

Instituto Tecnológico de Costa Rica

Escuela de Ingeniería Electrónica



Notas de Clase: Elementos Activos

Dr.-Ing. Paola Vega Castillo

2018

Edición: Dr.-Ing. Juan José Montero Rodríguez  
Semestre I 2019

# Índice general

<b>1. Fundamentos de semiconductores</b>	<b>5</b>
1.1. Introducción . . . . .	5
1.2. Estructura atómica . . . . .	5
1.3. Modelo de bandas de energía . . . . .	6
1.4. Clasificación de los materiales . . . . .	6
1.5. Cristalinidad . . . . .	8
1.6. Semiconductores . . . . .	9
1.7. Generación y Recombinación . . . . .	10
1.8. Dopado . . . . .	10
1.9. Portadores de carga . . . . .	13
1.9.1. El concepto de hueco . . . . .	14
1.9.2. Corriente de huecos . . . . .	14
1.10. Nivel de Fermi . . . . .	15
1.11. Transporte de portadores de carga . . . . .	18
1.11.1. Corriente de arrastre . . . . .	19
1.11.2. Corriente de difusión . . . . .	20
1.12. Diagramas de bandas de energía . . . . .	20
<b>2. Contactos</b>	<b>23</b>
2.1. Contactos metal-semiconductor . . . . .	23
2.1.1. Contacto Schottky . . . . .	23
2.1.2. Contacto Óhmico . . . . .	28
2.2. Contactos semiconductor-semiconductor . . . . .	29
2.2.1. Capacitancias de la junta PN: . . . . .	31
<b>3. El diodo</b>	<b>33</b>
3.1. Curva característica . . . . .	33
3.2. Modelos del diodo . . . . .	35
3.2.1. Modelos de gran señal . . . . .	35
3.3. Métodos de resolución de circuitos con diodos . . . . .	37
3.3.1. Solución analítica . . . . .	38
3.3.2. Solución gráfica: análisis por línea de carga . . . . .	38

<b>4. Aplicaciones del diodo</b>	<b>41</b>
4.1. Diodos en serie, diodos en paralelo y antiparalelo . . . . .	41
4.2. Rectificadores . . . . .	41
4.2.1. Rectificador de media onda . . . . .	41
4.2.2. Rectificador de onda completa . . . . .	43
4.2.3. Recortador serie . . . . .	45
4.2.4. Recortador paralelo . . . . .	46
4.2.5. Cambiador de nivel (sujetador) . . . . .	50
4.2.6. Doblador de voltaje . . . . .	51
4.2.7. Rectificador de media onda con filtro . . . . .	52
4.2.8. Rectificador de onda completa con filtro . . . . .	53
4.2.9. Análisis gráfico del regulador con diodo Zener . . . . .	53
<b>5. El transistor BJT</b>	<b>59</b>
5.1. Operación de Transistor Bipolar en Modo Activo . . . . .	60
5.2. Corriente de Colector – Modelo de Ebers-Moll . . . . .	62
5.3. Corrientes de base-emisor . . . . .	62
5.4. Regiones de operación del transistor . . . . .	63
5.5. Efecto Early en el transistor BJT . . . . .	66
<b>6. El transistor MOSFET</b>	<b>67</b>
6.1. El concepto de transistor y transistor MOSFET . . . . .	67
6.2. Estructura del MOSFET . . . . .	68
6.3. Principio de funcionamiento del MOSFET . . . . .	69
6.3.1. Análisis cualitativo . . . . .	69
6.3.2. Análisis cuantitativo . . . . .	71
6.3.3. Tensión de umbral $V_{TH}$ . . . . .	75
6.4. Clasificación y simbología del MOSFET . . . . .	75
6.5. Curvas características del MOSFET . . . . .	77
6.5.1. Región de corte . . . . .	79
6.5.2. Región lineal . . . . .	81
6.5.3. Región de saturación . . . . .	81
6.6. Polarización del substrato . . . . .	82
<b>7. Aplicaciones del MOSFET</b>	<b>87</b>
7.1. El MOSFET como interruptor . . . . .	87
7.2. Compuertas lógicas básicas . . . . .	90
7.2.1. El inversor CMOS . . . . .	90
7.2.2. Compuerta NAND . . . . .	93
7.2.3. Compuerta NOR . . . . .	94
7.3. Arreglos de transistores . . . . .	95
7.4. Consumo de potencia . . . . .	96
7.4.1. Potencia estática . . . . .	96
7.4.2. Potencia dinámica . . . . .	97
7.5. Memorias semiconductoras . . . . .	99

7.5.1. Organización de las memorias . . . . .	99
7.5.2. Memorias no volátiles . . . . .	101
7.5.3. Memoria SRAM . . . . .	103
7.5.4. Memoria DRAM . . . . .	106
7.6. Modelo de Pequeña Señal . . . . .	108
7.6.1. Reglas para cálculos en Gran Señal . . . . .	110
7.6.2. Reglas para cálculos en Pequeña Señal . . . . .	111
<b>8. Modelado del MOSFET</b>	<b>113</b>
8.1. Modelo analógico: gran señal . . . . .	113
8.2. Modelo analógico: pequeña señal . . . . .	115
8.2.1. Obtención del equivalente de pequeña señal . . . . .	116
8.2.2. Transconductancia . . . . .	116



# Capítulo 1

## Fundamentos de semiconductores

### 1.1. Introducción

Los semiconductores son el material base con el cual se construyen los elementos activos. Conviene en este momento recordar la clasificación de elementos en elementos pasivos y activos, para comprender el objeto de estudio de este curso.

Elementos activos son aquellos que cumplen alguna de las siguientes características:

- Pueden otorgar energía en un tiempo infinito, como ejemplo de estos elementos se tiene básicamente a las fuentes de tensión y corriente.
- Permiten la amplificación de voltaje o corriente. Como ejemplo de este caso se tiene a los transistores.

Por otra parte, los elementos pasivos son aquellos que:

- No pueden otorgar energía en un tiempo infinito, sino sólo consumirla o almacenarla durante un tiempo limitado.
- No permiten la amplificación de voltaje o corriente.

En este curso de estudiarán los elementos activos más importantes, así como otros dispositivos electrónicos basados en semiconductores. Además se presentará un resumen de los conceptos básicos de física de semiconductores, así como de los principales dispositivos basados en estos materiales y sus aplicaciones básicas. Así mismo, se discutirá la fabricación de circuitos integrados y el flujo de diseño físico de circuitos analógicos (flujo analógico de back-end).

### 1.2. Estructura atómica

El átomo posee un núcleo en el cual se concentran los protones y los neutrones, y una nube electrónica, en la cercanía del núcleo, en la cual se encuentran los electrones. Esta nube electrónica es en realidad una organización espacial de los electrones caracterizada por niveles de energía, los cuales se representan con el número cuántico principal  $n$ . El número cuántico principal indica la energía del nivel y la distancia de este nivel con respecto al núcleo.

Estos niveles de energía contienen orbitales, los cuales son regiones de alta probabilidad de encontrar al electrón. Los orbitales son zonas espaciales cuya forma está descrita por las soluciones de la ecuación de Schrödinger. La orientación de dichos orbitales en el espacio está dada por el número cuántico magnético  $m$ . El cuarto y último número cuántico, el spin, está relacionado con la rotación del electrón sobre su propio eje.

De acuerdo con el principio de exclusión de Pauli, los electrones de un átomo ocupan estados electrónicos -descritos por los cuatro números cuánticos- de forma tal que dos electrones no ocupan un mismo estado electrónico, es decir, no existen electrones que tengan los mismos números cuánticos. Este principio, junto con el principio de construcción (Aufbau Prinzip) determina la forma en que se ocupan los estados electrónicos definidos por los números cuánticos. El principio de construcción establece que los electrones ocupan primero los niveles de menor energía, dado que éstos son los niveles más estables.

De acuerdo con estos números, los electrones se han ordenado en la tabla periódica de los elementos. En esta pueden distinguirse, entre muchas otras formas de clasificación de los elementos, la división de los elementos en metales, metaloides y no metales. De los elementos de la tabla periódica interesa, para los objetivos de este curso, los elementos metálicos y metaloides de los grupos II, III, IV, V y VI, pero especialmente aquellos del grupo IV.

### 1.3. Modelo de bandas de energía

Cuando los átomos se aproximan unos a otros, la energía de los niveles individuales se modifica de manera que los electrones del material obedezcan al principio de exclusión de Pauli. Este fenómeno se conoce como desdoblamiento de los niveles de energía. Este desdoblamiento, es decir, reorganización de niveles de energía debido a la proximidad de átomos, da origen a dos bandas de energía: la banda de valencia y la banda de conducción, separadas por una banda prohibida en la cual no existen estados electrónicos que puedan ser ocupados por los electrones.

La banda de valencia es el estado de energía más alto que está lleno a cero grados Kelvin. Por otra parte, la banda de conducción, está vacía a cero grados Kelvin. Para que un electrón participe en el proceso de conducción, debe contar con una energía mínima que está dada por la energía del límite inferior de la banda de conducción.

La banda de valencia y la banda de conducción están separadas por un rango de energía sin estados electrónicos asociados; este rango de energía se conoce como la banda prohibida.

Los electrones en la banda de valencia no participan en el proceso de conducción. Se requiere entonces aplicar una energía mínima igual a la de la banda prohibida para lograr que un electrón de la banda de valencia pase a la banda de conducción y participe en el proceso de conducción.

### 1.4. Clasificación de los materiales

La clasificación de los materiales según su conductividad eléctrica se deriva del modelo de bandas de energía, ya que la relación de estas tres bandas determina las características eléctricas, ópticas y térmicas del material. Los materiales pueden clasificarse, de acuerdo con su conductividad eléctrica, en tres tipos:

- Aislantes:

Los aislantes se caracterizan por tener electrones fuertemente ligados al núcleo, a los cuales se les debe aplicar una gran cantidad de energía para desligar al electrón del núcleo de manera que éste pueda participar en el proceso de conducción eléctrica, es decir, se requiere mucha energía para mover un electrón de la banda de valencia a la banda de conducción. Estos materiales presentan, por lo tanto, una baja conducción eléctrica.

En el caso de los materiales aislantes, la banda prohibida es del orden de los 8-9 eV. Ejemplos de estos materiales son el dióxido de silicio ( $\text{SiO}_2$ ) y el nitruro de silicio ( $\text{Si}_3\text{N}_4$ ), como dos aislantes de gran importancia en la microelectrónica, así como el diamante.

El electrón-Volt (eV) es una medida de energía que equivale a  $1,6 \times 10^{-19}$  J.

b) Semiconductores:

Estos materiales se caracterizan por poseer una conductividad media. Se caracterizan por formar enlaces covalentes y obedecen a la regla del octecto, que establece que el número de átomos vecinos de un átomo está dado por  $8-N$ , donde  $N$  es el número de valencia del elemento. Con esto, se ocupan los estados electrónicos de valencia, logrando llenar el último subnivel y alcanzando una configuración estable. En los gases, esta configuración corresponde a la configuración de gas inerte.

Una característica clave de los materiales semiconductores es la posibilidad de controlar la resistividad del material mediante el proceso conocido como dopado, el cual se definirá más adelante. Esta característica es la base del desarrollo de la electrónica de estado sólido.

En los materiales semiconductores, existe una banda prohibida del orden 1-3 eV.

Entre estos materiales, el silicio tiene una banda prohibida de 1,12 eV.

El silicio y el germanio son ejemplos de materiales semiconductores.

c) Conductores:

En los conductores los electrones de valencia están débilmente ligados al núcleo, de manera que requieren de muy poca energía para contribuir al proceso de conducción eléctrica. Esto hace que los conductores posean una alta conductividad. Se dice que los metales son, en realidad, un arreglo de cationes sumergido en un mar de electrones. Estos electrones son los electrones de valencia, los cuales son sumamente móviles, y requieren de poca energía para cambiar su posición manteniendo unido el sólido. Esta es la razón por la cual los metales son dúctiles: en la deformación del material los electrones se redistribuyen logrando una configuración estable que mantiene el sólido. Por el contrario, los semiconductores son quebradizos, dado que en la deformación del material se rompen enlaces covalentes, logrando la separación del sólido, pues los enlaces covalentes son muy direccionales, y los electrones no son tan móviles para redistribuirse logrando mantener el sólido unido.

En los materiales conductores la separación entre bandas de valencia y conducción es muy pequeña, cero, o bien existe un traslape de bandas. Es decir, la banda prohibida es prácticamente inexistente o del todo no existe. Ejemplo de estos materiales son, en orden decreciente de conductividad: la plata, el cobre, el oro y el aluminio.

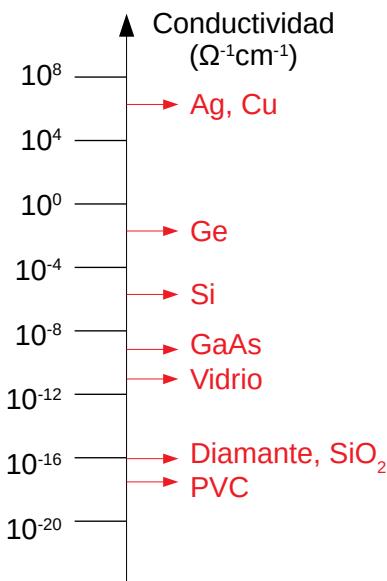


Figura 1.1: Clasificación de los materiales según su conductividad

## 1.5. Cristalinidad

De acuerdo con el grado de ordenamiento de los átomos de un material sólido (cristalinidad), los materiales pueden clasificarse en amorfos, policristalinos y cristalinos.

En los sólidos amorfos, los átomos están dispuestos sin seguir ningún orden en particular. No puede distinguirse ningún patrón de ordenamiento.

Los sólidos policristalinos presentan una estructura regular por tramos, es decir, pueden distinguirse patrones de ordenamiento por regiones. Sin embargo, estos patrones de ordenamiento no se extienden a todo el material.

En cambio, en los sólidos cristalinos los átomos siguen una estructura de ordenamiento regular en todo el material.

El grado de cristalinidad determina tanto las propiedades del material como su aplicación.

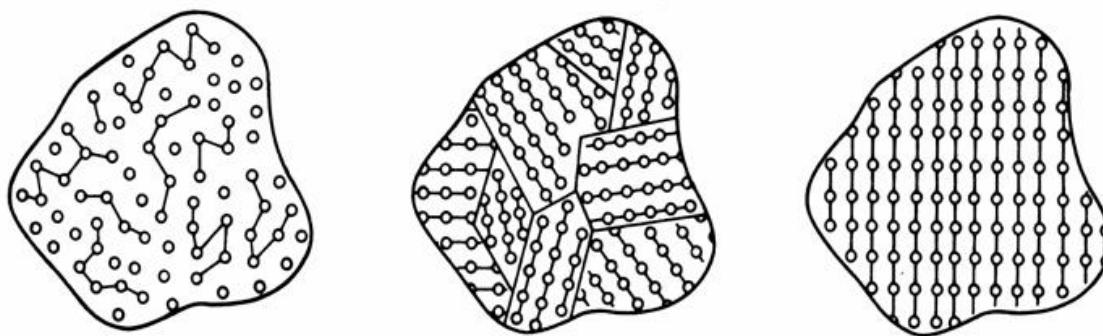


Figura 1.2: Clasificación de materiales según su grado de cristalinidad

## 1.6. Semiconductores

Los elementos semiconductores más importantes son el silicio y el germanio, del grupo IV de la tabla periódica, siendo el silicio uno de los elementos más abundantes en la naturaleza.

También existen semiconductores compuestos binarios, terciarios y cuaternarios.

Los semiconductores binarios se componen de los elementos de los grupos III y V o II y VI o bien se utiliza la mezcla SiGe (Silicio-Germanio). El arseniuro de galio (GaAs), es el compuesto binario de mayor aplicación en la industria microelectrónica.

Existen también compuestos semiconductores ternarios y cuaternarios tales como AlGaAs y InGaAsP, además de aleaciones como  $\text{Al}_x\text{Ga}_{1-x}\text{As}_y$ ,  $\text{Ga}_x\text{In}_{1-x}\text{As}_{1-y}\text{P}_y$ .

Estos materiales semiconductores se utilizan en la fabricación de circuitos integrados, diodos emisores de luz (LED), dispositivos láser y detectores lumínicos.

El silicio y el germanio tienen una estructura cristalina del tipo fcc (faced-centered cubic), es decir, cúbica centrada en las caras. Puesto que el diamante presenta la misma estructura, la estructura fcc se conoce como “estructura de diamante”.

El arseniuro de galio (GaAs) presenta también una estructura fcc, que se diferencia de la estructura del diamante por el hecho de incluir dos tipos de átomos (Ga y As). Esta estructura se conoce como estructura de zinc blenda (zinc blende).

El silicio cristalino posee una estructura de número de coordinación 4, es decir, cada átomo está rodeado por cuatro átomos con los cuales forma enlaces covalentes con los cuatro electrones de valencia, según la regla del octeto.

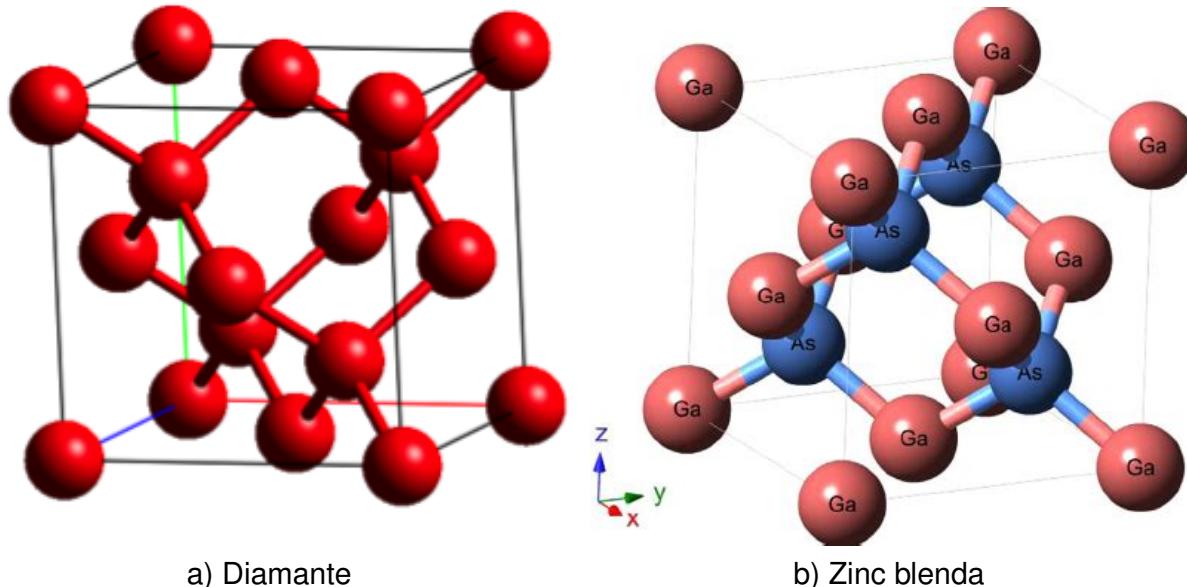


Figura 1.3: Estructuras cristalinas más comunes en los semiconductores.

## 1.7. Generación y Recombinación

A cero grados Kelvin, la estructura cristalina está totalmente quieta y no existen electrones libres. Conforme la temperatura aumenta, la energía térmica causa vibración de la estructura cristalina, lo cual rompe algunos enlaces, excitando (liberando) electrones que pueden contribuir al flujo de corriente. Al romperse un enlace se crea una vacante en el material debido a la ausencia del electrón excitado en el enlace. Esta vacante se conoce como hueco.

Los electrones y huecos interactúan en el semiconductor, dando lugar a un flujo de corriente. Los dos procesos básicos de interacción son la generación y la recombinación.

El proceso de generación consiste en la transición de un electrón de la banda de valencia a la banda de conducción. Esto genera un hueco (vacante) en la banda de valencia, dando lugar a la creación de pares electrón-hueco. La energía mínima para el proceso de generación es la energía de la banda prohibida.

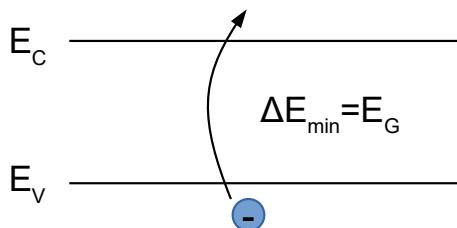


Figura 1.4: Proceso de generación.

El proceso de recombinación consiste en la transición de un electrón de la banda de conducción a la banda de valencia. Con esto un hueco es ocupado por un electrón, de manera que el hueco se elimina de la banda de valencia, es decir, hay una aniquilación de pares electrón-hueco.

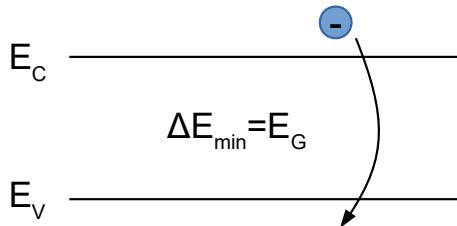


Figura 1.5: Proceso de recombinación.

Los procesos de recombinación y generación pueden originarse por luz o calor. En el caso de estímulos lumínicos, estos procesos se conocen como fotogeneración y fotorecombinación, respectivamente. En el caso de estímulos térmicos, los términos correspondientes son generación térmica directa y recombinación térmica.

## 1.8. Dopado

Como se mencionó anteriormente, la característica clave de los semiconductores es la posibilidad de variar su conductividad por medio de la introducción de impurezas en el material. Este

proceso se conoce como dopado.

Se puede distinguir entre dos tipos de semiconductores:

- Intrínsecos, o semiconductores puros.
- Extrínsecos, o semiconductores dopados.

El dopado consiste en tomar un material intrínseco e introducir intencionalmente impurezas.

Las impurezas se conocen como especies dopantes, y son, simplemente, átomos de un material diferente al intrínseco.

Al incorporarse en la estructura cristalina, los átomos dopantes reemplazan átomos del material intrínseco, por lo que estos átomos se conocen como impurezas substitucionales. El material resultante, que es un material dopado, se conoce como extrínseco.

Existen dos tipos de especies dopantes: donadores y aceptores.

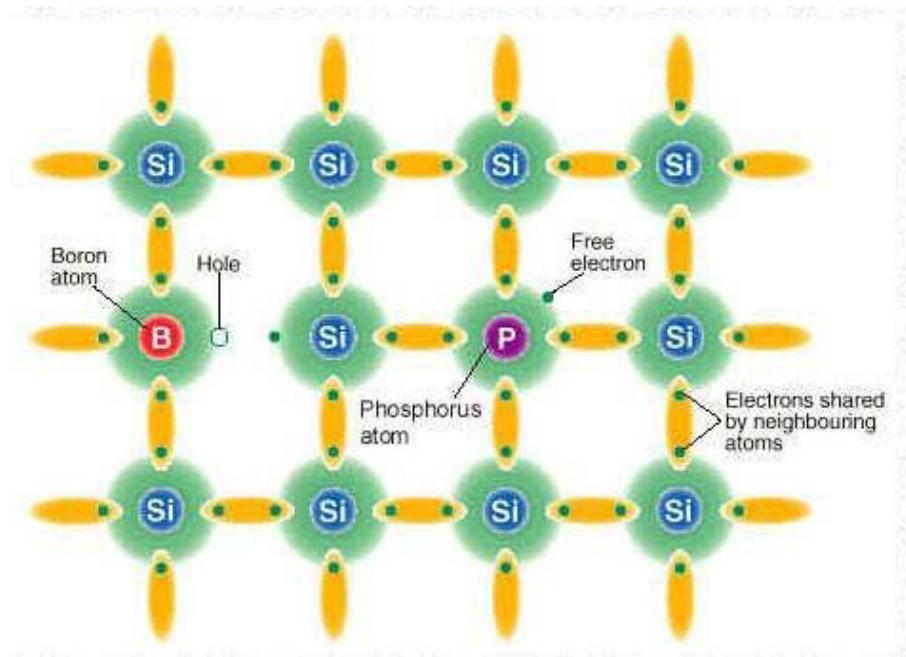


Figura 1.6: Concepto de dopado.

Las especies dopantes donadoras son aquellos elementos que tienen más electrones de valencia que los necesarios para establecer enlaces con sus átomos vecinos más próximos en el material intrínseco. Ejemplos de estos elementos son el arsénico (As), el fósforo (P) y el antimonio (Sb), los cuales tienen 5 electrones de valencia. Al incorporar un átomo de estos elementos en el silicio, el donador establece 4 enlaces covalentes con los 4 átomos de silicio más próximos. Puesto que el donador tiene 5 electrones de valencia pero requiere sólo 4 de ellos para formar los enlaces covalentes para incorporarse a la estructura cristalina, el quinto electrón queda débilmente ligado al átomo dopante y puede, adquiriendo poca energía, contribuir al flujo de corriente. Es decir, la especie donadora efectivamente dona un electrón para la conducción eléctrica.

Las especies acceptoras son aquellos elementos que tienen menos electrones de valencia que los necesarios para establecer enlaces con sus átomos vecinos más próximos en el material intrínseco. Ejemplos de estos elementos son el boro (B) y el indio (In), los cuales tienen 3 electrones de

valencia. Cuando un átomo aceptor se incorpora en el silicio, el aceptor establece 3 enlaces covalentes con 3 átomos de silicio. Ya que el aceptor tiene 3 electrones de valencia pero tiene 4 átomos de silicio vecinos, existe un enlace faltante, dado que falta un electrón para formar el enlace. Esto es lo que se conoce como hueco. Así, el átomo aceptor contribuye con un hueco que puede aceptar un electrón durante el proceso de conducción. De ahí el nombre de aceptor.

De acuerdo con el tipo de dopado, puede distinguirse entre dos tipos de semiconductores extrínsecos, a saber:

- Semiconductor tipo n, dopado con donadores.

- Semiconductor tipo p, dopado con aceptores.

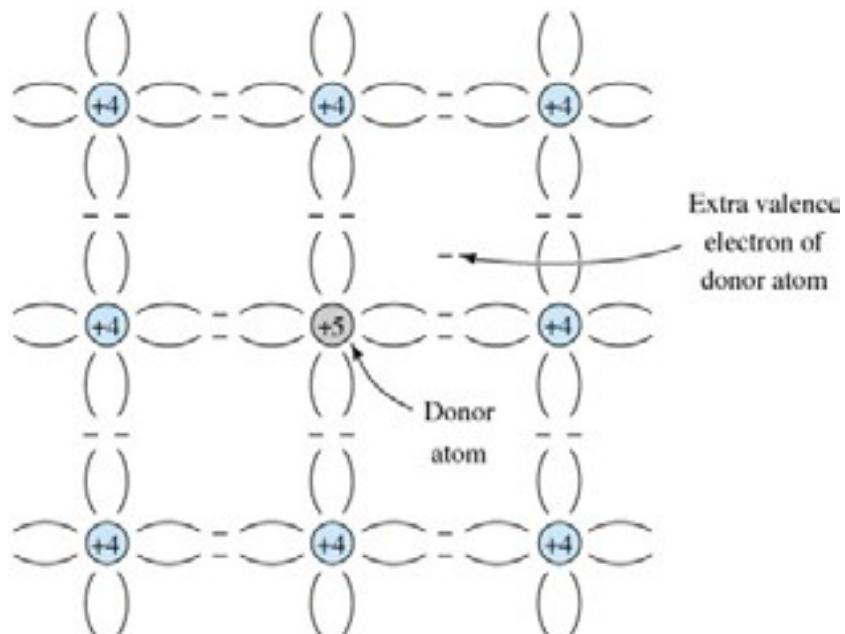


Figura 1.7: Dopado de un semiconductor N.

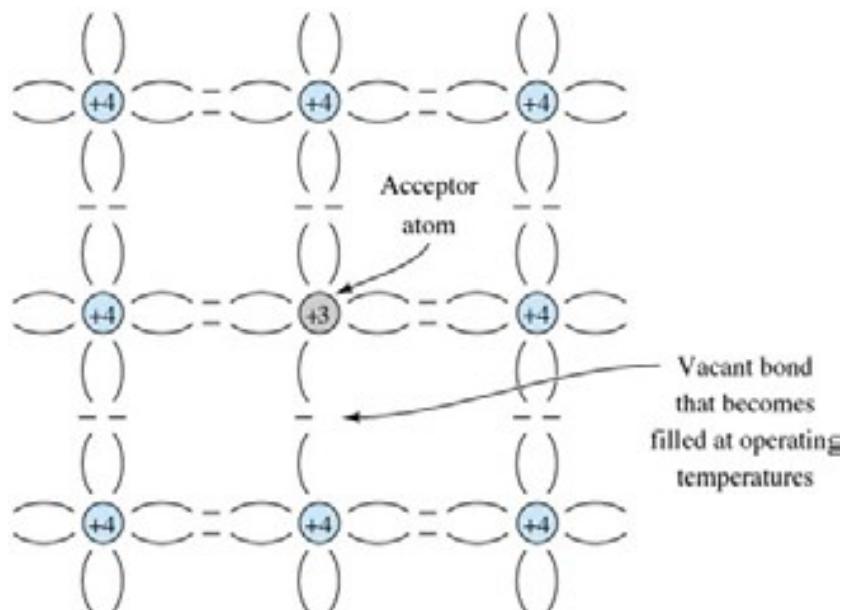


Figura 1.8: Dopado de un semiconductor P.

En cualquiera de los dos casos, el dopado no altera la neutralidad eléctrica del material semiconductor.

La introducción de donadores crea nuevos estados electrónicos en la muestra, los cuales se encuentran en la banda prohibida cercanos a la banda de conducción y a cero absoluto están ocupados por electrones. Conforme aumenta la temperatura, estos electrones pueden ganar suficiente energía para pasar a la banda de conducción, aumentando la cantidad de electrones libres en la muestra. La energía de ionización de los donadores en el silicio es mucho menor que la energía necesaria para el proceso de generación. De esta manera, a temperatura ambiente el dopado con donadores aumenta la cantidad de electrones libres en la banda de conducción.

Al introducir aceptores en un semiconductor intrínseco se crean nuevos estados electrónicos en la muestra, los cuales se encuentran en la banda prohibida cercanos a la banda de valencia y se encuentran vacíos a la temperatura de cero absoluto. Conforme aumenta la temperatura, electrones de la banda de valencia pueden ganar suficiente energía para ocupar estos estados electrónicos, aumentando la cantidad de huecos en la muestra. La energía de ionización de los aceptores en el silicio es mucho menor que la energía necesaria para el proceso de generación. De esta manera, a temperatura ambiente el dopado con aceptores aumenta la cantidad de huecos en la banda de valencia.

## 1.9. Portadores de carga

Existen dos tipos de portadores de carga: electrones (libres) y huecos. En un semiconductor extrínseco, se distingue entre portadores mayoritarios y portadores minoritarios, con base en la concentración de portadores.

Los portadores mayoritarios son aquellos que están presentes en mayor número en el semiconductor, es decir, huecos en semiconductores p y electrones en semiconductores n.

Los portadores minoritarios son aquellos que están presentes en menor número en el semiconductor, es decir, electrones en semiconductores p y huecos en semiconductores n.

Aquí conviene aclarar el origen de los portadores minoritarios. Es necesario recordar que debido a la energía térmica, se crea una cantidad necesaria de pares electrón-hueco en el semiconductor intrínseco. Esto hace que en un semiconductor siempre coexistan los dos tipos de portadores de carga. El dopado únicamente contribuye a aumentar el número de un tipo de portador de carga.

### 1.9.1. El concepto de hueco

El concepto de hueco es una representación de la banda de valencia con un estado electrónico vacío. El hueco se concibe como una partícula de carga positiva con una magnitud de carga igual a la del electrón y una masa efectiva que es mayor que la del electrón en un factor de 2 a 3. La masa efectiva es una masa equivalente que toma en cuenta el efecto del potencial del cristal en la partícula, sea ésta electrón o hueco. Con el uso de la masa efectiva, la partícula puede tratarse como si fuera una partícula libre en el vacío. El efecto de los átomos del cristal (potencial del cristal) en la partícula se considera tomando la masa efectiva de la partícula para realizar los cálculos. La masa efectiva tiene importantes consecuencias en la movilidad de los portadores de carga, como se discutirá más adelante.

Mientras que los electrones se mueven en la banda de conducción durante la conducción eléctrica, los huecos se mueven en la banda de valencia durante este proceso.

### 1.9.2. Corriente de huecos

El flujo de corriente en un semiconductor se debe al movimiento de carga libre, sean estos electrones o huecos. En el caso de la corriente de huecos, los electrones y los huecos interactúan para lograr un flujo de corriente.

Cuando un electrón se recombinan con un hueco, deja una vacante en otro punto de la red cristalina, es decir, generó un hueco en otra parte del material. Cuando otro electrón ocupa este hueco generado, dejó, igualmente que el primer electrón, una vacante (hueco) en otro punto del semiconductor. Esto puede verse como un desplazamiento de hueco a través del material, dando lugar a una corriente de huecos.

Puesto que el hueco se define como una partícula de carga positiva, la corriente de huecos tiene la misma dirección que la corriente técnica.

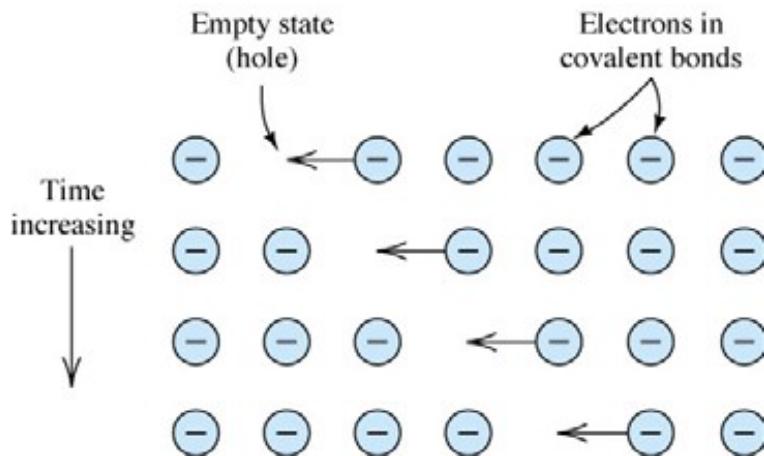


Figura 1.9: Corriente de huecos.

## 1.10. Nivel de Fermi

Uno de los conceptos más importantes de la teoría de bandas es el nivel de Fermi. Este se define como el nivel de energía en el cual la posibilidad de ocupación de un estado electrónico es de 50 %.

La probabilidad  $f(E)$  de ocupación de un estado electrónico en función de la energía  $E$  se describe matemáticamente con la distribución de Fermi-Dirac, la cual establece que:

$$f(E) = \frac{1}{1 + e^{\frac{E - E_F}{kT}}} \quad (1.1)$$

donde  $E_F$  denota el nivel de Fermi,  $k$  es la constante de Boltzmann y  $T$  es la temperatura en Kelvin.

La posibilidad de ocupación varía con la temperatura, siendo para 0K una función rectangular tal que:

$$f(E) = \begin{cases} 1 & \text{para } E < E_F \\ b & \text{para } E > E_F \end{cases} \quad (1.2)$$

Se cumple siempre que, independientemente de la temperatura,  $f(E_F) = 1/2$

El nivel de Fermi está íntimamente ligado a la concentración de portadores de carga libres en el semiconductor.

El nivel de Fermi del silicio intrínseco se encuentra aproximadamente a la mitad de la banda prohibida, como se muestra en la figura 1.10.

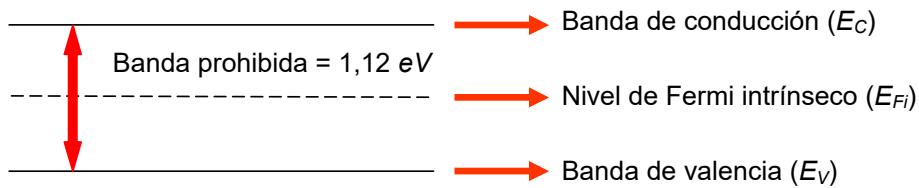


Figura 1.10: Diagrama de bandas de energía del silicio intrínseco.

La introducción de impurezas donadoras en el semiconductor crea nuevos estados electrónicos cuya energía es cercana a la de la banda de conducción, por lo cual, el nivel de Fermi extrínseco se desplaza a una energía mayor que la del nivel de Fermi intrínseco, como muestra la figura 1.11.

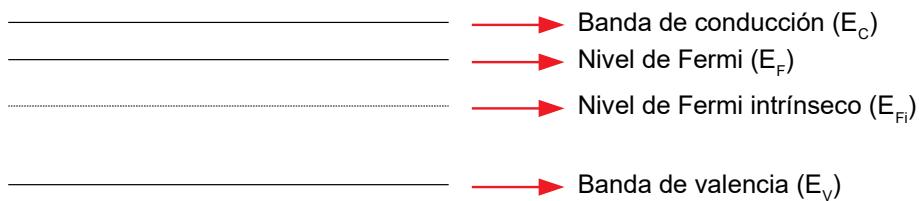


Figura 1.11: Diagrama de bandas de un semiconductor n.

Similarmente, al introducir impurezas aceptoras se crean nuevos estados electrónicos cuya energía es, en este caso, cercana a la banda de valencia, por lo que el nivel de Fermi extrínseco desplaza a una energía menor que la del nivel de Fermi intrínseco, como se muestra en la figura ??.

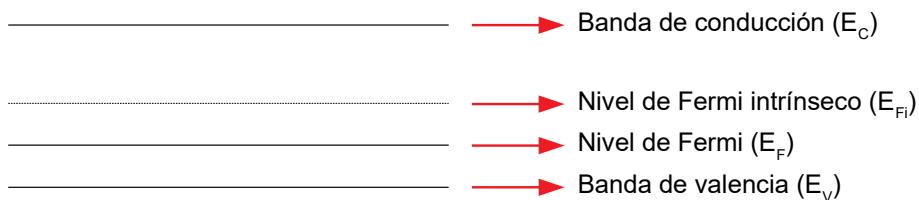


Figura 1.12: Diagrama de bandas de un semiconductor p.

La concentración de portadores de carga está determinada por la ley de acción de masas, que establece que en un semiconductor no degenerado el producto de la concentración de electrones y huecos es igual al cuadrado de la concentración intrínseca de portadores de carga. Es decir:

$$n_i^2 = n \cdot p \quad (1.3)$$

donde  $n_i$  es la concentración de portadores de carga en el semiconductor intrínseco en  $\text{cm}^{-3}$ ,  $p$  es la concentración de huecos ( $\text{cm}^{-3}$ ), y  $n$  es la concentración de electrones libres ( $\text{cm}^{-3}$ ). En el silicio, la concentración intrínseca de portadores de carga es de  $1,45 \times 10^{10} \text{ cm}^{-3}$ .

Se define como un semiconductor degenerado aquel en el cual la diferencia de energía entre el nivel de Fermi y la banda de conducción (semiconductores n) o la banda de valencia (semiconductores p), es estrechamente menor que  $3kT$ . En semiconductores degenerados, la energía de ionización de las especies dopantes es cero o prácticamente cero. Esto ocurre para niveles de dopado mayores que  $10^{18} - 10^{19} \text{ cm}^{-3}$ .

La concentración de portadores de carga y el nivel de Fermi están relacionados como sigue:

$$n = n_i \cdot e^{\frac{E_F - E_i}{kT}} \quad (1.4)$$

$$p = n_i \cdot e^{\frac{E_i - E_F}{kT}} \quad (1.5)$$

donde  $E_i$  denota el nivel de Fermi del semiconductor intrínseco,  $E_F$  denota el nivel de Fermi del semiconductor extrínseco.

En el proceso de conducción eléctrica los dopantes se ionizan como sigue:

- Los átomos donadores ceden un electrón, por lo cual se ionizan positivamente.
- Los átomos aceptores aceptan un electrón, por lo cual se ionizan negativamente.

Es decir:



donde  $N_D$  es la concentración de donadores,  $N_A$  es la concentración de aceptores,  $e$  representa un electrón libre y  $h$  representa un hueco.

A concentraciones de dopado mucho mayores que la concentración intrínseca de portadores de carga, puede despreciarse el efecto de los portadores mayoritarios intrínsecos y aproximar la concentración de portadores mayoritarios a la concentración del dopado. Así, en materiales p se tiene que la concentración de huecos es aproximadamente la concentración de aceptores:

$$p \approx N_A \quad (1.8)$$

mientras que la concentración de electrones (portadores de carga minoritarios) está dada por la ley de acción de masas:

$$n = \frac{n_i^2}{p} \approx \frac{n_i^2}{N_A} \quad (1.9)$$

En materiales n, la concentración de electrones es aproximadamente igual que la concentración de donadores:

$$n \approx N_D \quad (1.10)$$

y la concentración de huecos (portadores minoritarios) se calcula con la ley de acción de masas, como sigue:

$$p = \frac{n_i^2}{n} \approx \frac{n_i^2}{N_D} \quad (1.11)$$

El nivel de Fermi de un semiconductor puede expresarse también en términos de voltaje por medio del potencial de Fermi. Para semiconductores p:

$$\varphi_{F_p} = V_t \ln \left( \frac{n_i}{N_A} \right) \quad (1.12)$$

Para semiconductores n:

$$\varphi_{F_n} = V_t \ln \left( \frac{n_i}{N_D} \right) \quad (1.13)$$

donde  $V_t$  es el voltaje térmico, dado por:

$$V_t = \frac{kT}{q} \quad (1.14)$$

El nivel de Fermi y la concentración de portadores de carga dependen de la temperatura. A bajas temperaturas, la energía térmica es insuficiente para ionizar los átomos dopantes, por lo cual la concentración de portadores de carga mayoritarios es menor que la concentración del dopado:

$p < N_A$ , semiconductor p

$n < N_D$ , semiconductor n

A temperaturas medias, por ejemplo, a temperatura ambiente ( $25^\circ\text{C} = 298\text{ K}$ ), la energía térmica es suficiente para ionizar todos los átomos dopantes, y el semiconductor se comporta como un semiconductor extrínseco.

A altas temperaturas, el nivel de Fermi se acerca al nivel de Fermi intrínseco, debido a que la energía térmica contribuye a romper un número creciente de enlaces en la red cristalina, generando así un número creciente de pares electrón-hueco, de modo que el dopado tiene cada vez menos influencia en el material. Por esta razón, a altas temperaturas el semiconductor, aunque esté dopado, se comporta como un semiconductor intrínseco. Este comportamiento se manifiesta también en el diagrama de bandas de energía, donde se observa que al aumentar la temperatura el nivel de Fermi se acerca al nivel de Fermi intrínseco. Este acercamiento de niveles con el aumento de la temperatura se produce a mayor temperatura para niveles de dopado altos, pues debe generarse una mayor cantidad de portadores de carga intrínsecos para sobrepasar la cantidad de portadores de carga mayoritarios causada por el dopado.

## 1.11. Transporte de portadores de carga

La corriente eléctrica consiste en el movimiento de cargas, es decir, electrones, huecos o iones. En semiconductores, los dopantes ionizados son inmóviles y por lo tanto no contribuyen a la conducción, por lo tanto el flujo de corriente se debe a huecos y electrones libres.

Existen dos mecanismos de transporte de portadores de carga:

- Corriente de difusión, debido a gradientes de concentración de portadores de carga.
- Corriente de arrastre, debido a la aplicación de un campo eléctrico.

### 1.11.1. Corriente de arrastre

La corriente de arrastre está dada por la relación

$$\vec{J}_{\text{drift}} = \sigma \cdot \vec{E} \quad (1.15)$$

donde  $J_{\text{drift}}$  es densidad de corriente de arrastre,  $\sigma$  es conductividad del material y  $E$  es campo eléctrico aplicado.

La conductividad  $\sigma$  está determinada por la concentración de portadores de carga y su movilidad, de acuerdo con la relación

$$\sigma = (qn\mu_e + qp\mu_h) \quad (1.16)$$

donde  $q$  es la carga del electrón,  $\mu_e$  es la movilidad del electrón [cm<sup>2</sup> / Vs] y  $\mu_h$  es la movilidad del hueco [cm<sup>2</sup> / Vs].

El término “corriente de arrastre” se traduce muchas veces del término inglés “drift current” como “corriente de desplazamiento”, o “corriente de deriva”. Sin embargo, el término “corriente de desplazamiento” se refiere en realidad como aquella corriente originada por una variación de la densidad de flujo eléctrico en el tiempo ( $dD/dt$ ) en materiales aislantes. Por otra parte, el término “corriente de deriva” sugiere una corriente sin dirección definida, por lo que en este texto se utiliza el término “corriente de arrastre” por considerarse más apropiado, ya que refleja el hecho de que el campo eléctrico determina la magnitud y dirección de esta corriente.

La conductividad  $\sigma$  depende de la movilidad y la concentración de portadores de carga. La movilidad relaciona el campo eléctrico y la velocidad de arrastre según la siguiente ecuación:

$$\vec{v}_d = \mu \cdot \vec{E} \text{ para } |E| < 3 \times 10^3 \text{ V/cm} \quad (1.17)$$

donde  $v_d$  es la velocidad de arrastre de los portadores de carga [cm/Vs],  $E$  es campo eléctrico [V/cm] y  $\mu$  es movilidad [cm<sup>2</sup>/Vs].

En ausencia de un campo eléctrico, el electrón presenta un movimiento térmico aleatorio con una velocidad térmica promedio  $V_t$  y un desplazamiento efectivo de cero. Al aplicar un campo eléctrico, el electrón se desplaza con un desplazamiento efectivo diferente de cero y en la dirección determinada por el campo eléctrico.

Esta relación lineal entre velocidad de arrastre y campo eléctrico es válida para magnitudes de campo eléctrico  $E < 3 \times 10^3$  V/cm. Para  $E > 3 \times 10^3$  V/cm, la velocidad de arrastre permanece constante en  $v_d \approx 10^{-7}$  cm/s. Este fenómeno se conoce como saturación de la velocidad.

La movilidad está determinada por varios factores, entre ellos la masa efectiva del portador de carga, dispersión de portadores de carga debido a impurezas del material, y dispersión de portadores de carga causada por la estructura cristalina.

Debido a que los huecos tienen una masa efectiva que es 2 a 3 veces mayor que la de los electrones, la movilidad de los huecos es 2 a 3 veces menor que los electrones, es decir, los huecos se desplazan más lentamente que los electrones ante el mismo campo eléctrico aplicado.

La concentración de dopado también afecta la movilidad. A mayor concentración del dopado, menor movilidad, pues aumenta la dispersión debido a impurezas dopantes. La temperatura también disminuye la movilidad, dado que aumenta la vibración en la red cristalina.

### 1.11.2. Corriente de difusión

El fenómeno de difusión describe flujo de partículas debido a un gradiente de concentración, según la ley de Fick,

$$\mathbf{F} = -D \nabla \eta \quad (1.18)$$

La densidad de corriente de difusión para electrones y huecos se expresa como:

$$J_{\text{diff},n} = q \cdot D_n \cdot \nabla n \quad (1.19)$$

$$J_{\text{diff},p} = -q \cdot D_p \cdot \nabla p \quad (1.20)$$

donde  $D$  es el coeficiente de difusión [ $\text{cm}^2/\text{s}$ ],  $\nabla n$  y  $\nabla p$  son los gradientes de concentración de electrones y huecos, respectivamente.

El operador gradiente se define como

$$\nabla = \frac{\partial}{\partial x} \hat{x} + \frac{\partial}{\partial y} \hat{y} + \frac{\partial}{\partial z} \hat{z} \quad (1.21)$$

En análisis unidimensional, el operador gradiente se reduce a la derivada de la concentración con respecto a la posición de interés. La ausencia de signo en la ecuación de corriente de difusión de electrones obedece al hecho de que el gradiente da como resultado directo la dirección del flujo de la corriente técnica debido a la difusión de electrones.

El coeficiente de difusión se relaciona con la movilidad por medio de la relación de Einstein:

$$D = \frac{kT}{q} \mu = V_t \mu \quad (1.22)$$

donde  $V_t$  es voltaje térmico,  $\approx 25 \text{ mV}$  a  $300\text{K}$ .

### 1.12. Diagramas de bandas de energía

El diagrama de bandas de energía es una representación de la energía en función de la posición. Esta representación es de gran utilidad para el estudio de los dispositivos electrónicos y sistemas de materiales.

El diagrama consta de dos ejes: el eje de energía (abscisas) y el eje de posición (coordenadas). En muchos diagramas se asume que el lector está familiarizado con el sistema de coordenadas y se omiten los ejes. En otros, además de la energía, el eje "y" se interpreta como un eje de voltaje. En ese caso, la dirección de incremento del voltaje es opuesta a la de la energía, ya que se define:

$$V = \frac{E}{-q} \quad (1.23)$$

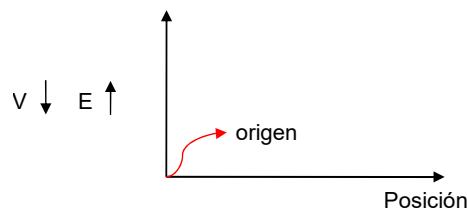


Figura 1.13: Ejes en un diagrama de bandas de energía.

La estructura completa de un diagrama de energía para un semiconductor intrínseco se presenta en la figura 1.14.

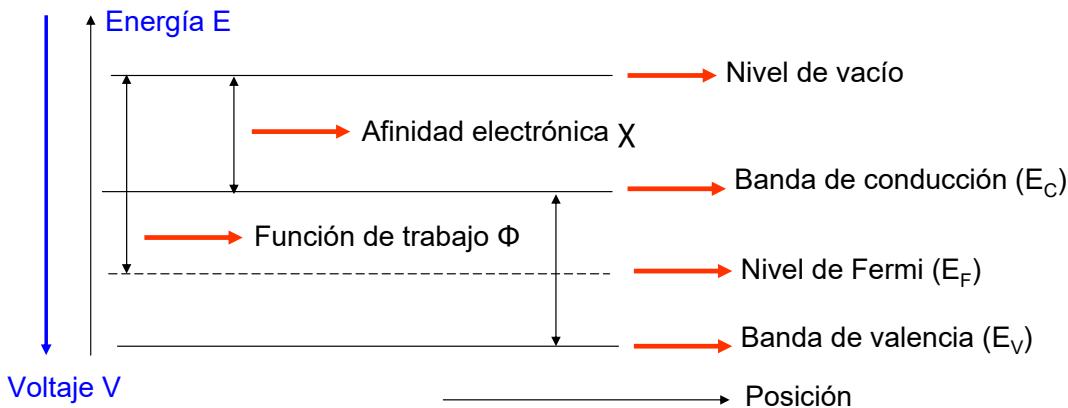


Figura 1.14: Diagrama de bandas de energía completo para un semiconductor intrínseco.

El nivel de vacío es un nivel de referencia, y corresponde a la energía a la cual un electrón se ha liberado del material, es decir, ya no se encuentra ligado a la estructura cristalina y por lo tanto, ha sido removido del material.

La afinidad electrónica es la energía que un electrón, en la banda de conducción, debe adquirir para convertirse en un electrón libre, es decir, para desprenderse del material. Esta cantidad no está definida para metales, para los cuales se utiliza sólo el concepto de función de trabajo. La función de trabajo es la diferencia de energía entre el nivel de vacío y el nivel de Fermi. En los metales, es la energía que debe aplicarse a un electrón para desprendérselo del material.

El nivel de Fermi intrínseco o nivel intrínseco, es el nivel de Fermi para un semiconductor intrínseco, y se encuentra localizado aproximadamente a la mitad de la banda prohibida. El dopado modifica el nivel de Fermi del semiconductor de acuerdo con la intensidad del dopado y el tipo de impureza dopante.



# Capítulo 2

## Contactos

### 2.1. Contactos metal-semiconductor

Cuando un trozo de metal y un material semiconductor entran en contacto físico, se forma una unión o contacto metal-semiconductor.

Los contactos metal-semiconductor se clasifican en contactos óhmicos y contactos Schottky, dependiendo de la diferencia de energía entre la función de trabajo del metal y la función de trabajo del semiconductor. Ambos tipos de contacto se discutirán a continuación.

#### 2.1.1. Contacto Schottky

El contacto Schottky, también conocido como contacto rectificante, se caracteriza por conducir corriente en una sola dirección.

La condición para obtener un contacto Schottky es:

- $\varphi_m > \varphi_s$  para semiconductores tipo n
- $\varphi_m < \varphi_s$  para semiconductores tipo p

A continuación se considerará el caso de un metal en contacto con semiconductor n.

Primeramente se toma el nivel de vacío como nivel de referencia para ambos materiales.

Puesto que los electrones tienden a ocupar estados de menor energía, con  $\varphi_m > \varphi_s$  la banda de conducción del metal se encuentra a un nivel de energía menor que el de la banda de conducción del semiconductor. Ya que  $E_{CS} > E_{CM}$ , electrones del semiconductor n pasan al metal. Estos electrones son cedidos por los átomos donadores del material n, dejando átomos donadores ionizados tras su paso al metal. Esto se produce en la cercanía de la interfaz (unión), de forma tal que el semiconductor n está cargado positivamente cerca de la interfaz.

Los electrones desplazados del material n se ubican en la banda de conducción del metal en la región cercana a la interfaz. El flujo de electrones del material n al metal continúa hasta que los niveles de Fermi del metal y del semiconductor se alinean, logrando así condiciones de equilibrio en el sistema metal-conductor. En condiciones de equilibrio (sin ninguna perturbación), el sistema alcanza una corriente neta de cero y los niveles de Fermi se alinean y están planos, es decir, no hay doblamiento de bandas.

Cuando los electrones se desplazaron al metal se redujo la concentración de electrones libres en la superficie del semiconductor en la interfaz. Este proceso se conoce como agotamiento (reducción

de la concentración de portadores de carga mayoritarios). De esta forma, en la región cercana a la interfaz se presenta una carga positiva debido a los dopantes ionizados inmóviles. Esta región se conoce como zona de carga espacial.

Al mismo tiempo, los electrones que emigraron al metal tienden a mantenerse en la superficie del metal cercana a la interfaz debido a la atracción de los donadores ionizados en el material n. Las cargas de la zona de carga espacial dan origen a un campo eléctrico dirigido del material n al metal. Este campo eléctrico genera una corriente que contrarresta el paso de electrones del semiconductor al metal, dando lugar a una corriente neta cero y estableciendo el equilibrio en el contacto. El campo eléctrico se manifiesta como una caída de tensión en el contacto, conocida como potencial de contacto. En el diagrama de bandas de energía del sistema, este campo se expresa como un doblamiento de bandas del semiconductor en la región cercana a la interfaz.

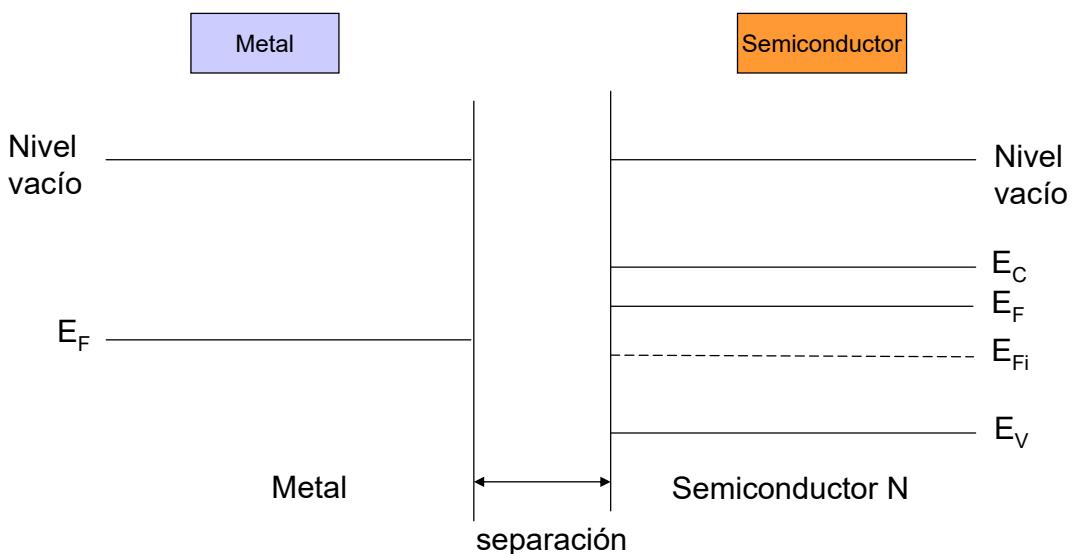


Figura 2.1: Diagrama de bandas de energía de un contacto Schottky antes del contacto.

Este doblamiento de bandas representa una barrera de potencial para los electrones del metal, impidiéndoles el paso hacia la región n. Esta barrera de potencial se conoce como barrera Schottky y se calcula como:

$$\varphi_B = \varphi_m - \chi \quad (2.1)$$

Igualmente los electrones del semiconductor encuentran una barrera de energía impidiendo su paso hacia el metal; esta se conoce como potencial de contacto. El potencial del contacto visto por los electrones del semiconductor es:

$$V_{bi} = \varphi_m - \varphi_s \quad (2.2)$$

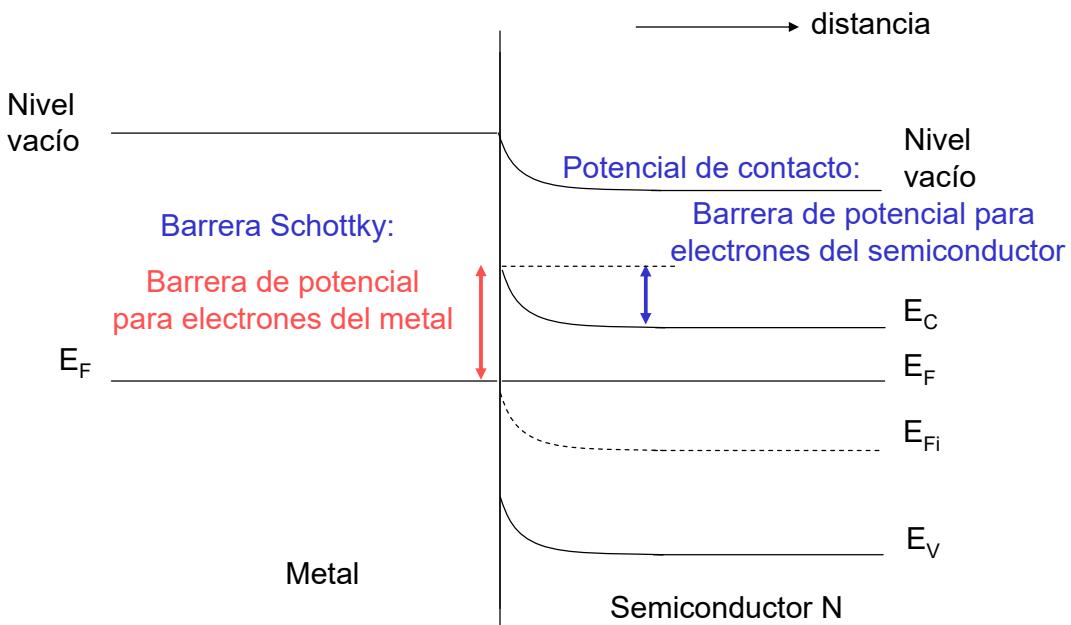


Figura 2.2: Diagrama de bandas de energía de un contacto Schottky después del contacto.

Puesto que en condiciones electrostáticas el campo en el interior de un metal es cero, se tiene que en el metal no hay doblamiento de bandas de energía, es decir, el metal es equipotencial. Adicionalmente, el campo tiende a cero fuera de la zona de carga espacial.

### Contacto Schottky con polarización inversa

La figura 2.3 muestra un contacto Schottky con polarización inversa.

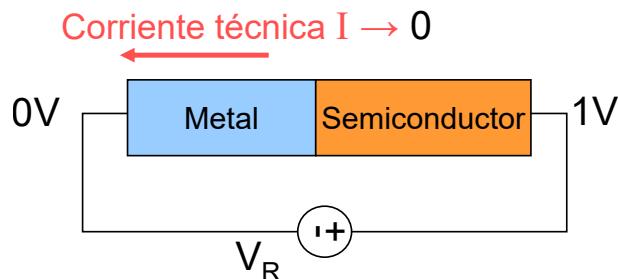


Figura 2.3: Contacto Schottky con polarización inversa.

Para comprender la influencia de la polarización inversa, deben tomarse en cuenta las siguientes consideraciones:

- Los niveles de Fermi se mantienen alineados en equilibrio.
- La afinidad electrónica del semiconductor no cambia, así como la función de trabajo del metal.
- El nivel de Fermi intrínseco, la banda de conducción y la banda de valencia pueden sufrir doblamiento.

La polarización inversa aumenta la barrera de potencial para los electrones y el ancho de la zona de carga espacial, ya que contribuye a agotar el material n cerca de la unión. Esto se refleja en el diagrama de bandas como un incremento en el doblamiento de bandas, como muestra la figura 2.4.

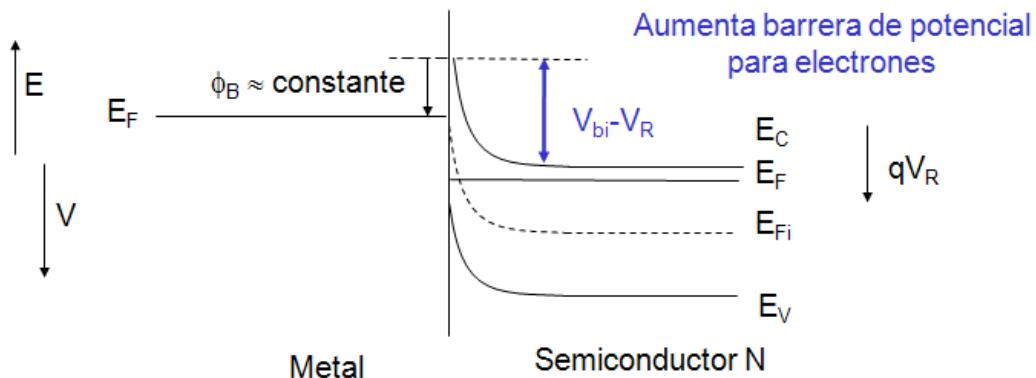


Figura 2.4: Diagrama de bandas de energía del contacto Schottky con polarización inversa.

La barrera Schottky se incrementa en  $V_R$ ; el potencial de reversa aumenta la barrera. Los electrones requieren entonces de mayor energía para desplazarse de una material a otro, por lo que solamente fluye una pequeña corriente de reversa ( $I \approx 0$ ) del semiconductor al metal a través de la unión. Sólo pueden fluir los electrones con energía mayor a la barrera del metal.

La característica de bloquear la corriente en un sentido, da al contacto Schottky el nombre de contacto rectificador.

### Contacto Schottky con polarización directa

La figura 2.5 muestra un contacto Schottky con polarización directa.

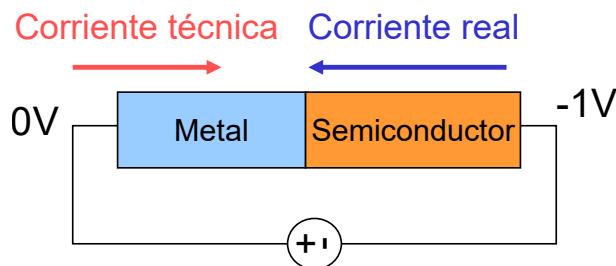


Figura 2.5: Contacto Schottky con polarización directa.

Como se muestra en el diagrama de bandas de energía de la figura 2.6, la polarización directa tiene el efecto de disminuir la barrera de potencial de la unión, atrayendo electrones del semiconductor n al metal y con ello estableciendo un flujo de corriente de polarización directa  $I > 0$ . Puesto que los electrones fluyen del semiconductor n al metal, la corriente técnica fluye del metal al semiconductor n. De esta manera, el contacto Schottky permite el paso de corriente técnica en la dirección de polarización directa pero bloquea el flujo de corriente en polarización inversa (efecto rectificador). Además, la polarización directa disminuye el ancho de la zona de carga espacial.

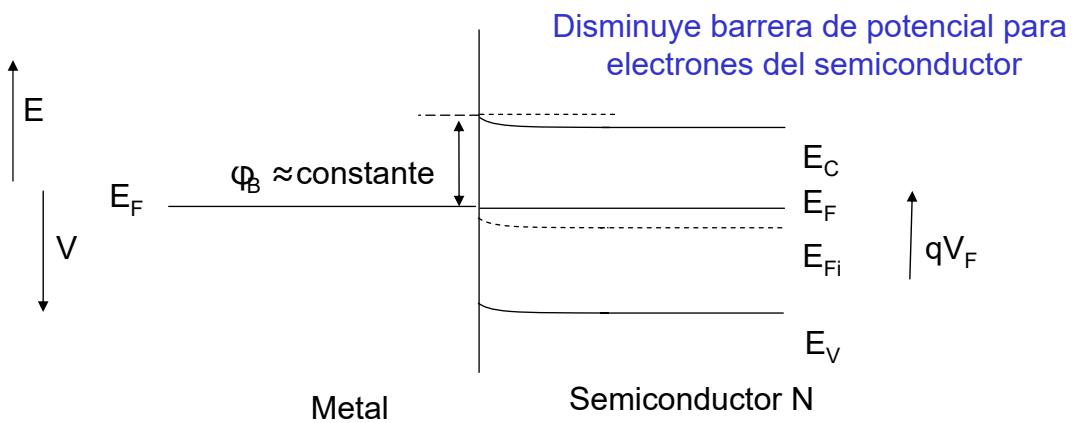


Figura 2.6: Diagrama de bandas de energía del contacto Schottky con polarización directa.

### Curva característica del contacto Schottky

La figura 2.7 muestra la curva característica (corriente contra voltaje) del contacto Schottky.

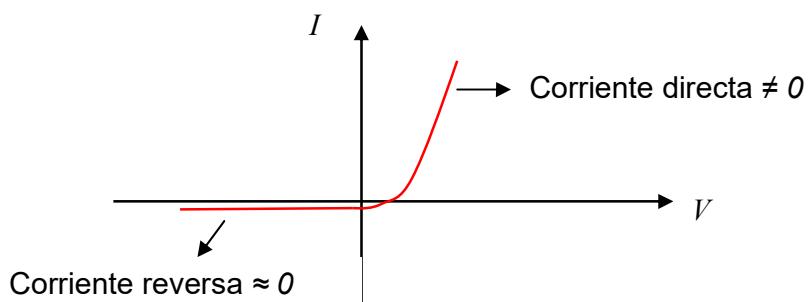


Figura 2.7: Curva característica del contacto Schottky.

La curva característica IV está descrita matemáticamente por la ecuación de Schockley, como sigue:

$$I_D = I_s(e^{(V_D/nV_t)} - 1) \quad (2.3)$$

donde

$I_s$ : corriente de saturación de reversa, depende de características físicas del diodo,  
 $V_t$  : voltaje térmico,  $n=1..2$ , depende de características constructivas del diodo.

### Aplicaciones del contacto Schottky

El contacto Schottky encuentra aplicación en el llamado diodo Schottky, el cual es un diodo de alta velocidad y bajo potencial de contacto utilizado en aplicaciones de alta frecuencia.

Un diodo se define como un dispositivo de dos terminales cuya característica corriente-voltaje es no lineal. La figura 2.8 muestra el símbolo del diodo Schottky, con sus dos terminales: ánodo y cátodo.

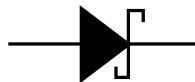


Figura 2.8: Símbolo del diodo Schottky.

La corriente de polarización directa fluye de ánodo y cátodo. La figura 2.9 muestra la dirección de la corriente de polarización directa para un diodo Schottky con semiconductor n y con semiconductor p, respectivamente.

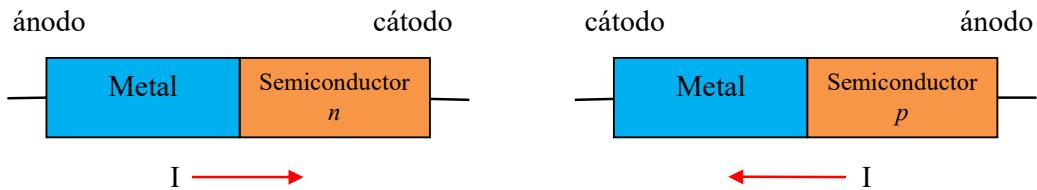


Figura 2.9: Dirección de la corriente de polarización directa en diodos Schottky.

### 2.1.2. Contacto Óhmico

El contacto óhmico se forma cuando un metal y un semiconductor entran en contacto físico y se cumple que:

$$\varphi_m < \varphi_s \text{ para semiconductores tipo n}$$

$$\varphi_m > \varphi_s \text{ para semiconductores tipo p}$$

El contacto óhmico, a diferencia del contacto Schottky, conduce corriente en ambas direcciones sin bloquearla. A continuación se analizará el caso de un metal en contacto con un semiconductor n.

Dado que  $\varphi_m < \varphi_s$ , inicialmente hay un flujo de electrones del metal al semiconductor hasta que los niveles de Fermi se alinean, haciendo que el sistema alcance el nivel de equilibrio. Como resultado de la migración de electrones del metal al semiconductor, el metal queda positivamente cargado en la región próxima a la interfaz, mientras que el semiconductor experimenta una acumulación de carga negativa cerca de la interfaz. Esto crea un campo eléctrico dirigido del metal al semiconductor. Así, el potencial de contacto causa un doblamiento de bandas tal que no existe barrera de potencial para el paso de electrones del semiconductor al metal, y la barrera para el paso del metal al semiconductor es muy pequeña. De esta manera, el flujo de corrientes es posible en ambas direcciones.

#### Curva característica del contacto óhmico

La curva característica del contacto óhmico se presenta en la figura 2.10. Como se puede observar, la curva característica corresponde a la de una resistencia, según se concluyó del análisis del diagrama de bandas de energía.

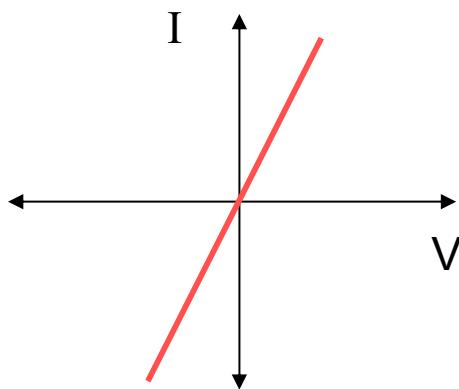


Figura 2.10: Curva característica de un contacto óhmico.

### Aplicaciones del contacto óhmico

El contacto óhmico es de gran importancia en la industria electrónica, puesto que se requiere para el contacto eléctrico (conexión) de los componentes semiconductores. Para lograr este tipo de contacto se enfocan grandes esfuerzos en la ciencia e ingeniería de materiales para lograr interfaces en las que la resistencia de contacto se minimice.

## 2.2. Contactos semiconductor-semiconductor

El contacto semiconductor-semiconductor se forma cuando dos materiales semiconductores, uno de tipo p y otro de tipo n, entran en contacto físico. Por este motivo, el contacto semiconductor-semiconductor se conoce como unión o junta PN.

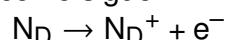
Para el análisis del funcionamiento de la unión PN deben tomarse en cuenta las siguientes consideraciones:

- El sistema está en equilibrio, es decir, no está afectado por ninguna perturbación. Ejemplo de perturbación son: voltaje, luz, gradientes térmicos, campos magnéticos.
- En la superficie de unión de los materiales (unión metalúrgica), hay un cambio abrupto de dopado, de tipo n a tipo p.

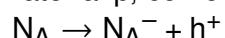
Debido al cambio de tipo de material, al unir dos materiales se presenta un gradiente de concentración de portadores: alta concentración de huecos en el material p y baja en el material n, y alta concentración de electrones en el material n y baja en el material p.

Este gradiente de concentración causa una corriente de difusión de huecos del lado p al lado n y una corriente de difusión de electrones del lado n al lado p.

El paso de electrones del material n al material p causa la ionización de donadores en el material n, como sigue:



Similarmente, el paso de huecos del material p al material n causa la ionización de aceptores en el material p, como sigue:



Esta redistribución de portadores de carga causa la formación de una zona de carga espacial (zona de agotamiento) en la cercanía de la interfaz, así como un campo eléctrico dirigido de la zona n a la zona p.

Las regiones semiconductores no agotadas se conocen como regiones cuasineutrales.

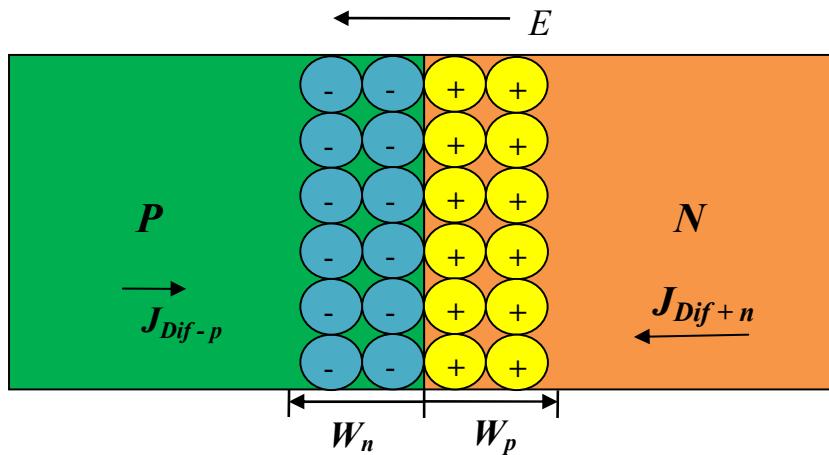


Figura 2.11: Difusión de portadores de carga en la junta PN.

Este campo eléctrico genera una corriente de arrastre de dirección contraria a la corriente de difusión, de forma que la contrarresta. La corriente neta a través de la unión es cero.

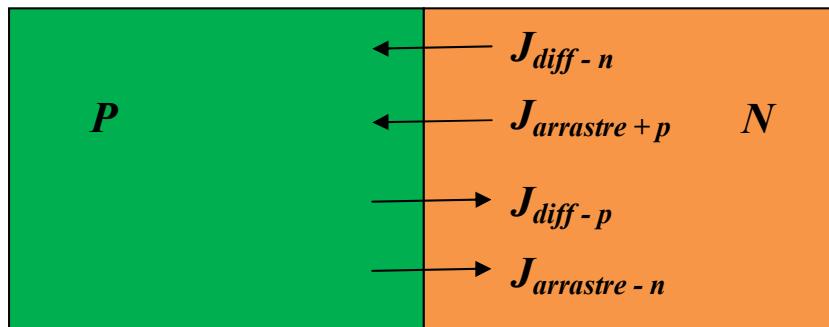


Figura 2.12: Difusión de portadores de carga en la junta PN.

El potencial de contacto  $V_{bi}$  a través de la unión se calcula como:

$$V_{bi} = V_t \ln \left( \frac{N_A N_D}{n_i^2} \right) \quad (2.4)$$

y depende de las concentraciones de portadores mayoritarios de cada material, dado que el potencial de contacto es la diferencia de potenciales de Fermi de los materiales p y n.

A temperatura ambiente las uniones PN (diodos) de silicio presentan un potencial de contacto de aproximadamente 0.7 V, los diodos de germanio de 0.3 V y los diodos Schottky aproximadamente 0.3 V.

### 2.2.1. Capacitancias de la junta PN:

En la junta PN existen dos capacitancias parásitas:

- Capacitancia de difusión
- Capacitancia de agotamiento

La capacitancia de agotamiento predomina cuando el diodo opera en la región de reversa, y depende de la tensión aplicada.

Sin tensión aplicada, se tiene el valor máximo de la capacitancia de agotamiento. Esta se debe a la existencia de la zona de agotamiento. Un aumento en la tensión de polarización inversa atrae a los portadores de carga mayoritarios hacia la región cuasineutral, es decir, deja a los portadores de carga de los extremos de la zona de carga espacial. Esta carga desplazada causa la ionización de las regiones aledañas a la zona de carga especial, de forma que el ancho de esta zona se amplía.

Dado que la tensión aplicada afecta solamente los extremos de la zona de agotamiento, puede concebirse que la zona de agotamiento se comporta como un dieléctrico de ancho variable y las regiones cuasineutrales, al no estar agotadas, presentan una conductividad mucho mayor, haciéndolas equivalentes a las placas conductoras de un capacitor. De esta forma, la zona de agotamiento y las regiones cuasineutrales forman un capacitor parásito en el cual el espesor del dieléctrico aumenta conforme aumenta la polarización inversa.

Este efecto es aprovechado para la implementación de capacitancias contraladas por tensión, también llamadas varactores (variable capacitor), y es utilizado en circuitos sintonizadores.

Por otra parte, la capacitancia de difusión predomina cuando el diodo opera en la región de polarización directa, aun cuando en estas condiciones la capacitancia de agotamiento aumenta.

Debido a la corriente de difusión, los portadores de carga mayoritarios en las regiones cuasineutrales cruzan la zona de agotamiento hasta encontrarse en el material en el que se convierten en portadores de carga minoritarios. Gracias a esto existe un gradiente de concentración de portadores de carga minoritarios en las regiones aledañas a la zona de agotamiento, donde el pico de concentración se da en la interfaz entre la zona de agotamiento y las regiones cuasineutrales. Al alejarse de la zona de agotamiento, esta concentración disminuye exponencialmente hasta alcanzar los valores de concentración en equilibrio. El exceso de portadores minoritarios es proporcional a la corriente y por ello aumenta conforme aumenta la polarización directa.

Esta carga almacenada cambia en respuesta al voltaje aplicado, creando un efecto capacitivo en el proceso de alcanzar las concentraciones de equilibrio, dado que

$$C = \frac{dQ}{dV} \quad (2.5)$$

Los portadores de carga minoritarios no continúan acumulándose en toda la región cuasineutral, dado que, por la continuidad de corriente, los portadores minoritarios se inyectan en la misma tasa en la que son extraídos, manteniendo constante la carga acumulada en la región cuasineutral. Esta carga afecta el tiempo de recuperación de reversa del diodo, es decir, el tiempo en que el diodo pasa del estado encendido al apagado. Este tiempo depende de:

- La cantidad de portadores de carga minoritarios en la región cuasineutral aledaña a la zona de agotamiento, es decir, los portadores de carga que se encontraban desplazándose por difusión cuando el diodo estaba polarizado en directa.

- El tiempo medio de vida de los portadores de carga minoritarios, es decir, el tiempo durante el cual los portadores de carga pueden desplazarse sin recombinarse.

# Capítulo 3

## El diodo

### 3.1. Curva característica

La junta PN dio lugar a uno de los elementos semiconductores más comunes: el diodo. El diodo es un dispositivo de dos terminales que bloquea la corriente en una dirección y permite su paso en la otra dirección. La figura 3.1 muestra el símbolo del diodo y sus terminales:

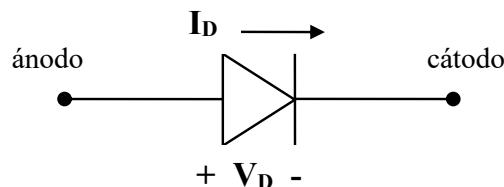


Figura 3.1: Símbolo del diodo y sus terminales.

Como se aprecia en la figura, el triángulo del símbolo del diodo es como una flecha que apunta en la dirección en la que fluye la corriente técnica.

El diodo se polariza en directo aplicando un voltaje positivo al ánodo y negativo al cátodo. El ánodo es la terminal en contacto con el lado  $p$  de la unión, mientras que el cátodo contacta el lado  $n$ . Puesto que los huecos fluyen de  $p$  a  $n$ , el cátodo atrae a los huecos, de ahí su nombre. En el caso de los electrones, éstos son atraídos al material  $p$  y recolectados en el ánodo.

El diodo ideal puede modelarse como un cortocircuito para corrientes de polarización directa y como un circuito abierto para corrientes de polarización inversa. Esto se muestra en la figura 3.2.

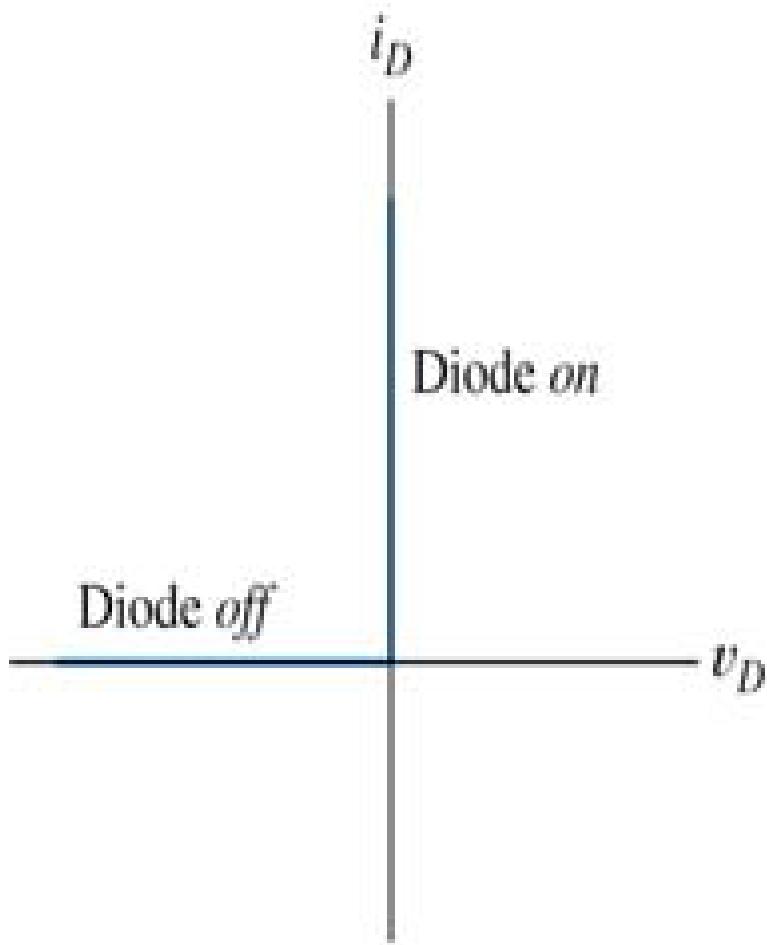


Figura 3.2: Curva característica del diodo ideal.

La característica IV del diodo real está descrita por la ecuación de Schockley:

$$I_D = I_S \left( e^{\frac{nV_D}{V_t}} - 1 \right) \quad (3.1)$$

donde  $I_D$  es la corriente del diodo,  $I_S$  es la corriente de saturación de reversa,  $V_D$  es el voltaje aplicado al diodo,  $V_t$  el voltaje térmico y  $n$  es un exponente que varía entre 1 y 2, y está determinado por las características constructivas del diodo.

La corriente de saturación de reversa depende de las características físicas del diodo, como sigue:

$$I_S = qA \left[ \frac{D_n}{L_n} n_{p0} + \frac{D_p}{L_p} p_{n0} \right] \quad (3.2)$$

donde  $A$  es el área transversal del diodo,  $D$  es el coeficiente de difusión,  $L$  es la longitud de difusión de los portadores minoritarios,  $n_{p0}$  es la concentración de electrones en el lado p en condiciones de equilibrio y  $p_{n0}$  es la concentración de huecos del lado n en condiciones de equilibrio.

La característica IV del diodo descrita por la ecuación de Schockley se presenta en la figura 3.3. Esta curva puede extenderse para mostrar el fenómeno de ruptura. Para voltajes de reversa más altos, el diodo entra en la región de ruptura y presenta una corriente de reversa no despreciable que crece rápidamente, como se muestra en la figura 3.4.

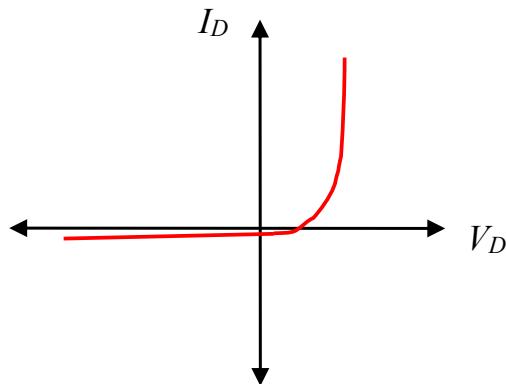


Figura 3.3: Curva característica del diodo real.

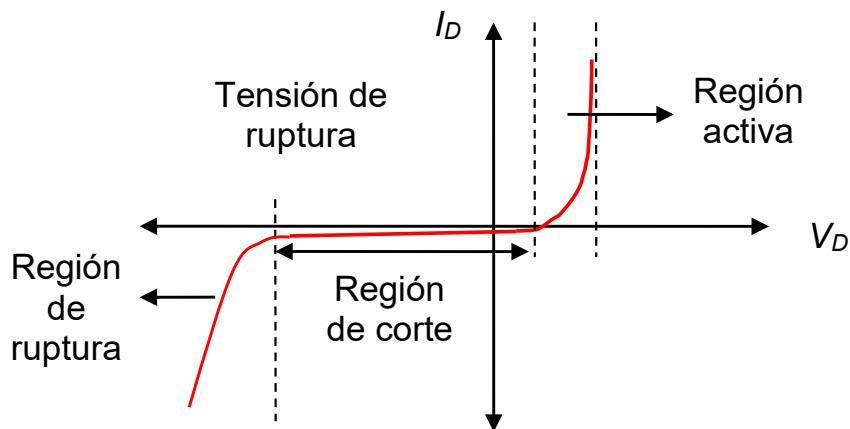


Figura 3.4: Curva característica del diodo real incluyendo la región de ruptura.

Las causas de este aumento de corriente en reversa se discutirán más adelante cuando se consideren el efecto tener y el efecto de avalancha.

## 3.2. Modelos del diodo

### 3.2.1. Modelos de gran señal

Los modelos de gran señal se utilizan para aplicaciones que no involucran amplificación o atenuación de señales. En gran señal, se distinguen cuatro modelos del diodo, mostradas en la figura 3.5.

- **Modelo del diodo ideal:** el diodo se modela como un corto circuito en polarización directa

- Modelo de caída de tensión constante: el diodo se modela como una fuente de tensión con la polaridad y magnitud de la tensión de activación del diodo
- Modelo lineal: agrega la resistencia dinámica del diodo al modelo de caída de tensión constante
- Modelo del diodo real, con la ecuación de Shockley

En la región de polarización inversa, todos los modelos representan al diodo como un circuito abierto.

### Modelo lineal por tramos

El diodo puede modelarse por tramos considerando tres regiones: región activa, región de corte y región de ruptura.

En la región activa el diodo puede modelarse como una fuente de tensión en serie con una resistencia. El valor de la fuente de tensión está dado por el potencial de contacto de la unión, y la resistencia está definida por la resistencia del diodo entre ánodo y cátodo, es decir, la pendiente de la curva característica en la región en la que el diodo se encuentra completamente activo. El circuito equivalente se muestra en la figura 3.6.

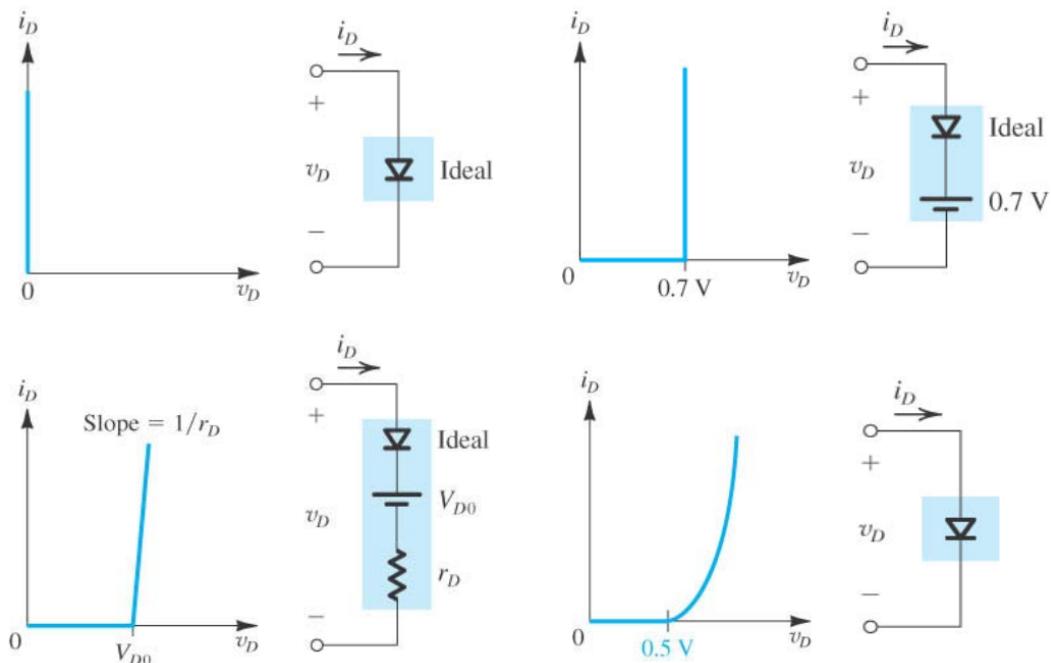


Figura 3.5: Modelos del diodo en la región de polarización directa: a) modelo ideal, b) modelo de caída de tensión constante, c) modelo lineal, d) modelo real.

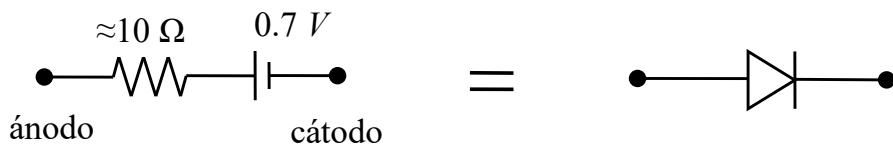


Figura 3.6: Modelo del diodo en la región activa.

En la región de corte, el diodo se modela como un circuito abierto, ya que la resistencia del diodo en reversa es muy alta (en el orden de los  $\text{G}\Omega$ ). Por otra parte, el comportamiento en la región de ruptura se modela con una fuente de tensión en serie con una resistencia. El valor de la fuente de tensión está definida por la tensión de ruptura del diodo. Para diodos de baja potencia, esta tensión ronda típicamente los 6 V. En la región de ruptura, la resistencia del diodo es sólo ligeramente más alta que en la región activa y corresponde a la pendiente de la curva característica en la región de ruptura. El circuito equivalente se muestra en la figura 3.7.

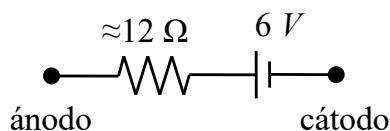


Figura 3.7: Modelo del diodo en la región de ruptura.

### 3.3. Métodos de resolución de circuitos con diodos

Primero se considerará la operación del diodo fuera de la región de ruptura. Considérese el circuito serie en la figura 3.8:

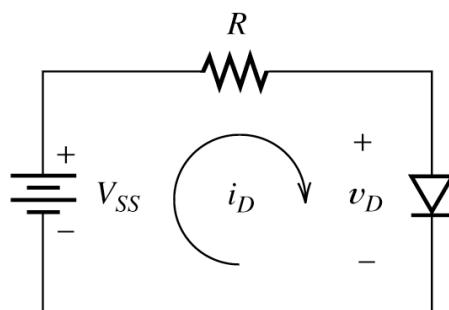


Figura 3.8: Ejemplo de circuito para estudio del método de línea de carga.

El voltaje y la corriente a la que opera un elemento del circuito se conoce como punto de operación. Existen dos métodos para encontrar el punto de operación:

- La solución analítica del circuito, aplicando leyes de Kirchhoff, ecuaciones características de los dispositivos, etc.
- La solución gráfica, también llamada análisis de líneas de carga.

### 3.3.1. Solución analítica

La solución analítica del circuito se obtiene aplicando la ley de voltajes de Kirchhoff. Para este caso, se tiene que:

$$V_S = V_R + V_D = IR + V_D \quad (3.3)$$

Sabemos que el diodo de silicio se activa aplicando una tensión de polarización directa, que está dada por el potencial de contacto de la unión PN. Este voltaje es de aproximadamente 0.7 V para que establezca una corriente en el circuito, es decir, se utiliza el modelo de caída de tensión constante. Conociendo un conjunto de parámetros de la ecuación puede entonces decirse que:

$$I = \begin{cases} 0 & \text{para } V_S \leq V_D \\ \frac{V_S - V_D}{R} & \text{para } V_S > V_D \end{cases} \quad (3.4)$$

R también puede dimensionarse conociendo cuánta corriente debe fluir en el circuito.

Por supuesto el punto de operación exacto del circuito puede resolverse también tomando la ecuación de Shockley y la malla del circuito.

### 3.3.2. Solución gráfica: análisis por línea de carga

El análisis de línea de carga consiste en resolver gráficamente las ecuaciones de la solución numérica. Para ello se procede como sigue:

1. Separar el circuito en una red de entrada, que incluye la fuente de entrada y una red de salida, que incluye la sección de la red en la que se definió la salida.
2. Graficar las características IV de los elementos del circuito de cada red en un mismo sistema de coordenadas voltaje-corriente y encontrar la característica IV resultante para cada red.
3. Encontrar la intersección de las curvas características de la red de entrada y la red de salida.

Para el ejemplo del circuito serie, el gráfico de líneas de carga queda como muestra la figura 3.9, donde se muestra la curva IV del diodo (red de salida) y la curva IV del equivalente de Thévenin del resto del circuito (red de entrada). El punto de intersección es el punto de operación, también llamado punto quiescente (punto Q).

$$\begin{cases} V = V_D + I_D R \\ I_D = I_S (e^{V_D/V_t} - 1) \end{cases} \quad (3.5)$$

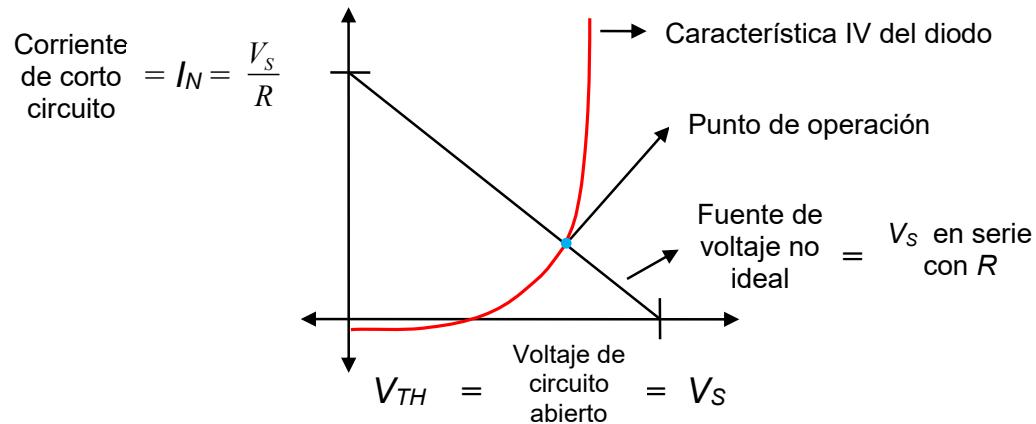


Figura 3.9: Solución del circuito utilizando el método de línea de carga.



# Capítulo 4

## Aplicaciones del diodo

### 4.1. Diodos en serie, diodos en paralelo y antiparalelo

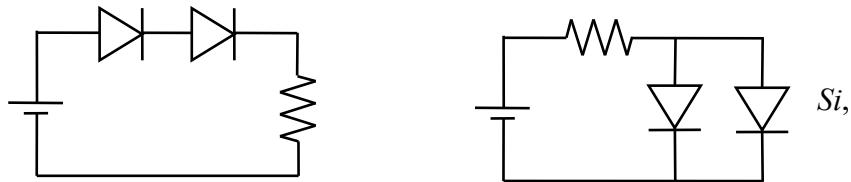


Figura 4.1: Diodos en serie y en paralelo.

En el caso de diodos en serie, si ambos conducen en la misma dirección, sus caídas de potencial se suman, y si tienen direcciones de conducción opuestas, no fluirá corriente por ellos, excepto que se den las condiciones para que uno de ellos entre en la región de ruptura.

En caso de diodos en paralelo, el diodo con menor potencial de contacto se activará, fijando la tensión del diodo que se encuentra en paralelo e inactivándolo.

Para diodos en antiparalelo (direcciones de conducción opuestas), es claro que sólo se activará el diodo que esté polarizado en directa.

### 4.2. Rectificadores

#### 4.2.1. Rectificador de media onda

El rectificador de media onda es un circuito que permite obtener una tensión positiva a partir de una tensión de corriente alterna (AC). Esto se logra eliminando la componente negativa de la señal.

Considere el circuito de la figura 4.2(a) y la tensión de entrada de la figura 4.2(b):

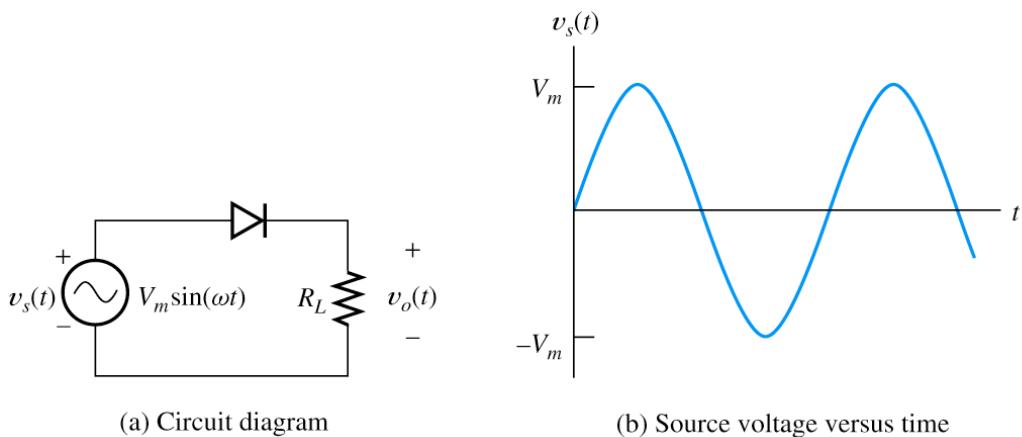


Figura 4.2: Rectificador de media onda: (a) circuito y (b) tensión de entrada.

La tensión de salida del rectificador de media onda se observa a continuación:

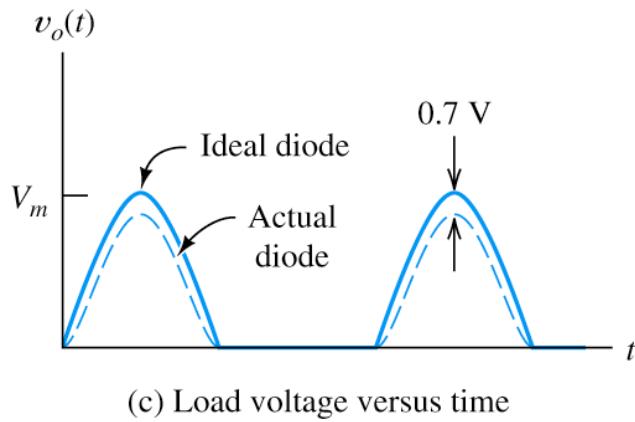


Figura 4.3: Tensión de salida para el caso de un diodo ideal y para un diodo real.

Definición de valor medio:

$$V_m = \frac{1}{T} \int_0^T f(t) dt \quad (4.1)$$

Caso ideal, en términos del ángulo:

$$V_m = \frac{1}{2\pi} \int_0^{2\pi} V_p \sin \theta d\theta \quad (4.2)$$

Caso ideal, en términos del tiempo:

$$V_m = \frac{1}{T} \int_0^T V_p \sin \omega t dt \quad (4.3)$$

pero como sólo hay media onda se integra hasta  $T/2$ :

$$V_m = \frac{1}{T} \int_0^{T/2} V_p \sin \omega t dt \quad (4.4)$$

donde  $\omega = 2\pi f = 2\pi/T$ :

$$V_m = \frac{V_p}{T} \int_0^{T/2} \sin \left( \frac{2\pi}{T} t \right) dt \quad (4.5)$$

$$V_m = \frac{V_p}{T} \cdot \frac{T}{2\pi} \cdot -\cos \left( \frac{2\pi}{T} \cdot t \right) \Big|_0^{T/2} \quad (4.6)$$

$$V_m = \frac{V_p}{2\pi} \cdot - \left( \cos \frac{2\pi T}{T} \frac{1}{2} - \cos 0^\circ \right) \quad (4.7)$$

$$V_m = -\frac{V_p}{2\pi} (-1 - 1) \quad (4.8)$$

$$V_m = \frac{V_p}{\pi} \approx 0.318 V_p \quad (4.9)$$

Como se observa en la ecuación anterior, el valor medio es independiente de la frecuencia. Esta ecuación es válida solamente cuando  $V_D \ll V_p$ . En caso contrario, se debe evaluar la integral considerando las tensiones del diodo.

¿Qué pasa si no se tiene el dato del período de la señal, pero se desea calcular el voltaje promedio de manera exacta? En ese caso, debe integrarse en términos de ángulo. El valor del ángulo se determina por evaluación de la función de la tensión de entrada:

### 4.2.2. Rectificador de onda completa

El rectificador de onda completa permite aprovechar el periodo completo de la señal de entrada. El circuito de este rectificador se muestra en la figura 4.4:

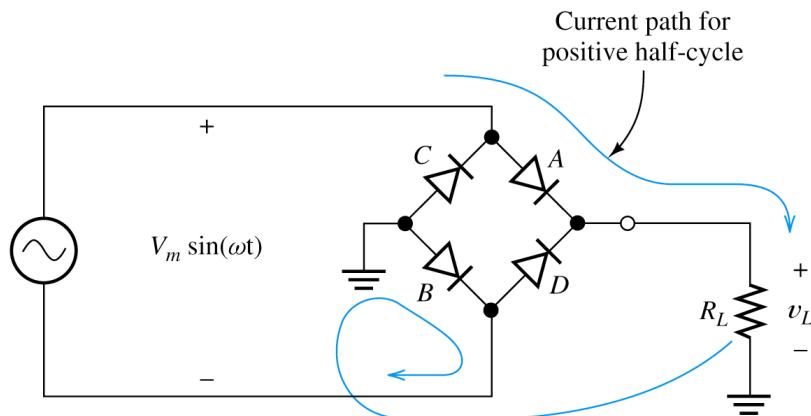


Figura 4.4: Rectificador de onda completa.

En el semiciclo positivo conducen los diodos  $D_A$  y  $D_B$ .

En el semiciclo negativo conducen los diodos  $D_C$  y  $D_D$ .

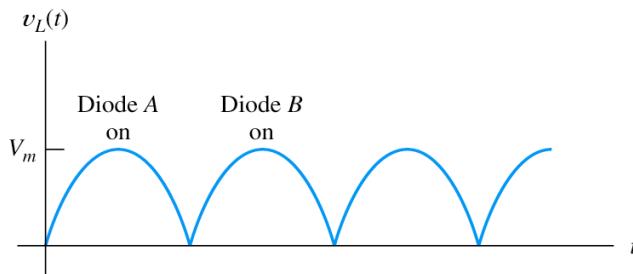


Figura 4.5: Tensión de salida para el rectificador de onda completa.

Nótese que la polaridad de la tensión en la carga no cambia, a pesar de la variación de polaridad de la fuente, es decir, se rectifica la señal.

¿Cuánto es el valor medio de la tensión de salida? Dos veces el área en el caso de media onda para el mismo período:

$$V_m = 2V_{\text{mediaonda}} \quad (4.10)$$

$$V_m = 2 \cdot 0.318V_p = 0.636V_p \quad (4.11)$$

Véase por medio de la integral:

$$V_m = \frac{1}{T'} \int_0^{T'} V_p \sin \omega t dt \quad (4.12)$$

Donde

$$T' = \frac{T}{2} \quad (4.13)$$

Por lo tanto:

$$V_m = \frac{1}{T/2} \int_0^{T/2} V_p \sin \omega t dt \quad (4.14)$$

$$V_m = \frac{V_p}{T/2} \int_0^{T/2} V_p \sin \left( \frac{2\pi}{T} t \right) dt \quad (4.15)$$

$$V_m = \frac{V_p}{T/2} \cdot \frac{T}{2\pi} \cdot -\cos \left( \frac{2\pi}{T} t \right) \Big|_0^{T/2} \quad (4.16)$$

$$V_m = \frac{-V_p}{\pi} \left( \cos \left( \frac{2\pi}{T} \cdot \frac{T}{2} \right) - \cos 0 \right) \quad (4.17)$$

$$V_m = \frac{-V_p(-1 - 1)}{\pi} = \frac{2V_p}{\pi} \quad (4.18)$$

Igual a dos veces el valor del caso de media onda e independiente de la frecuencia. Nótese que las fórmulas de valor medio de los rectificadores de media onda y onda completa se derivan suponiendo diodos ideales.

### 4.2.3. Recortador serie

Diodo en serie con la salida. Es en realidad un rectificador de media onda que puede tener una fuente de tensión de CC en serie.

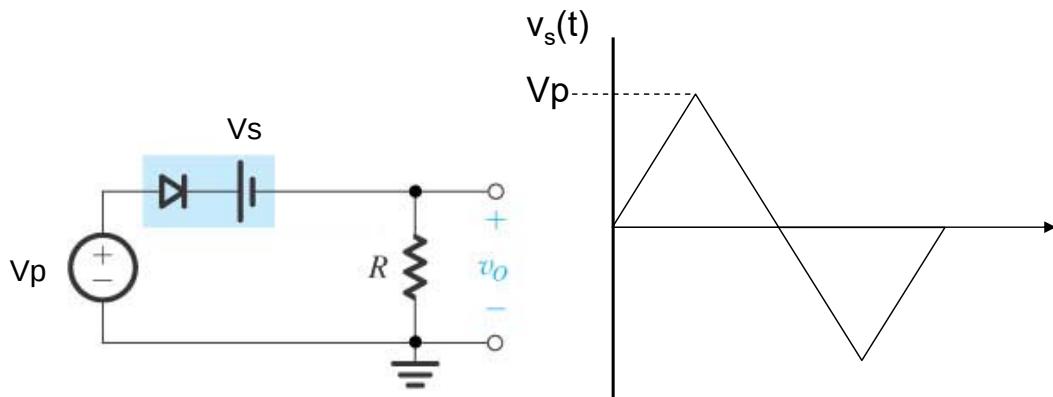
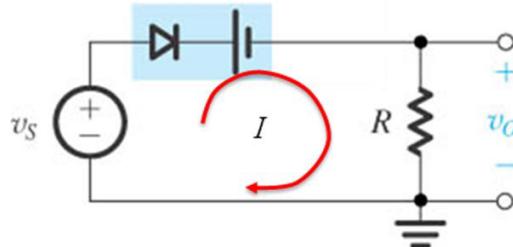
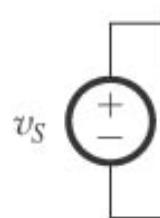


Figura 4.6: Recortador serie: (a) circuito y (b) tensión de entrada.

1. Si el diodo conduce, ¿Qué dirección tendría la corriente?



2. ¿Qué fuente originaría una corriente en esa dirección?



3. ¿En qué condiciones de  $V_s$  hay una polarización adecuada para que el diodo conduzca?

$$V_s > V_I + V_D \quad (4.19)$$

4. ¿Qué pasa en la salida cuando el diodo conduce y cuando no conduce?

No conduce  $\Rightarrow I = 0, V_O = 0$   
 Conduce  $\Rightarrow V_O = V_S - V_I - V_D$

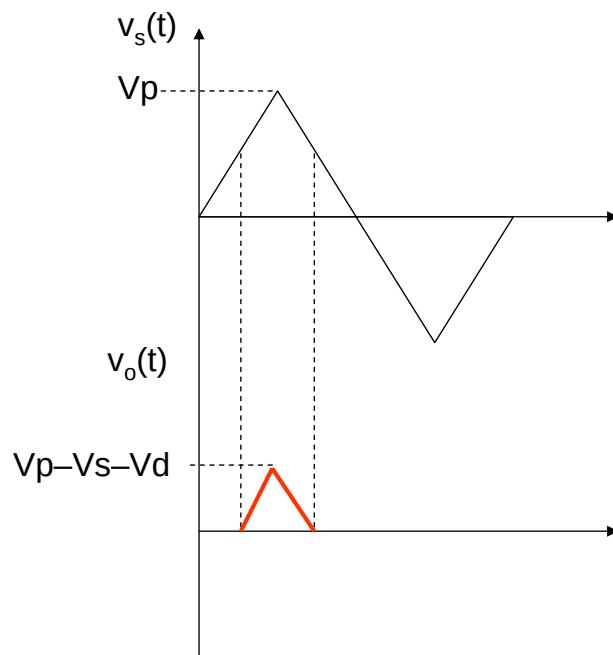


Figura 4.7: Tensión de salida del recortador serie.

#### 4.2.4. Recortador paralelo

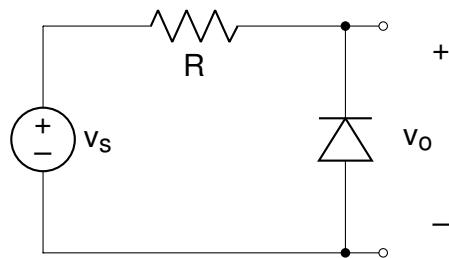


Figura 4.8: Recortador paralelo.

Diodo en paralelo (o en la rama en paralelo) con la salida.

Si el diodo conduce,  $V_O = -0.7V$

Si el diodo no conduce,  $V_O = V_S$

Diodo conduce en el semiciclo negativo

$V_S < 0.7V$  diodo activo

$V_S > 0.7V$  diodo inactivo

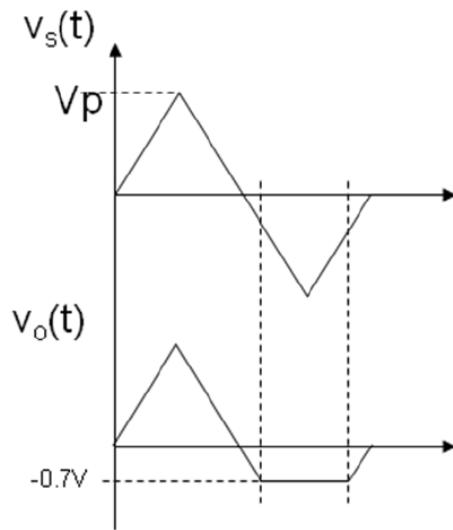


Figura 4.9: Tensión de salida del recortador paralelo.

Puede agregarse fuentes en serie (recortador polarizado).

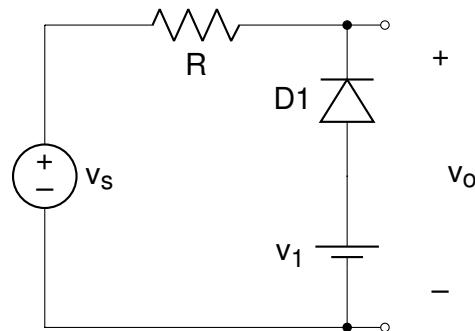


Figura 4.10: Recortador paralelo con fuente.

En el circuito de la figura 4.10, el diodo conduce en el semiciclo negativo. En el semiciclo positivo, conduce si  $V_S$  puede mantenerlo polarizado en directa. El resto de la tensión cae en la resistencia  $R$ .

Conduce para  $V_S < V_1$

No conduce para  $V_S > V_1$

El diodo tiene exactamente la tensión para operar y  $I = 0 \Rightarrow$

$$-V_S + V_R - V_D + V_1 = 0 \quad (4.20)$$

$$V_S = V_R - V_D + V_1 \quad (4.21)$$

Dado que con  $I = 0$ , la tensión en la resistencia es cero:

$$V_S = V_1 - V_D \quad (4.22)$$

Si  $V_S$  es mayor que  $V_1$  el diodo queda polarizado inversamente.

$\Rightarrow$  Conduce si  $V_S < V_1 - V_D$

$$\Rightarrow -V_D - V_1 - V_O = 0$$

$$V_O = V_1 - V_D$$

No conduce si  $V_S \geq V_1 - V_D$

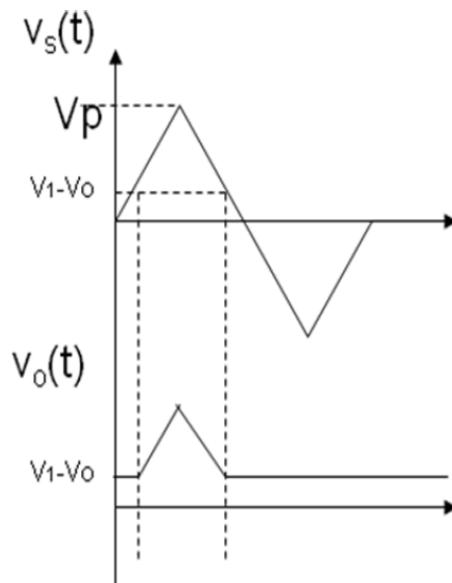


Figura 4.11: Tensión de salida del recortador paralelo con fuente.

$V_1$  debe contrarrestar  $V_S$  y proveer polarización para el diodo. Ahora, invirtiendo la polaridad de  $V_1$ , se tiene:

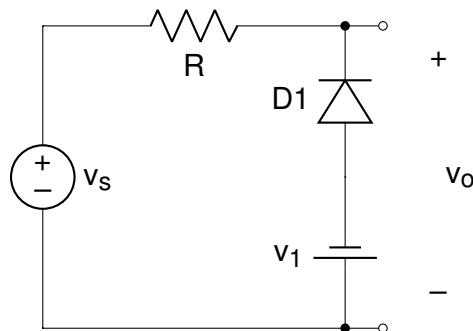


Figura 4.12: Recortador paralelo con fuente invertida.

$V_1$  polariza el diodo inversamente.

$V_S > 0$  polariza el diodo inversamente

$V_S < 0$  puede polarizar el diodo directamente si contrarresta  $V_1$ .

$\Rightarrow V_S > 0 =$  diodo no conduce,  $V_O = V_S$

En el semiciclo negativo ( $V_S < 0$ ) debe revisarse cuándo  $V_S$  tiene una magnitud tal que hace conducir el diodo. Cuando eso ocurre, la magnitud de la salida es  $V_1 + V_D$ , y la polaridad es negativa.

Conduce para  $V_S$  negativo

$$|V_S| > V_D + V_1$$

$V_S$  debe contrarrestar  $V_1$  y polarizar el diodo

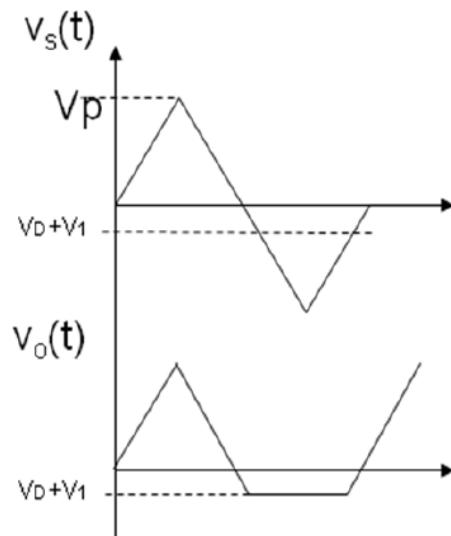


Figura 4.13: Tensión de salida del recortador paralelo con fuente invertida.

Ahora se analizará el caso para un recortador paralelo con dos ramas.

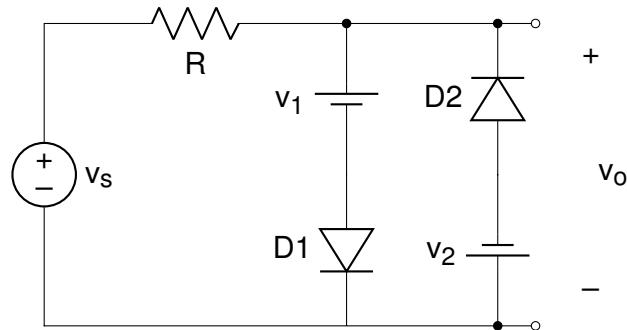


Figura 4.14: Recortador paralelo con dos ramas.

$D_1$  conduce en el semiciclo positivo.

$$V_S > V_D + V_1$$

$D_2$  conduce en el semiciclo negativo.

$$V_S < V_D + V_2$$

$$\Rightarrow |V_S| > |V_2 + V_O|$$

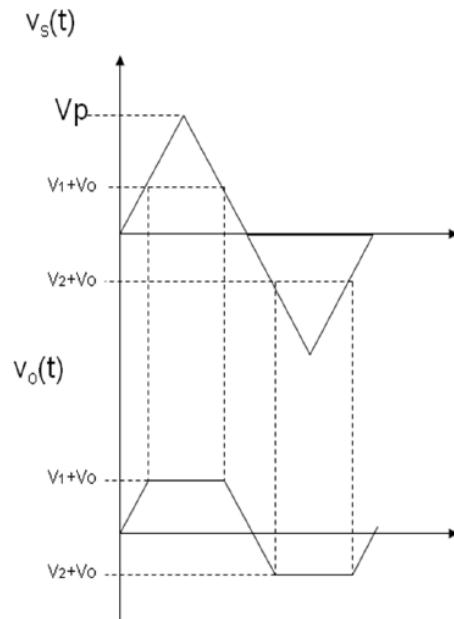


Figura 4.15: Tensión de salida del recortador paralelo con dos ramas.

#### 4.2.5. Cambiador de nivel (sujetador)

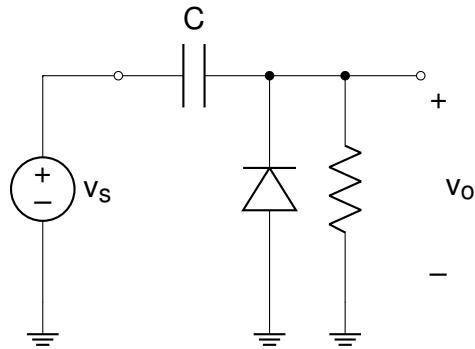


Figura 4.16: Cambiador de nivel.

$\Rightarrow V_S > 0$  polariza el diodo en directa

$r_d \ll R$ ,  $\tau = r_d C$  Carga rápida del diodo

C se carga hasta  $V_S - V_D$ ,  $V_O = V_D$

$\Rightarrow V_S < 0$  polariza el diodo en reversa

C se descarga a través de R,  $\tau = RC$

$$V_O = -V_S - (V_C - V_D) \approx -2V_S \quad (4.23)$$

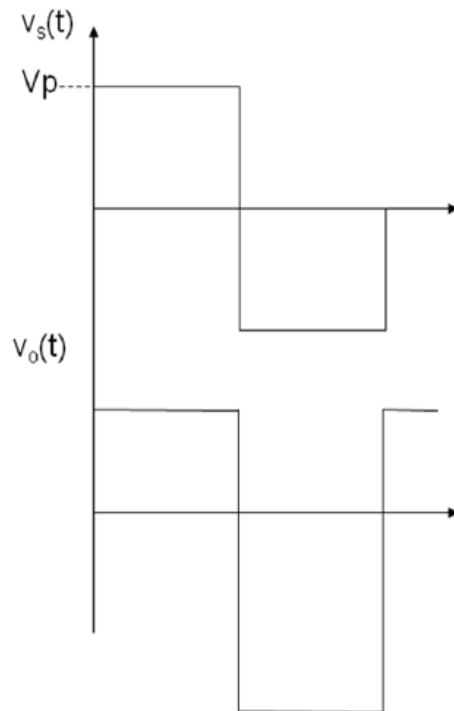


Figura 4.17: Tensión de salida del cambiador de nivel.

R y C deben escogerse de forma tal que:

$$RC >> T/2 \quad (4.24)$$

#### 4.2.6. Doblador de voltaje

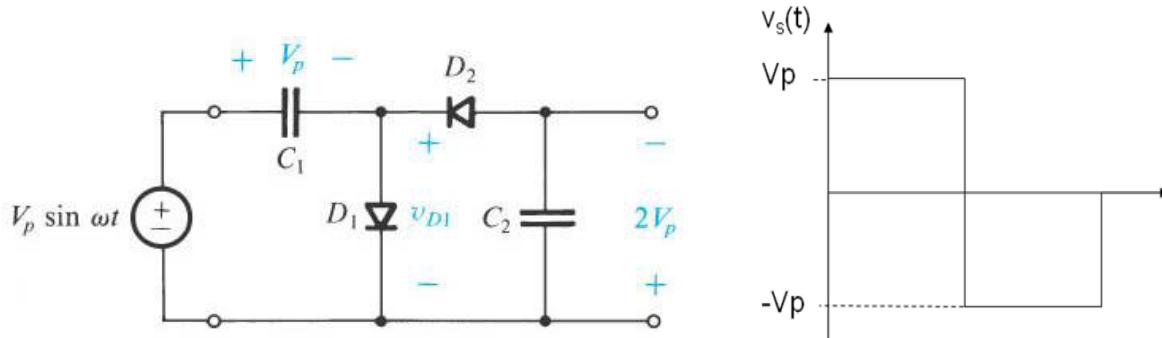


Figura 4.18: Doblador de tensión: (a) circuito y (b) tensión de entrada.

$\Rightarrow V_S > 0$ :  $D_1$  conduce,  $D_2$  en abierto,  $C_1$  se carga,  $V_O = V_{C2}$

$$\begin{array}{c} V_{IN}-V_D \\ \parallel \\ C_1 \end{array}$$

$\Rightarrow V_S < 0$ :  $D_1$  no conduce,  $D_2$  conduce

$$V_O \approx 2V_{IN} = V_{IN} + V_{C1} \quad (4.25)$$

$$V_O = V_{IN} + (V_{IN} - V_D) \quad (4.26)$$

$C_2$  se carga por medio de  $D_2$ ,  $\tau_2 = r_{d2}C_2$

$C_1$  se carga por medio de  $D_2$ ,  $\tau_1 = r_{d1}C_1$

Igualmente,  $\tau_1$  y  $\tau_2 \gg T/2$

Semiciclo positivo  $\rightarrow$  para proveer carga a  $C_1$ ,  $C_2$  se descarga a través de  $R_L$ .

Semiciclo negativo  $\rightarrow$  recarga de  $C_2$ .

#### 4.2.7. Rectificador de media onda con filtro

Es un rectificador de media onda con un capacitor en paralelo con la salida, con el fin de disminuir la variación de la tensión de salida (rizo) y proveer de energía a la carga durante el semiciclo en que el diodo no conduce.

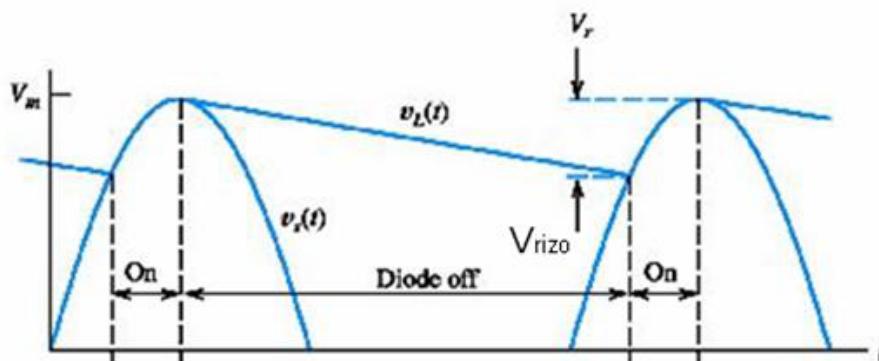


Figura 4.19: Tensión de salida del rectificador de media onda con filtro.

¿De qué valor es el capacitor?

Asumir diodo ideal

Rizo: variación de  $V_O$

$$i_C = \frac{CdV}{dt} \quad (4.27)$$

Si  $i_C$  es constante,  $i_R$  es constante y  $V_{OUT}$  es constante:

$$i_L = \frac{CdV}{dt} \quad (4.28)$$

$$\Rightarrow \frac{i_L \Delta t}{\Delta V} \approx C \quad (4.29)$$

Donde  $\Delta t \approx T$ ,  $\Delta V = V_{rizo}$

$$C \approx \frac{i_L T}{V_{rizo}} \quad (4.30)$$

#### 4.2.8. Rectificador de onda completa con filtro

Es un rectificador de onda completa con un capacitor en paralelo con la salida, con el fin de disminuir la variación de la tensión de salida (rizo).

$$\frac{i_L \cdot T/2}{V_{\text{rizo}}} \approx C \quad (4.31)$$

T = período de onda sin rectificar

#### 4.2.9. Análisis gráfico del regulador con diodo Zener

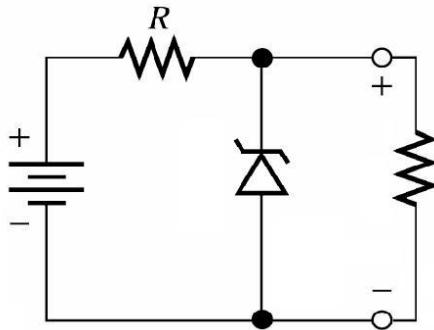


Figura 4.20: Regulador con diodo Zener.

R: resistencia de limitación de corriente

Corriente de operación del diodo Zener:  $I_{Z\min} < I_Z < I_{Z\max}$ , con  $I_{Z\min} = 0.1I_{Z\max}$ ,  $I_{Z\max}$  determinada por máxima potencia del diodo:

$$I_{Z\max} = \frac{P_{Z\max}}{V_Z} \quad (4.32)$$

En este regulador puede variar la tensión de entrada y la resistencia de carga. A pesar de eso, la tensión de salida debe mantenerse constante.

Siguiendo el método de la línea de carga, dividimos el circuito en una red de entrada y una red de salida y obtendremos la curva característica de cada red.

La red de entrada es una fuente de tensión real:

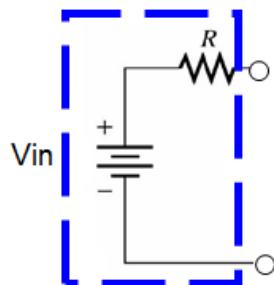


Figura 4.21: Red de entrada del regulador Zener.

La curva característica de la red de entrada se define con dos puntos: la corriente de Norton y la tensión de Thevenin. Esta curva característica resultante es:

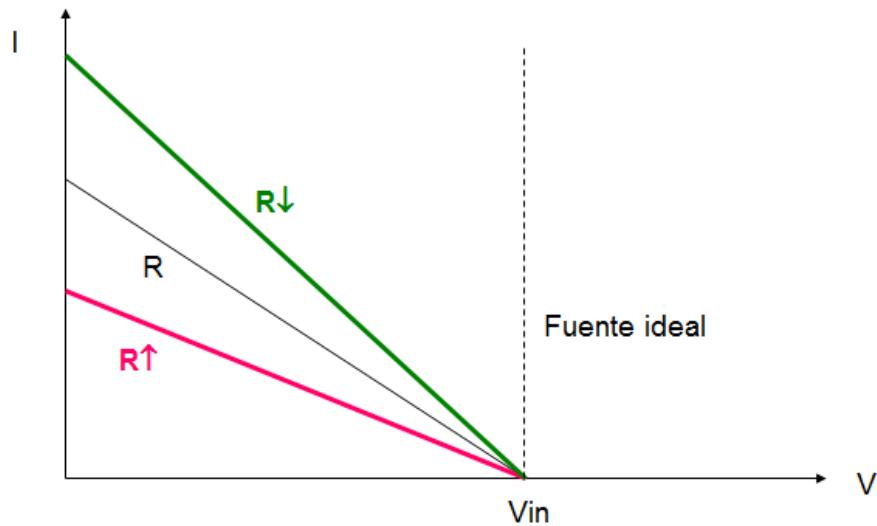


Figura 4.22: Curva característica de la red de entrada del regulador Zener.

La red de salida incluye el diodo Zener y la resistencia de carga:

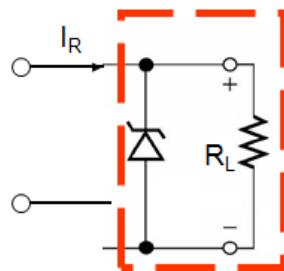


Figura 4.23: Red de salida del regulador Zener.

La curva característica de la red de salida se obtiene al tomar en cuenta lo siguiente:

- con  $V < V_Z$ , el diodo Zener no está activo, sólo se ve el comportamiento de R.
- con  $V = V_Z$ , el punto de intersección de la curva característica de la resistencia con la curva característica del diodo Zener indica la corriente en la resistencia cuando su tensión es  $V_{RL} = V_Z$

La corriente que entra al nodo de la red de salida, es decir, la corriente provista por la fuente, se distribuye entre el diodo Zener y la carga.

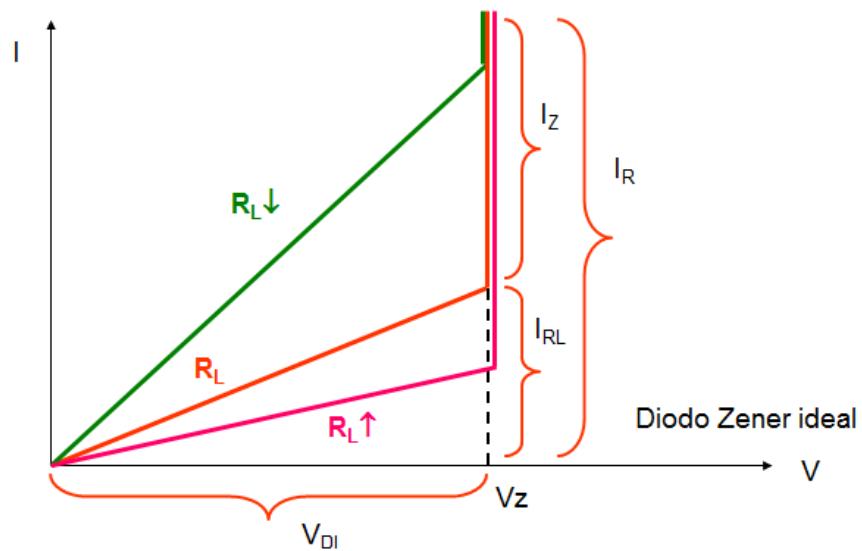


Figura 4.24: Curva característica de la red de entrada del regulador Zener.

Para encontrar el punto de operación del circuito completo, debe encontrarse la intersección de las dos curvas resultantes de las redes de entrada y salida.

Para operar, el regulador debe cumplir los siguientes criterios:

1. La tensión de la fuente debe ser suficiente para polarizar el Zener, y el divisor de tensión entre la resistencia de carga y la resistencia limitadora debe dar como resultado una tensión mayor o igual a la tensión Zener.
2. La resistencia limitadora debe proteger al diodo Zener, de forma que cuando la resistencia de carga aumenta (y por lo tanto  $I_{RL}$  disminuye), la corriente que fluye por el diodo Zener no debe sobrepasar  $I_{Zmax}$ .
3. El diodo Zener requiere una corriente mínima  $I_{Zmin}$  para operar en ruptura.

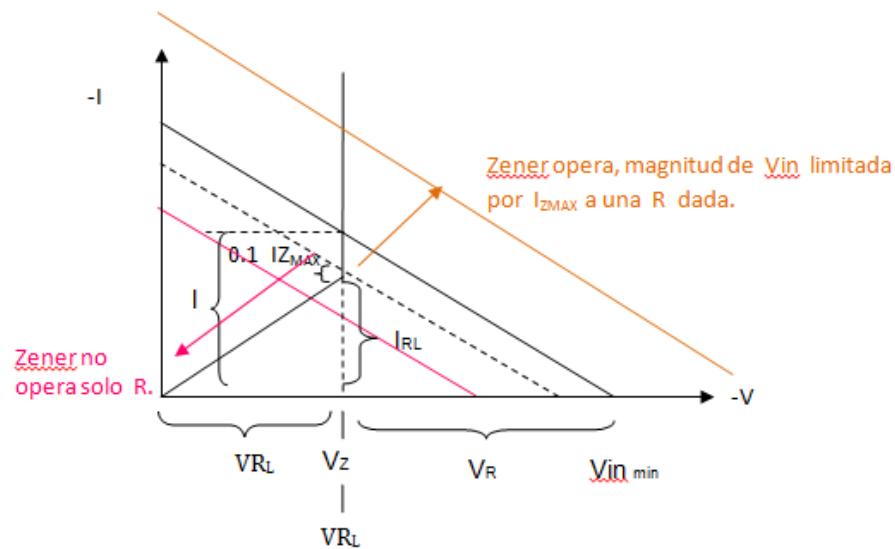


Figura 4.25: Curva característica de la red de entrada del regulador Zener.

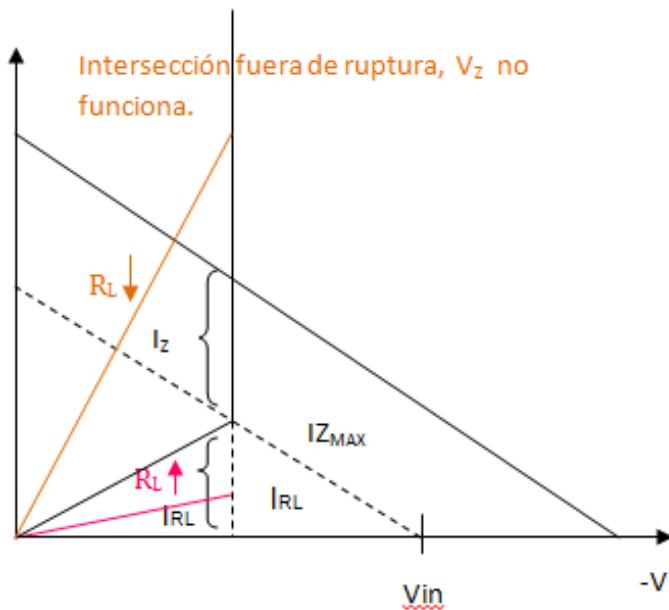


Figura 4.26: Curva característica de la red de entrada del regulador Zener.

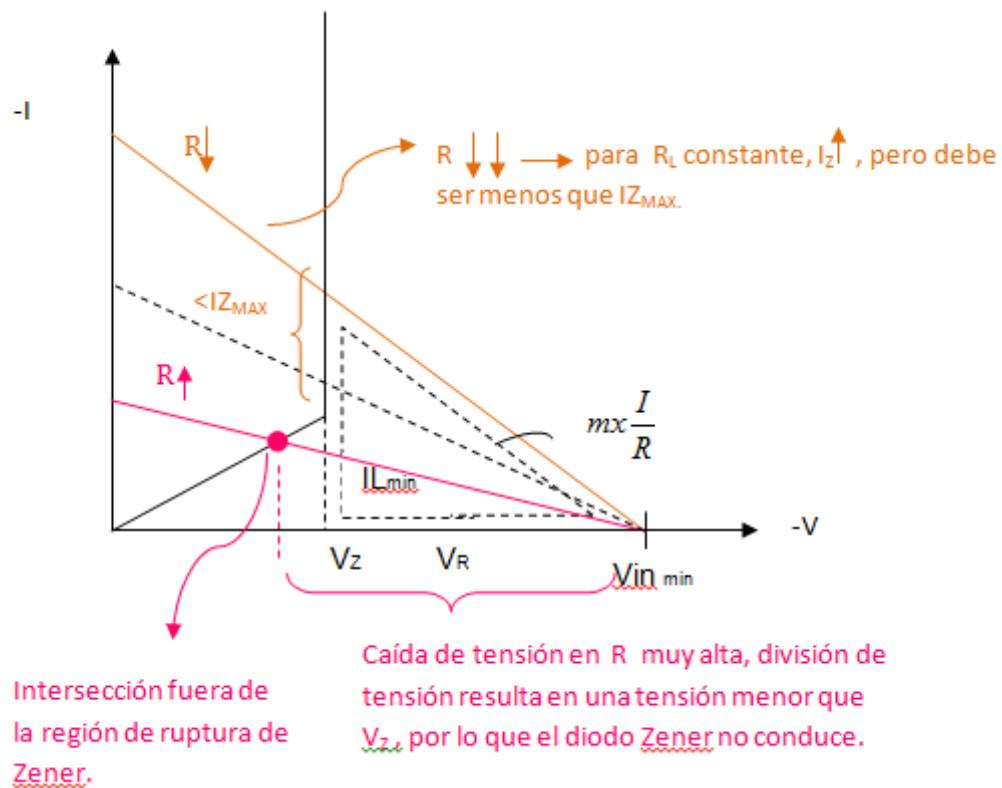


Figura 4.27: Curva característica de la red de entrada del regulador Zener.



# Capítulo 5

## El transistor BJT

El transistor de unión bipolar (BJT) consiste en tres regiones dopadas que forman capas. En la Fig. 5.1(a) se muestra un ejemplo que comprende una capa de p intercalada entre dos regiones n y se llama un transistor "NPN". Las tres terminales se denominan: "base", "emisor", y "colector". El emisor "emite" portadores de carga y el colector "recoge" mientras la base controla el número de portadores que participan en el proceso. El símbolo eléctrico del transistor npn se muestra en la Fig. 5.1(b). Se indican las tensiones en los terminales  $V_E$ ,  $V_B$ , y  $V_C$ , y las diferencias de voltaje por  $V_{BE}$ ,  $V_{CB}$ , y  $V_{CE}$ . El transistor se etiquetó con el nombre  $Q_1$ .

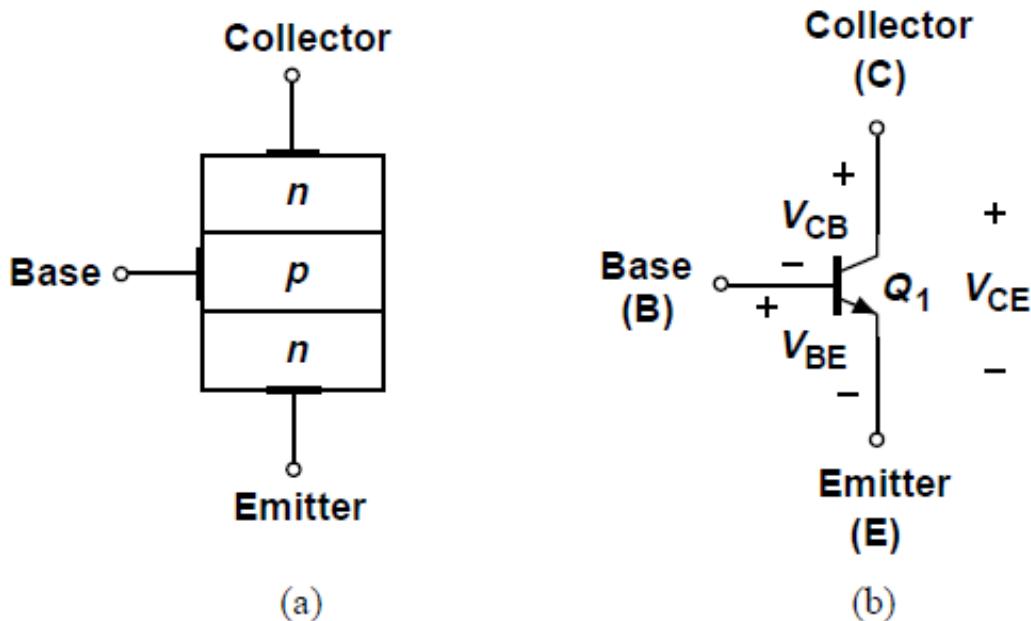


Figura 5.1: (a) Estructura y (b) símbolo eléctrico del transistor bipolar.

Se observa a partir de la Fig. 5.1(a) que el dispositivo contiene dos diodos de unión pn: uno entre la base y el emisor y el otro entre la base y el colector. Por ejemplo, si la base es más positiva que el emisor,  $V_{BE} > 0$ , entonces esta unión es de polarización directa. Aunque este diagrama simple puede sugerir que el dispositivo es simétrico con respecto al emisor y el colector, en la realidad, las dimensiones y los niveles de dopaje de estas dos regiones son muy diferentes. En otras palabras, E y C no pueden ser intercambiados. También vamos a ver que el funcionamiento correcto requiere

una región fina base, por ejemplo, aproximadamente 100 Å en transistores bipolares integrados modernos. Para el dispositivo de la Fig. 1 (a),  $V_{BE}$ ,  $V_{BC}$ , y  $V_{CE}$  se puede asumir valores positivos o negativos, lo que lleva a 23 posibilidades para los voltajes terminales del transistor. Afortunadamente, sólo uno de estos ocho combinaciones encuentra valor práctico.

Considere la unión polarizada inversamente de la Fig. 2. En la región de agotamiento se produce un campo eléctrico debido a la ionización. Supongamos ahora que un electrón es de alguna manera "inyectado" desde el exterior en esa región. Sirviendo como un portador minoritario en el lado p, el electrón experimenta el campo eléctrico y es barrido rápidamente desde la zona de agotamiento hacia el lado n. La capacidad de una unión pn polarizada inversamente para recoger eficientemente electrones injectados externamente resulta imprescindible para el funcionamiento del transistor bipolar.

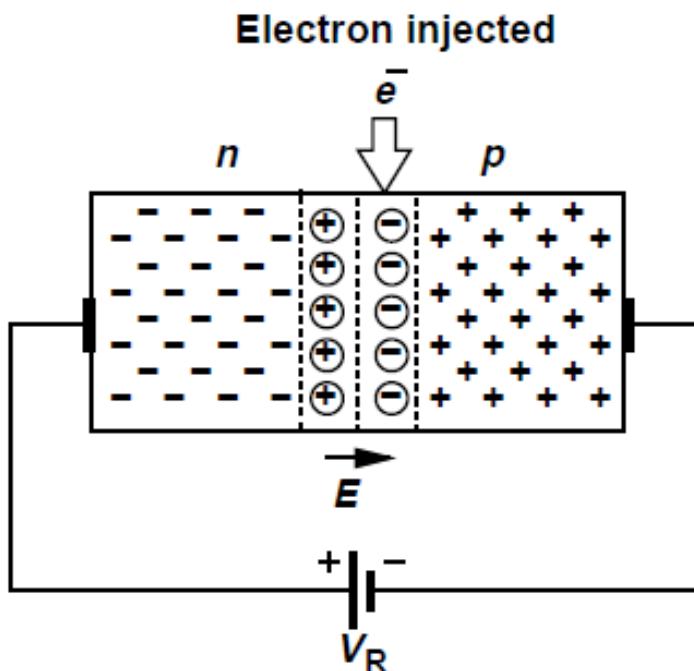


Figura 5.2: Inyección de electrones en la región de agotamiento.

## 5.1. Operación de Transistor Bipolar en Modo Activo

En determinadas condiciones el BJT actúa como una fuente de corriente controlada por voltaje. Se puede demostrar que (a) el flujo de corriente desde el emisor al colector puede ser visto como una fuente de corriente atada entre estos dos terminales, y (b) esta corriente es controlada por la diferencia de voltaje entre la base y el emisor  $V_{BE}$ . Para trabajar en modo activo, la unión base-emisor se polariza directamente ( $V_{BE} > 0$ ) y la unión base-colector es polarización inversa ( $V_{BC} < 0$ ). Por ejemplo, con el emisor conectado a tierra, la tensión de base se establece en alrededor de 0,8 V y la tensión de colector de un valor más alto, por ejemplo, 1 V [Fig. 3(a)]. Por tanto, la unión base-colector experimenta una polarización inversa de 0,2 V.

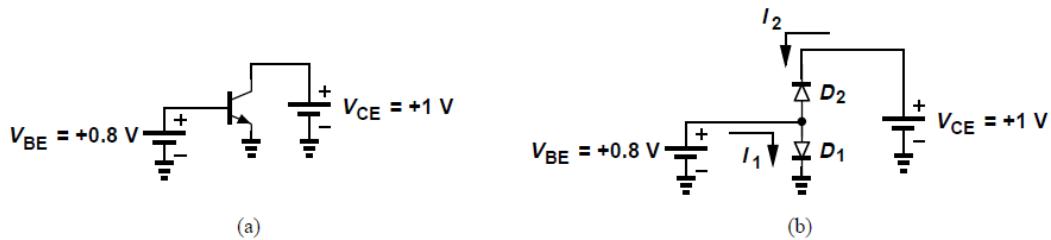


Figura 5.3: (a) transistor bipolar con tensiones de polarización de base y de colector, (b) visión simplista del transistor bipolar.

Consideremos ahora el funcionamiento del transistor en el modo activo. Podemos tener la tentación de simplificar el ejemplo de la Fig. 3 (a) para el circuito equivalente se muestra en la Fig. 3 (b). Después de todo, parece que el transistor bipolar simplemente consiste en dos diodos que comparten sus ánodos en la terminal de la base. Este punto de vista implica que  $D_1$  lleva una corriente y  $D_2$  no lo hace; si esto fuera cierto, el transistor no funcionaría como una fuente de corriente controlada por voltaje y resultaría de poco valor.

Para entender por qué el transistor no puede ser modelado como simplemente dos diodos espalda con espalda, debemos examinar el flujo de carga en el interior del dispositivo, teniendo en cuenta que la zona tipo p es muy delgada. Como la unión base-emisor está polarizada directamente, los electrones fluyen desde el emisor a la base y los agujeros de la base al emisor. El nivel de dopaje de emisor es mucho mayor que el de la base, por lo tanto, denotamos la región de emisor con  $n^+$ , donde el superíndice hace indicar mayor nivel de dopaje. La Fig 4 (a) se resumen estas observaciones, lo que indica que el emisor inyecta un gran número de electrones en la base, mientras que la base inyecta un pequeño número de huecos en el emisor.

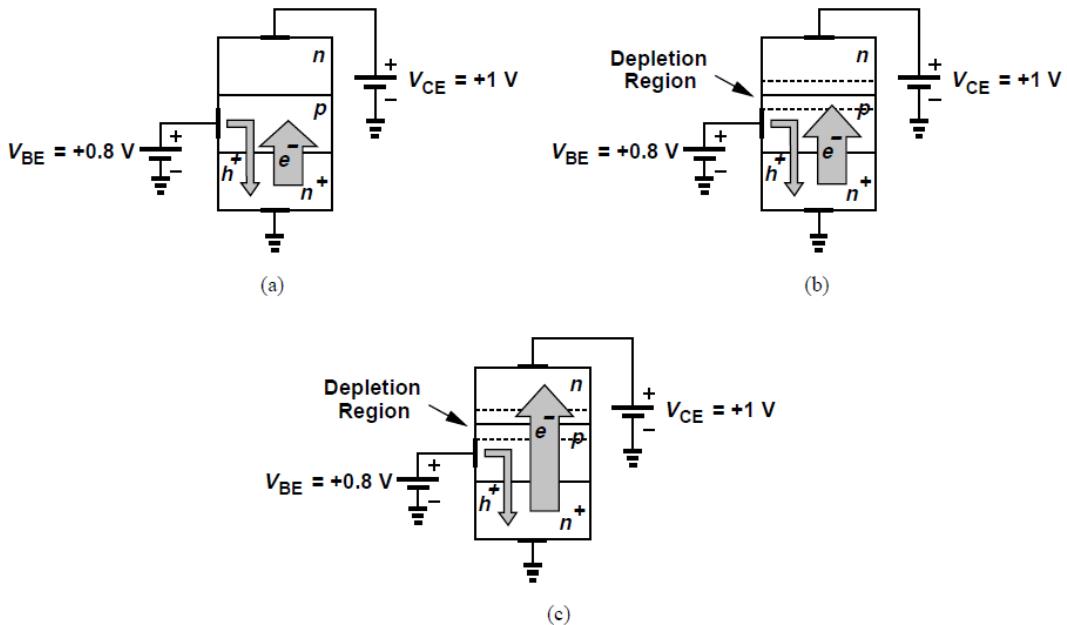


Figura 5.4: (a) flujo de electrones y agujeros a través de la unión base-emisor, (b) los electrones se acercan a la unión de base-colector, (c) los electrones pasan a través de la unión de base-colector.

Como la región de base es muy delgada, la mayoría de los electrones alcanzan el borde de la región de agotamiento colector-base, comenzando a experimentar la fuerza del campo eléctrico de esa región (Fig. 4 (b)). Como consecuencia, los electrones son barridos hacia la región del colector (Fig. 4.5 (c)) y absorbidos por el terminal positivo de la batería. Por lo tanto, observamos que la unión colector-base con polarización inversa lleva una corriente porque los portadores minoritarios ( $e^-$  acumulados en la región p de la base) se inyectan.<sup>en</sup> la región de agotamiento.

## 5.2. Corriente de Colector – Modelo de Ebers-Moll

En analogía con la ecuación de corriente del diodo, tenemos

$$I_C = I_S e^{V_{BE}/V_T} \quad (5.1)$$

Donde

$$I_S = \frac{A_E q D_n n_i^2}{N_B W_B} \quad (5.2)$$

$A_E$ : área de la unión base-emisor

$q$ : carga del portador

$D_n$ : Coeficiente de difusión de los portadores del emisor

$N_B$ : Dopado de la base

$W_B$ : espesor del semiconductor de la base

$V_T$ : voltaje térmico.

La ecuación 5.1 implica que el transistor bipolar funciona como una fuente de corriente controlada por voltaje, lo que demuestra un buen candidato para la amplificación. Podemos decir, alternativamente, el transistor realiza conversión "tensión-a-corriente".

## 5.3. Corrientes de base-emisor

Puesto que el transistor bipolar debe satisfacer la ley de Kirchoff, del cálculo de la corriente de base se obtiene fácilmente la corriente de emisor también.

En el transistor npn de la Fig. 5. (a), la corriente de base,  $I_B$ , es el resultado del flujo de huecos desde la terminal. Las corrientes de huecos y electrones en una unión pn de polarización tienen una relación constante determinada principalmente por los niveles de dopaje. Por lo tanto, el número de huecos que entran desde la base al emisor es una fracción constante de la cantidad de electrones que viajan desde el emisor a la base. A modo de ejemplo, por cada 200 electrones inyectados por el emisor, un hueco es suministrado por la base.

En la práctica, la corriente de base contiene un componente adicional de huecos. Como los electrones inyectados por el emisor de los viajes a través de la base, algunos pueden recombinar con los huecos que se inyectan a la base [Fig. 5. (b)]; en esencia, algunos electrones y agujeros se pierden como resultado de la recombinación. Por ejemplo, en promedio, de cada 200 electrones inyectados por el emisor, uno se recombinó con un agujero.

En resumen, la corriente de base debe proporcionar huecos tanto para inyección inversa en el emisor como para la recombinación con los electrones viajan hacia el colector. Por tanto, podemos

ver  $I_B$  como una fracción constante de  $I_E$  o una fracción constante del  $I_C$ . La corriente de colector se puede describir como

$$I_C = \beta I_B \quad (5.3)$$

donde  $\beta$  es la ganancia de corriente del transistor, ya que muestra la cantidad de la corriente de base es amplificada para lograr una corriente de colector.

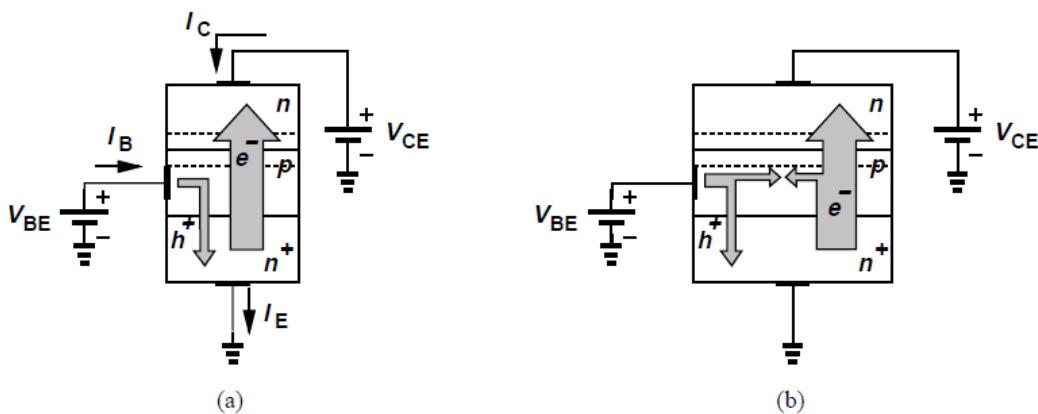
En transistores NPN, la ganancia de corriente  $\beta$  oscila entre 50-200.

Aplicando KCL, podemos calcular la corriente del emisor con las direcciones de las corrientes representadas en la Fig. 5. (a) como

$$I_E = I_C + I_B \quad (5.4)$$

$$I_C = \alpha I_E \quad (5.5)$$

$$\alpha = \frac{\beta}{\beta + 1} \quad (5.6)$$



## 5.4. Regiones de operación del transistor

A diferencia del transistor MOSFET, las conexiones del transistor BJT están muy bien definidas por las características constructivas, por lo que no se pueden intercambiar el emisor y el colector. Por esta razón, las posibles combinaciones de tensión en las terminales del BJT permiten diferencias cuatro regiones de operación, tal como los indica la Figura 5.6.

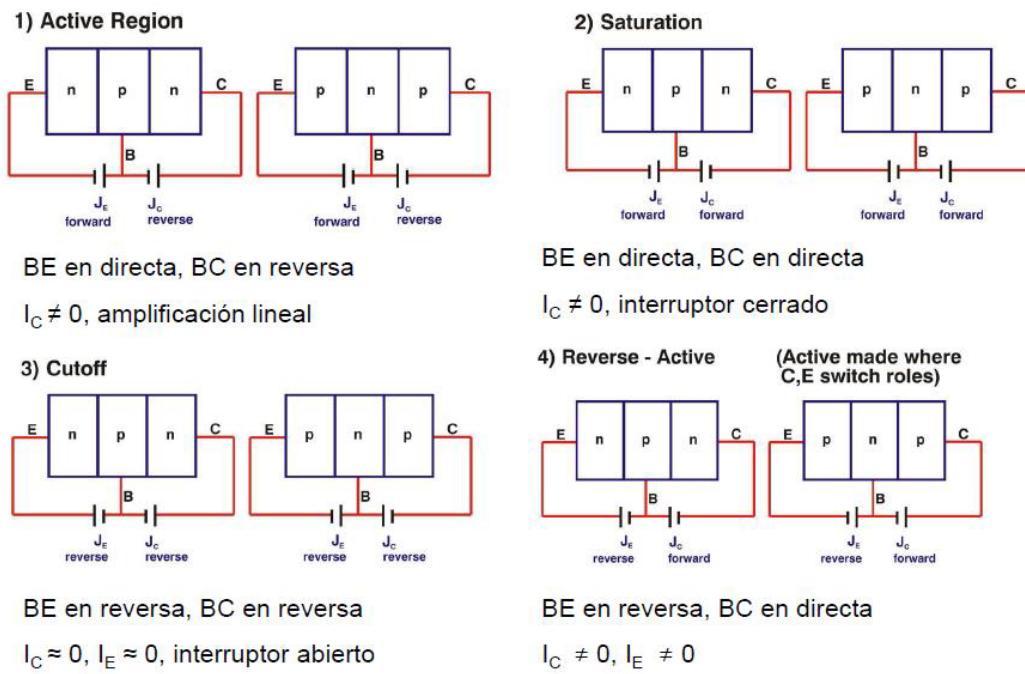


Figura 5.6: Regiones de operación del BJT (1) región activa, (2) región de saturación, (3) región de corte, (4) región de reversa activa.

En la región activa el diodo formado por la base y el emisor esta polarizado en directa y el diodo formado por la base y el colector esta polarizado en inversa, o expresado en términos de tensión: para un NPN  $V_{BE} > 0$  y  $V_{CE} \geq V_{BE}$ , para un PNP  $V_{EB} > 0$  y  $V_{EC} \geq V_{EB}$ .

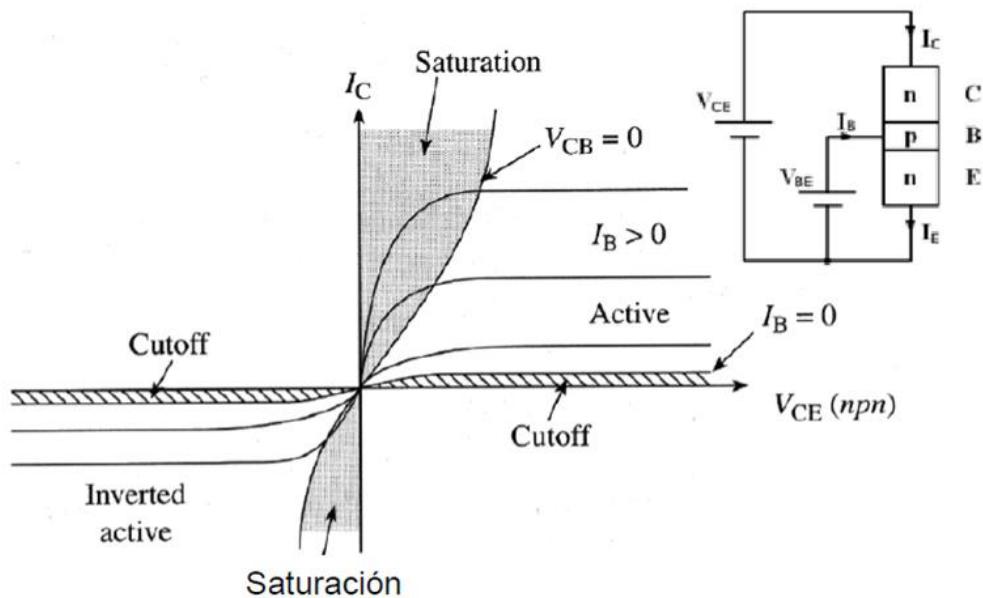
En la región de saturación el diodo formado por la base y el emisor esta polarizado en directa y el diodo formado por la base y el colector esta polarizado también en directa, o expresado en términos de tensión: para un NPN  $V_{BE} > 0$  y  $V_{CE} < V_{BE}$ , para un PNP  $V_{EB} > 0$  y  $V_{EC} < V_{EB}$ .

En la región de corte tanto el diodo formado por la base y el emisor como el diodo formado por la base y el colector están polarizados en inversa, o expresado en términos de tensión: para un NPN  $V_{BE} \leq 0$  y  $V_{CE} \geq V_{BE}$ , para un PNP  $V_{EB} \leq 0$  y  $V_{EC} \geq V_{EB}$ .

En la región de reversa activa el diodo formado por la base y el emisor esta polarizado en inversa y el diodo formado por la base y el colector estapolarizado en directa, o expresado en términos de tensión: para un NPN  $V_{BE} \leq 0$  y  $V_{CE} < V_{BE}$ , para un PNP  $V_{EB} \leq 0$  y  $V_{EC} < V_{EB}$ .

Nótese que en la región de reversa activa se intercambia la función del emisor y el colector, por lo que el funcionamiento del transistor no es óptimo, aún cuando se tiene una corriente entre el colector y el emisor.

Las curvas característica de las regiones de operación del BJT se detallan en la figura 5.7.



$$V_{CB}=0 \text{ implica que } V_{CE}=V_{BE}$$

Figura 5.7: Curvas características de las regiones de operación del BJT.

Cuando se polariza el transistor BJT de manera correcta, se puede definir una curva de transferencia y una curva característica de salida. Ambas curvas se observan en la figura 5.8.

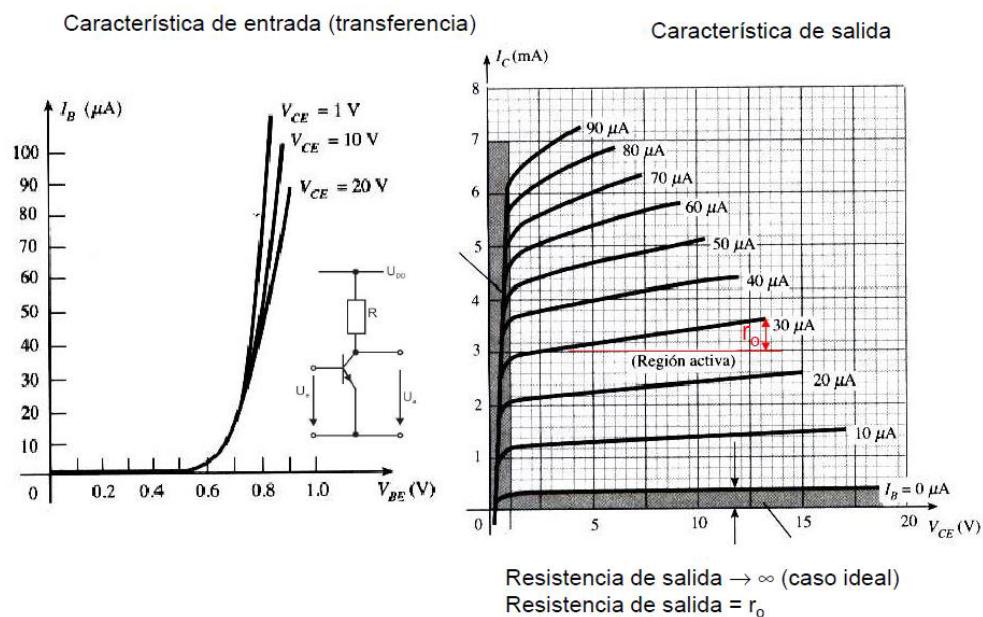


Figura 5.8: Curva de transferencia y curva característica del BJT.

## 5.5. Efecto Early en el transistor BJT

Para comprender este efecto, volvemos a analizar la operación interna del transistor. Considere la región de agotamiento en las áreas colector y la base. Inicialmente el exceso de carga de electrones injectados desde el emisor se ubica en la superficie de la unión entre la base y el colector, donde el campo eléctrico generado por el voltaje en el colector con respecto a la base  $V_{CB}$  (polarización inversa del diodo base colector) dispara los electrones desde la base hasta el colector, generando la corriente de colector.

Si el voltaje  $V_{CB}$  aumenta, entonces se aumenta el tamaño de la región de agotamiento, que se traslapa con la región del semiconductor de la base y de esta forma se reduce el ancho de la base, lo que disminuye la resistencia entre el colector y el emisor. Por lo tanto, al aumentar el voltaje  $V_{CE}$  se aumenta la corriente del colector  $I_C$  por efecto Early.

# Capítulo 6

## El transistor MOSFET

### 6.1. El concepto de transistor y transistor MOSFET

Un transistor se define como un dispositivo de al menos tres terminales en el cual una terminal controla el flujo de corriente entre las otras dos. La palabra transistor proviene de la combinación de los términos resistor (resistencia) y transfer (transferencia), siendo una resistencia controlada. El término Transfer indica que hay una variable de entrada y una variable de salida; una variable de control que se aplica a una terminal controlando así la corriente que fluye entre las otras 2.

El transistor MOSFET es un dispositivo electrónico semiconductor de cuatro terminales. El acrónimo MOSFET se deriva del nombre "*metal oxide semiconductor field effect transistor*". Este acrónimo se explica al descomponerlo en dos partes: MOS, por las siglas en inglés de metal-óxido-semiconductor y se refiere al sistema de materiales que al estar en contacto hace posible el efecto de campo, y FET es el acrónimo en inglés de field effect transistor. Como el nombre lo indica, el MOSFET es un dispositivo basado en el efecto de campo, es decir, se utiliza un campo eléctrico para controlar la cantidad de corriente que fluye entre dos terminales. En el caso del MOSFET, el transistor es una resistencia controlada por voltaje.

El MOSFET es un dispositivo unipolar; esto significa que el flujo de corriente se debe solamente a un tipo de portadores de carga, o más exactamente, está dominado por un tipo de portadores de carga. Contrariamente, en el caso bipolar, como en el transistor bipolar de junta, se necesita una corriente de electrones y una corriente de huecos para su operación.

El MOSFET es el dispositivo semiconductor más importante debido a que los dispositivos de efecto de campo son los que dominan la industria electrónica, representando un 80 % del mercado, por ello cualquier nueva tecnología será comparada con el MOSFET.

El principio de funcionamiento del MOSFET, llamado efecto de campo, fue patentado en 1925 por Julius Lillienfeld. El transistor MOSFET fue concebida en 1928 también por Lillienfeld; la patente le fue otorgada en 1930. Sin embargo, no fue sino hasta muchos años después que la tecnología de materiales logró producir un material aislante ( $\text{SiO}_2$ ) de la calidad necesaria para implementar el transistor MOSFET. Se logró fabricar una capa de dióxido de silicio suficientemente delgada para lograr el efecto de campo pero que a pesar de su reducido espesor se comportara como un buen aislante eléctrico y con poca contaminación de iones metálicos del proceso de fabricación, pues esta contaminación altera la tensión de activación del transistor ( $V_{TH}$ ).

La implementación del MOSFET ocasionó una revolución en la industria microelectrónica, hasta entonces dominada por el BJT (bipolar junction transistor). Esta revolución tomó especial fuerza con la introducción de la tecnología CMOS (complementary metal oxide semiconductor) en 1960,

así como con la introducción de la teoría de escalamiento de Robert H. Dennard en los años 60's, haciendo posible la integración a gran escala que no hubiera sido posible utilizando transistores BJT.

La tecnología CMOS se originó cuando se logró fabricar una tina N en un substrato P. Esto permitió fabricar transistores NMOS y PMOS en una sola oblea de silicio, y es la tecnología que utilizamos actualmente.

Debido al continuo escalamiento del MOSFET, para indicar el grado de avance de una tecnología se habla de generaciones tecnológicas, las cuales se diferencian por el concepto de dimensión crítica F (feature size). La dimensión crítica es la dimensión más pequeña que puede fabricarse con el silicio policristalino de compuerta en el plano de la oblea. Así, por ejemplo, en la generación de 14nm el valor de F es de 14nm.

## 6.2. Estructura del MOSFET

El MOSFET es un dispositivo de cuatro terminales: compuerta, drenador, surtidor (o fuente) y substrato. Estas terminales se representan con las letras G (gate), D (drain), S (Source) y B (bulk), respectivamente. La compuerta permite controlar, por medio del efecto de campo, la cantidad de corriente que fluye entre drenador y surtidor. La estructura del MOSFET se presenta en la figura 6.1.

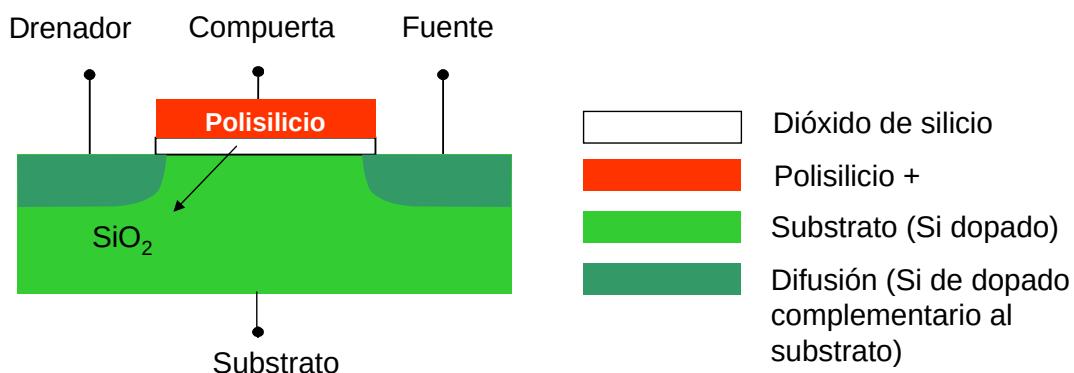


Figura 6.1: Sección transversal de un transistor MOSFET

El substrato en el cual se fabrica el MOSFET es una oblea de silicio con espesor de 100 a 800  $\mu\text{m}$ . Normalmente se utilizan obleas tipo P debido a que los transistores NMOS, que son los más rápidos, requieren un substrato P. En la figura 6.1 el substrato se representa con el color verde claro.

El MOSFET consiste en un substrato de silicio dopado, con dos regiones de dopado complementario a del substrato a cada lado. Estas regiones constituyen la fuente y el drenado del transistor y se muestran con color verde oscuro en la figura 6.1. El color verde significa que las regiones de difusión, al igual del substrato, son de silicio. El tono más oscuro indica que el dopado es mayor. Como la figura 6.1 sugiere, la fuente y el drenador son idénticos desde el punto de vista de su construcción. La definición de cuál de las regiones constituye la fuente y cuál el drenador se obtiene al aplicar una diferencia de potencial entre ambas terminales.

Sobre el substrato entre las regiones de difusión se coloca una capa delgada de  $\text{SiO}_2$ , representado en la figura 6.1 con la capa blanca sobre el substrato. El espesor de la capa de óxido es menor

que 10nm y disminuye con cada generación tecnológica. El espesor mínimo de dióxido de silicio es de 1.2nm. Desde hace unos 7 años se reemplazó el SiO<sub>2</sub> por HfO<sub>2</sub> en los transistores de última generación.

Sobre la capa de SiO<sub>2</sub> se encuentra una capa de silicio policristalino, formando la compuerta del transistor; esta se representa con la capa de color rojo en la figura 6.1. La compuerta tiene un espesor típico de 300nm. Aunque originalmente se concibió que la compuerta del MOSFET fuera de metal, en la práctica se utiliza silicio policristalino altamente dopado, de manera que sus propiedades se acerquen a las de un metal.

La capa de óxido (SiO<sub>2</sub>) aísla eléctricamente la superficie del silicio de la capa de silicio policristalino que constituye la compuerta, de forma que el substrato bajo la compuerta experimente el efecto del campo eléctrico sin que haya conducción de corriente entre compuerta y substrato. Esta capa de óxido se refiere comúnmente como "óxido de compuerta"(gate oxide), mientras que el término "polisilicio"(polysilicon), o simplemente "poli"(poly), es una abreviación comúnmente adoptada en la literatura para el término silicio policristalino.

La figura 6.2 muestra las dos posibles variaciones en el dopado complementario en la implementación del MOSFET.

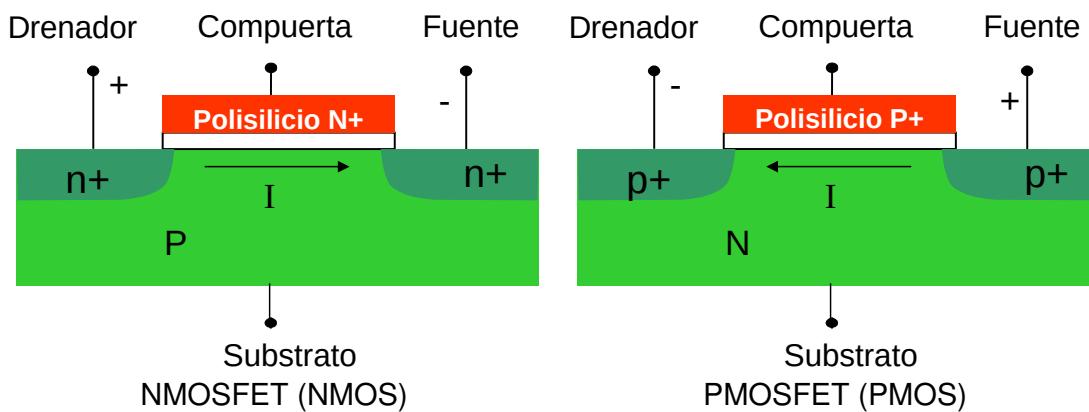


Figura 6.2: Sección transversal de transistores NMOS y PMOS

Como muestra la figura 6.2, el substrato puede tener dopado p o dopado n, mientras que las regiones de drenador y fuente están fuertemente dopadas (pero sin llegar a ser degeneradas) complementariamente al dopado del substrato. Estas dos variaciones dan origen a los dos tipos posibles de transistores MOSFET: NMOS y PMOS, donde el substrato está dopado respectivamente, con aceptores y donadores, mientras que al activar los transistores, el canal formado entre drenador y fuente es de tipo n en el NMOS y de tipo p en el PMOS, como se verá más adelante.

## 6.3. Principio de funcionamiento del MOSFET

El principio de funcionamiento del MOSFET se analizará para el caso del transistor NMOS. Este análisis podrá ser luego fácilmente extensivo al transistor PMOS.

### 6.3.1. Análisis cualitativo

Primeramente se examinará el transistor sin aplicar ninguna tensión en sus terminales. Como se muestra en la figura 6.3, existen uniones PN entre drenador y substrato y entre fuente y substrato. Estos diodos están conectados en serie por medio del substrato, pero dado que presentan direcciones de conducción contrarias, impiden el flujo de corriente entre drenador y surtidor.

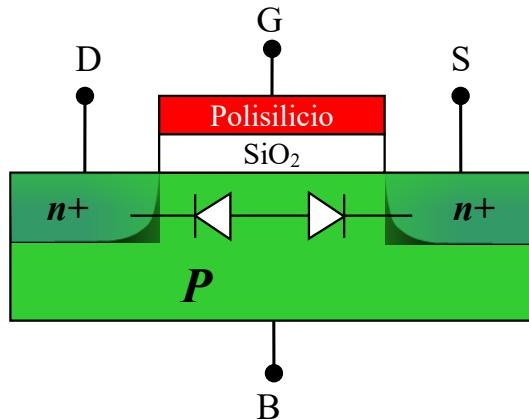


Figura 6.3: Diodos entre substrato-drenador y substrato-surtidor.

Al aplicar una tensión de compuerta positiva, y con  $V_{GS} > 0$ , la influencia del campo eléctrico de la compuerta repele a los huecos cercanos a la superficie del semiconductor, como se muestra en la figura 6.4.

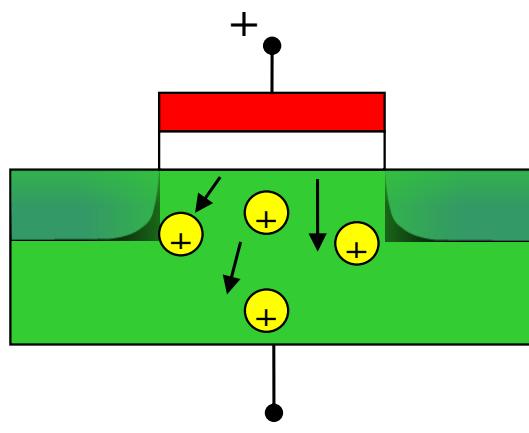


Figura 6.4: Principio de activación del MOSFET.

De esta manera, se agotan los huecos en la superficie del substrato en la región entre el drenador y la fuente. Este estado se conoce como agotamiento, debido al agotamiento de portadores de carga mayoritarios en la superficie del silicio entre drenador y fuente.

Conforme se continúa aumentando el voltaje de compuerta, electrones libres del substrato se ven atraídos a la superficie. Cuando el voltaje de compuerta ha alcanzado un valor conocido como el voltaje de umbral del transistor, la concentración de portadores minoritarios en la superficie del substrato iguala a la concentración de portadores mayoritarios con la que fue dopado el substrato.

Cuando esto ocurre, la superficie del substrato entre las regiones de difusión, que fue dopada de tipo p, se comporta como un material de tipo n, debido a la concentración de electrones en la superficie. Este fenómeno se conoce como inversión, puesto que ahora la superficie se comporta como un material de dopado opuesto al del substrato.

En estas condiciones, los electrones acumulados en la superficie del substrato entre drenador y fuente forman un canal entre drenador y fuente, permitiendo el flujo de corriente al aplicar un voltaje entre drenador y surtidor. El canal, es decir, la superficie en estado de inversión, se comporta entonces como una superficie del mismo tipo que las regiones de drenador y surtidor, por lo que ya no existe impedimento para el flujo de corriente entre ambas terminales, como se muestra en la figura 6.5. Una vez formado el canal, sólo se necesita una diferencia de tensión entre drenador y surtidor para que fluya corriente.

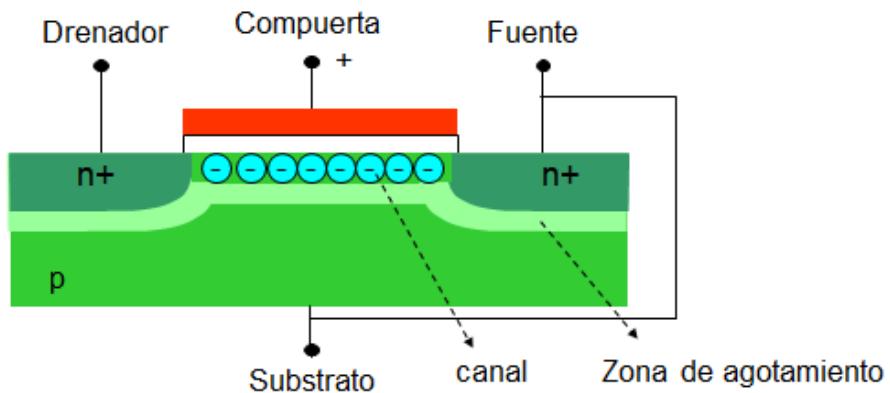


Figura 6.5: NMOS con el canal formado

Por otra parte, si el voltaje de compuerta aplicado es negativo, los huecos del substrato (portadores de carga mayoritarios) son atraídos a la superficie del substrato, aumentando la concentración de huecos en la superficie. Este fenómeno se conoce como acumulación. Puesto que no existe un canal entre drenador y fuente, la corriente entre fuente y drenador es prácticamente cero.

### 6.3.2. Análisis cuantitativo

Para analizar el comportamiento del MOSFET desde el punto de vista cuantitativo, debe considerarse primero el diagrama de bandas de energía del sistema metal-óxido-semiconductor (MOS), como se muestra en la figura 6.6.

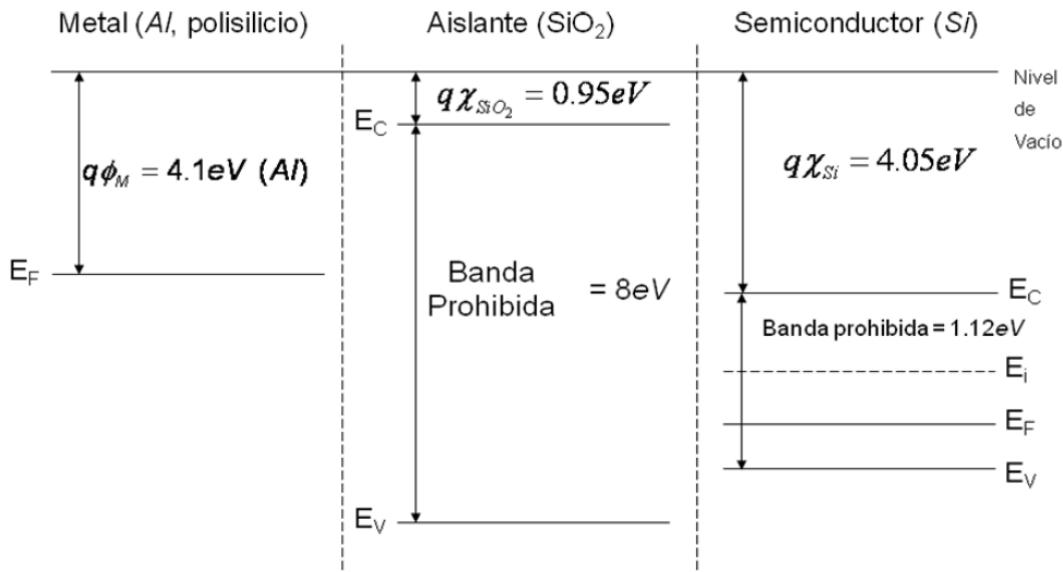


Figura 6.6: Diagrama de bandas de energía del sistema de materiales MOS antes del contacto.

Al poner en contacto el metal, el óxido y el semiconductor, los niveles de Fermi no se alinean debido a que el óxido impide el flujo de portadores de carga entre el metal y el substrato. En los MOSFET modernos, el metal de la compuerta ha sido reemplazado por silicio policristalino degenerado. En otras palabras, el silicio policristalino ha sido dopado tan fuertemente, que su comportamiento se aproxima al de un metal. Recuérdese aquí, que un semiconductor degenerado es aquel en el cual el nivel de dopado es tan alto que el nivel de Fermi a menos de  $3kT$  de la banda de conducción (para dopado n) o de la banda de valencia (para dopado p).

Debido al potencial de contacto de la banda MOS, es decir, la diferencia de funciones de trabajo entre el material de la compuerta (metal o silicio policristalino) y el substrato, el campo eléctrico resultante conlleva a una caída de tensión en el óxido y en la superficie del semiconductor, la cual se manifiesta en el diagrama de bandas como una deformación de bandas. El potencial intrínseco de Fermi en la cercanía de la interfaz substrato-óxido se conoce como potencial de superficie  $\varphi$ , y es la magnitud con la cual se describe el comportamiento del MOSFET.

Desde el punto de vista de  $V_{GS}$  y más exactamente del potencial de superficie  $\varphi$ , se distinguen 4 regiones de operación:

- Acumulación
- Banda plana
- Agotamiento
- Inversión, en la cual se distinguen
  - Inversión débil
  - Inversión fuerte

### Región de acumulación

Si se aplica una tensión  $V_{GS}$  en un transistor, tal que se atraen los portadores mayoritarios del substrato a la región donde se formaría el canal, se obtiene la condición conocida como acumulación. El canal no puede formarse y con ello el transistor está apagado. Esto se muestra en la figura 6.7.

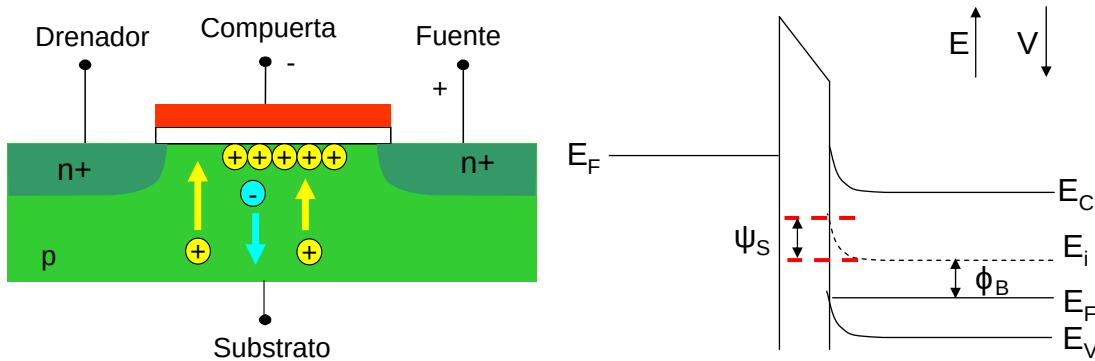


Figura 6.7: Diagrama de bandas de energía de un NMOS en acumulación

### Región de banda plana

Para compensar la caída de tensión debido a la diferencia de funciones de trabajo, debe aplicarse una cierta tensión de compuerta, conocida como tensión de banda plana  $V_{FB}$  (flat band voltage). Al aplicar este voltaje de compuerta, la deformación de bandas desaparece. Este estado se conoce como condición de banda plana (flat band condition) y se ilustra en la figura 6.8.

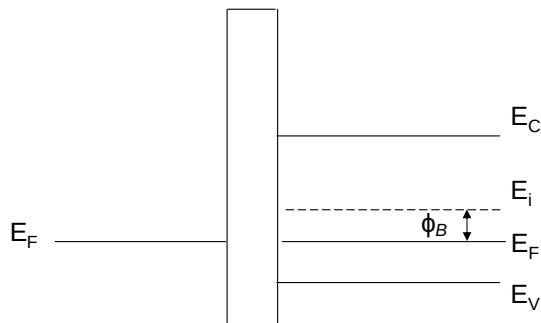


Figura 6.8: Diagrama de bandas de energía de un NMOS en banda plana

La caída de tensión también se debe, en parte, a cargas parásitas en el óxido. El voltaje de banda plana se calcula como:

$$V_{FB} = \phi_M - \phi_S - \frac{Q_{ox}}{C_{ox}'} \quad (6.1)$$

donde  $\phi_M$  es la función de trabajo del material de la compuerta,  $\phi_S$  es la función de trabajo del substrato,  $Q_{ox}$  es la densidad de carga superficial de cargas parásitas en el óxido (en general, en el aislante) y  $C_{ox}'$  es la capacitancia de compuerta por unidad de área ( $F/m^2$ ).

## Región de agotamiento

Al aumentar el voltaje de compuerta, se forma una zona de agotamiento en la superficie del substrato cercana a la interfaz Si-SiO<sub>2</sub>. Al aumentar el voltaje de compuerta, los huecos son repelidos de la superficie, lo cual se manifiesta en el diagrama de bandas como un doblamiento de bandas en el que el nivel intrínseco de Fermi se acerca al nivel de Fermi, como se muestra en la figura 6.9.

Para  $|\varphi_S| < |\phi_B|$ , el MOSFET se encuentra en la condición de acumulación.

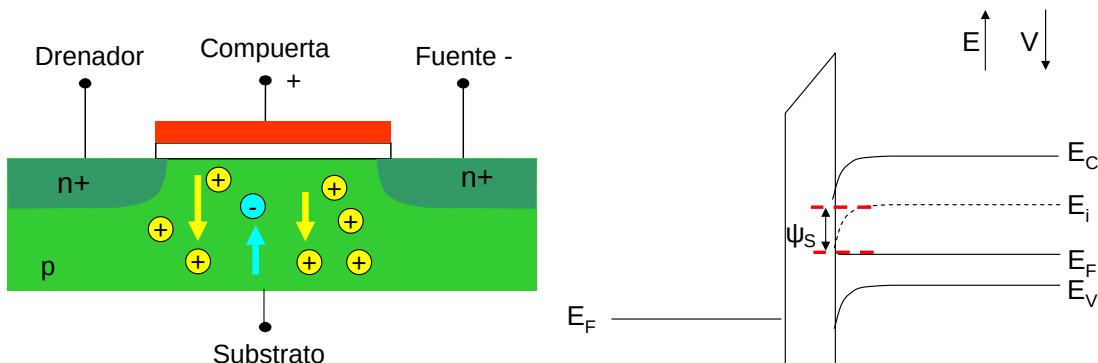


Figura 6.9: Diagrama de bandas de energía de un NMOS en agotamiento

## Región de inversión

Si el voltaje  $V_{GS}$  continúa aumentando, el transistor entra en la condición de inversión débil, en la que  $|\phi_B| \leq |\varphi_S| < |2\phi_B|$ .

Una vez que el potencial de superficie ha alcanzado la condición  $|\varphi_S| = |2\phi_B|$ , se dice que el transistor está en la condición de inicio de la inversión fuerte.

Para  $|\varphi_S| > |2\phi_B|$ , el transistor está en la condición de inversión fuerte. En el diagrama de bandas puede verse que en esta condición el nivel intrínseco de Fermi en la cercanía de la superficie está ubicado debajo del nivel de Fermi. Esto corresponde al diagrama de bandas de un material n, por lo que el substrato, en la cercanía de la interfaz, se comporta como un material n. Esto se ilustra en la figura 6.10.

Por definición, la condición de inversión es aquel estado en el cual la concentración de portadores minoritarios en la superficie del substrato iguala la concentración de portadores mayoritarios del substrato.

Con creciente tensión de compuerta la concentración de electrones en la superficie puede aumentar, pero la profundidad de la zona de agotamiento se mantiene aproximadamente constante.

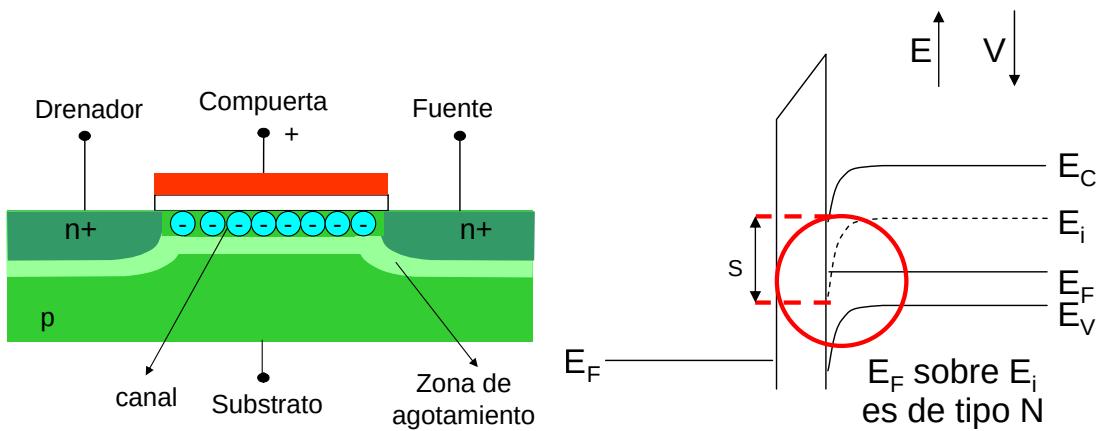


Figura 6.10: Diagrama de bandas de energía de un NMOS en inversión.

### 6.3.3. Tensión de umbral $V_{TH}$

Por definición todos los voltajes del MOSFET toman como referencia la tensión del surtidor. En términos generales, se define la tensión de umbral  $V_{TH}$  como la tensión entre compuerta y surtidor  $V_{GS}$  necesaria para causar inversión de la superficie del substrato, es decir, la formación del canal.

La tensión de umbral está determinada por las siguientes componentes:

1. El voltaje necesario para invertir la superficie (cambiando el potencial de superficie),  $2\phi_B$
2. El voltaje necesario para compensar la carga de la zona de carga espacial,  $Q_{BO}/C_{OX}'$
3. La diferencia de función de trabajo entre la compuerta y el substrato,  $\phi_M - \phi_S$
4. El voltaje necesario para compensar cargas parásitas en el óxido, así como en la interfaz Si-SiO<sub>2</sub>,  $Q_{OX}/C_{OX}'$

Los componentes 3 y 4 fueron previamente contemplados al definir el voltaje de banda plana  $V_{FB}$ .

Matemáticamente, el voltaje de umbral se expresa como:

$$V_{TH} = V_{FB} - \frac{Q_{BO}}{C_{OX}'} - \frac{Q_{OX}}{C_{OX}'} \quad (6.2)$$

$$V_{TH} = \phi_M - \phi_S - 2\phi_B - \frac{Q_{BO}}{C_{OX}'} - \frac{Q_{OX}}{C_{OX}'} \quad (6.3)$$

cuando el substrato y la fuente están conectados al mismo potencial, es decir  $V_{BS} = 0$ .

## 6.4. Clasificación y simbología del MOSFET

El transistor MOSFET puede clasificarse de varias maneras, de acuerdo con:

- El tipo de canal, en PMOS (canal p, substrato n) y NMOS (canal n, substrato p).

- El modo de operación, en transistor de enriquecimiento (normalmente inactivo) y de empobrecimiento (normalmente activo).

El MOSFET de enriquecimiento es aquel en el cual el canal debe ser formado aplicando una tensión de compuerta. Esto implica para un NMOS, una tensión de umbral positiva, y para el PMOS, una tensión de umbral negativa.

Por el contrario, el MOSFET de empobrecimiento es aquel en el cual el canal ya está formado, y se aplica una tensión de compuerta para remover el canal. Esto implica, para NMOS una tensión de umbral negativa y para PMOS una tensión de umbral positiva. Esto se ilustra en la figura 6.11. La tensión de umbral de un transistor puede ajustarse con un perfil de dopado especial en la región del canal aplicando implantación iónica.

El MOSFET puede representarse con símbolos de tres o cuatro terminales. En el símbolo de tres terminales, el substrato y el surtidor están conectados al mismo potencial, razón por la cual no se muestra explícitamente la conexión al substrato. En este símbolo, la flecha indica dirección de corriente técnica entre drenador y surtidor. Si no se usan flechas, un círculo sin rellenar se añade a la compuerta de los transistores PMOS. Los símbolos de tres terminales se encuentran en la figura 6.12. El transistor de empobrecimiento se distingue por tener una línea más gruesa entre drenador y surtidor.

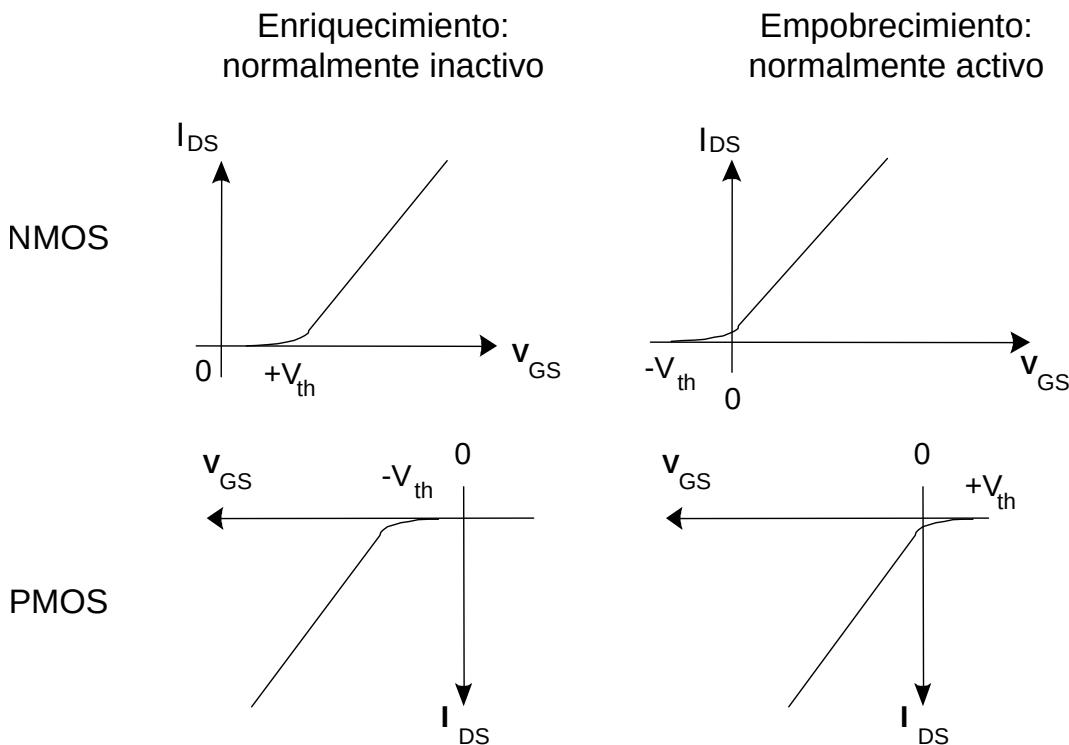


Figura 6.11: Curvas características de entrada de los transistores NMOS y PMOS de enriquecimiento y empobrecimiento

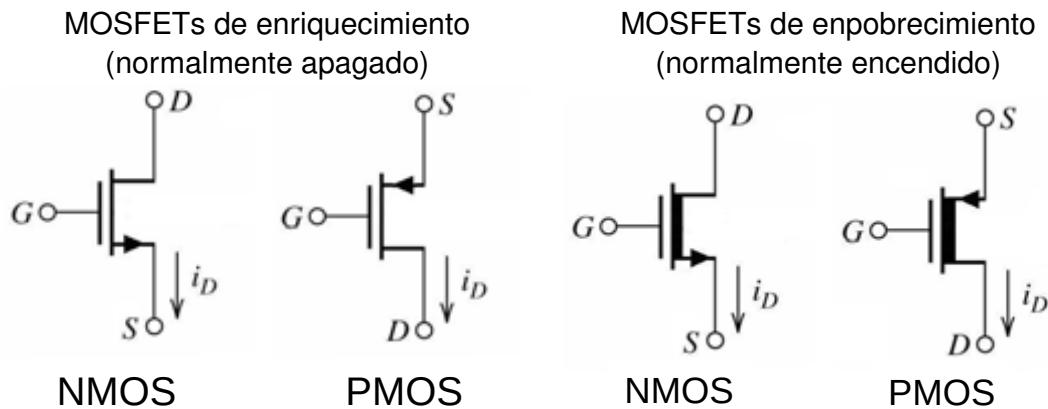


Figura 6.12: Símbolos de tres terminales del transistor MOSFET

Por otra parte, el símbolo de cuatro terminales muestra explícitamente la conexión del substrato. En este caso, la flecha en la terminal de substrato apunta de la región P a la región N. Los símbolos de cuatro terminales se encuentran en la figura 6.13. El transistor de empobrecimiento se distingue por tener una línea más continua entre drenador y surtidor.

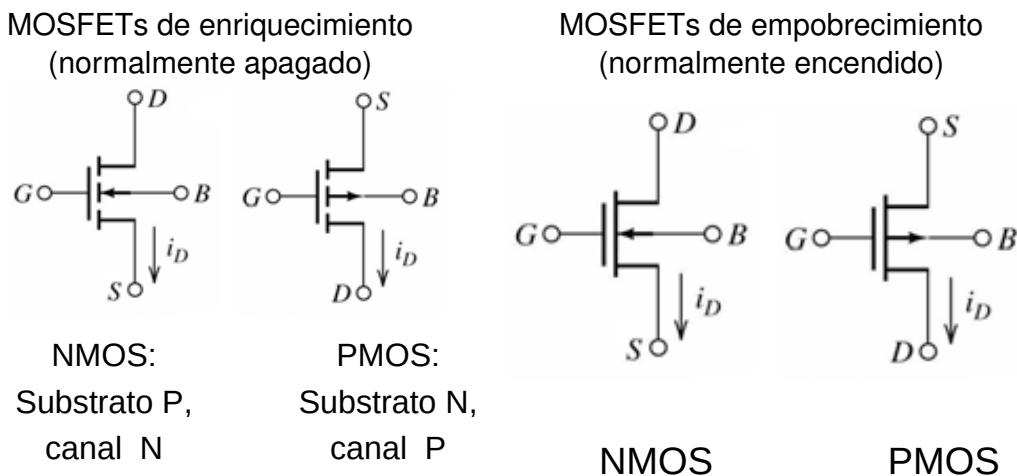


Figura 6.13: Símbolos de cuatro terminales del transistor MOSFET

## 6.5. Curvas características del MOSFET

Para estudiar las curvas características del MOSFET es necesario definir primero las dimensiones involucradas en las ecuaciones. Estas dimensiones se muestran en la figura 6.14 el largo del canal L, el ancho del canal W y el espesor del dieléctrico de compuerta (óxido)  $t_{ox}$ . Cuando se habla de generaciones tecnológicas del MOSFET, la dimensión crítica F es, como una primera aproximación, el valor del largo del canal L.

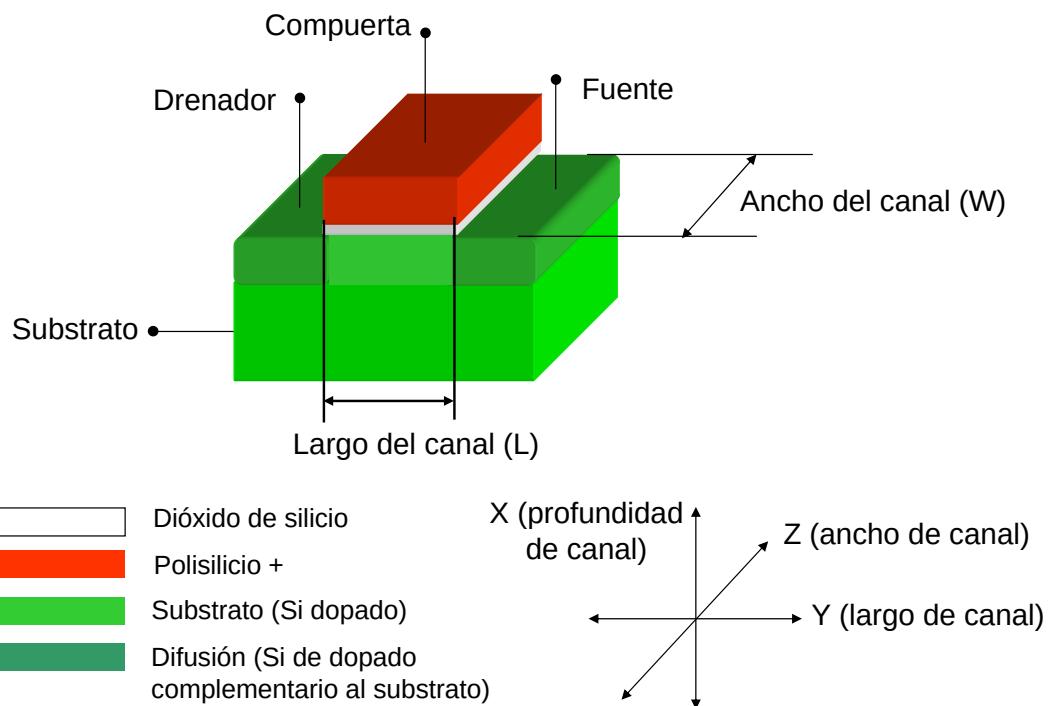


Figura 6.14: Dimensiones importantes en un transistor MOSFET

Con base en la curva IV de salida del MOSFET, pueden distinguirse tres importantes regiones de operación para aplicaciones analógicas, como se muestra en la figura 6.15:

- La región de corte
- La región lineal
- La región de saturación

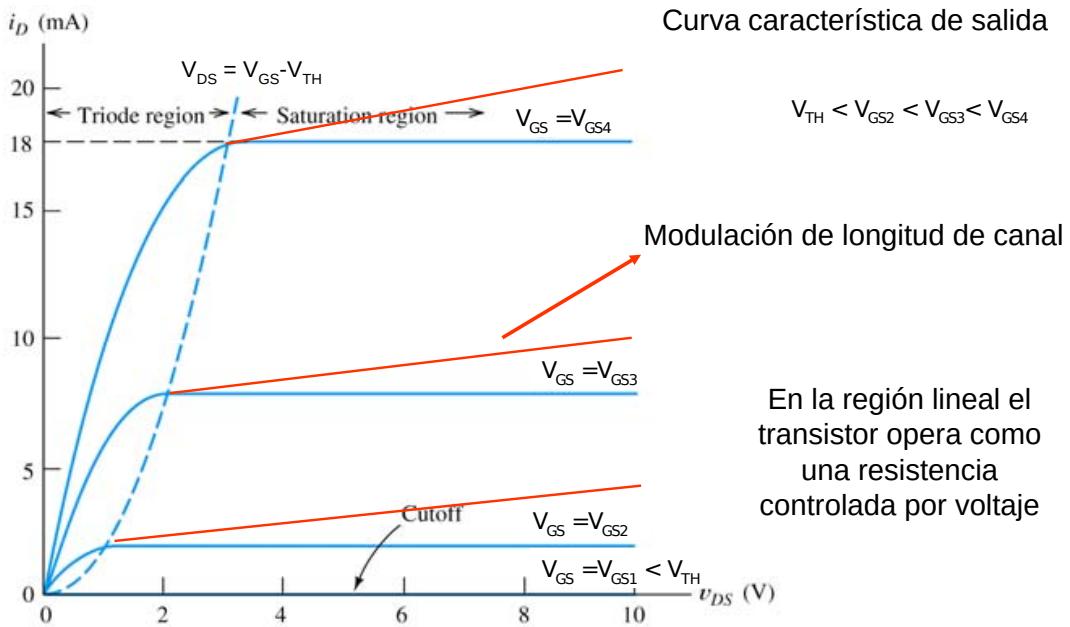


Figure 12.6 Characteristic curves for an NMOS transistor.

Figura 6.15: Curva característica de salida del MOSFET

### 6.5.1. Región de corte

Cuando  $V_{GS} < V_{TH}$ , el MOSFET se encuentra operando en la región de corte o subumbral. Aunque idealmente la corriente  $I_{DS}$  es cero cuando el transistor está en la región de corte, en realidad existe un flujo de corriente de fuga entre drenador y surtidor. Esta corriente se conoce como corriente de subumbral y se calcula como sigue:

$$I_D = I_{D0} \cdot e^{\frac{(V_{GS} - V_{TH})}{mV_t}} \cdot \left[ 1 - e^{\frac{-V_{DS}}{V_t}} \right] \quad (6.4)$$

donde  $I_{D0} = I_{DS}|_{(V_{GS}=V_{TH})} \cdot W/L \approx 0.1\mu A \cdot W/L$  y  $m = 1 + C_{dep}/C_{ox}$ , donde  $C_{dep}$  es la capacitancia de agotamiento.

Para  $V_{DS} > 4V_t$ , la corriente de subumbral puede aproximarse como sigue:

$$I_{DS} \approx I_{D0} e^{\frac{(V_{GS} - V_{TH})}{mV_t}} \quad (6.5)$$

La curva característica de entrada se muestra en la figura 6.16. Esta misma curva, pero ahora mostrando la región de subumbral, se muestra en la figura 6.17. Un parámetro de gran importancia relacionado con la corriente de subumbral es la llamada pendiente de subumbral  $S$ , que se define como:

$$S = \left[ \frac{d(\log I_{DS})}{dV_{GS}} \right]^{-1} = \ln 10 \cdot V_t \cdot m \quad (6.6)$$

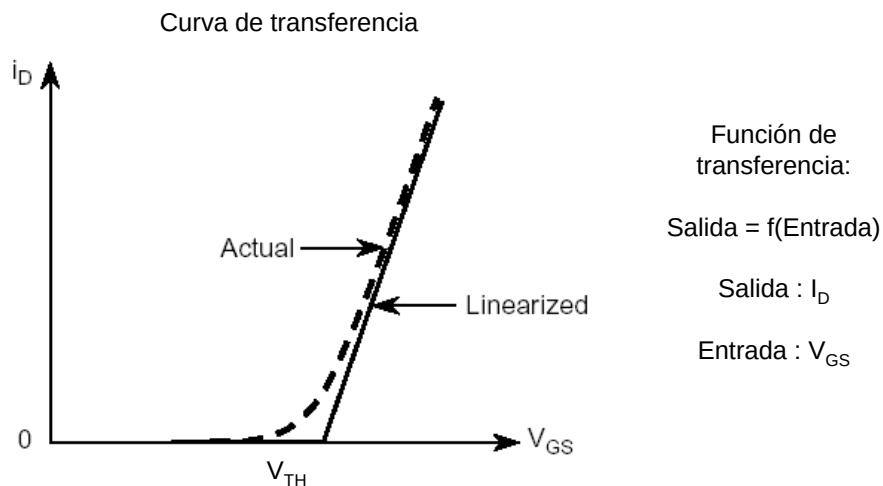


Figura 6.16: Curva característica de entrada del MOSFET

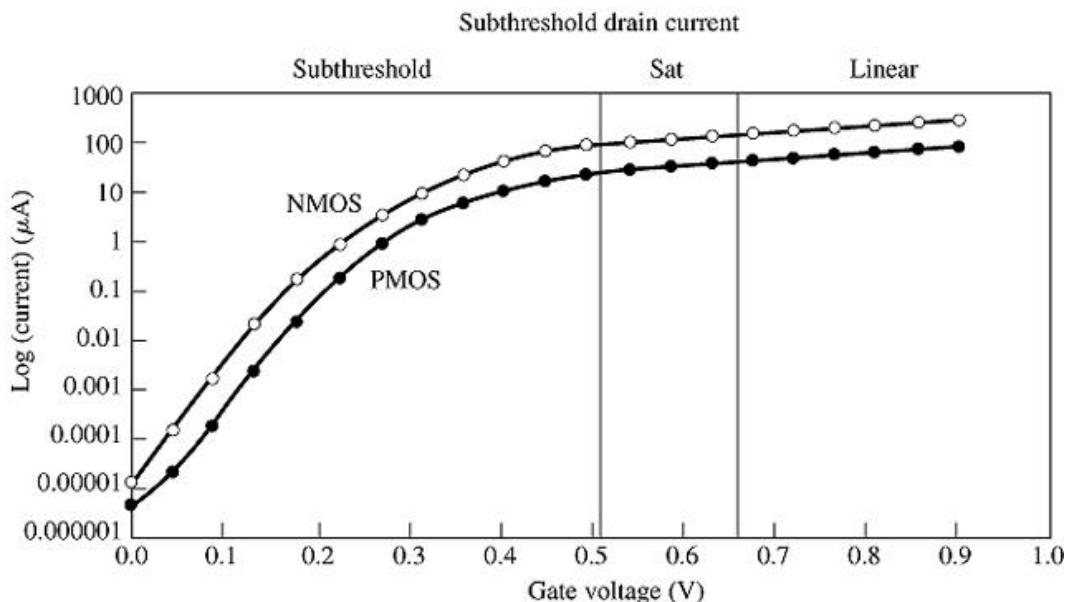


Figura 6.17: Curva característica de entrada del MOSFET mostrando la región de subumbral

La pendiente de subumbral S representa el cambio en  $V_{GS}$  necesario para obtener una variación de una década (un factor de 10) en  $I_{DS}$ . Las unidades de S son mV/dec. S varía con la temperatura. Valores típicos de S a temperatura ambiente (=300 K) son de 80 a 85 mV/ dec. Por otra parte, S es aproximadamente 100 mV/ dec a 100°C.

El valor ideal de S se obtiene al reducir la capacitancia de agotamiento a cero.

En este caso,

$$S = \ln 10 \cdot V_t \cdot m = \ln 10 \cdot V_t \cdot 1 = 60 \text{mV/dec} \quad (6.7)$$

con lo cual el valor ideal de S es de 60 mV/dec.

### 6.5.2. Región lineal

En la región lineal el transistor se comporta como una resistencia (controlada), como lo muestra la curva característica de salida ( $I_{DS}$  vs  $V_{DS}$ ) de la figura 6.15.

En MOSFET opera en la región lineal en la condición

$$V_{GS} - V_{TH} > V_{DS}, V_{GS} \geq V_{TH} \quad (6.8)$$

La ecuación que describe  $I_{DS}$  en la región lineal es

$$I_D = K \left( V_{GS} - V_{TH} - \frac{V_{DS}}{2} \right) V_{DS} \quad (6.9)$$

El valor de la resistencia del MOSFET en la región lineal se obtiene con la definición de resistencia:

$$R = \frac{\partial V}{\partial I} \quad (6.10)$$

en este caso

$$R_{ch} = \left( \frac{\partial I_{DS}}{\partial V_{DS}} \right)^{-1} \quad (6.11)$$

donde  $R_{ch}$  es la resistencia del MOSFET en la zona lineal y es llamada resistencia de canal. Aplicando la definición anterior,

$$\frac{\partial I_{DS}}{\partial V_{DS}} = K(V_{GS} - V_{TH} - V_{DS}) \quad (6.12)$$

Por lo tanto:

$$R_{ch} = \frac{1}{K(V_{GS} - V_{TH} - V_{DS})} \quad (6.13)$$

lo cual para  $V_{DS} \ll V_{DS,sat}$  puede aproximarse a:

$$R_{ch} \approx \frac{1}{K(V_{GS} - V_{TH})} \quad (6.14)$$

### 6.5.3. Región de saturación

Por otra parte, en la región de saturación la corriente es, en el caso ideal, independiente del voltaje, manteniendo un valor constante. Este comportamiento corresponde al de una fuente de corriente. El paso de la región lineal a la región de saturación está determinado por el voltaje de saturación  $V_{DS,sat}$ :

$$V_{DS,sat} = V_{GS} - V_{TH}, V_{GS} \geq V_{TH} \quad (6.15)$$

El MOSFET se encuentra en la región de saturación cuando  $V_{DS} \geq V_{DS,sat}$ .

En la práctica, la modulación de ancho de canal se manifiesta como un incremento de la corriente de drenador con valores crecientes de  $V_{DS}$ . Este comportamiento puede modelarse por medio de

una fuente de corriente real. La modulación de largo de canal consiste en el acortamiento del canal al aumentar  $V_{DS}$ , y dicho acortamiento disminuye la resistencia del canal, aumentando la corriente del transistor.

Comparando el modelo de una fuente de corriente real y la ecuación de la corriente de drenador en la región de saturación, se tiene:

$$I_D = \frac{1}{2}K(V_{GS} - V_{TH})^2 + \frac{1}{2}K(V_{GS} - V_{Th})^2 (\lambda V_{DS}) \quad (6.16)$$

donde el primer término es una constante respecto a  $V_{DS}$ , y el segundo término es una función de  $V_{DS}$ . El término independiente de  $V_{DS}$  representa una fuente de corriente ideal, mientras que el término dependiente de  $V_{DS}$  representa una resistencia y permite modelar el efecto de largo de canal. En la ecuación, esto se manifiesta en el término que contiene  $V_{DS}$  y el coeficiente de modulación de largo de canal. Como puede observarse en la ecuación, las unidades de  $\lambda$  son  $V^{-1}$ .

En el circuito equivalente de la figura 6.18, el efecto de modulación de largo de canal está representado por la resistencia de salida:

$$r_O = \frac{1}{\lambda I_{DS,sat}} \quad (6.17)$$

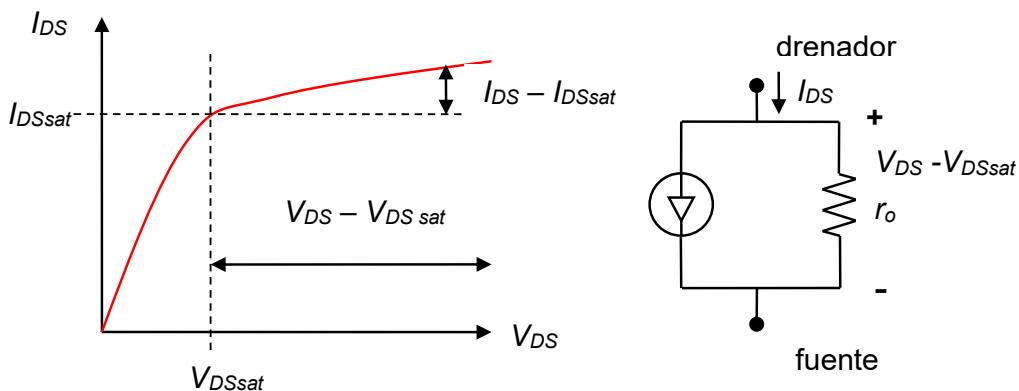


Figura 6.18: Modelo analógico en la región de saturación

## 6.6. Polarización del substrato

Puesto que el MOSFET es un dispositivo de cuatro terminales, debe considerarse también la polarización del substrato. Como puede apreciarse en la figura 6.19, la presencia de las regiones de difusión en el substrato da origen a diodos parásitos, los cuales deben mantenerse inactivos para operar el transistor. Para ello, debe conectarse el substrato P en un NMOS al potencial más bajo disponible en el circuito (por ejemplo, GND), y el substrato N en un PMOS al potencial más alto disponible en el circuito (por ejemplo,  $V_{DD}$ ).

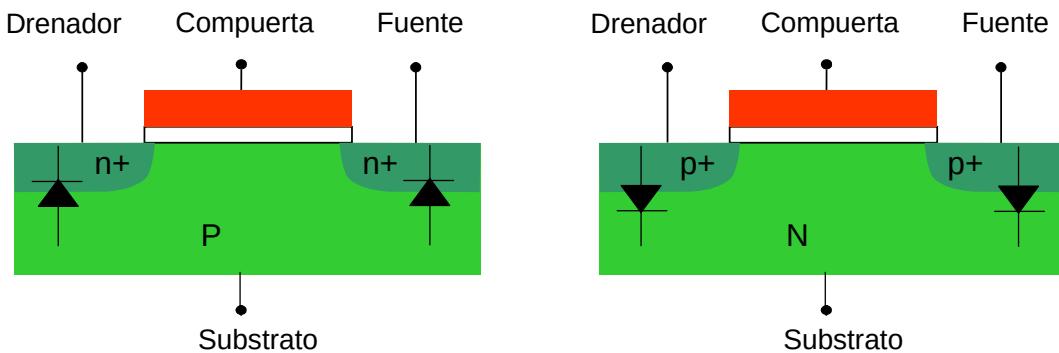


Figura 6.19: Diodos parásitos difusión-substrato

La figura 6.20 muestra el diagrama de bandas de energía drenador-substrato-surtidor de un transistor NMOS sin tensión aplicada y con el substrato conectado al mismo potencial que el surtidor. Como se observa en la figura, existen barreras de energía (potencial de contacto) que impide el flujo de portadores de carga.

Al aplicar una tensión  $V_{GS}$  para lograr la inversión del substrato, el diagrama de bandas de energía drenador-substrato-surtidor de un transistor NMOS se modifica como se ilustra en la figura 6.21. Al aplicar una tensión  $V_{DS}$  diferente de cero se posibilita el flujo de portadores de carga. En este diagrama el substrato está conectado al mismo potencial que el surtidor.

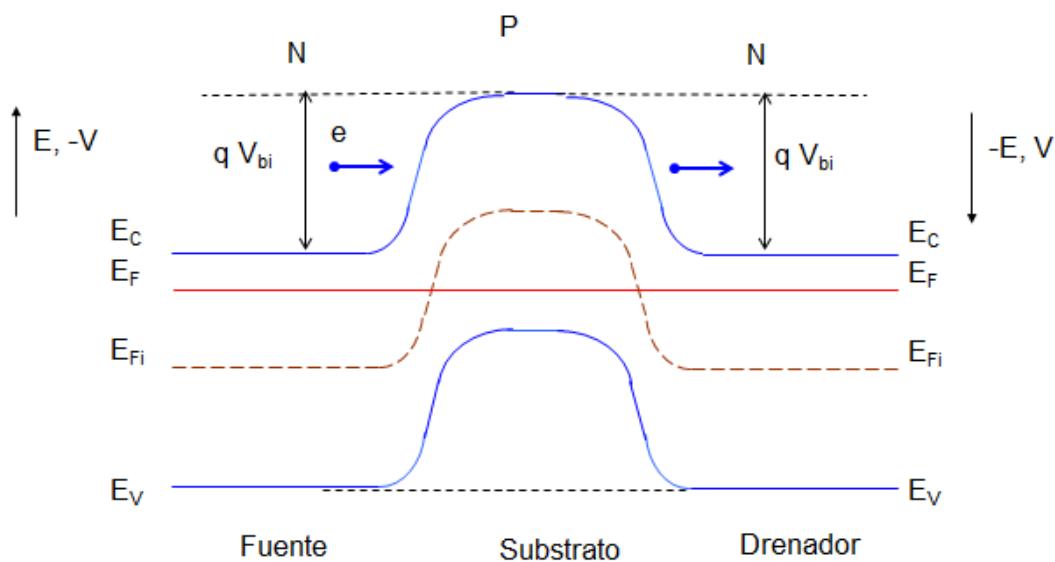


Figura 6.20: Diagrama de bandas drenador-substrato-surtidor de un transistor NMOS sin tensión aplicada

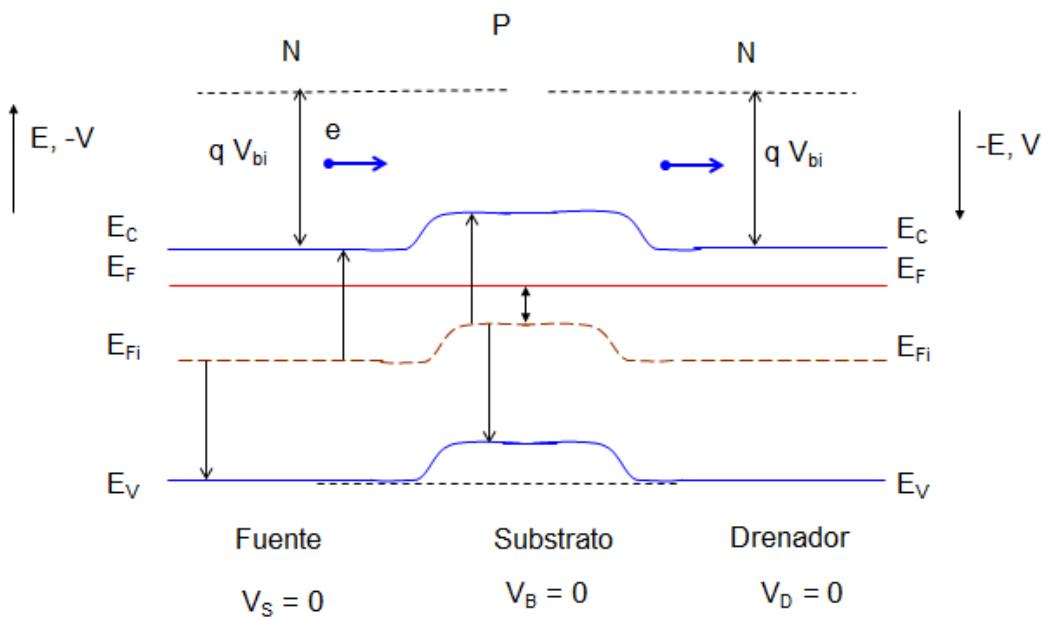


Figura 6.21: Diagrama de bandas drenador-substrato-surtidor de un transistor NMOS con tensión  $V_{GS}$  positiva aplicada

Hasta ahora se ha estudiado el funcionamiento del MOSFET asumiendo que la tensión del substrato es igual a la tensión de surtidor. Al aplicar una tensión  $V_{BS} \neq 0$  se altera la tensión de umbral del transistor. Si la tensión aplicada al substrato favorece la inversión del substrato, la tensión de umbral disminuye. Por el contrario, si la tensión aplicada al substrato favorece la acumulación, la tensión de umbral aumenta. Esto se ilustra en las figuras 6.22 y 6.23, respectivamente. Esta alteración de la tensión de umbral se conoce como efecto de cuerpo o efecto de substrato.

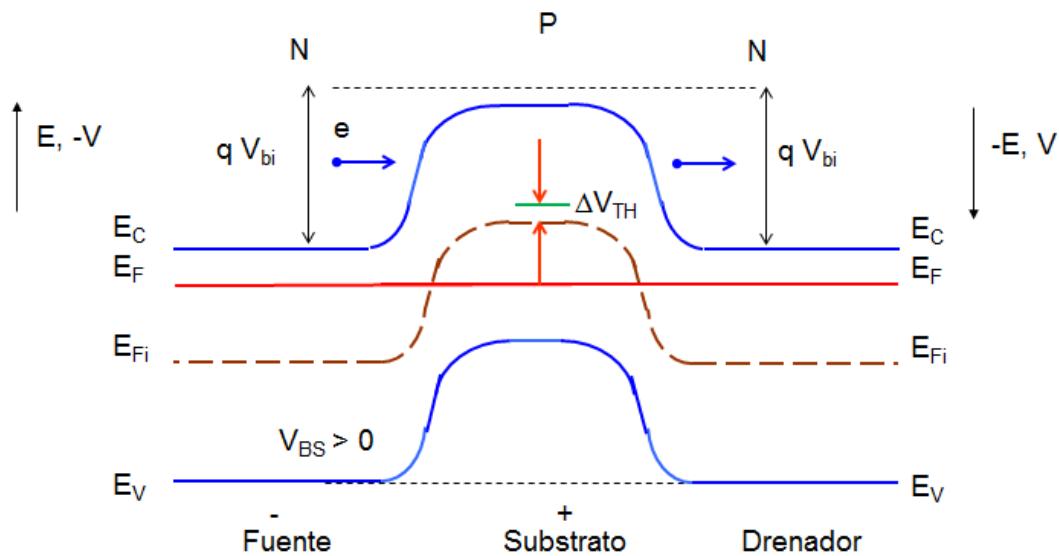


Figura 6.22: Diagrama de bandas drenador-substrato-surtidor de un transistor NMOS con efecto de cuerpo disminuyendo la tensión de umbral

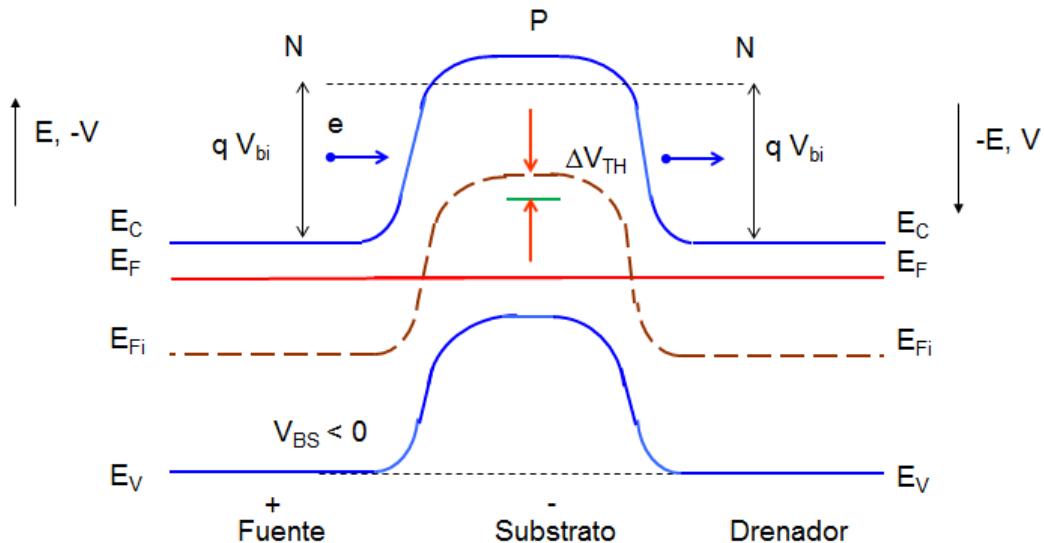


Figura 6.23: Diagrama de bandas drenador-substrato-surtidor de un transistor NMOS con efecto de aumentando la tensión de umbral

La curva característica de entrada en la figura 6.24 muestra el efecto de cuerpo para un transistor NMOS de enriquecimiento. El transistor opera en saturación. En este caso, al tener una tensión  $V_{BS}$  negativa, la tensión de umbral aumenta y la curva característica se desplaza hacia la derecha, es decir, a valores más positivos de  $V_{GS}$ .

El voltaje de umbral tomando en cuenta el efecto de cuerpo se calcula como:

$$V_{TH} = V_{TH0} + \gamma(\sqrt{2\phi_B - V_{BS}} - \sqrt{2\phi - B}) \quad (6.18)$$

Donde el factor  $\gamma$  es el coeficiente de efecto de cuerpo (coeficiente de efecto de substrato) y  $\phi_B$  es la diferencia entre el nivel de Fermi intrínseco y el nivel de Fermi del substrato.

La permitividad del silicio  $\epsilon_{Si}$  es  $11.7\epsilon_0$ .

$$\gamma = \frac{\sqrt{2qN_A\epsilon_{Si}}}{C_{OX}'} \quad (6.19)$$

$$\phi_B = V_t \cdot \ln \left( \frac{N_A}{n_i} \right) \quad (6.20)$$

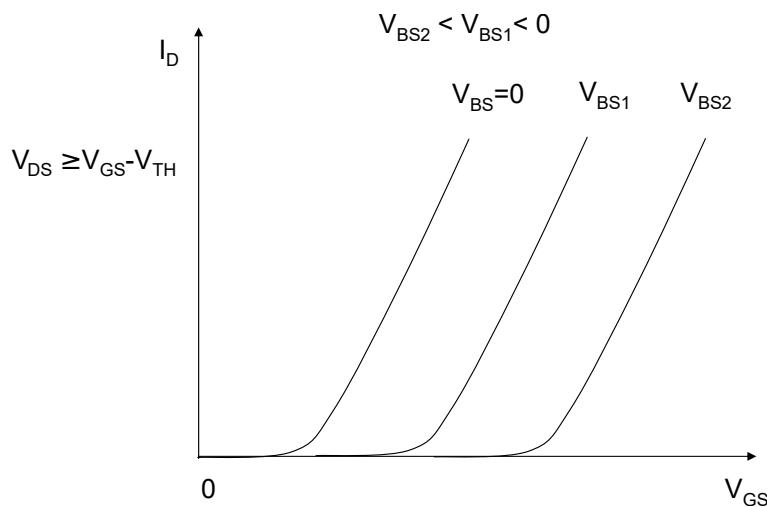


Figura 6.24: Curva característica de entrada de un NMOS mostrando el efecto de cuerpo

# Capítulo 7

## Aplicaciones del MOSFET

### 7.1. El MOSFET como interruptor

En la lógica digital se representan solamente dos estados, correspondientes al 1 y 0 en el sistema binario. El nivel uno corresponde a una tensión de  $V_{DD}$ , mientras que el nivel cero corresponde a una tensión de 0 V.

Una de las aplicaciones básicas del MOSFET es la implementación de circuitos digitales, en los cuales una vez transcurrido el transiente del circuito, puede modelarse el transistor como un interruptor abierto o un interruptor cerrado, según la tensión aplicada a su compuerta. En el caso del interruptor abierto, el transistor opera en la región de subumbral, por lo cual su comportamiento es bastante cercano al ideal. En el caso del interruptor cerrado, deben analizarse dos casos: la transmisión de un 1 lógico y la de un 0 lógico.

Consideremos primeramente la transmisión de un 0 lógico. Iniciemos con el caso de un transistor NMOS, con ayuda del circuito de la figura 7.1, el cual tiene un capacitor de carga  $C_L$  conectado a la salida. El capacitor de carga incluye la capacitancia parásita de las interconexiones, las capacitancias del transistor MOSFET y las capacitancias de cualquier otro circuito que esté conectado a terminal de salida (no mostrado en la figura 7.1). Aplicaremos una tensión de  $V_{DD}$  (1 lógico) a la compuerta de transistor con el fin de formar el canal. Asuma que la tensión de entrada pasa de  $V_{DD}$  a 0 en tiempo cero, y que el capacitor inicialmente se encuentra cargado a  $V_{DD}$ . Ahora se desea fijar un 0 lógico a la salida del circuito, es decir, debemos descargar el capacitor  $C_L$ .

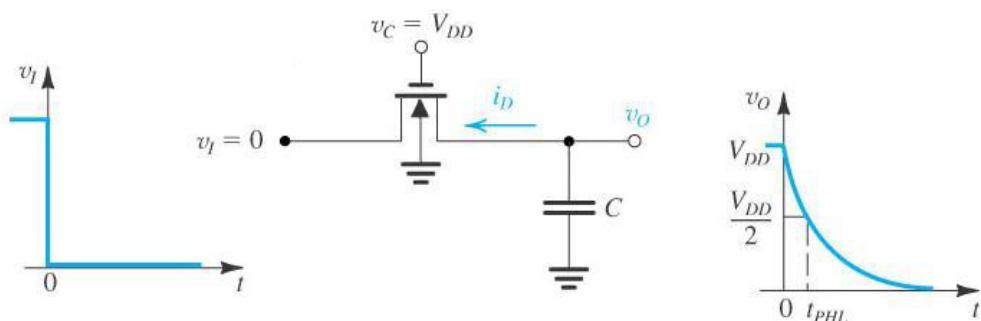


Figura 7.1: Transmisión de un cero lógico utilizando un transistor NMOS.

Debido a la tensión  $V_{DD}$  aplicada a la compuerta del transistor, en  $t < 0$  la corriente en el circuito

es cero, puesto que el capacitor estaba cargado a  $V_{DD}$  y además  $V_{in} = V_{DD}$ , por lo que el transistor opera en la región de corte con  $V_{GS} = 0$ . En  $t > 0$ , puede ahora distinguirse entre drenador y surtidor, debido a que la tensión aplicada a cada terminal es distinta.

De acuerdo con la figura 7.1,  $V_{GS} = V_{DD} = V_{DS}$  y al inicio de la descarga el transistor opera en saturación. Ahora, puesto que  $V_G$  es constante y  $V_S$  es constante, el capacitor puede descargarse completamente y al final de la descarga el transistor opera en la región lineal, es decir, el canal está formado, sin embargo, con  $V_C = 0$ ,  $V_{DS} = 0$  y no fluye corriente en el transistor, dado que el capacitor ya no entrega carga. De esta forma, el transistor NMOS fija de forma óptima un cero lógico a la salida del circuito.

Considere ahora el mismo caso, pero sustituyendo el transistor NMOS por un transistor PMOS, como se muestra en la figura 7.2, en el cual se aplica una tensión de 0 V en la compuerta del transistor.

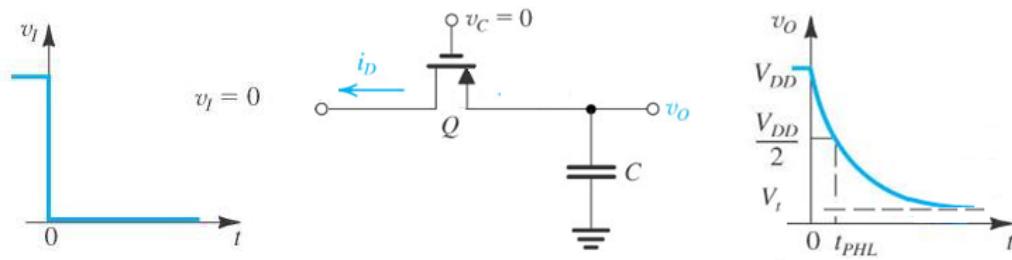


Figura 7.2: Transmisión de un cero lógico utilizando un transistor PMOS

En  $t < 0$ , con  $V_{in} = V_{DD}$  aún no puede distinguirse entre drenador y surtidor del transistor. Sin embargo, puede verse que  $V_{SG} = V_{DD}$  y el transistor opera en la región lineal. No fluye corriente en el transistor debido a que  $V_{SD} = 0$ . En  $t = 0^+$ , dado que  $V_{in} = 0$ , puede distinguirse entre drenador y surtidor, con  $V_{SG} = V_{DD}$  y  $V_{SD} = V_{DD}$ .

Observamos también que  $V_S = V_C(t)$  y por lo tanto  $V_{SG} = V_C(t) - 0$ . Así cuando el capacitor se descarga hasta alcanzar  $V_C = V_{TH}$ , el transistor está aún encendido y en el umbral de conducción, pero si la tensión disminuye un poco más, el transistor se apaga. Por ello el transistor PMOS no transmite (fija) de manera óptima un 0 lógico a la salida, dado que la tensión más baja a la que puede descargar el capacitor es  $V_{C,min} = V_{TH}$ .

Una ventaja adicional del transistor NMOS para la transmisión de un cero lógico es el hecho de que durante dicha operación no sufre de efecto de cuerpo, a diferencia del transistor PMOS.

Veamos ahora la transmisión de un 1 lógico, es decir, deseamos fijar un 1 lógico a la salida del circuito, y para eso debemos cargar el capacitor  $C_L$ . Iniciemos con el caso de un transistor NMOS, con ayuda del circuito de la figura 7.3. La tensión inicial en el capacitor de carga es 0 V. Se aplica una tensión de  $V_{DD}$  (1 lógico) a la compuerta de transistor con el fin de formar el canal. Asuma que la tensión de entrada pasa de 0 a  $V_{DD}$  en tiempo cero.

Debido a la tensión aplicada a la compuerta del transistor, en  $t < 0$  la corriente en el circuito es cero, puesto que el capacitor estaba cargado a 0 y además  $V_{in} = 0$ , por lo que el transistor opera en la región de lineal; en este caso el canal está formado pero no fluye corriente porque  $V_{DS} = 0$ . En  $t > 0$ , puede ahora distinguirse entre drenador y surtidor.

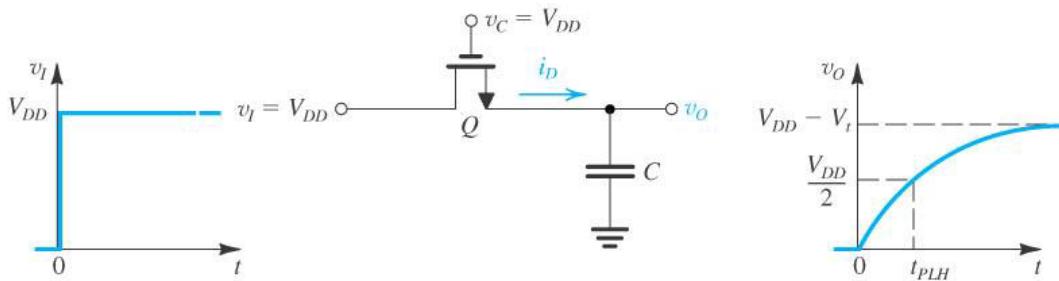


Figura 7.3: Transmisión de un uno lógico utilizando un transistor NMOS

De acuerdo con la figura 7.3,  $V_{GS} = V_{DD} = V_{DS}$  y al inicio de la carga el transistor opera en saturación. Ahora, puesto que  $V_G$  es constante y  $V_S = V_C(t)$ , el capacitor se empieza a cargar. Sin embargo, cuando el capacitor alcanza  $V_c(t) = V_{DD} - V_{TH}$ , la tensión  $V_{GS}$  correspondiente es  $V_{TH}$ . Si la tensión del capacitor aumenta un poco más, el transistor NMOS entra a operar en la región de corte. De esta forma, el transistor NMOS no puede fijar de forma óptima un uno lógico a la salida del circuito, puesto que la tensión máxima a la que puede cargar el capacitor es  $V_{DD} - V_{TH}$ .

Considere ahora el mismo caso, pero sustituyendo el transistor NMOS por un transistor PMOS, como se muestra en la figura 7.4, en el cual se aplica una tensión de 0V en la compuerta del transistor. F

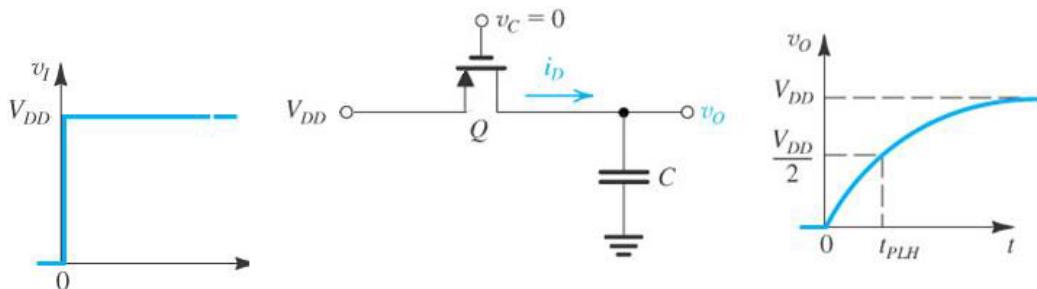


Figura 7.4: Transmisión de un uno lógico utilizando un transistor PMOS

En  $t < 0$ , con  $V_{in} = 0$ ,  $V_{SG} = 0$  y el transistor opera en la región de corte. En  $t = 0^+$ , dado que  $V_{in} = V_{DD}$ , puede distinguirse entre drenador y surtidor. Aquí  $V_{SG} = V_{DD}$  y  $V_{SD} = V_{DD}$  y el transistor opera en la región de saturación.

Observamos también que  $V_S = V_{DD}$  y por lo tanto  $V_{SG}$  es constante, de forma que el transistor puede permanecer encendido y puede completarse la carga del capacitor hasta el valor de  $V_{DD}$ . Por ello el transistor PMOS transmite (fija) de manera óptima un 1 lógico a la salida.

Una ventaja adicional del transistor PMOS para la transmisión de un uno lógico es que, a diferencia del transistor NMOS, no sufre de efecto de cuerpo durante esta operación.

## 7.2. Compuertas lógicas básicas

### 7.2.1. El inversor CMOS

El inversor CMOS es la compuerta lógica básica más sencilla que puede implementarse. La figura 7.5 muestra el esquema de un inversor CMOS, es decir, un inversor implementado con transistores MOSFET complementarios, ya que consta de un transistor PMOS y un transistor NMOS. Observe que debido al comportamiento de los transistores PMOS y NMOS como interruptores, el transistor PMOS se conecta entre la  $V_{DD}$  y la salida para transmitir un 1 lógico a la salida, mientras que el transistor NMOS se conecta entre la salida y tierra para transmitir un 0 lógico a la salida.

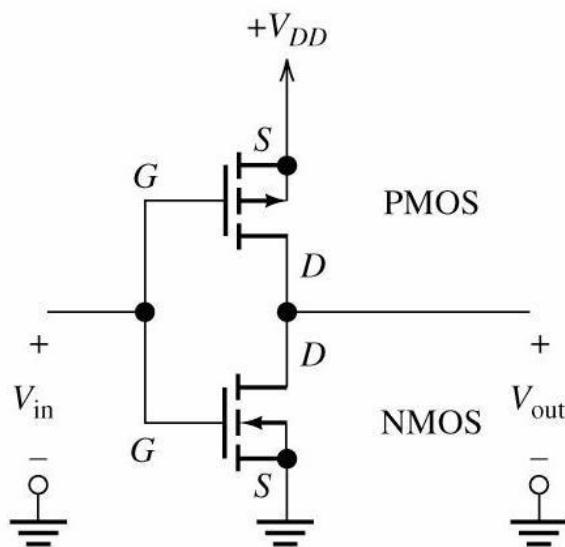


Figura 7.5: Inversor CMOS.

La tabla de verdad de un circuito lógico muestra el comportamiento del circuito, es decir, sus valores de salida, ante las diferentes combinaciones de señales de entrada. Los valores mostrados en las tablas de verdad son binarios. La Tabla 7.1 muestra la tabla de verdad de un inversor.

Tabla 7.1: Tabla de verdad de un inversor lógico

Entrada	Salida
0	1
1	0

Observe que la función de inversor es transmitir a la salida el valor contrario al aplicado a la entrada, es decir, invertir la señal. Desde el punto de vista lógico, esta función se conoce también como negación, es decir, NOT.

Para obtener la tabla de verdad de un circuito lógico debe considerarse el valor final de la salida, es decir, el valor en estado estable. Al analizar el circuito debe observarse si la tensión de entrada contribuye a activar o desactivar los transistores a los cuales está conectada. Es importante recordar que, desde el punto de vista de un circuito digital, las señales solamente puede tomar dos valores:

0 lógico (tierra) o 1 lógico ( $V_{DD}$ ). Además, la tensión de alimentación  $V_{DD}$  es mayor que la magnitud de los voltajes de umbral de los transistores.

Con estas consideraciones en mente, procederemos a analizar el comportamiento del inversor CMOS. Para ello, en el estado estable representaremos a un transistor en corte como un interruptor abierto, y a un transistor activo como un interruptor cerrado.

Con  $V_{in} = 0$  lógico, el transistor NMOS presenta una tensión  $V_{GS}$  de 0 V, por lo cual opera en la región de corte, mientras que la tensión  $V_{SG}$  del transistor PMOS es  $V_S - V_G = V_{DD}$ , por lo que el transistor PMOS se activa. De esta forma, en el estado estable, podemos representar al transistor NMOS como un interruptor abierto y al PMOS como un interruptor cerrado. De esta forma, el transistor PMOS conecta la salida  $V_{out}$  a  $V_{DD}$ , definiendo el valor de la salida como un 1 lógico. Dado que en estado estable la tensión  $V_{SD}$  es de 0 V, el transistor PMOS mantiene el canal formado pero no fluye corriente entre sus terminales, y el transistor opera en la región lineal.

Si ahora  $V_{in} = 1$  lógico, el transistor PMOS presenta una tensión  $V_{SG}$  de 0 V, por lo cual opera en la región de corte, mientras que la tensión  $V_{GS}$  del transistor NMOS es  $V_G - V_S = V_{DD}$ , por lo que el transistor NMOS se activa. Representamos entonces al transistor NMOS como un interruptor cerrado y al PMOS como un interruptor abierto. De esta forma, el transistor NMOS conecta la salida  $V_{out}$  a tierra, definiendo el valor de la salida como un 0 lógico. En estado estable la tensión  $V_{DS}$  del transistor NMOS es de 0 V, por lo que el transistor mantiene el canal formado pero no fluye corriente entre sus terminales, y el transistor opera en la región lineal.

El comportamiento del inversor CMOS para ambos casos de la tensión de entrada se muestra en la figura 7.6. El símbolo lógico de un inversor se muestra en la figura 7.7. El triángulo representa la transmisión de la señal de la entrada a la salida, mientras que el círculo representa inversión.

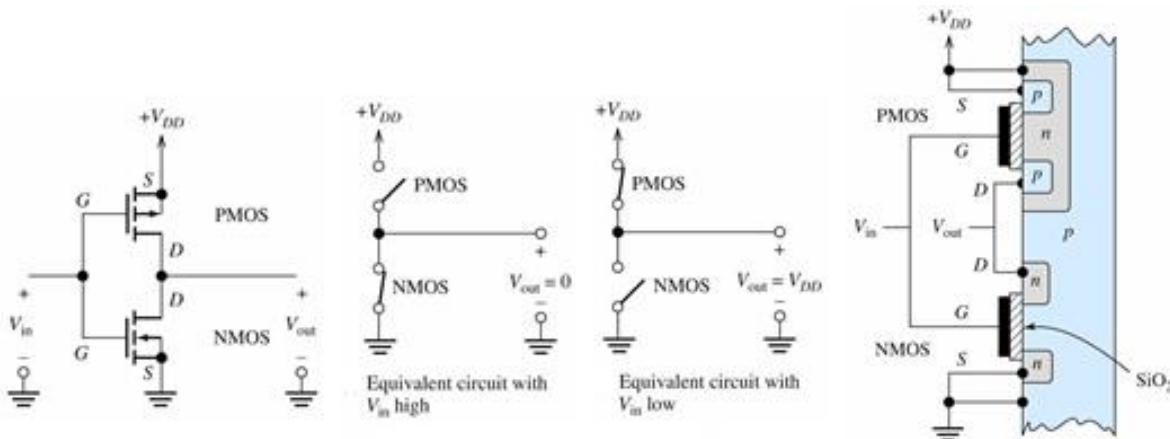


Figura 7.6: Comportamiento del inversor CMOS.

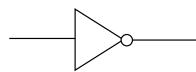


Figura 7.7: Símbolo lógico de un inversor.

En un circuito, el comportamiento de la tensión de salida como función de la tensión de entrada se conoce como curva de transferencia (función de transferencia). La función de transferencia del

inversor CMOS se muestra en la figura 7.8; se observa que existe un rango de tensiones de entrada para las cuales ambos transistores se encuentran activos, y por lo tanto existe un flujo de corriente entre  $V_{DD}$  y tierra a través de ambos transistores. Esta corriente se conoce como corriente de corto circuito, pues los transistores forman una red de baja impedancia entre  $V_{DD}$  y tierra.

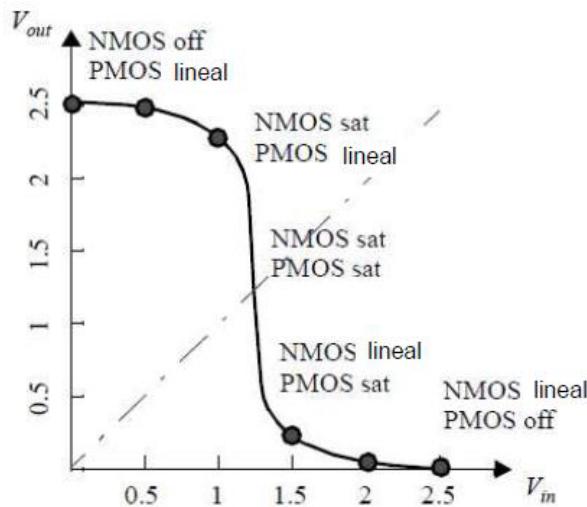


Figura 7.8: Curva de transferencia de un inversor CMOS.

En la curva característica existe un valor de tensión de entrada tal que  $V_{in} = V_{out}$ . Este valor se conoce como punto de disparo  $V_{SP}$  (switching point) de una compuerta lógica. Cuando el punto de disparo de un inversor es  $V_{DD}/2$ , se dice que el inversor es simétrico.

La ecuación del punto de disparo de un inversor CMOS es

$$V_{SP} = \frac{V_{TH,N} + \sqrt{\frac{K_P}{K_N}(V_{DD} - |V_{TH,P}|)}}{1 + \sqrt{\frac{K_P}{K_N}}} \quad (7.1)$$

Donde

$V_{TH,N}$ : tensión de umbral del transistor NMOS

$V_{TH,P}$ : tensión de umbral del transistor PMOS

$K_N$ : parámetro de transconductancia del transistor NMOS

$K_P$ : parámetro de transconductancia del transistor PMOS

Como puede observarse de la ecuación del punto de disparo, para que un inversor sea simétrico es necesario que  $V_{TH,N} = V_{TH,P}$  y  $K_P/K_N = 1$ . Esto implica que la diferencia de movilidades de huecos y electrones debe compensarse ajustando el ancho del transistor PMOS de forma que  $K_P = K_N$ .

Así, cuando un inversor es simétrico, cumple las siguientes características:

- $V_{TH,N} = V_{TH,P}$

- $K_P = K_N$

- $V_{SP} = V_{DD}/2$
- Su tiempo de subida  $t_{rise}$  y su tiempo de bajada  $t_{fall}$  son iguales

Es importante considerar que aunque analizamos las salidas de los circuitos digitales en estado estable, en realidad las señales de entrada y salida tienen un tiempo de transición diferente de cero. El tiempo en el cual una señal pasa de 0 a 1 lógico se conoce como tiempo de subida  $t_{rise}$ , y el tiempo en el cual pasa de 1 a 0 se conoce como tiempo de bajada  $t_{fall}$ .

### 7.2.2. Compuerta NAND

La compuerta NAND CMOS de dos entradas se compone de cuatro transistores: dos PMOS y dos NMOS, como se muestra en la figura 7.9. Es importante recalcar que, dado que los transistores PMOS se utilizan para transmitir el valor de 1 lógico a la salida, están conectados entre  $V_{DD}$  y la salida, mientras que los NMOS deben estar conectados entre la salida y a tierra, porque se utilizan para transmitir el valor de 0 lógico a la salida.

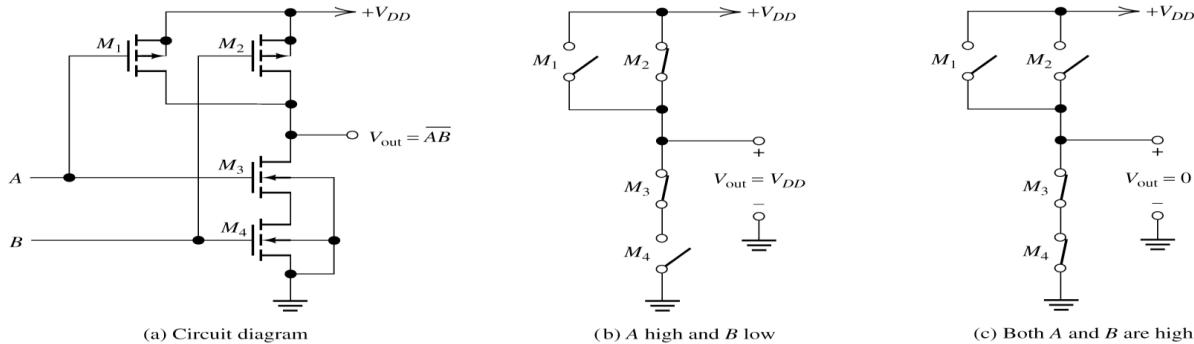


Figura 7.9: Compuerta NAND CMOS de dos entradas.

Como se observa en la figura 7.9, en una compuerta NAND CMOS se necesitan dos transistores por cada entrada, uno NMOS y otro PMOS.

La función lógica NAND corresponde a una función AND negada. Para comprender esto, definamos primero la función lógica AND, cuyo comportamiento se muestra en la Tabla 7.2.

Tabla 7.2: Comportamiento de una compuerta AND.

Entrada lógica A	Entrada lógica B	Salida lógica
0	0	0
0	1	0
1	0	0
1	1	1

Como podemos observar en la Tabla 7.2, la salida tendrá un valor de 1 lógico si y solo si A y B son 1, de allí el nombre AND. La compuerta NAND es una AND seguida de un inversor (NAND = NOT AND), como se resume en la Tabla 7.3.

Las primeras tres columnas de la Tabla 7.3 corresponden a la tabla de verdad de la compuerta NAND, mientras que las últimas dos muestran los valores de tensión para cada fila de la tabla de verdad. Los símbolos lógicos de las compuertas AND y NAND se muestran en la figura 7.10.

Tabla 7.3: Tabla de verdad de una compuerta NAND.

Entrada lógica A	Entrada lógica B	Salida lógica	Voltaje de entrada	Voltaje de salida
0	0	1	$V_A = 0, V_B = 0$	$V_{DD}$
0	1	1	$V_A = 0, V_B = V_{DD}$	$V_{DD}$
1	0	1	$V_A = V_{DD}, V_B = 0$	$V_{DD}$
1	1	0	$V_A = V_{DD}, V_B = V_{DD}$	0

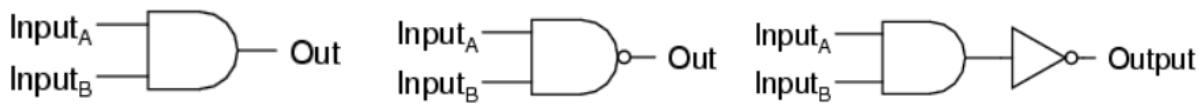


Figura 7.10: Símbolos lógicos de las compuertas AND, NAND y equivalencia de la compuerta NAND.

### 7.2.3. Compuerta NOR

Al igual que la compuerta NAND CMOS, la compuerta NOR CMOS de dos entradas se compone de cuatro transistores, dos por cada entrada: dos PMOS y dos NMOS, como se muestra en la figura 7.11.

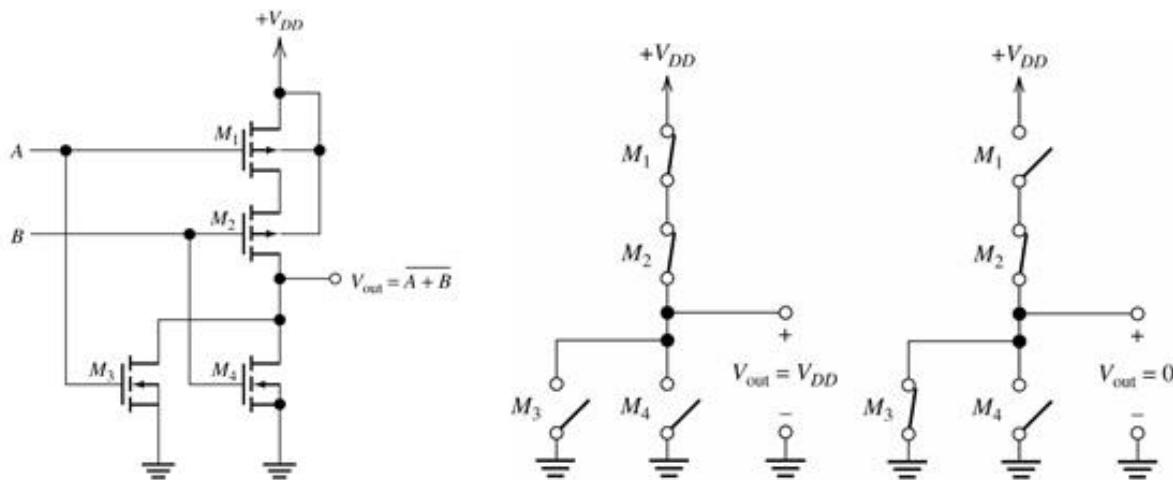


Figure 12.33 Two-input CMOS NOR gate.

(a) A and B both low

(b) A high and B low

Figura 7.11: Compuerta NOR CMOS de dos entradas.

La función lógica NOR corresponde a una función OR negada. Para comprender esto, definamos primero la función lógica OR, cuyo comportamiento se muestra en la Tabla 7.4.

Tabla 7.4: Comportamiento de una compuerta OR.

Entrada lógica A	Entrada lógica B	Salida lógica
0	0	0
0	1	1
1	0	1
1	1	1

Como podemos observar en la Tabla 7.4, la salida tendrá un valor de 1 lógico si A o B son 1, de allí el nombre OR. La compuerta NOR es una OR seguida de un inversor (NOR = NOT OR), como se resume en la Tabla 7.5. Las primeras tres columnas de la Tabla 7.5 corresponden a la tabla de verdad de la compuerta NOR, mientras que las últimas dos muestran los valores de tensión para cada fila de la tabla de verdad. Los símbolos lógicos de las compuertas OR y NOR se muestran en la figura 7.12.

Tabla 7.5: Tabla de verdad de una compuerta NOR.

Entrada lógica A	Entrada lógica B	Salida lógica	Voltaje de entrada	Voltaje de salida
0	0	1	$V_A = 0, V_B = 0$	$V_{DD}$
0	1	0	$V_A = 0, V_B = V_{DD}$	0
1	0	0	$V_A = V_{DD}, V_B = 0$	0
1	1	0	$V_A = V_{DD}, V_B = V_{DD}$	0

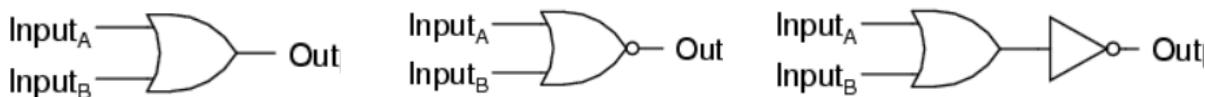


Figura 7.12: Símbolos lógicos de las compuertas OR, NOR y equivalencia de la compuerta NOR.

### 7.3. Arreglos de transistores

Todo transistor MOSFET tiene una resistencia asociada, la cual se refleja en las regiones lineal y de saturación por medio del parámetro de transconductancia y la resistencia de salida del transistor. Cuando se conectan transistores MOSFET en serie, estos pueden reemplazarse por un único transistor equivalente si su parámetro de transconductancia da como resultado la misma corriente que todos los transistores sustituidos juntos. El transistor equivalente tiene la misma tensión de umbral que la de los transistores a los que sustituye. En el caso de transistores conectados en paralelo, podemos encontrar el parámetro de transconductancia del transistor equivalente la suma de los parámetros de transconductancias individuales. Para transistores conectados en serie, el parámetro de transconductancia del transistor equivalente es el inverso de la suma de los inversos de los parámetros de transconductancia individuales. Dicho de otra forma, se utilizan las fórmulas para el cálculo de una conductancia equivalente, tomando para ello los valores de parámetro de transconductancia.

## 7.4. Consumo de potencia

El consumo de potencia de un circuito es uno de las especificaciones de diseño más importantes en la Electrónica. Para un circuito dado, debe considerarse tanto la potencia en estado activo como inactivo. De acuerdo con esto, existen dos tipos de consumo de potencia:

a) Potencia dinámica, que es la potencia consumida por los circuitos en estado activo, por ejemplo, cuando una compuerta lógica cambia de estado

b) Potencia estática, que es la potencia consumida por los circuitos en estado inactivo, por ejemplo, un dispositivo en "stand by" o una compuerta lógica en espera de un cambio en su señal de entrada

La potencia estática se consideró despreciable durante mucho tiempo, pero con el advenimiento de los dispositivos portátiles y el aumento en la densidad de integración de los circuitos microelectrónicos, su contribución no puede despreciarse, ya que determina en buena parte la duración de la carga en las baterías de computadoras portátiles, teléfonos celulares, y en general, de dispositivos móviles.

### 7.4.1. Potencia estática

La potencia estática está causada por:

- Corriente de fuga de compuerta
- Corriente de reversa de juntas PN
- Corriente de subumbral

La corriente de fuga de compuerta es aquella corriente que fluye entrando o saliendo de la compuerta del transistor. Se debe a que el material dieléctrico de la compuerta no ofrece condiciones de aislamiento 100 % ideales, aunque se aproximan mucho a esto. Además de que el dieléctrico de la compuerta tiene una conductividad diferente de cero, la reducción del espesor del dieléctrico depositado contribuye a aumentar la corriente de fuga de compuerta por efecto túnel.

La corriente de reversa de las juntas PN difusión-substrato de los transistores MOSFET son el siguiente factor causante de la potencia estática. Durante la operación de los transistores MOSFET el substrato se polariza de forma que estas juntas PN se mantengan operando en la región de corte, donde presentan una corriente que, aunque muy baja, es diferente de cero, y por lo tanto, disipan potencia.

La corriente de subumbral es la causa más importante de la potencia estática, y ocurre cuando un transistor opera en la región de corte (región de subumbral). Su valor está determinado tanto por la tensión aplicada a los transistores que estén operando en la región de corte, como por el valor de la tensión de umbral y la pendiente de subumbral. La pendiente de subumbral, a su vez, depende de la temperatura, de forma que cuanto mayor temperatura, tanto mayor será la corriente de subumbral.

Para el cálculo exacto de la potencia estática, se requieren las tres corrientes de fuga: de compuerta, de reversa de las juntas PN y de subumbral. Sin embargo, dado que la corriente de subumbral es la mayor de todas, la potencia estática puede aproximarse utilizando para el cálculo solamente la corriente de subumbral.

### 7.4.2. Potencia dinámica

La potencia dinámica se presenta cuando existen transiciones de un estado lógico a otro, es decir, cuando la salida de un circuito digital cambia en respuesta a una tensión aplicada a la entrada.

La potencia dinámica está causada por:

- Corriente de corto circuito
- Corriente debido a cargas capacitivas

Para comprender estas componentes de la potencia dinámica, así como obtener ecuaciones para calcularlas, considere un inversor CMOS.

Cuando cambia la tensión aplicada a la entrada del inversor, el cambio no se produce instantáneamente, sino que la tensión de entrada presenta una pendiente, llamada flanco de subida (para transiciones de 0 a 1 lógico) o flanco de bajada (para transiciones de 1 a 0 lógico). Esto origina un consumo de potencia dinámica durante las transiciones.

#### Potencia dinámica de corto circuito

Durante la transición de la salida de un estado lógico a otro, la tensión de entrada toma por un tiempo determinado valores de voltaje entre 0 y  $V_{DD}$  que hacen que el transistor PMOS y NMOS se encuentren activos simultáneamente, causando una trayectoria de baja impedancia entre  $V_{DD}$  y tierra (GND) y con ello un alto consumo de corriente en el inversor. La corriente que fluye en esta condición se conoce como corriente de corto circuito, y es la causa de la potencia dinámica de corto circuito.

El valor máximo de la corriente de corto circuito se obtiene en la condición  $V_{in} = V_{out}$ , es decir, en la tensión de disparo del inversor. La potencia dinámica de corto circuito para un inversor simétrico se calcula como:

$$P_{SC} = I_{SC} \cdot V_{DD} = \frac{2}{3} \cdot K \cdot \frac{t_r}{T} \left( \frac{V_{DD}}{2} - V_{TH} \right)^3 \quad (7.2)$$

Donde

$P_{SC}$ : potencia de corto circuito

$I_{SC}$ : corriente de corto circuito

$K$ : parámetro del transconductancia de los transistores del inversor simétrico

$t_r$ : tiempo de subida

$T$ : período de la señal de entrada

Esta fórmula es aplicable bajo las siguientes condiciones:

- El inversor es simétrico
- El factor de actividad es 1
- El tiempo de subida  $t_{rise}$  y el tiempo de bajada  $t_{fall}$  son iguales

## Potencia dinámica de carga capacitiva

La potencia dinámica de carga capacitiva se debe a las capacitancias parásitas presentes en el circuito, y que pueden representarse como una sola capacitancia de carga  $C_L$ , como se muestra en la figura 7.13.

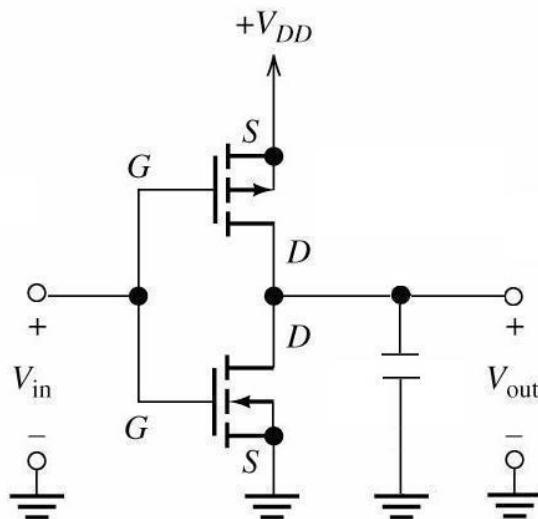


Figura 7.13: Capacitancia de carga en un inversor CMOS.

Esta capacitancia de carga se debe a la suma de:

- La capacitancia de compuerta  $C_{OX}$  de las compuertas lógicas conectadas a la salida
  - La capacitancia de compuerta  $C_{OX}$  propia de la compuerta lógica
  - La capacitancia parásita  $C_W$  de las interconexiones conectadas a la salida

La potencia dinámica de carga capacitiva se debe a la potencia consumida por la compuerta lógica en el proceso de carga y descarga del capacitor al cambiar el estado lógico de la salida. Esta potencia se calcula como:

$$P_I = A \cdot f \cdot C_I \cdot V_{DD}^2 \quad (7.3)$$

Donde

$f$ : frecuencia de conmutación.

$C_L$ : capacitancia de carga.

A: factor de actividad

El factor de actividad A es la probabilidad de conmutación de la salida (cambio de estado) debido a cambios en la entrada, en un período. En el caso de un inversor al que se aplica una señal cuadrada en la entrada, el factor de actividad es 1.

## 7.5. Memorias semiconductoras

Las memorias electrónicas comerciales actuales se basan principalmente en el uso de transistores MOSFET, debido a la alta densidad de integración que puede alcanzarse con la tecnología CMOS.

Las memorias semiconductoras pueden clasificarse en memorias volátiles y no volátiles, como se muestra en la figura 7.14. Las memorias volátiles no pueden almacenar los datos una vez que se ha removido la alimentación del circuito. Las memorias no volátiles, por el contrario, pueden retenerlos hasta por 10 años, de acuerdo con los estándares comerciales.

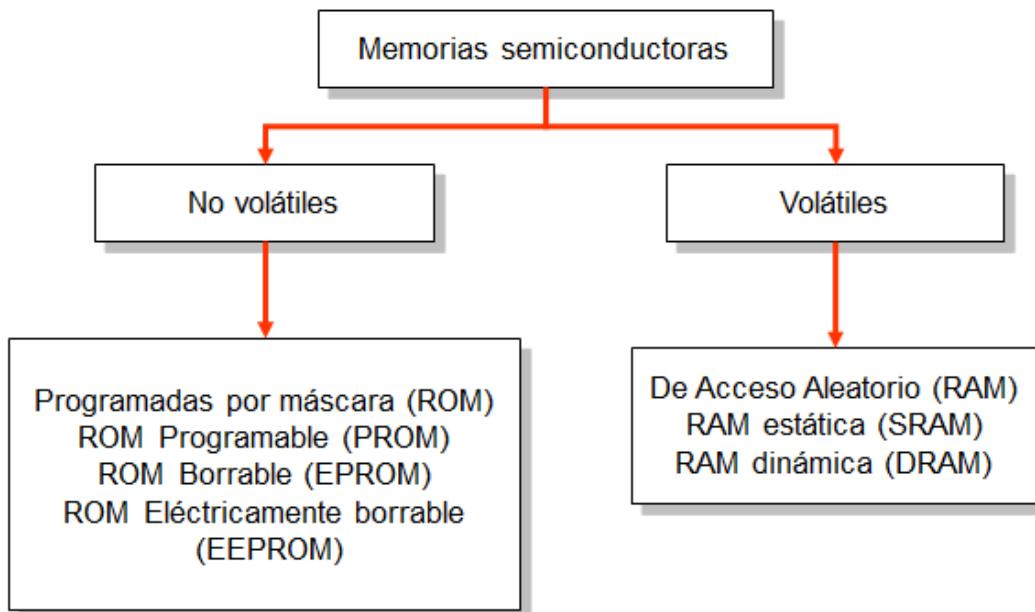


Figura 7.14: Clasificación de memorias semiconductoras.

### 7.5.1. Organización de las memorias

Para almacenar los datos, las memorias se organizan como una matriz, como muestra la figura 7.15. Cada bit se almacena en la intersección de una fila y una columna, es decir, en cada intersección se localiza un elemento de almacenamiento de datos. A las columnas se les conoce como bitlines (líneas de bit) y a las filas se les conoce como wordlines (líneas de palabra). Para leer un bit individual debe indicarse la posición del dato utilizando la fila y la columna correspondiente. Para leer una palabra (un conjunto de bits), debe indicarse la fila correspondiente y activar todas las líneas de bit de esa fila. Así, la dirección completa de un dato en una memoria se compone de bits de fila y bits de columna. Para formar palabras, se utiliza un conjunto de matrices y de cada matriz se lee un bit de la palabra, como muestra la figura 7.16.

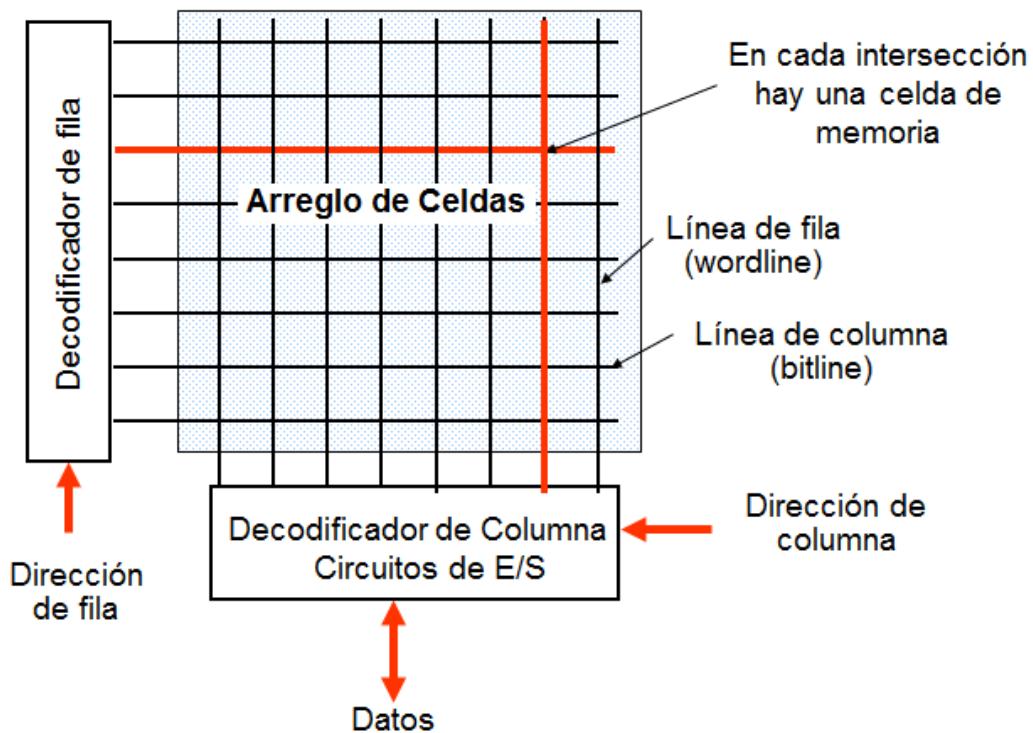


Figura 7.15: Organización de un arreglo de memoria.

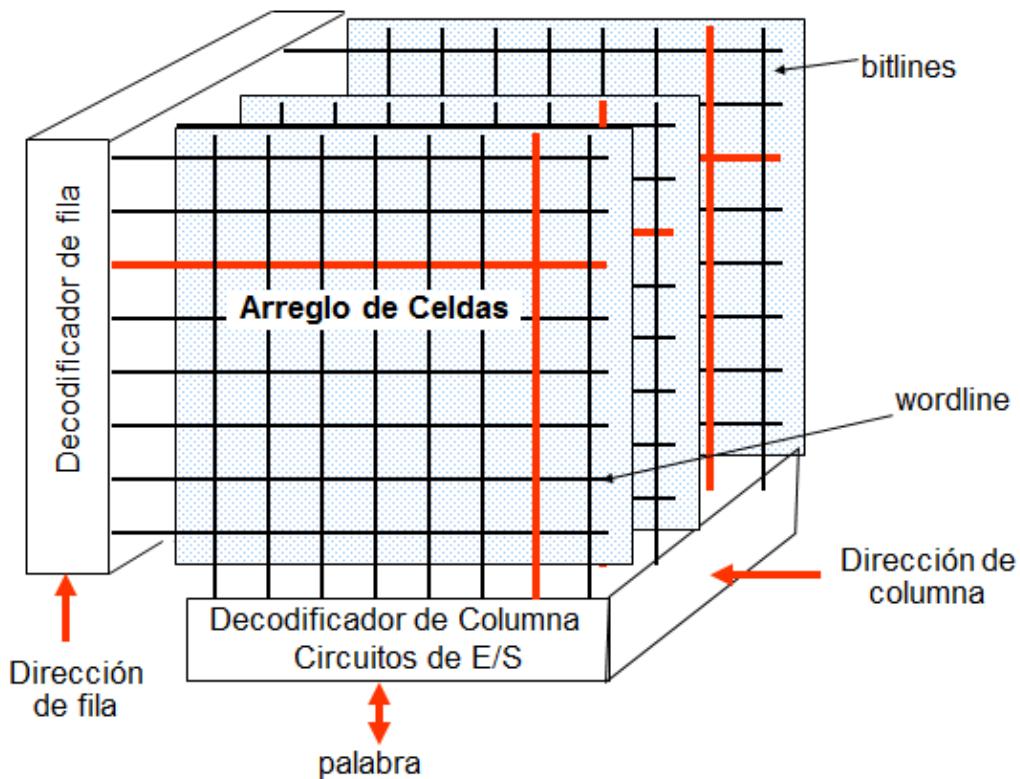


Figura 7.16: Organización de una memoria para formar una palabra.

### 7.5.2. Memorias no volátiles

Existen varios tipos de memorias no volátiles, de acuerdo con las posibilidades de programación y borrado. Las memorias no volátiles se identificaron inicialmente con el acrónimo ROM (read only memory), pues el usuario final no podía borrarlas ni escribirlas, sino sólo leerlas. Con el tiempo se implementaron mecanismos de escritura eléctrica y borrado eléctrico u óptico de estas memorias, aunque el acrónimo ROM se mantiene como parte del nombre de este tipo de memorias para indicar su característica de no volatilidad. Estas son las memorias que se utilizan en USB/memory sticks (llaves mayas), memorias de cámaras y otros aparatos electrónicos, almacenamiento de BIOS y similares.

Las memorias no volátiles se clasifican de acuerdo con sus métodos de programación y borrado, como sigue:

- Memorias programadas por máscara (ROM)
- Memorias ROM programables eléctricamente (PROM)
- Memorias ROM programables eléctricamente y borrables ópticamente (EPROM)
- Memorias ROM programables y borrables eléctricamente (EEPROM)

Las memorias programadas por máscara se programan durante el proceso de fabricación. Una forma de diferenciar entre los estados de cero y uno lógico consiste en fabricar transistores MOS-FET de canal N, uno de mayor tensión de umbral para uno de los estados y otro con menor tensión

de umbral para el otro estado. Esto puede lograrse creando transistores con diferente espesor del aislante de la compuerta o bien ajustando las tensiones de umbral por medio de implantación iónica. En el primer caso, un transistor con aislante más grueso presenta una mayor tensión de umbral que un transistor de aislante delgado. En el segundo caso, se utilizan transistores de enriquecimiento y de empobrecimiento para representar los estados lógicos. Para crear el transistor de empobrecimiento, se dopa la región bajo la compuerta para crear el canal y con ello obtener una tensión de umbral negativa.

Otra forma de obtener memorias no volátiles programadas por máscara es contar con una matriz de transistores MOSFET en los cuales los dos estados lógicos se representan por la presencia/ausencia de una interconexión al drenador del transistor, permitiendo o impidiendo el flujo de corriente, respectivamente.

En las memorias programables eléctricamente se cuenta con transistores MOSFET de canal N con doble compuerta a los cuales se inyectan huecos o electrones los cuales quedan atrapados en el óxido de compuerta. La inyección de huecos permite disminuir la tensión de umbral y eventualmente volverla negativa, mientras que la inyección de electrones aumenta la tensión de umbral. Así puede distinguirse entre dos estados lógicos. La inyección de portadores de carga en la compuerta se lleva a cabo aplicando una diferencia de tensión de diferencia compuerta y drenador/surtidor que genere un campo eléctrico suficientemente alto para observar el efecto cuántico llamado efecto túnel (Fowler-Norheim o directo). Dado que los transistores tienen doble compuerta, una de ellas es flotante y permite que los portadores de carga queden atrapados en vez de fluir por el transistor como una corriente de compuerta.

Sin embargo, cada vez que se inyectan portadores de carga en el óxido de la compuerta flotante, dicho óxido sufre daños. Estos daños limitan el número de veces que la memoria se puede programar y borrar (ciclos de programación-borrado), así como el tiempo durante el cual los transistores pueden almacenar los datos. Para toda memoria programable y borrible los fabricantes deben garantizar, al menos, 100 mil ciclos de programación-borrado y una permanencia de datos en la memoria de al menos 10 años. Estas son las memorias PROM, EPROM, EEPROM y Flash, la cual es una memoria EEPROM borrible por sectores.

Para borrar los datos de una memoria no volátil puede recurrirse a luz ultravioleta o a métodos eléctricos. En caso de las memorias EPROM se utiliza luz ultravioleta para excitar los electrones inyectados en la compuerta flotante de forma que puedan regresar al substrato del transistor; este proceso de borrado dura algunos minutos (hasta 20 minutos) y consiste en exponer la memoria a luz ultravioleta, al cual radia sobre el transistor a través de una ventana de vidrio en el encapsulado de la memoria. Cuanto mayor sea el número de ciclos de escritura-borrado que haya experimentado la memoria, tanto más tiempo se requiere para completar un proceso de borrado. En las memorias borrables eléctricamente, se aplica un campo eléctrico entre compuerta y drenador/surtidor que extraiga los electrones de la compuerta flotante por medio de efecto túnel.

### Lectura de datos en memorias no volátiles

Las memorias no volátiles pueden presentar dos configuraciones: NOR y NAND. La más utilizada es la NAND debido a que permite implementar memorias más densas. En la configuración NAND, los transistores que almacenan la información están conectados en serie. Para leer el dato de un transistor en particular, se aplica a todos los demás transistores de la rama en serie una tensión  $V_G = V_{DD}$ , mientras que al transistor por leer se le aplica 0 V en la compuerta. Para determinar el

dato almacenado se sensa la corriente por la rama: si el transistor es de enriquecimiento, la corriente en la rama será cero, pues  $V_G = 0V$  no será suficiente para encender el transistor; por el contrario, si el transistor es de empobrecimiento se podrá sentir una corriente diferente de cero en la rama. De esta forma puede determinarse la presencia de un cero o un uno lógico almacenado.

### 7.5.3. Memoria SRAM

La memoria SRAM es volátil. La figura 7.17 muestra una celda de memoria SRAM. Esta consiste en dos inversores mutuamente acoplados (cerrojo) cuyas salidas están conectadas a transistores de acceso, para un total de 6 transistores. Las compuertas de los transistores de acceso están conectadas a la línea de palabra (wordline), y permiten acceder a los nodos internos de la memoria donde se leen los datos, es decir, las salidas de los inversores, conectando estos nodos a las líneas de bit (bitlines).

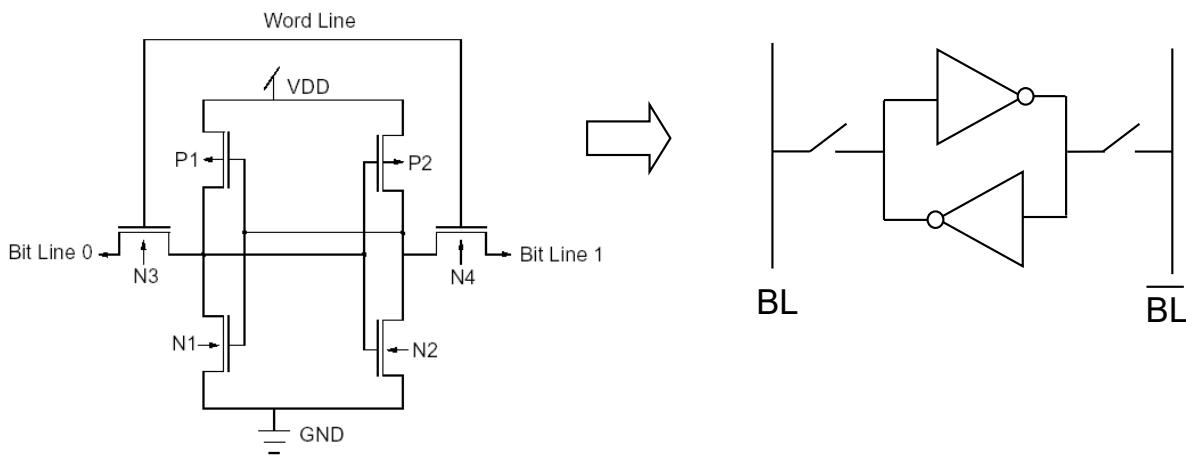


Figura 7.17: Celda de memoria SRAM.

#### Lectura de un dato de una celda de memoria SRAM

¿Cómo se lee un dato de una celda de memoria SRAM? Para este ejemplo, asumamos que tenemos un 1 lógico almacenado en el lado izquierdo de la celda y un 0 al lado derecho, de acuerdo con la figura 7.18.

El proceso es el siguiente:

a) Preparación para la lectura

Los transistores de acceso de la celda SRAM están en corte ( $WL=0$ ) para mantener los nodos internos de la celda aislados de la línea de bit. Se precargan las líneas de bit a  $V_{DD}$  (y con ello los capacitores parásitos de bitline). Para esto se activan los transistores de precarga para que las líneas de bitline se conecten a  $V_{DD}$ . Se activa el transistor de ecualización para igualar las tensiones de la línea bit. Pasado el tiempo de ecualización, se desactivan los transistores de precarga y ecualización.

b) Lectura de dato

Se activan los transistores de acceso de la celda SRAM. Se espera un tiempo determinado para que las tensiones de línea bit cambien debido a la tensión de los nodos internos de la celda,

obteniendo una diferencia de tensión de 100 mV. Puesto que hay un 0 en uno de los nodos de la celda, una línea bit empezará a descargarse. El transistor M2 está activo y empieza a descargar el capacitor de la de la línea de bit derecha. En el lado izquierdo de la celda, tanto en el nodo interno de la celda como la línea de bit están a  $V_{DD}$ , pues M5 está encendido y en saturación. En este lado la tensión no cambia. Después de un tiempo determinado, la diferencia de tensión entre la línea de bit izquierda y derecha alcanza 100 mV, como sigue:

$$\text{Lado izquierdo} = V_{DD} - 100 \text{ mV}$$

$$\text{Lado derecho} = V_{DD}$$

Cuando esto sucede se desactivan los transistores de acceso de la celda SRAM.

### c) Amplificación del dato

En esta etapa se activa el transistor de activación de sensado de la red NMOS del amplificador de sensado (Activación de sensado = 1). En el lado derecho del amplificador tiene una tensión de  $V_{DD} - 100\text{mV}$ , y el lado izquierdo a  $V_{DD}$ .

El transistor M8 está activo a ( $V_{GS8} = V_{DD}$ ) y empieza a descargar la capacitancia de bitline del lado derecho a través de M8. Conforme la tensión del lado derecho disminuye, M7 se empieza a desactivar. Ahora se activa el transistor de activación de sensado de la red PMOS del amplificador de sensado (Activación de sensado = 0) y M9 se empieza a activar.

Esto mantiene la tensión del lado izquierdo en  $V_{DD}$ , y con ello M8 se mantiene activo completando la descarga del lado derecho. Esto completa la lectura del dato, y a partir de aquí se activan circuitos de salida (no mostrados) que trasladan el dato a las patillas del circuito integrado.

Siempre se traslada sólo un dato (0 o 1) a la salida de la memoria, es decir, el dato que está conectado a Bitline (no a Bitline). El diseñador debe leer siempre del mismo lado en todas las celdas.

El mismo proceso se sigue si los voltajes están intercambiados ( $V_{DD}$  a la derecha y 0 a la izquierda), con la diferencia de que los transistores que descargan la línea de bit son los del lado izquierdo.

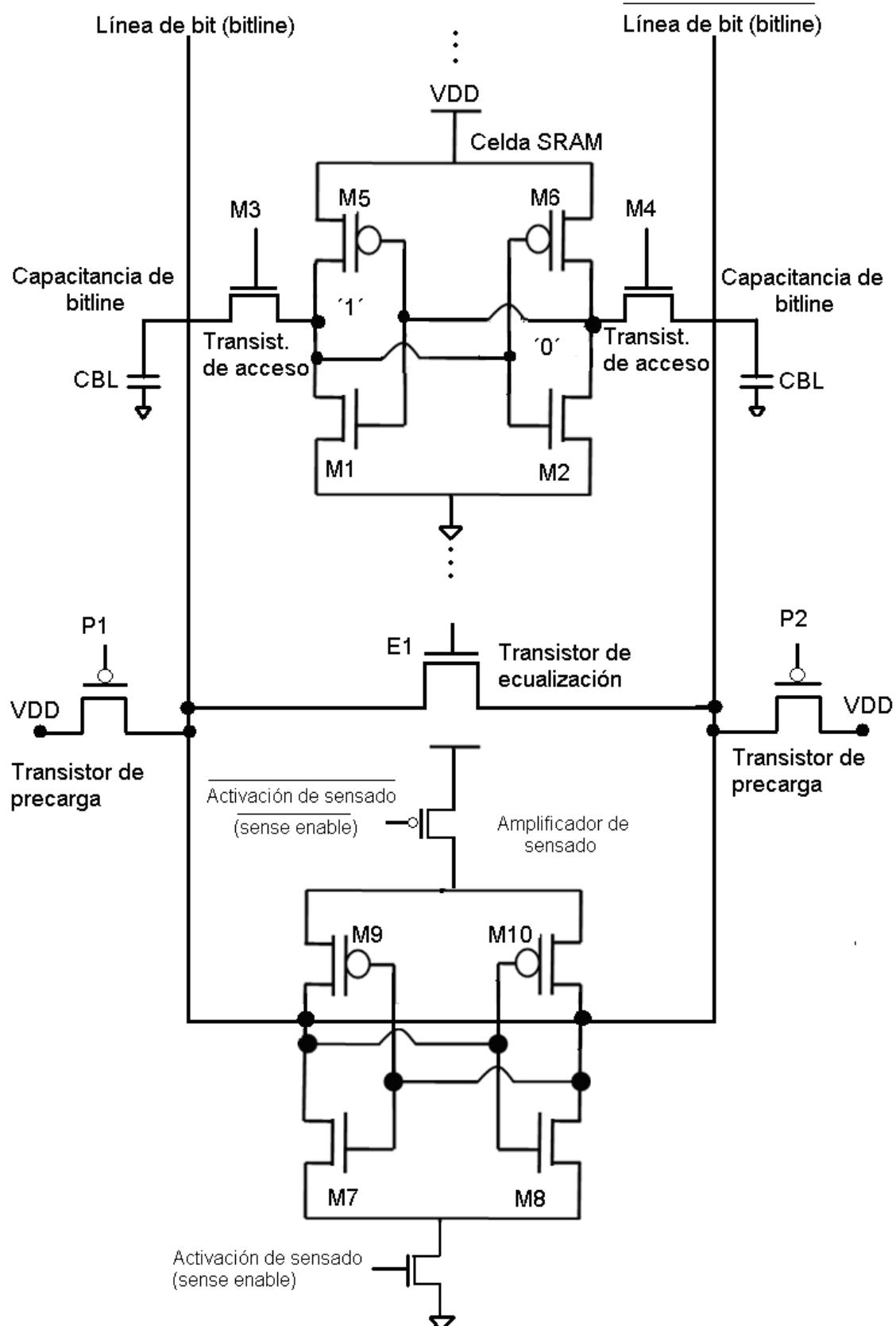


Figura 7.18: Lectura de una memoria SRAM.

### 7.5.4. Memoria DRAM

La memoria DRAM es volátil. La figura 7.19 muestra una celda de memoria DRAM. Esta consiste en un transistor de acceso y un capacitor de almacenamiento  $C_{st}$ . La compuerta del transistor de acceso está conectada a la línea de palabra (wordline), y permite acceder al capacitor de almacenamiento, conectando este capacitor a la línea de bit (bitline).

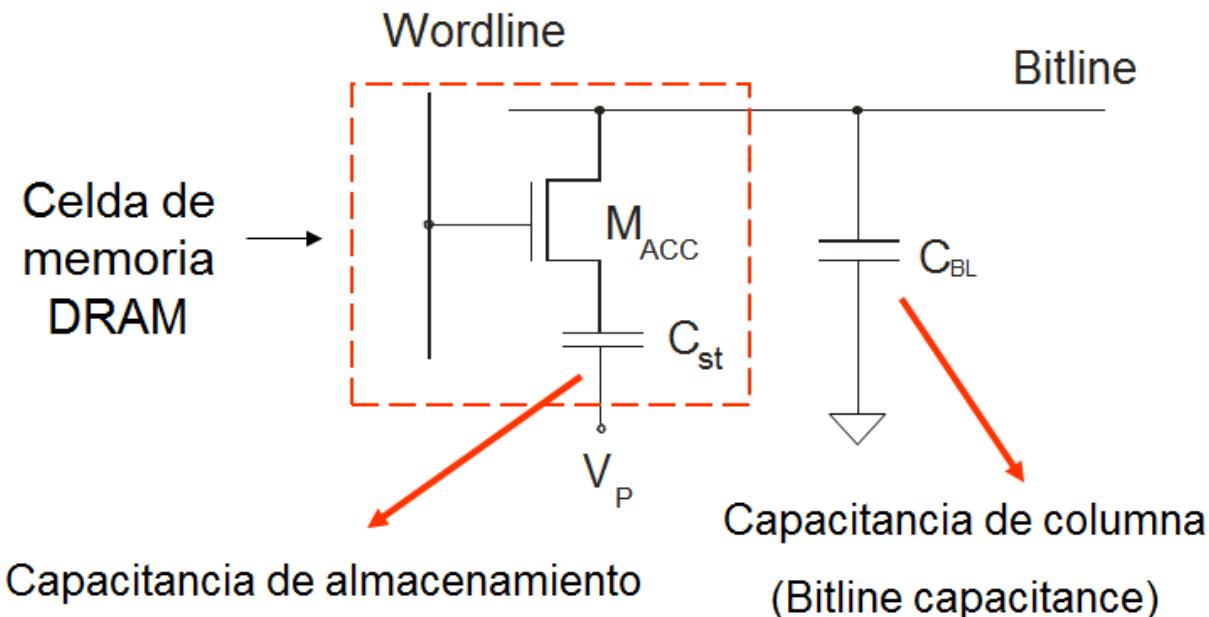


Figura 7.19: Celda de memoria DRAM.

#### Lectura de un dato de una celda de memoria DRAM

¿Cómo se lee un dato en una memoria DRAM? Para este ejemplo, asumamos que tenemos un 1 lógico almacenado en la celda de memoria conectada en Bitline, de acuerdo con la figura 7.20. La operación de lectura ocurre en tres etapas, explicadas a continuación.

##### a) Preparación para la lectura

En esta etapa se inicia con la activación de los transistores de precarga P1 y P2. Con esto, las líneas de bit se precargan a  $V_{DD}/2$ . Después de esto se desactivan los transistores de precarga y se activa el transistor de ecualización, para igualar ambas tensiones de la línea de bit a  $V_{DD}/2$ .

##### b) Lectura del dato

Se activa solamente la línea de palabra de la celda de memoria que contiene el dato a leer. La otra línea de palabra (Wordline) se deja en 0 para que el transistor de acceso de esa celda opere en corte, pero que la tensión de precarga en Bitline sirva como tensión de comparación para el amplificador de sensado.

Al activar la línea de palabra para leer el dato, el transistor de acceso se activa, conectando la línea de bit precargada a  $V_{DD}/2$  con el capacitor de almacenamiento  $C_{ST}$ . Si el transistor de acceso activo se considera como un corto circuito, inmediatamente después de activar el transistor de acceso se obtiene el circuito de la figura 7.21. Si el dato almacenado era un cero lógico, el capacitor de bitline se descarga 100mV, en el caso de un uno lógico se carga 100 mV.

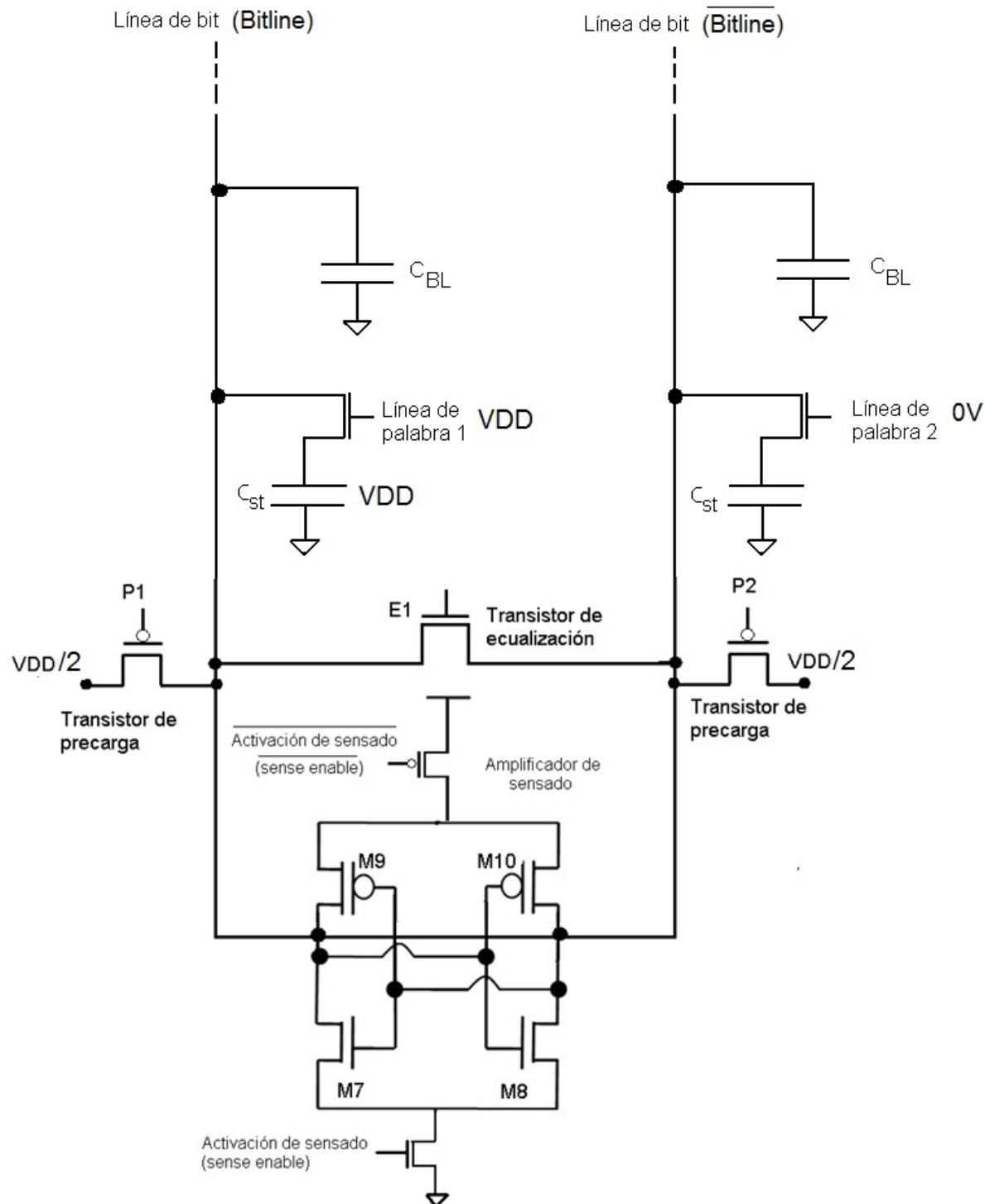


Figura 7.20: Lectura de una memoria DRAM.

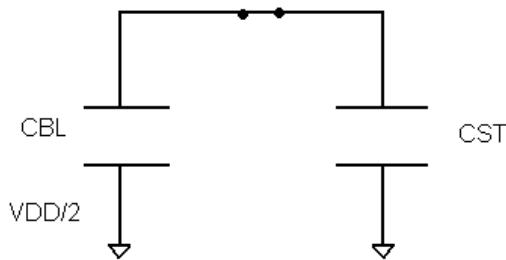


Figura 7.21: Circuito equivalente para la lectura de un dato de una celda de memoria DRAM.

$$Q_{\text{initial}} = Q_{\text{final}} \quad (7.4)$$

$$Q_{\text{final}} = (C_{\text{ST}} + C_{\text{BL}}) \cdot V_{\text{ST}} \quad (7.5)$$

$$Q_{\text{initial}} = C_{\text{BL}} \cdot V_{\text{precharge}} \quad (7.6)$$

$$Q_{\text{final}} = C_{\text{BL}} \cdot V_{\text{BL}} = C_{\text{BL}} \left( \frac{V_{\text{DD}}}{2} \pm 100\text{mV} \right) \quad (7.7)$$

### c) Amplificación del dato

En esta etapa se activa el transistor de activación de sensado de la red NMOS del amplificador de sensado (Activación de sensado = 1). En un lado del amplificador se tiene una tensión de  $V_{\text{DD}}$  100 mV, y en el otro  $V_{\text{DD}}/2$ . Para continuar esta explicación, asuma que la celda que se está leyendo está conectada al lado izquierdo del amplificador de sensado y que el dato almacenado es un 1 lógico.

La compuerta del transistor M8 tiene entonces una tensión de  $V_{\text{DD}}/2 + 100\text{mV}$  ( $V_{\text{GS}8} = V_{\text{DD}}/2 + 100\text{mV}$ ) y la capacitancia de Bitline se descarga más rápido la que la capacitancia de Bitline. Conforme la tensión del lado derecho disminuye, M7 se empieza a desactivar. Ahora se activa el transistor de activación de sensado de la red PMOS del amplificador de sensado, es decir, (Activación de sensado = 0) y M9 conduce cada vez más.

Esto empieza a aumentar la tensión del lado izquierdo hacia  $V_{\text{DD}}$ , y con ello M8 se mantiene activo completando la descarga del lado derecho, mientras que M9 completa la carga del lado izquierdo. Esto completa la lectura del dato, y a partir de aquí se activan circuitos de salida (no mostrados) que trasladan el dato de Bitline a las patillas del circuito integrado.

El mismo proceso se sigue si los voltajes están intercambiados ( $V_{\text{DD}}$  a la derecha y 0 a la izquierda), con la diferencia de que los transistores que descargan la línea de bit son los del lado izquierdo y los que la cargan son los del lado derecho.

## 7.6. Modelo de Pequeña Señal

El modelo de pequeña señal se utiliza para hacer cálculos analíticos del funcionamiento analógico de circuitos de amplificación de señales. Se puede decir que el modelo de pequeña señal es el cálculo de señales AC de pequeña amplitud.

Para hacer un modelo de pequeña señal, primero se debe determinar si el modelo aplica a las características del circuito que se va a analizar. Esto quiere decir que el punto de operación del circuito debe corresponder a la polarización que se requiere para el tipo de transistor que se analiza, tal como se muestra en la tabla 1. Este primer análisis es un análisis de Gran Señal donde se contemplan las características de CD que permiten determinar cómo está polarizado el circuito.

Tabla 1. Análisis de Gran Señal (CD)

Tipo de Transistor	Zona de Operación	Especificaciones de la zona de operación	Ecuaciones aplicables
MOSFET	Saturación	$V_{GS} \geq V_{TH}$ $V_{DS} \geq V_{GS} - V_{TH}$	$I_D = \frac{K}{2} (V_{GS} - V_{TH})^2$ $I_D = \frac{K}{2} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS})$ - Leyes de Kirchoff - Thevenin/Norton
BJT	Activa	$V_{BE} \geq V_{DIODO}$ $V_{CE} \geq V_{BE}$	$I_C = \beta I_B$ $I_E = I_C + I_B = (\beta + 1)I_B$ $I_C = \alpha I_E$ $\alpha = \frac{\beta}{\beta + 1}$ - Leyes de Kirchoff - Thevenin/Norton

Si con el análisis de gran señal se determina el punto de operación del circuito es el requerido para el modelo, entonces se procede a calcular los parámetros de pequeña señal, como se muestra en la tabla 2.

Tabla 2. Cálculo de parámetros para el modelo de pequeña señal

Transistor	Parámetro	Ecuación
MOSFET	$g_m$ (Transconductancia)	$g_m = K(V_{GS} - V_{TH})$ $g_m = \sqrt{2KI_D}$ $g_m = \frac{2I_D}{V_{GS} - V_{TH}}$
	$r_o$ (resistencia de salida)	$r_o = \frac{1}{\lambda I_{DS,SAT}} = \frac{V_A}{I_{DS,SAT}}$
BJT	$g_m$ (Transconductancia)	$g_m = \frac{I'_C}{V_T} = \frac{\beta I_B}{V_T}$
	$r_o$ (resistencia de salida)	$r_o = \frac{V_A}{I'_C}$
	$r_\pi$ (resistencia de la base)	$r_\pi = \frac{\beta}{g_m}$

Recuerde que:  $I_{DS,SAT}$ ,  $I_C$  son corrientes sin tomar en cuenta el efecto Early sobre el circuito, o sea, sin modulación del canal (MOSFET) ni de la base (BJT).

Una vez calculados los parámetros de pequeña señal, puede sustituir el modelo del transistor equivalente dentro del circuito, como se muestra en la figura 7.22.

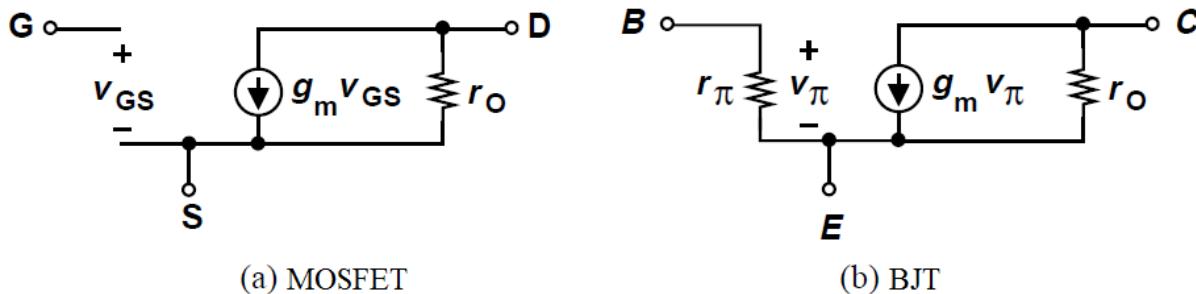


Figura 7.22: Modelo de pequeña señal (a) MOSFET y (b) BJT.

### 7.6.1. Reglas para cálculos en Gran Señal

1. Los capacitores son circuitos abiertos para CD.

2. Se contemplan las fuentes de voltaje CD ( $V_{DD}$ ,  $V_{CE}$ ).
3. Se contemplan las fuentes de corriente CD ( $I_D$ ,  $I_C$ ).
4. Se determina si el punto de operación del transistor es el correcto para el modelo (saturación para MOSFET, activo para BJT).
5. Si el punto de operación es correcto, se determina la corriente del circuito sin efecto Early ( $I_{DS,SAT}$  para el MOSFET,  $I_C$  para el BJT).
6. Calcular los parámetros de pequeña señal.

### 7.6.2. Reglas para cálculos en Pequeña Señal

1. Los capacitores son corto circuitos para CA.
2. Las fuentes de voltaje de CD se igualan a cero V ( $V_{DD} = 0V$ ,  $V_{CE} = 0V$ ), lo que equivale a convertirlas en corto circuitos.
3. Las fuentes de corriente CD se igualan a cero A ( $I_D = 0A$ ,  $I_C = 0A$ ), lo que equivale a convertirlas en circuitos abiertos.
4. Si se desea calcular la ganancia de voltaje  $A_V$  del amplificador, se hace plantea una ecuación que relacione la entrada de voltaje de pequeña señal  $v_{in}$  con el voltaje de entrada del transistor ( $v_{gs}$  o  $v_{be}$ ); luego se plantea una ecuación que relacione el voltaje de salida  $v_{out}$  con la fuente dependiente del transistor ( $g_m v_{gs}$  o  $g_m v_{be}$ ). Luego se despeja ( $v_{gs}$  o  $v_{be}$ ) en ambas ecuaciones y se igualan, de manera que se puede encontrar la razón  $v_{in}/v_{out}$ .
5. Si se desea calcular la resistencia de entrada  $R_{in}$ , ésta puede determinarse poniendo en corto circuito la salida del circuito, de esta forma puede calcular  $R_{in} = V_{in}/I_{in}$ .
6. Si se desea determinar la resistencia de salida  $R_{out}$ , calcule el voltaje de circuito abierto  $V_{OC}$  y la corriente de corto circuito  $I_{SC}$  en la salida, de esta forma puede calcular la resistencia de Thevenin  $R_{TH} = V_{OC}/I_{SC}$ .



# Capítulo 8

## Modelado del MOSFET

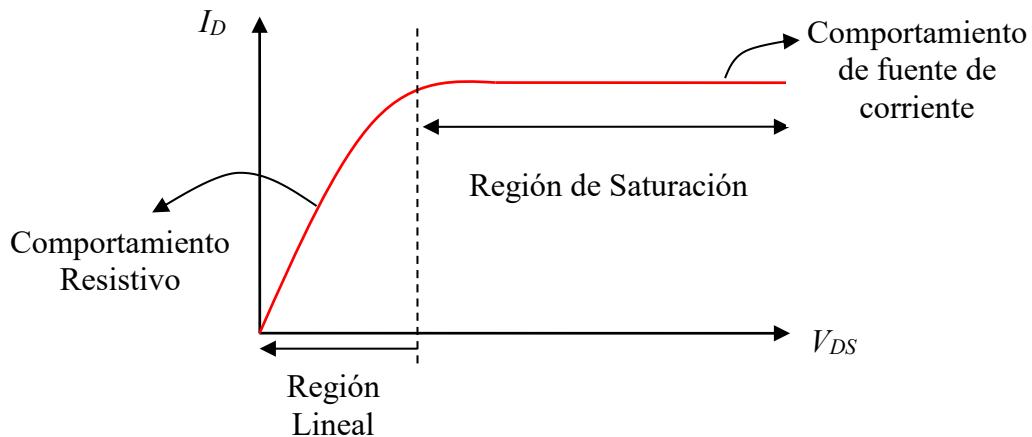
### 8.1. Modelo analógico: gran señal

Con base en la curva IV de salida del MOSFET, pueden distinguirse dos importantes regiones de operación para aplicaciones analógicas:

- La región lineal
- La región de saturación

#### a) Región lineal

En la región lineal el transistor se comporta como una resistencia (controlada), como lo muestra la curva IV de Salida ( $I_D$  vrs  $V_{DS}$ ).



El valor de la resistencia del MOSFET en la región lineal se obtiene con la definición de resistencia:

$$R = \frac{\partial V}{\partial I} \quad (8.1)$$

en este caso

$$R_{ch} = \left( \frac{\partial I_D}{\partial V_{DS}} \right)^{-1} \quad (8.2)$$

donde  $R_{ch}$  es la resistencia del MOSFET en la zona lineal y es llamada resistencia de canal. Por lo tanto, con la ecuación de  $I_D$  en la región lineal

$$I_D = K \left( V_{GS} - V_{TH} - \frac{V_{DS}}{2} \right) V_{DS} \quad (8.3)$$

se tiene que:

$$\frac{\partial I_D}{\partial V_{DS}} = K(V_{GS} - V_{TH} - V_{DS}) \quad (8.4)$$

por lo tanto:

$$R_{ch} = \frac{1}{K(V_{GS} - V_{TH} - V_{DS})} \quad (8.5)$$

lo cual para  $V_{DS} \ll V_{DS,sat}$  puede aproximarse a:

$$R_{ch} \approx \frac{1}{K(V_{GS} - V_{TH})} \quad (8.6)$$

### b) Región de saturación

En la región de saturación la corriente es, en el caso ideal, independiente del voltaje entre drenador y surtidor, manteniendo un valor constante. Este comportamiento corresponde al de una fuente de corriente. En la práctica, la modulación de ancho de canal se manifiesta como un incremento de la corriente de drenador con valores crecientes de  $V_{DS}$ . Este comportamiento puede modelarse por medio de una fuente de corriente real.

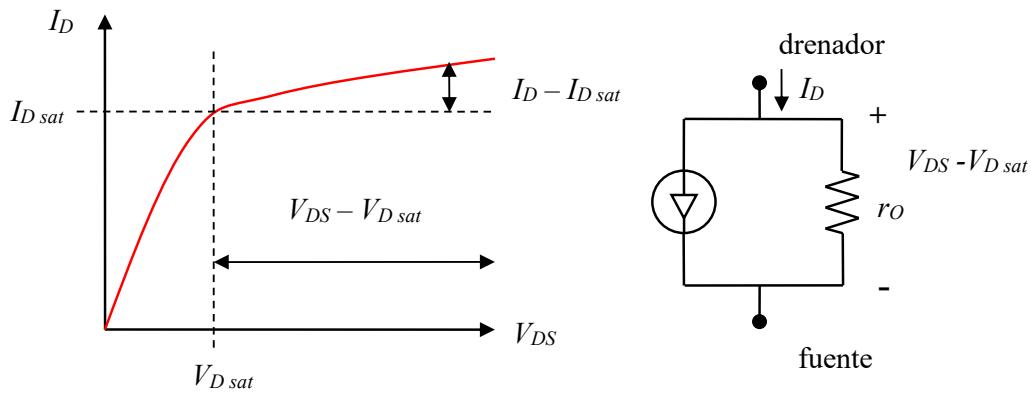
El paso de la región lineal a la región de saturación está determinado por el voltaje de saturación  $V_{D,sat}$ :

$$V_{D,sat} = V_{GS} - V_{TH}, V_{GS} \geq V_{TH} \quad (8.7)$$

Comparando el modelo de una fuente de corriente real y la ecuación de la corriente de drenador en la región de saturación, se tiene:

$$I_D = \underbrace{\frac{1}{2} K(V_{GS} - V_{TH})^2}_{\text{constante}} + \underbrace{\frac{1}{2} K(V_{GS} - V_{TH})^2 \square (V_{DS} - V_{D,sat})}_{I = f(V_{DS}) \rightarrow \text{resistencia}}$$

El término independiente de  $V_{DS}$  representa una fuente de corriente ideal, mientras que el término dependiente de  $V_{DS}$  representa una resistencia, según el siguiente circuito equivalente:



donde

$$r_o = \frac{1}{2I_{D,sat}} \quad (8.8)$$

con

$$I_{D,sat} = \frac{1}{2}K(V_{GS} - V_{TH})^2 \quad (8.9)$$

## 8.2. Modelo analógico: pequeña señal

El modelo de pequeña señal es una representación de un dispositivo operando en condiciones de corriente alterna, omitiendo las componentes de corriente continua utilizadas para la polarización del dispositivo. En el modelo de pequeña señal, interesa solamente la respuesta del dispositivo ante fuentes de excitación alterna.

Al polarizar un dispositivo para que opere en una región determinada de su característica IV, se establecen condiciones de corriente continua conocidas como el punto de operación. La superposición de fuentes de excitación de corriente alterna genera variaciones del voltaje y corriente de operación del dispositivo alrededor del punto de operación.

El efecto de estas variaciones como por ejemplo, ganancias de voltaje, se estudia con el modelo de pequeña señal. El concepto de punto de operación (punto de polarización) y la superposición de señales alternas se ilustra en la figura 8.1.

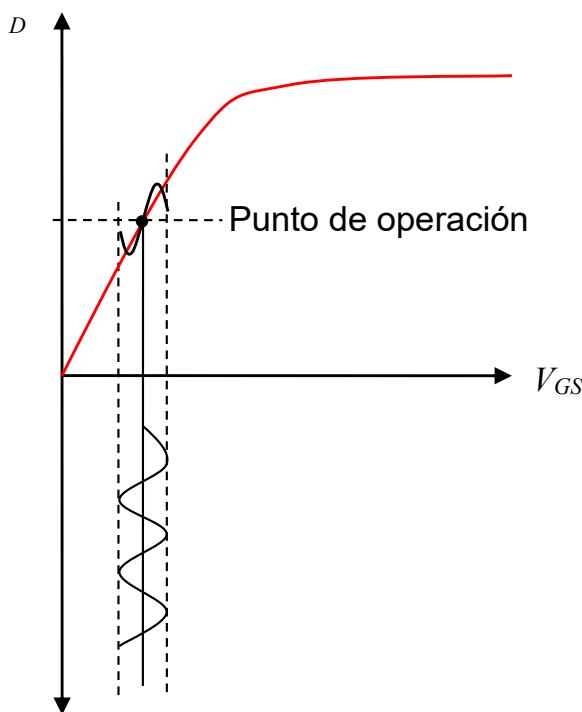


Figura 8.1: Punto de operación.

### 8.2.1. Obtención del equivalente de pequeña señal

Los pasos a seguir para obtener el equivalente de pequeña señal son:

1. Determinar el punto de operación de los elementos del circuito utilizando análisis de circuitos y las ecuaciones características de los dispositivos del circuito.
2. Con base en los puntos de operación obtenidos, calculen los parámetros de corriente alterna de los dispositivos que forman el circuito.
3. Reemplazar las fuentes de corriente directa (CD) por cortocircuitos y las fuentes de corriente CD por circuitos abiertos.
4. Reemplazar los elementos activos del circuito por sus equivalentes de pequeña señal de acuerdo con los parámetros calculados en el punto 2.

### 8.2.2. Transconductancia

La transconductancia de un MOSFET se define como:

$$g_m = \frac{i_d}{v_{gs}} \quad (8.10)$$

evaluado a  $i_D$  y  $v_{GS}$  constantes.

Aplicando la definición:

$$g_m = \sqrt{2KI_D} \quad (8.11)$$