS de canal P son los huecos (en lugar de los electrones) los portadores que forman el canal y, por ello, sus tensiones e intensidades tienen signo opuesto al que presentan en el transistor NMOS: el transistor PMOS requiere tensiones de polarización V_D y de puerta V_G negativas respecto al sustrato y a la fuente, siendo también negativa su tensión umbral V_{TO} .



Además, la movilidad de los huecos es unas tres veces inferior a la de los electrones:

- Los electrones de la banda de conducción se encuentran efectivamente libres, han dejado su correspondiente enlace de valencia y se mueven por efecto del campo eléctrico con una cierta movilidad μ_e .
- En cambio, los huecos se encuentran en la banda de valencia y no se mueven por sí mismos sino como resultado de que un electrón ligado, que se encontraba en un enlace, pasa a cubrir el hueco y deja un nuevo hueco en su enlace anterior; de manera que el hueco se mueve en sentido contrario a como lo hacen los electrones ligados que lo rellenan.
- Siempre es más costoso movilizar un electrón ligado (que forma parte de un enlace en la banda de valencia) que un electrón libre (que se encuentra ya suelto en la banda de conducción); la movilidad de los huecos es muy inferior a la de los electrones.

En la práctica, la movilidad de los huecos μ_h resulta del orden de un tercio de la movilidad de los electrones μ_e ; por ello, a igualdad de dimensiones (a igualdad de factor de forma W / L), la intensidad conducida por un transistor PMOS será muy inferior a la de un transistor NMOS y la resistencia que presenta en zona lineal será considerablemente mayor para el transistor PMOS respecto al NMOS.

Modelos SPICE

El simulador eléctrico **SPICE** *Simulation Program With Integrated Circuits Emphasis* permite diversos niveles para caracterizar los transistores MOS.

El modelo de nivel 1, modelo de Shichman - Hodges, corresponde a las funciones desarrolladas anteriormente:

$$\begin{cases} I_D = K_p \frac{W}{L} \left(V_{GS} - V_{TO} - \frac{V_{DS}}{2} \right) V_{DS} & \text{canal no saturado} \\ I_D = \frac{K_p}{2} (V_{GS} - V_{TO})^2 & \text{canal saturado} \end{cases}$$

En este modelo la descripción SPICE de un transistor MOS requiere solamente tres parámetros:

- KP** K_p parámetro de transconductancia
- VTO** V_{TO} tensión umbral del transistor
- TOX** t_{ox} espesor del óxido de puerta

TOX es necesario para tener en cuenta la capacidad de puerta; caso de no utilizarlo, se supone capacidad de puerta nula.

Ejemplos del modelo de transistores MOS (nivel 1):

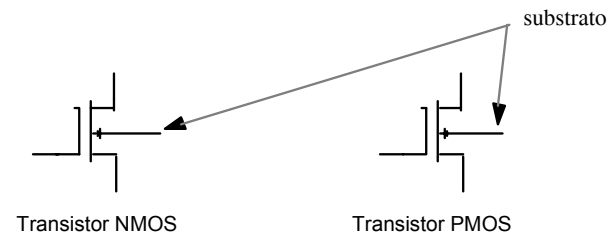
```
MODEL "nombre" NMOS LEVEL=1 KP=40U VTO=1 TOX=50E-9
```

```
MODEL "nombre" PMOS LEVEL=1 KP=15U VTO=-1 TOX=50E-9
```

Es necesario indicar también la dimensión de los transistores **L**, **W** bien dentro del modelo o en la declaración de cada transistor; por defecto configura las dimensiones de los transistores de 1 metro (dichos transistores de un metro cuadrado son tan grandes que dan lugar a resultados de simulación erróneos).

Nota aclaratoria sobre los símbolos utilizados para los transistores MOS

Los transistores MOS son dispositivos de 4 terminales: fuente, drenaje, puerta y sustrato y sus símbolos normalizados son:



En el caso de transistores MOS individuales, ciertamente hay ejemplares comerciales con los cuatro terminales disponibles, pero es más frecuente que tengan solamente tres terminales externos, con el sustrato conectado a la fuente:



Aún más, en el caso de circuitos integrados, el sustrato de todos los transistores NMOS es común y lo mismo sucede con el sustrato de los transistores PMOS, por lo que no resulta necesario dibujar el terminal de sustrato para cada transistor individual.

Por ello, los siguientes símbolos MOS con tres terminales (fuente, drenaje y puerta) resultan sumamente adecuados para representar el funcionamiento de los transistores MOS en los circuitos digitales y son los que se utilizan en este texto:



Estos símbolos no incluyen el terminal correspondiente al sustrato y, por tanto, no reflejan la polarización del sustrato de los transistores (necesaria para que las uniones de la difusión y del canal con el sustrato estén en polarización inversa y el transistor quede aislado del sustrato).

Habrà de tenerse en cuenta en forma implícita que el sustrato de los transistores NMOS ha de conectarse a la tensión más negativa de las de alimentación del circuito (que, generalmente, será el terminal correspondiente a 0 V), mientras que el sustrato de los transistores PMOS ha de estar conectado a la tensión más positiva de ellas (que suele ser el terminal correspondiente a V_{CC}).