

POLITEXT

Lluís Prat Viñas
Josep Calderer Cardona

Dispositivos electrónicos y fotónicos. Fundamentos

EDICIONS UPC

Dispositivos electrónicos y fotónicos. Fundamentos

POLITEXT

Lluís Prat Viñas
Josep Calderer Cardona

Dispositivos electrónicos y fotónicos. Fundamentos

EDICIONS UPC

Primera edición: enero de 2003
Segunda edición: marzo de 2006

Diseño de la cubierta: Manuel Andreu

© Los autores, 2003

© Edicions UPC, 2003
Edicions de la Universitat Politècnica de Catalunya, SL
Jordi Girona Salgado 31, 08034 Barcelona
Tel.: 93 401 68 83 Fax: 93 401 58 85
Edicions Virtuals: www.edicionsupc.es
A/e: edicions-upc@upc.edu

Producción: Ediciones Gráficas Rey
Albert Einstein 54, C/B, nave 15,
08940 Cornellà de Llobregat

Depósito legal: B-15933-2006
ISBN: 84-8301-854-3

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

Índice

1 Propiedades eléctricas de los semiconductores

1.1 Materiales semiconductores	11
1.2 Concentración de portadores en el equilibrio. El nivel de Fermi	26
1.3 Corrientes en los semiconductores	37
1.4 Ecuaciones de continuidad	47
1.5 Campos y cargas en un semiconductor	57
1.6 Diagrama de bandas de energía de un semiconductor	63
Apéndice 1.1 Efectos de alto dopaje	68
Apéndice 1.2 Algunas implicaciones de la naturaleza cuántica del electrón	70
Problemas propuestos	73
Formulario del capítulo 1	75

2 La unión PN

2.1 La unión PN: bandas de energía y efecto rectificador	77
2.2 Análisis de la zona de carga espacial de la unión PN	85
2.3 Distribución de portadores y de corrientes en régimen permanente	94
2.4 Característica corriente-tensión del diodo	104
2.5 Modelo dinámico del diodo	114

2.6	El diodo en conmutación y en pequeña señal.....	123
2.7	Contactos metal-semiconductor	128
2.8	El diodo de heterounión	138
Apéndice 2.1	La velocidad de recombinación superficial	142
	Problemas propuestos	143
	Formulario del capítulo 2	145
3	Tecnología de fabricación	
3.1	Integración de un circuito en silicio	147
3.2	Etapas de los procesos de fabricación	157
3.3	Secuencia de las etapas de fabricación de un circuito integrado	174
	Notas para una perspectiva histórica	179
4	Dispositivos optoelectrónicos	
4.1	Radiación electromagnética y semiconductores.....	181
4.2	Dispositivos receptores de radiación.....	193
4.3	Dispositivos emisores de radiación	210
4.4	Otros dispositivos optoelectrónicos	224
	Problemas propuestos	232
	Formulario del capítulo 4	233
5	El transistor bipolar	
5.1	Introducción	235
5.2	El transistor bipolar ideal en régimen permanente.....	245
5.3	El transistor bipolar real	258

5.4 El transistor bipolar en régimen dinámico	270
5.5 El transistor bipolar como amplificador	281
5.6 Otros transistores bipolares	292
Problemas propuestos	298
Formulario del capítulo 5	299
6 Transistores de efecto de campo	
6.1 Electrostática del sistema metal-óxido-semiconductor	302
6.2 El transistor MOS en polarización continua.....	323
6.3 Símbolos circuitales	332
6.4 Tecnología MOS	335
6.5 Efectos no ideales en transistores MOS	341
6.6 Capacidades en el transistor MOS.....	349
6.7 Modelos dinámicos del transistor MOS	354
6.8 Otros transistores de efecto de campo: el JFET y el MESFET	359
Apéndice 6.1 Relación entre carga y potencial en el semiconductor.....	374
Apéndice 6.2 Modelos SPICE para transistores MOS.....	377
Problemas propuestos	379
Formulario del capítulo 6	380
Apéndice A. Resolución de ecuaciones diferenciales.....	381
Apéndice B. Constantes, unidades y parámetros.....	386
Apéndice C. El cuestionario interactivo DELFOS	387
Índice alfabético.....	391

1 Propiedades eléctricas de los semiconductores

El objetivo de este capítulo es la introducción del lector al conocimiento de los semiconductores y de sus propiedades eléctricas fundamentales, que le permita emprender en los capítulos siguientes el estudio de los dispositivos electrónicos realizados con semiconductores, lo cual, en definitiva, es el objetivo de este libro. Se comienza con una breve descripción de los semiconductores y de su dopaje, haciendo especial hincapié en el silicio. Se estudia después una propiedad de importancia fundamental, como es la cantidad de cargas móviles que pueden originar corriente en el semiconductor, denominadas *portadores*. Se analizan los mecanismos por los cuales esos portadores inducen una corriente eléctrica, y se llega a la formulación de una ecuación de importancia clave para los dispositivos semiconductores: la ecuación de continuidad. Finalmente se presenta la relación entre cargas, campos eléctricos, potencial y bandas de energía, que permite emprender en el capítulo siguiente el estudio de la unión PN, la estructura básica para fabricar dispositivos.

1.1 Materiales semiconductores

1.1.1 Introducción

Los materiales semiconductores ocupan una posición intermedia en la escala de conductividades entre los conductores y los aislantes. La resistividad de los buenos conductores, como el cobre, es del orden de $10^{-6} \Omega\text{-cm}$, y la de los buenos aislantes supera los $10^{12} \Omega\text{-cm}$, mientras que la de los semiconductores ocupa prácticamente todo el intervalo limitado por los dos valores anteriores. Los primeros estudios sobre los semiconductores fueron realizados por Tomas Seebeck en 1821, y las primeras aplicaciones se deben a Werner von Siemens (1875, fotómetro de selenio) y a Alexander Graham Bell (1878, sistema de comunicación telefónica). No obstante, esos materiales no tuvieron un papel importante en el mundo de la electrónica hasta 1947, cuando se descubrió el transistor bipolar. Desde entonces, los nombres *electrónica* y *semiconductores* han quedado unidos de forma indisoluble.

En la tabla 1.1 se muestra una parte de la tabla periódica en la que aparecen los principales semiconductores. Se indica en cada celda el número atómico del elemento. Debe recordarse que los elementos de la columna II tienen dos electrones de valencia, mientras que los de la columna III tienen tres, y así sucesivamente.

En la tabla 1.2 se muestran los principales semiconductores utilizados actualmente en aplicaciones electrónicas. Obsérvese que aparecen semiconductores simples como el silicio (Si) y el germanio (Ge) y semiconductores compuestos. Entre éstos se hallan los binarios IV-IV, III-V y II-VI —formados por parejas de elementos procedentes cada uno de ellos de las columnas indicadas— y las aleaciones constituidas por tres o más elementos, como son los compuestos ternarios y cuaternarios. En esos compuestos, x e y indican el tanto por uno del elemento considerado.

Tabla 1.1 Parte de la tabla periódica en la que figuran los elementos que juegan un papel importante en la electrónica de los semiconductores

II	III	IV	V	VI
⁴ Be	⁵ B	⁶ C	⁷ N	⁸ O
¹² Mg	¹³ Al	¹⁴ Si	¹⁵ P	¹⁶ S
³⁰ Zn	³¹ Ga	³² Ge	³³ As	³⁴ Se
⁴⁸ Cd	⁴⁹ In	⁵⁰ Sn	⁵¹ Sb	⁵² Te
⁸⁰ Hg	⁸¹ Tl	⁸² Pb	⁸³ Bi	⁸⁴ Po

Tabla 1.2 Tipos de semiconductores utilizados en aplicaciones electrónicas

Tipos de semiconductores	Ejemplos
Semiconductores simples	Si, Ge
Semiconductores compuestos IV-IV	SiC, SiGe
Semiconductores compuestos III-V	GaAs, GaP, GaSb, AlAs, AlP, AlSb, InAs, InP, InSb
Semiconductores compuestos II-VI	ZnS, ZnSe, ZnTe, CdS, CdSe, CdTe
Aleaciones	$Al_xGa_{1-x}As$, $GaAs_{1-x}P_x$, $Hg_{1-x}Cd_xTe$, $Ga_xIn_{1-x}As_{1-y}P_y$

El semiconductor más utilizado en la electrónica actual es el silicio —en un porcentaje superior al 95 %— pero los semiconductores compuestos empiezan a jugar un papel cada vez más significativo en aplicaciones de alta velocidad y en la optoelectrónica. Por ese motivo, en este libro se considera al silicio como semiconductor de referencia, si no se indica otra cosa explícitamente.

1.1.2 La estructura cristalina

Un semiconductor se denomina *amorfo* cuando sus átomos no siguen una ordenación espacial más allá de unos pocos átomos. Contrariamente, cuando todos los átomos están perfectamente ordenados, siguiendo una estructura básica que se repite indefinidamente en las tres direcciones del espacio, se dice que es un *monocristal*. Cuando el material está constituido por un aglomerado de granos cristalinos, se dice que presenta una estructura *poli-cristalina*.

El silicio es un elemento con 14 electrones, como indica su número atómico. Diez de esos electrones ocupan órbitas muy próximas al núcleo, y están tan ligados a éste que prácticamente no cambian su estado durante las interacciones normales entre átomos. No sucede lo mismo con los cuatro más externos, llamados *electrones de valencia*, los cuales participan activamente en las interacciones con los demás átomos. Por ese motivo, se dice que el silicio es un átomo *tetravalente*.

Cuando el silicio forma un monocristal, cada átomo se une a cuatro átomos contiguos mediante cuatro enlaces covalentes, en las direcciones que se muestran en la figura 1.1.a. Un enlace covalente está formado por un par de electrones compartidos por los dos átomos. En el silicio, cada uno de los átomos unidos aporta un electrón de valencia para formar un enlace. Esa estructura básica se repite por todo el espacio, tal y como se muestra en la figura 1.1.b, en la que se representa la celda básica del silicio. Como se puede observar, cada átomo de la celda básica está unido a cuatro átomos contiguos mediante cuatro enlaces covalentes, siguiendo la estructura de la figura 1.1.a.

La celda básica es el menor volumen representativo del cristal, que repetido indefinidamente en las tres direcciones del espacio origina el monocristal. Para describir esa celda se emplean dos elementos: la red cristalina y el grupo atómico. La red es un conjunto de puntos, denominados *nodos*; vinculado a

cada nodo se halla el grupo atómico. De ese modo, el grupo atómico se repite en el espacio siguiendo la distribución marcada por la red.

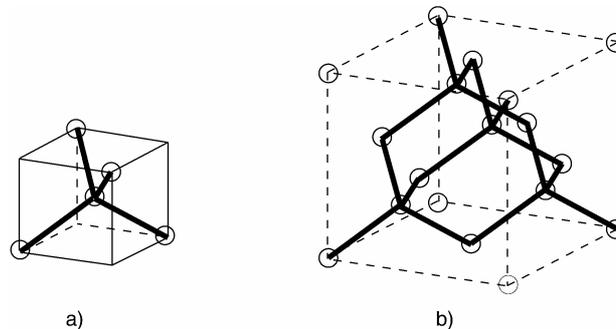


Fig. 1.1 a) Estructura de enlaces entre átomos. b) Celda unitaria del silicio. En la figura, cada esfera representa un átomo, y cada segmento entre esferas un enlace covalente.

La red cristalina del silicio es un cubo con nodos en los ocho vértices y en el centro de cada una de las caras, denominada *estructura cúbica centrada en las caras* (en inglés, *fcc, face centered cube*). El grupo atómico está constituido por dos átomos de silicio: uno de ellos situado en el nodo, y el segundo separado del primero según la dirección marcada por la diagonal principal del cubo, a una distancia igual a $1/4$ de esa diagonal (v. fig. 1.1.b). Esa estructura cristalina se denomina *diamante*, ya que es la misma que presenta el diamante (cristal de carbono). La longitud de una arista del cubo se denomina *constante de red* (en inglés, *lattice constant*), y para el silicio es de 5.43 \AA . Teniendo en cuenta ese dato, es fácil comprobar que hay 5×10^{22} átomos de silicio por cada centímetro cúbico.

El GaAs y otros semiconductores compuestos tienen una estructura cristalina ligeramente diferente, denominada *zincblenda*. La red cristalina es también cúbica centrada en las caras, pero el grupo atómico está constituido por un átomo de galio y uno de arsénico, con la misma distribución espacial que la descrita anteriormente para el silicio. La constante de red de ese semiconductor es de 5.65 \AA . Los enlaces entre átomos son covalentes y parcialmente iónicos, dado que los átomos pentavalentes de As ceden un electrón a los átomos trivalentes de Ga para formar el enlace.

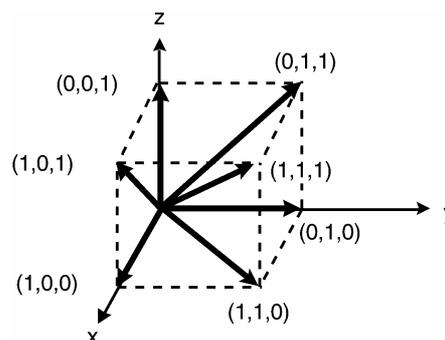


Fig. 1.2 Principales direcciones cristalinicas en el silicio. El cubo representa la celda básica.

Se pueden agrupar todos los átomos de un cristal en un conjunto de planos paralelos y equidistantes, denominados *planos cristalográficos*. Existen infinitas familias de planos cristalográficos en un

semiconductor, y cada una viene especificada por un conjunto de tres enteros denominados *índices de Miller*. En los cristales cúbicos, esos índices son proporcionales a los componentes de un vector perpendicular a los planos. En la figura 1.2 se representa la celda básica de silicio junto con algunas direcciones cristalinas. Las direcciones cristalinas son importantes para determinadas propiedades físicas y tecnológicas del semiconductor.

1.1.3 Los portadores de corriente. Modelos de enlaces y de bandas

La corriente eléctrica a través de una sección de un material se define como la carga que atraviesa esa sección por unidad de tiempo. Para que haya corriente debe haber, por lo tanto, partículas que transporten la carga; esas partículas móviles con carga eléctrica se denominan *portadores de corriente*. En los semiconductores existen dos tipos de portadores de corriente: los electrones de conducción, que tienen carga negativa, y los huecos, con carga positiva.

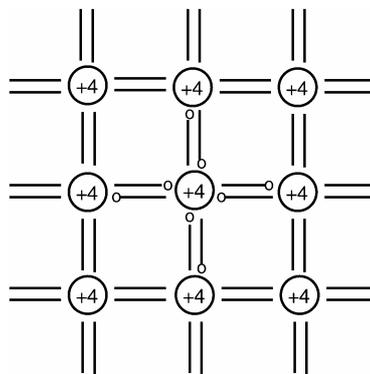


Fig. 1.3 Modelo de enlaces del semiconductor. Los electrones de valencia, representados por pequeños círculos, se dibujan solo para los enlaces del átomo central, con indicación del átomo que los aporta.

Para estudiar la naturaleza de esos portadores de corriente se debe partir de la estructura cristalina del semiconductor, descrita en el apartado anterior. Para evitar la complicación que comporta el carácter tridimensional de esa estructura, se realiza una representación simplificada bidimensional que se denomina *modelo de enlaces*, la cual se muestra en la figura 1.3. En ese modelo, cada círculo representa el núcleo más los electrones internos de un átomo del semiconductor, y cada línea entre los círculos un electrón de valencia compartido por los átomos. Nótese que la carga +4 indicada en cada círculo se neutraliza con la carga negativa de los cuatro electrones de valencia que el átomo aporta para realizar los cuatro enlaces covalentes con los átomos contiguos. Debe tenerse en cuenta que en la realidad los cuatro enlaces no están en un mismo plano, sino que se disponen en el espacio según se ha mostrado en la figura 1.1.

El electrón de valencia que forma parte de un enlace covalente está fuertemente ligado a los átomos que une. Si se aplica un campo eléctrico al semiconductor, los electrones de valencia siguen ligados al enlace, y los de las capas más internas al núcleo aun con más fuerza, por lo cual no se produce movimiento de cargas alguno. La corriente es, por tanto, nula, y el semiconductor es consecuentemente un aislante. Tal es el comportamiento de un semiconductor a 0 K.

Un estudio riguroso del semiconductor exige la aplicación de la mecánica cuántica. No obstante, muchas de las propiedades de esos dispositivos se pueden entender utilizando una aproximación

basada en la física clásica, que considera a los electrones como partículas materiales que obedecen a las leyes de Newton. Esa aproximación didáctica posee indudables ventajas, dado que evita la demora que supondría el aprendizaje previo de los conocimientos sobre mecánica cuántica y física del estado sólido que resultan necesarios para emprender el estudio riguroso de los semiconductores. Pero, lamentablemente, la aproximación clásica requiere ocasionalmente la aportación «externa» de resultados de la física cuántica que permitan superar sus limitaciones. Una de esas aportaciones es la *cuantificación* de la energía: la energía que un electrón puede absorber o emitir está constituida por *paquetes* de energía indivisibles denominados cada uno de ellos *cuanto* de energía. El cuanto de energía electromagnética se denomina *fotón*, y el de energía térmica, *fonón*.

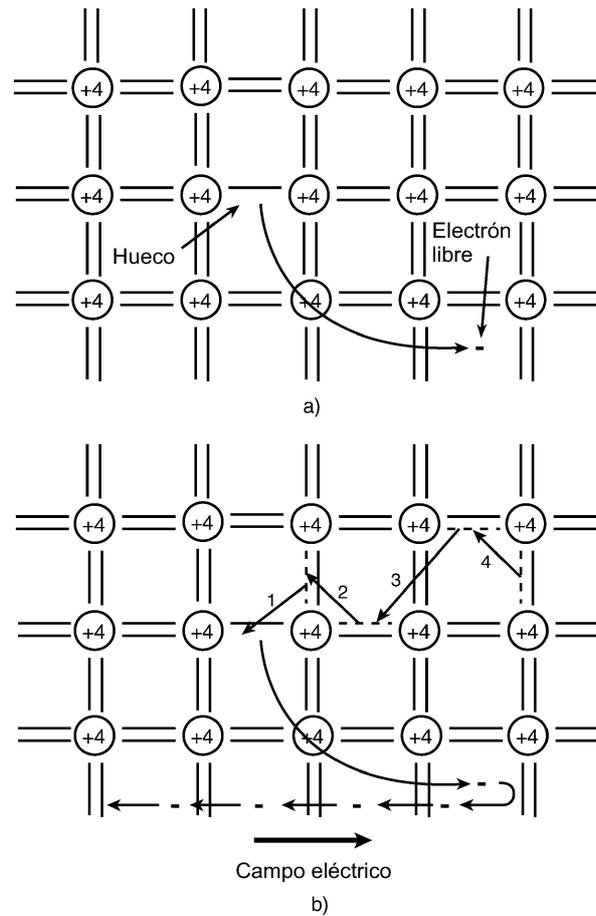


Fig. 1.4 a) Generación de un par electrón-hueco por la ruptura de un enlace covalente. b) Desplazamiento del electrón libre y del hueco por la acción de un campo eléctrico aplicado sobre el semiconductor.

Si un electrón de valencia que forma parte de un enlace covalente absorbe un cuanto de energía del valor adecuado, puede romper la ligazón con el enlace covalente y moverse libremente por el semiconductor. Ese electrón libre, desligado del enlace, se denomina *electrón de conducción* y es un portador de corriente, ya que es una carga que se mueve en sentido contrario al campo eléctrico aplicado al semiconductor debido a su carga negativa.

El enlace covalente roto produce un desequilibrio en la red cristalina, el cual «reclama» la presencia de un electrón para su reconstrucción. Un electrón de valencia de un enlace próximo puede verse afectado por ese desequilibrio, abandonar su enlace y reconstruir el enlace roto. Pero esa acción significa, simplemente, que el enlace covalente roto ha cambiado de sitio, por lo cual se repite la acción anterior. Se vería, por tanto, que el enlace covalente roto —el *hueco*— se va moviendo por el cristal. Cuando se aplica un campo eléctrico se favorece el desplazamiento de los electrones de valencia que ocupan sucesivamente el enlace covalente roto, en sentidos contrarios al campo eléctrico (v. fig. 1.4.b). El resultado es que el hueco, es decir, el enlace covalente roto, se mueve en el mismo sentido que el campo eléctrico, como si de una carga positiva se tratara. La física cuántica demuestra que se puede considerar al enlace covalente roto como a una partícula positiva del mismo valor absoluto que el electrón, y con una masa específica. Es, por lo tanto, un portador de corriente con carga positiva.

Para generar un electrón de conducción y un hueco se debe proporcionar un cuanto de energía a un electrón de valencia. El electrón desligado adquiere una energía mayor que cuando formaba parte del enlace covalente. La representación de la energía de los electrones en los distintos puntos del semiconductor se denomina *modelo de bandas de energía*.

Para alcanzar una comprensión cualitativa de ese modelo debe considerarse en primer lugar el modelo atómico de Bohr para un átomo aislado. Como es bien sabido, Bohr propuso para el átomo un modelo planetario corregido en algunos aspectos: los electrones giran en órbitas circulares alrededor del núcleo, de tal forma que la fuerza eléctrica de atracción entre el electrón negativo y el núcleo positivo se neutralizan exactamente con la fuerza centrífuga. No obstante, solo se permiten aquellas órbitas cuyo momento angular sea un múltiplo entero de $h/2\pi$. El electrón que se halla en una órbita permitida no emite energía. En cada órbita, el electrón tiene una energía total bien definida, que viene dada por la suma de las energías potencial y cinética. El electrón puede saltar de una órbita a otra absorbiendo o emitiendo un cuanto de energía igual a la diferencia entre las respectivas energías totales que tiene en cada una de las órbitas consideradas. En la figura 1.5 se muestra el radio y la energía total del electrón en las órbitas permitidas. El electrón puede por tanto tener en el átomo aislado unos *niveles* de energía permitidos, mientras que el resto de energías no le están permitidas.

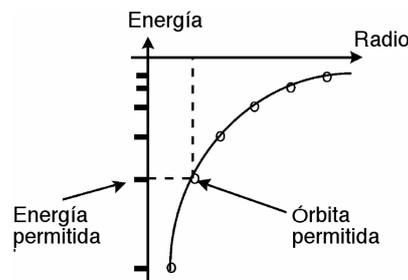


Fig. 1.5 Modelo de Bohr para el átomo aislado: el electrón tiene en cada órbita permitida una energía bien definida

Cuando se forma el cristal, hay muchos átomos que interactúan entre sí, de manera que el modelo del átomo aislado deja de ser válido. El principio de exclusión de Pauli establece que no puede haber dos electrones de un mismo sistema con el mismo estado cuántico. Por esa razón, cuando dos átomos se aproximan y los electrones comienzan a interactuar entre sí y a formar un mismo sistema, los niveles

de energía del átomo aislado tienen que desdoblarse, dado que en caso contrario podría haber dos electrones con el mismo estado cuántico (v. fig. 1.6). Cuando, en lugar de dos átomos, son muchos más los que interactúan, como en el caso de un cristal, el nivel original se debe subdividir en tantos niveles como átomos interactúen. Aparecen así intervalos de energía en los cuales hay una gran densidad de niveles permitidos, y da la impresión de que hay una continuidad de energías permitidas. Se dice entonces que ese intervalo constituye una *banda de energía permitida*. Cuando en un intervalo de energías no hay nivel permitido, se le denomina *banda prohibida (BP)*.

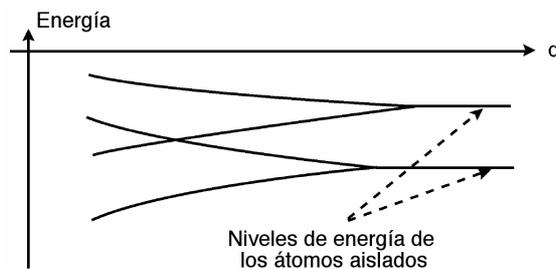


Fig. 1.6 Desdoblamiento de los niveles permitidos cuando la distancia d entre los dos átomos se reduce

Ejercicio 1.1

Calcúlese el radio y la energía total del electrón para las dos órbitas más próximas al núcleo del átomo de hidrógeno. Datos: el radio de las órbitas permitidas es $r = (h^2 \epsilon_0 / \pi m q^2) n^2 = 0,53 n^2 \text{ \AA}$; la velocidad $v = (q^2 / h \epsilon_0) (1/n)$, donde n es entero.

Las dos primeras órbitas son para $n = 1$ y $n = 2$. Los radios son $r_1 = 0,53 \text{ \AA}$ y $r_2 = 2,12 \text{ \AA}$.

La energía total del electrón en una órbita es $E = E_{\text{cin}} + E_{\text{pot}} = (1/2)mv^2 + (-q^2/4\pi\epsilon_0 r) = -q^2/8\pi\epsilon_0 r = -13,6/n^2 \text{ eV}$

Por lo tanto, $E_1 = -13,6 \text{ eV}$ y $E_2 = -3,4 \text{ eV}$. El signo negativo de la energía total indica que el electrón está ligado al núcleo —para liberarse es necesario un radio infinito, lo cual significa un valor de n infinito y, por lo tanto, una energía total nula—.

Las energías de los electrones de valencia se agrupan en un intervalo que se denomina *banda de valencia*; el límite superior de ese intervalo es E_v . Las energías que pueden tener los electrones que se han desligado de los enlaces covalentes se agrupan también en un intervalo, denominado *banda de conducción*; su límite inferior es E_c . Entre las dos bandas se extiende la *banda prohibida*: ningún electrón puede tener una energía de ese margen. Véase la figura 1.7.

La amplitud de la banda prohibida se denomina E_g (en inglés, la banda prohibida se conoce como *band gap*), y es uno de los parámetros más importantes de un semiconductor. Esa energía, $E_g = E_c - E_v$, es la energía mínima que se debe proporcionar a un electrón de valencia para desligarlo del enlace covalente. Cuanto mayor es el valor de E_g , más fuerte es el enlace covalente y menos electrones lo pueden romper, por lo cual el semiconductor es más aislante. A temperatura ambiente el silicio tiene una $E_g = 1,1 \text{ eV}$, el GaAs de $1,43 \text{ eV}$ y el Ge de $0,68 \text{ eV}$. Los semiconductores que tienen

2 La unión PN

La mayoría de los dispositivos semiconductores utilizados en la electrónica contienen regiones del tipo P y regiones del tipo N. Las propiedades de los contactos entre los dos tipos de zona, denominados *uniones PN*, son fundamentales para el funcionamiento del dispositivo. Como ejemplos, un diodo es un dispositivo formado por una sola unión; un transistor, tanto si es bipolar como MOS, contiene al menos dos uniones, etcétera. Se dedica este capítulo al estudio del comportamiento de la unión PN.

Después de una descripción cualitativa del funcionamiento de una unión PN, se dedica la mayor parte del capítulo al análisis cuantitativo empleando un modelo simple. Se comienza por el estudio del sistema en equilibrio, se continúa con la evaluación de las corrientes bajo la aplicación de tensiones continuas, y se pasa luego a ver el efecto de las tensiones dependientes del tiempo —los efectos dinámicos—. Aunque este estudio se centra en las uniones en las que la región P y la región N son del mismo semiconductor —homouniones—, se discute también las características de las heterouniones, formadas por dos semiconductores distintos. Se acaba el capítulo presentando el comportamiento de los contactos entre un metal y un semiconductor, y discutiendo las características que deben tener los contactos que actúan como terminales de un dispositivo —los contactos óhmicos—.

2.1 La unión PN: bandas de energía y efecto rectificador

2.1.1 Hipótesis iniciales del modelo

Una unión PN es un cristal semiconductor único, con una región dopada con impurezas aceptoras y otra con impurezas donadoras. Cuando los dopajes son homogéneos en el interior de cada región y existe un plano de separación entre ambas, se habla de *unión abrupta*. Si el cambio de dopaje es progresivo, se habla de *unión gradual*. Este estudio se centra en las uniones abruptas; se denomina N_A la concentración de impurezas aceptoras de la región P, mientras que N_D es el nivel de dopaje de la región N. Se supone que para esas impurezas se cumple la condición de ionización total.

Se consideran válidas las principales hipótesis que han permitido llevar a cabo el estudio de los semiconductores del primer capítulo; particularmente, que los semiconductores son no degenerados y que se mantienen las condiciones de baja inyección. Por otro lado, este trabajo se simplifica empleando un modelo unidimensional: las variables que se utilizan dependen de una sola coordenada, medida en la perpendicular al plano de la unión, pero no de las otras dos coordenadas.

2.1.2 El diagrama de bandas

La figura 2.1 representa el símbolo, la estructura y el diagrama de bandas en equilibrio de la unión PN. Debe recordarse aquí que el nivel de Fermi, E_f , es constante en equilibrio térmico. La deformación de

los niveles E_c y E_v —y junto con ellos, de E_f — indica que hay un campo eléctrico en sentido de derecha a izquierda en la región de transición, es decir, un campo que va de la región N a la P. Tal como se ha descrito en el capítulo 1, ese campo eléctrico confina los portadores mayoritarios en las respectivas regiones: los electrones en la región N y los huecos en la P. Solo pueden pasar a la otra región los portadores que tengan suficiente energía como para superar el escalón representado por la curvatura de las bandas —en equilibrio, una energía cinética superior a qV_{bi} —. Dado el valor que suele tener esa barrera, el porcentaje de portadores que la pueden superar es muy pequeño.

Se distinguen tres regiones en la unión PN: la *región neutra P*, en la que el campo eléctrico es nulo y en la que hay, por lo tanto, neutralidad de carga; la *zona de carga espacial (ZCE) o región de transición entre P y N*, en la que hay un campo eléctrico intenso producido por un dipolo de carga espacial; y, finalmente, la *zona neutra N*. Se supone que en la zona neutra P la concentración de huecos es N_A y $n \ll p$, mientras que para la región neutra N se supone que $n = N_D$ y $p \ll n$.

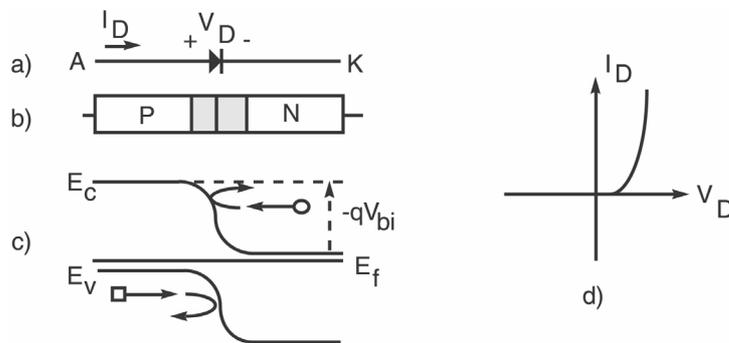


Fig. 2.1 La unión PN: a) Símbolo circuital. b) Estructura física. c) Diagrama de bandas en equilibrio térmico. d) Característica $I(V)$ mostrando el efecto rectificador.

2.1.3 El efecto rectificador

El efecto rectificador consiste en la propiedad de permitir el paso de la corriente en un sentido —de P a N, en el caso de una unión PN— y de bloquearlo en el sentido contrario (v. fig. 2.1.d). El dispositivo que presenta ese efecto se denomina *diodo*.

Más adelante se justifica que si se aplica una tensión positiva de valor V_D en la región P con respecto a la región N —se denomina *polarización directa*—, el escalón de energía en la ZCE pasa a valer $q(V_{bi} - V_D)$. La polarización directa causa la disminución del campo eléctrico en la región de transición, porque reduce la pendiente del perfil de E_f . En polarización inversa —región N positiva con respecto a la P— la altura del escalón $q(V_{bi} - V_D)$ aumenta.

En polarización directa, la disminución del escalón de energía en la región de transición permite el paso de muchos portadores mayoritarios de cada una de las regiones a la otra, porque precisan menos energía cinética para ello. En efecto: según la ley de Fermi, la distribución energética de los electrones dentro de la banda de conducción disminuye exponencialmente a medida que su energía se aleja de E_c . De la misma manera, el número de huecos dentro de la banda de valencia disminuye exponencialmente a medida que su energía se aleja de E_v . Así pues, en polarización directa hay un flujo muy intenso de huecos de P a N, y de electrones de N a P. En consecuencia, la corriente a través del diodo —en sentido de P a N— aumenta exponencialmente con la tensión de polarización.

En polarización inversa, el aumento de la altura del escalón de energía produce un confinamiento todavía mayor de los portadores: es todavía más difícil para los huecos ir de P a N, y para los electrones ir de N a P. La corriente es casi nula. Nótese que el aumento de la altura del escalón implica un aumento del campo eléctrico en la ZCE porque la pendiente de E_{fi} aumenta.

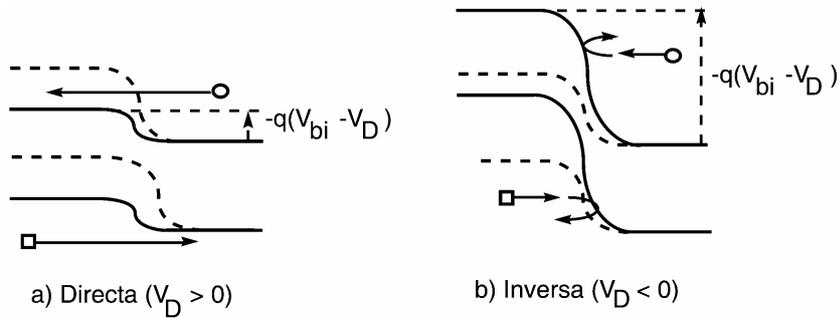


Fig. 2.2 Diagrama de bandas: a) En polarización directa. b) En polarización inversa.

2.1.4 Potencial, campo eléctrico y carga espacial en la región de transición

El diagrama de bandas nos muestra que hay una diferencia de potencial, un campo eléctrico y un dipolo de carga en la región de transición entre P y N. En efecto, recordando las relaciones halladas en el capítulo 1,

$$\theta_i(x) = V(x) = -\frac{E_{fi}(x) - E_f}{q} \quad E_{el}(x) = -\frac{dV(x)}{dx} \quad \rho(x) = \epsilon \frac{dE_{el}}{dx} \quad (2.1)$$

Estas funciones se hallan representadas, junto con la curvatura de las bandas, en la figura 2.3, para condiciones de equilibrio térmico y para polarización directa. Se observa que en equilibrio la región N está un potencial V_{bi} por encima de la región P. Ese potencial, que se denomina *potencial de contacto* o *potencial de difusión*, tiene como valor

$$\begin{aligned} V_{bi} = V_N - V_P = \phi_{iN} - \phi_{iP} &= \frac{E_f - E_{fi}(x)}{q} \Big|_N - \frac{E_f - E_{fi}(x)}{q} \Big|_P = \\ &= \frac{k_B T}{q} \ln \frac{N_D}{n_i} + \frac{k_B T}{q} \ln \frac{N_A}{n_i} = \frac{k_B T}{q} \ln \frac{N_D N_A}{n_i^2} \end{aligned} \quad (2.2)$$

donde se admite que la concentración de mayoritarios en cada región es aproximadamente igual a su nivel de dopaje.

La aplicación de una tensión de polarización directa V_D reduce la altura de la barrera de potencial entre la parte N y la P, que pasa a valer $V_{bi} - V_D$; ello implica una disminución de la intensidad del campo eléctrico y de la densidad de carga. En cambio, en polarización inversa V_D es negativa, y por lo tanto la barrera de potencial se hace más elevada, el campo eléctrico más intenso y la densidad de carga mayor.

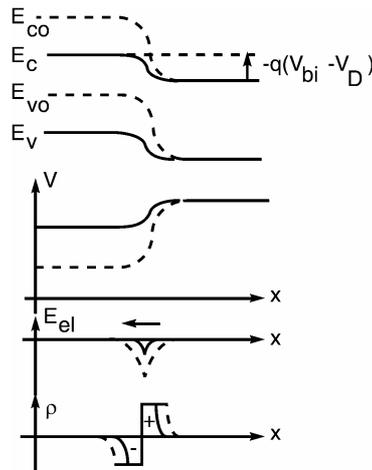


Fig. 2.3 Diagrama de bandas, potencial, campo eléctrico y densidad de carga en equilibrio térmico y en polarización directa

Si se va polarizando cada vez más la unión en inversa, el aumento del campo eléctrico llega a provocar la *ruptura de la unión PN*. Cuando el campo eléctrico alcanza un valor crítico —cercano a 3×10^5 V/cm en el caso del silicio—, comienza a circular una corriente inversa muy intensa, porque se genera un elevado número de portadores.

Ejercicio 2.1

Suponiendo que el valor máximo que puede adoptar el dopaje en un semiconductor es de 10^{21} cm⁻³, estime el valor máximo del potencial de difusión a 300 K en: a) Si, b) GaAs, c) Ge.

Aplicando la expresión 2.2 y empleando los valores de n_i del apartado 1.3, resulta

$$a) V_{bi}(300 \text{ K}, \text{Si}) = 25 \times 10^{-3} \cdot \ln(10^{42}/2.25 \times 10^{20}) = 1.25 \text{ V}$$

$$b) V_{bi}(300 \text{ K}, \text{GaAs}) = 25 \times 10^{-3} \cdot \ln(10^{42}/4 \times 10^{12}) = 1.69 \text{ V}$$

$$c) V_{bi}(300 \text{ K}, \text{Ge}) = 25 \times 10^{-3} \cdot \ln(10^{42}/6.25 \times 10^{26}) = 0.87 \text{ V}$$

Ejercicio 2.2

Estímesese el valor máximo de V_{bi} según el diagrama de bandas. Compárense los resultados con los del ejercicio anterior, y justifíquense las diferencias.

Si se supone que para dopajes muy elevados se da que $E_f = E_c$ en la región N, y $E_f = E_v$ en la región P, la barrera de potencial es E_g , es decir, $V_{bi}(300 \text{ K}, \text{Si}) = 1.1 \text{ V}$, $V_{bi}(300 \text{ K}, \text{GaAs}) = 1.43 \text{ V}$, y $V_{bi}(300 \text{ K}, \text{Ge}) = 0.68 \text{ V}$. Los valores obtenidos en el ejercicio anterior son más elevados, y ello exige que el nivel de Fermi en un semiconductor muy dopado se halle dentro de la banda de conducción si es del tipo N, y dentro de la de valencia si es del tipo P. Son referidos con la denominación de semiconductores degenerados.

Comentario: las concentraciones de mayoritarios y de minoritarios calculadas en el capítulo 1, donde se ha aplicado la aproximación de Boltzmann, no son válidas para los semiconductores degenerados. Las conclusiones de este ejercicio son cualitativamente correctas, pero los valores numéricos no son exactos.

2.1.5 La generación del dipolo de carga en la región de transición

Según la ley de Gauss, la variación del campo eléctrico se debe a la presencia de carga eléctrica distribuida en la región. La estructura de bandas pone de manifiesto que hay un dipolo de carga en la región de transición de P a N. Aparece la cuestión sobre cuál es el origen de esa carga, su procedencia.

Considérese un semiconductor del tipo P homogéneo, que se pone en contacto con uno del tipo N (se trata de un proceso conceptual, no de una técnica de fabricación). Los huecos comenzarán entonces a difundirse desde la región P, donde su concentración es elevada, hacia la región N, donde casi no existen. Como consecuencia de esa difusión, la concentración de huecos de la región P en las cercanías de la interfaz con la región N disminuirá. Antes de realizar la unión, el semiconductor P tiene aproximadamente tantos huecos como impurezas N_A^- ionizadas negativamente. Esa igualdad garantiza la neutralidad de carga en todos los puntos. Después de la unión desaparecen los huecos de esa región porque se desplazan hacia la región N, de manera que N_A^- resulta mayor que la concentración de huecos —positivos— causando que en la región P —en las cercanías de la interfaz— haya una carga neta negativa. Un razonamiento paralelo explica que aparezca una carga neta positiva en la región N, cerca de la interfaz con la región P, a causa de los electrones que han abandonado por difusión la región N, dejando atrás impurezas donadoras positivas no neutralizadas (v. fig. 2.4).

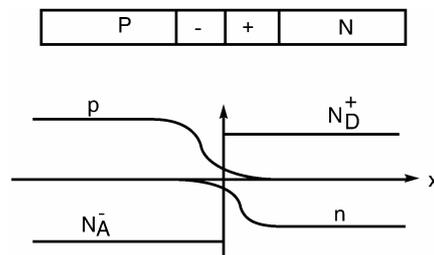


Fig. 2.4 Generación de un dipolo de carga en la región de transición de la unión PN. Los huecos abandonan la región P por difusión, dejando en su lugar iones negativos N_A^- no neutralizados. De la misma manera, los electrones se desplazan de la región N dejando iones positivos N_A^+ sin neutralizar.

De esa manera se forma un dipolo de carga que genera un campo eléctrico en sentido de N hacia P. Ese campo eléctrico confina los portadores mayoritarios en las respectivas regiones de origen, porque «se opone» a la difusión, «retornando» huecos a la región P y electrones a la región N por arrastre. Durante el dominio de la difusión disminuyen las concentraciones de mayoritarios en sus respectivas regiones, con lo que se refuerza el dipolo de carga y se aumenta el campo eléctrico. Al cabo de cierto tiempo se llega a un equilibrio entre la difusión y el arrastre, de manera que ambas corrientes se neutralizan exactamente en cada punto, y se alcanza así el equilibrio térmico, representado en el diagrama de bandas del apartado 2.1.1.

2.1.6 Los contactos entre metal y semiconductor

Los terminales de un dispositivo que contiene una unión PN, como es el caso de un diodo, deben ser metálicos, tal como se indica en la figura 2.5. En el contacto entre un metal y un semiconductor también aparece un dipolo de carga espacial, similar al de una unión PN. Esta afirmación se justifica al final de este capítulo. Dicho dipolo da lugar a un potencial de contacto entre el metal y el semiconductor.

Entre el terminal de ánodo (A) y el de cátodo (K) hay, en el equilibrio térmico, tres potenciales de contacto: V_{c1} entre metal y el semiconductor P, V_{bi} en la unión PN, y V_{c2} entre el semiconductor N y el metal. La suma de los tres potenciales debe ser cero, de manera que $V_A = V_K$, dado que de otro modo un diodo en equilibrio térmico se comportaría como una fuente de tensión, lo cual es incompatible con los principios de la termodinámica.

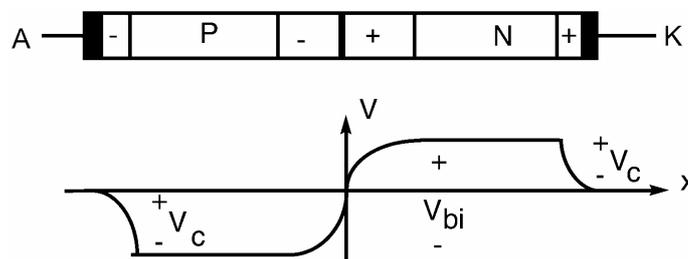


Fig. 2.5 Potencial de contacto entre ánodo y cátodo de una unión PN en equilibrio

En el citado apartado se demuestra que cuando se polariza un dispositivo como el de la figura 2.5 se pueden dar dos situaciones distintas en los terminales: que los potenciales de contacto entre metal y semiconductor se mantengan constantes en su valor de equilibrio, o que, por el contrario, una parte de la tensión $V_A - V_K$ caiga en esos contactos. Únicamente el primer tipo de contacto resulta útil para construir los terminales de una unión PN. En ese caso, toda la tensión aplicada aparece en la ZCE de la unión PN, suponiendo que en las zonas neutras el campo eléctrico sea inapreciable. Para justificar esa hipótesis, se observa que en las zonas neutras no hay carga neta localizada: según la ley de Poisson, no hay campo. Se supone que las posibles caídas de tensión en las zonas neutras asociadas al paso de corriente son pequeñas; este punto, no obstante, se discute en el apartado 2.5, en el que se ve cómo se corrige el modelo de dispositivo en caso de que esta suposición no sea aceptable.

CUESTIONARIO 2.1

1. Considérese una unión entre una región de un semiconductor del tipo P y una región intrínseca. Existen cargas localizadas en cada lado del plano de la unión —un dipolo de carga—. Razónese qué cargas —portadores de corriente, impurezas ionizadas, etcétera— constituyen ese dipolo; determínese en consecuencia cuál de las siguientes proposiciones no es correcta.

- Una zona de la región P tiene carga negativa porque ha perdido portadores mayoritarios.
- Una zona de la región intrínseca tiene carga positiva porque ha ganado huecos.

c) La densidad de carga de la región P en cada punto es proporcional a la concentración de impurezas ionizadas no neutralizadas por portadores libres.

d) La densidad de carga de la región intrínseca en cada punto es proporcional a la concentración de átomos del semiconductor no neutralizados por portadores libres.

2. Considérese el diagrama de bandas de la unión de la cuestión anterior. Para simplificar, se supone que en la región P el nivel de Fermi se encuentra en el punto más elevado de la banda de valencia ($E_f - E_v = 0$). ¿Cuanta energía se debe proporcionar a un hueco para hacerlo pasar de la región P a la intrínseca? Dato: la anchura de banda del semiconductor es $E_g = 1.5$ eV. Nota: el signo negativo significa que el hueco cede energía.

- a) 1.5 eV b) 0.75 eV c) -1.5 eV d) -0.75 eV

3. ¿Qué valor tiene la tensión de difusión —la barrera de potencial— V_{bi} en la unión de las cuestiones anteriores? Supóngase que se pueda introducir impurezas donadoras en la región intrínseca hasta que el nivel de Fermi alcance el fondo de la banda de conducción ($E_c - E_f = 0$). ¿Qué valor adopta V_{bi} en esta ocasión?

- a) $V_{bi} = 0.75$ V en el primer caso y $V_{bi} = 1.5$ V en el segundo.
 b) $V_{bi} = 1.5$ V en el primer caso y $V_{bi} = 0.75$ V en el segundo.
 c) $V_{bi} = 0.75$ V en el primer caso y $V_{bi} = -1.5$ V en el segundo.
 d) $V_{bi} = 1.5$ V en el primer caso y $V_{bi} = -0.75$ V en el segundo.

4. Vuélvase a la unión de la cuestión 2. Sabiendo ya que la energía necesaria para superar la barrera de potencial es igual para los dos tipos de portadores de corriente, la cuestión es sobre el número de portadores que atraviesan el plano de la unión en equilibrio térmico. ¿Cuál de las siguientes proposiciones es falsa?

- a) El número de huecos que pasan de la región P a la intrínseca es igual al número de los que se desplazan en el sentido opuesto. Lo mismo sucede con el número de electrones.
 b) El número de huecos que pasan de la región P a la intrínseca es mayor que el de electrones que se desplazan en el sentido opuesto.
 c) El número de electrones que van de la región intrínseca a la P es superior al de los que van en sentido opuesto.
 d) El número de huecos y el de electrones que atraviesan el plano de la unión deben ser iguales.

5. Calcúlese el potencial de difusión de una unión abrupta en los dos casos siguientes:

- 1) $N_A = N_D = N_1 = 10^{15} \text{ cm}^{-3}$
 2) $N_A = N_D = N_3 = 10^{21} \text{ cm}^{-3}$.

Para la respuesta, revísense las hipótesis necesarias para llegar a la fórmula empleada, y decidase qué respuesta es correcta. Datos: $n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$, $k_B T/q = 0.025$ eV.

- a) 0.275 V, 0.622 V. El primer resultado no es válido porque con dopajes muy bajos no se puede aplicar la aproximación de vaciamiento.
- b) 0.55 V, 1.25 V. El segundo resultado no es válido porque con dopajes muy elevados no se puede aplicar la aproximación de Boltzmann.
- c) 0.275 V, 0.622 V. Ambos resultados son válidos.
- d) 0.55 V, 1.25 V. El segundo resultado no es válido porque con dopajes muy elevados el semiconductor degenera y no se puede hacer referencia al nivel de Fermi.
6. De los siguientes argumentos que justifican que la tensión de difusión V_{bi} no es directamente medible por la aplicación de un voltímetro entre los terminales de un diodo, ¿cuál es correcto?
- a) La corriente que el voltímetro hace pasar por la unión es incompatible con las condiciones de equilibrio térmico.
- b) Es un potencial virtual.
- c) Las caídas de tensión en los terminales metálicos suman un valor igual a V_{bi} pero con signo contrario, y proporcionan un resultado total de la medición igual a cero.
- d) Por el voltímetro circulan electrones pero no huecos.

3 Tecnología de fabricación

3.1 Integración de un circuito en silicio

3.1.1 Introducción

Este capítulo presenta una introducción a las técnicas empleadas en la fabricación de dispositivos semiconductores y de circuitos integrados. La finalidad del presente estudio es doble: por una parte, mostrar el proceso de producción como un posible campo de actividad propio de la ingeniería electrónica; por otra, mostrar la estructura real de los dispositivos, que se debe tomar en cuenta en un estudio de su funcionamiento que vaya más allá de los modelos simples que se exponen en este curso.

La extensión disponible para tratar este tema impone algunas limitaciones. La primera es que se trata casi exclusivamente la tecnología del silicio, y solo ocasionalmente se hace referencia a la de los semiconductores III-V. La segunda es que se hace mayor hincapié en aquellos puntos más necesarios para la comprensión del funcionamiento del dispositivo —por ejemplo, los perfiles de dopaje que se obtienen— en detrimento de otros que son importantes en el ámbito de la producción, como el encapsulado o la verificación de los dispositivos, pero que no lo son tanto en el contexto del presente estudio.

Se empieza examinando la estructura real de algunos dispositivos que ponen en evidencia la necesidad de disponer de determinadas operaciones —que se denominan etapas— en el proceso de fabricación. Se continúa examinando las características de cada una de esas etapas, dando más importancia a los aspectos conceptuales que a los desarrollos matemáticos o a los problemas prácticos. Finalmente, se presentan las etapas en forma de una secuencia que integra el ciclo completo de fabricación. Cabe mencionar aquí que se trata de procesos genéricos, con aspectos comunes a casi todos los fabricantes. Los detalles difieren en cada línea de producción.

3.1.2 La estructura de un dispositivo y las técnicas de fabricación

Caso 1: un diodo discreto

Considérese uno de los dispositivos más simples que se puede imaginar: un diodo de unión PN con los terminales unidos a hilos de conexión (dispositivo discreto). La figura 3.1.a representa un posible corte esquemático de este dispositivo. En este caso se consideran las regiones P y N con niveles de dopaje moderados, de manera que para obtener contactos óhmicos con el metal de los terminales se debe disponer de las respectivas regiones P^+ y N^+ . Las capas del metal y los hilos conductores soldados completan el dispositivo. El encapsulado se pasa por alto. Para fabricar este dispositivo se necesita como mínimo dos grupos de técnicas:

- Técnicas para dopar los semiconductores: partiendo de un material con un único tipo de dopaje, por ejemplo del tipo P, se tiene que crear la región N introduciéndole átomos donadores hasta una cierta profundidad, en la concentración suficiente como para compensar el dopaje P y crear una región N. Las regiones P⁺ y N⁺ se obtienen de manera similar, mediante una aportación adicional de aceptores y donadores respectivamente. Hay dos técnicas principales para dopar: la difusión térmica de impurezas, y la implantación anódica. Ambas se describen en los próximos apartados.
- Técnicas para metalizar las superficies de contacto: más adelante se presentan las técnicas de evaporación térmica y de pulverización catódica (en inglés, *sputtering*) para realizar esa operación. No se contempla la soldadura de los hilos de contacto.

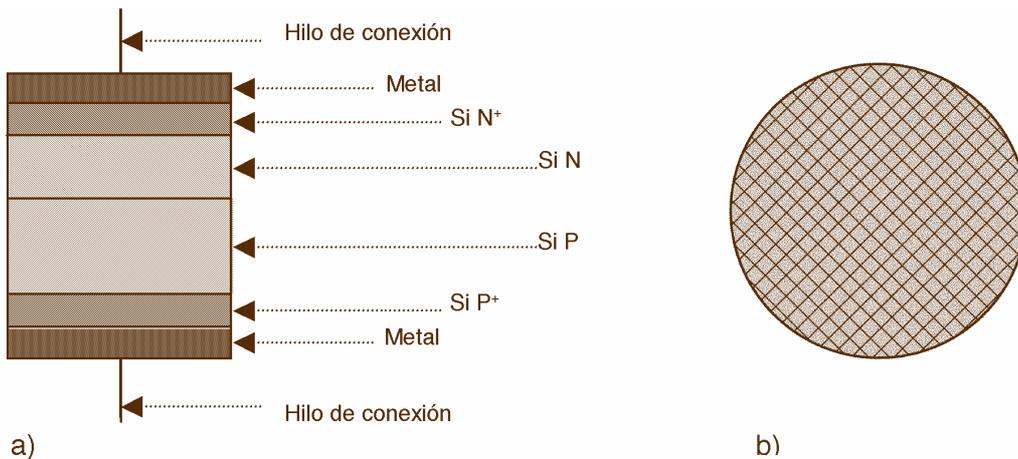


Fig. 3.1 a) Sección esquemática de un diodo de unión PN discreto. Por simplicidad, no se representan los recubrimientos de protección ni el encapsulado. b) Superficie de la oblea. Cada cuadrado (un dado) se convierte en un diodo, una vez cortado y montado.

Todas esas operaciones no se realizan para cada unidad que se vaya a fabricar, sino para una cantidad grande en una sola vez; se separan cortando el bloque de material, denominado *oblea*. La figura 3.1.b representa la superficie de la oblea, en la que se ha fabricado un elevado número de diodos. Los planos de las figuras 3.1.a y 3.1.b son perpendiculares entre ellos.

EJEMPLO 3.1

Una célula solar de silicio monocristalino es un diodo PN⁺, en el que se puede hallar la región P constituida por la oblea de silicio, de espesor entre 300 y 500 μm y dopaje moderado (10^{16} cm^{-3} , por ejemplo). La región N⁺, muy dopada, puede tener una profundidad del orden de 0.5 μm .

EJEMPLO 3.2

Un diodo rectificador de potencia tiene las dos regiones poco dopadas, para reducir el campo eléctrico máximo de la unión a una tensión dada; asimismo, ambas regiones deben tener un espesor considerable (hasta decenas de micrómetros) para que puedan contener una ZCE muy extensa.

Ejercicio 3.1

En el caso de un diodo rectificador en el que los dopajes de las regiones P y N son respectivamente 10^{15} aceptores/cm³ y 10^{16} donadores/cm³, determínese qué espesor ocupa la ZCE cuando el dispositivo soporta una tensión inversa de 200 V.

Solución: 16.6 μm .

Caso 2: un diodo integrado

Supóngase ahora que el diodo debe formar parte de un circuito integrado. En ese caso, los dos terminales de contacto deben hallarse en una misma cara, la superficie del chip (u oblea). La estructura representada en la figura 3.2 constituye una alternativa a la de la figura 3.1. Para simplificar el dibujo, se considera la región N como muy dopada (N^+), de manera que no es necesario un dopaje adicional para realizar el contacto óhmico sobre la región N.

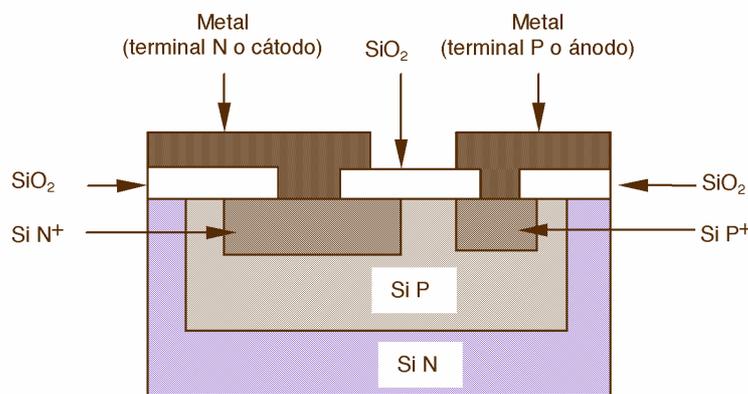


Fig. 3.2 Sección esquemática de un diodo de unión PN^+ integrado

Nótese que se ha situado la región P “incrustada” dentro de un sustrato N. La finalidad es que dispositivos adyacentes (dos diodos en este caso) no tengan las respectivas regiones P en común, es decir, en cortocircuito. De este modo, para ir desde la región P de un diodo hasta la región P de otro es necesario atravesar dos uniones PN con polaridades opuestas. Es una técnica habitual el aislar un dispositivo de otro perteneciente al mismo circuito integrado.

Los elementos que aparecen como nuevos en la figura 3.2 con respecto a la 3.1 son los siguientes:

a) Los dopajes que se introducen ya no afectan toda la anchura del dispositivo, sino únicamente ciertas regiones. La consecuencia es que se deben hallar técnicas para definir las áreas seleccionadas para dopar. Se realizará mediante dos operaciones encadenadas:

- La oxidación térmica del silicio, que crea una capa protectora contra la entrada de impurezas.
- La fotolitografía —fotograbado, más precisamente—, para eliminar esa capa protectora de las regiones en las que se desea introducir los dopantes.

Estas operaciones se describen en los correspondientes apartados.

Obsérvese al mismo tiempo que se ha creado una región P, la del ánodo, dentro de un bloque N —el sustrato—, y que se ha vuelto a invertir el tipo de conductividad de una parte de esa región P para crear la región N⁺ del cátodo. Estos procesos de doble dopaje son habituales, pero establecen restricciones sobre la gama de valores de las concentraciones de impurezas que se pueden obtener.

b) La creación de pistas conductoras para las conexiones entre los dispositivos del mismo circuito, en sustitución de los hilos soldados a los terminales. Para que esas pistas no provoquen cortocircuitos entre las regiones subyacentes del semiconductor, es necesario que se pueda crear capas aislantes, constituidas en este caso por dióxido de silicio (SiO₂). Esa capa aislante tiene unas perforaciones denominadas *aberturas de contacto*, que permiten que la pista metálica llegue a la superficie del silicio. Para crear esas regiones de conductores y aislantes se debe disponer de más etapas de proceso:

- Obtención de capas de dieléctricos. Se consideran dos grupos de operaciones: la oxidación térmica del silicio, ya mencionada, y el depósito de materiales, mediante técnicas denominadas *de CVD* (del inglés *chemical vapor deposition*: depósito químico en fase vapor).
- Apertura de contactos a través de esas capas por la fotolitografía ya mencionada.
- Definición de la forma de las pistas en una capa metálica depositada, mediante una técnica de grabado similar a la utilizada para los dieléctricos.

EJEMPLO 3.3

Supóngase que el dispositivo de la figura 3.2 es un diodo rectificador. En el proceso de diseño se decide la separación de las regiones N⁺ y P⁺ sobre la base de una serie de criterios; entre ellos, que sea lo suficientemente grande como para permitir la extensión de la ZCE en polarización inversa, y lo suficientemente pequeña como para no desperdiciar superficie de silicio ni introducir una resistencia en serie innecesaria.

Ejercicio 3.2

La región N⁺ de la figura 3.2 tiene una sección de 25 μm × 25 μm, y una profundidad de 2 μm. Propóngase un valor para asignar a la sección del diodo, si se desea estudiar este dispositivo con el modelo unidimensional de unión PN presentado en el capítulo 2.

$$A = 25 \mu\text{m} \cdot 25 \mu\text{m} + 4 \cdot (2 \mu\text{m} \cdot 25 \mu\text{m}) = 825 \mu\text{m}^2 = 8.25 \times 10^{-6} \text{ cm}^2.$$

Una vez se ha llegado a este punto, se observa que la representación de una sección es insuficiente para conocer la geometría del dispositivo, y que se debe completar con una imagen de la superficie a la manera de un plano: se denomina *composición en planta* (en inglés, *layout*). Esta representación recoge los perímetros de las distintas regiones que componen la estructura. La figura 3.3 presenta una de las posibles composiciones en planta compatibles con el corte de la figura 3.2.

Se puede observar que las aberturas de contacto a través del óxido de silicio no ocupan la totalidad de las respectivas regiones N⁺ y P⁺, sino que se deja un margen de tolerancia para que no se produzca un

cortocircuito en el caso de que haya un cierto error en el proceso de apertura. Por la misma razón, la pista metálica recubre «generosamente» la abertura de contacto. Esas tolerancias son esenciales para garantizar el éxito del proceso de fabricación.

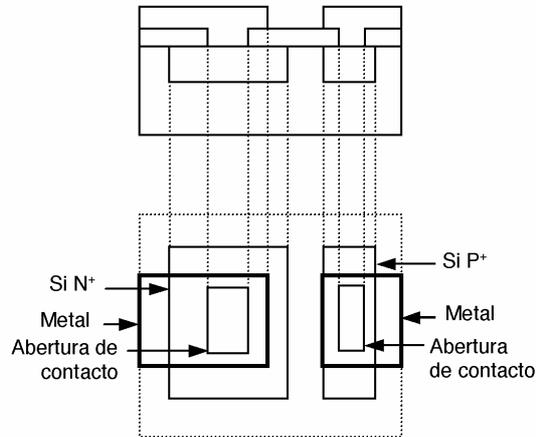


Fig. 3.3 Composición en planta del diodo de la figura 3.2

Si la estructura de la figura 3.2 se halla incrustada en un bloque de silicio del tipo N, se obtiene un transistor bipolar NPN, como se muestra en la figura 3.4. Este dispositivo es discreto porque requiere un contacto por la cara posterior. Se estudia en el capítulo 5 del presente volumen.

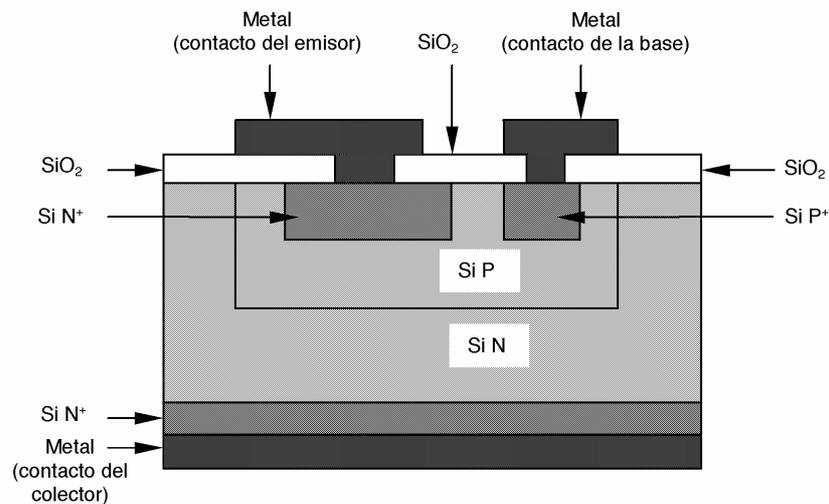


Fig. 3.4 Sección esquemática de un transistor bipolar NPN discreto

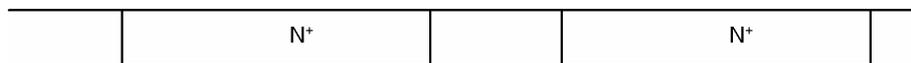
Un circuito integrado constituido por dispositivos como el de la figura 3.2 tiene un solo nivel de conectividad, es decir, todas las pistas metálicas se hallan en el mismo plano, y no pueden por tanto cruzarse entre ellas. Evidentemente, esta es una limitación demasiado rígida para el diseñador de un circuito complejo.

Lo que sucede en la práctica es que hay diversos pisos o niveles de pistas de conductores separados entre ellos por capas de dieléctricos. Cada capa de aislamiento tiene aberturas en los puntos en los que es necesario realizar contactos entre pistas alojadas en distintos niveles. En los circuitos integrados avanzados, el número de capas necesarias para las conexiones es mayor que el de las contenidas en el interior del cristal semiconductor.

Dado que el funcionamiento de los dispositivos depende esencialmente de los elementos del interior del semiconductor, se dedica la mayor parte de este estudio a las operaciones necesarias para crearlos, dejando de lado los problemas de la interconectividad.

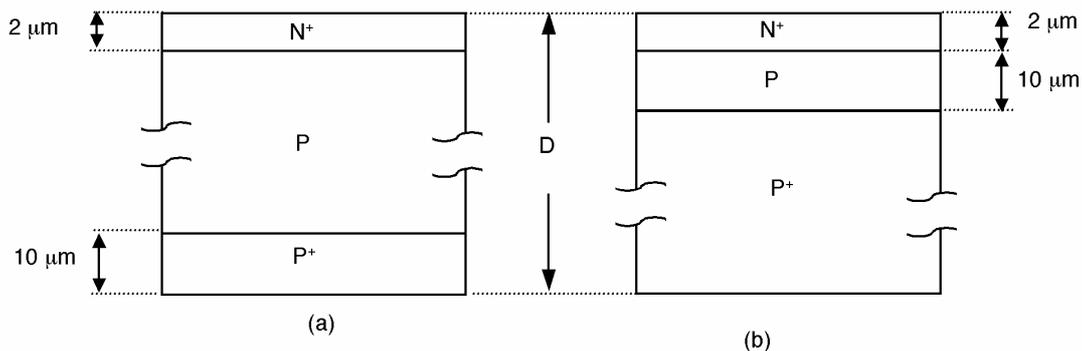
CUESTIONARIO 3.1.a

1. Se consideran dos dispositivos, que por simplicidad son resistencias que forman parte de un circuito integrado realizado sobre una oblea del tipo P, tal y como se indica en la figura. ¿Cuál será la corriente parásita entre los dos dispositivos?



- Cero.
- Proporcional a la diferencia de tensión entre las regiones N^+ .
- La corriente inversa de saturación del diodo N^+P .
- La que circula por el diodo N^+P polarizado directamente con la diferencia de las tensiones aplicadas a las dos regiones N^+ .

2. Considérense los dos diodos discretos de la figura, donde D , el espesor de la oblea, vale $500\ \mu\text{m}$ en este caso. No se representan las metalizaciones por simplicidad. ¿Cuál de las siguientes afirmaciones no es correcta?

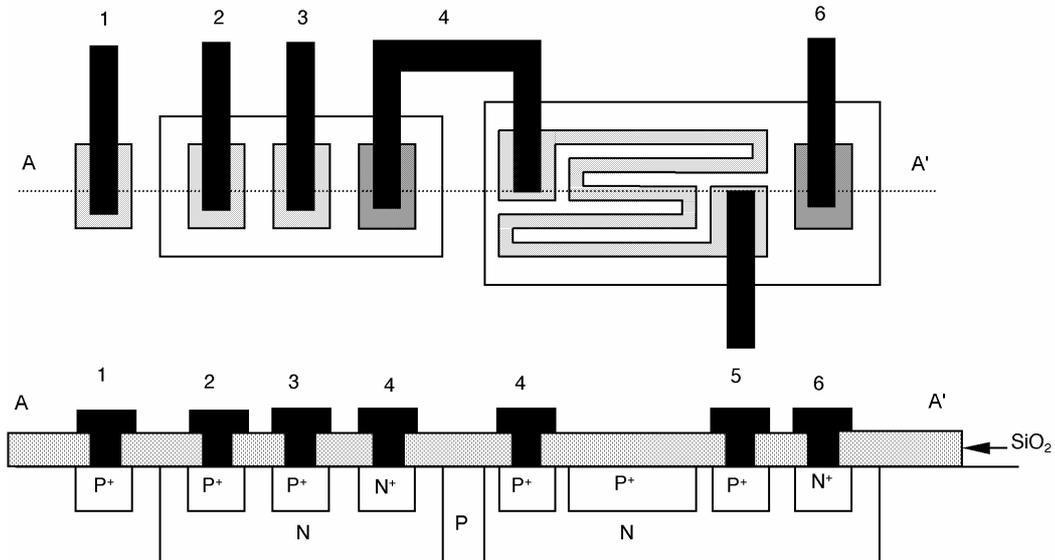


- El diseño a) tiene una resistencia serie muy elevada.
- La resistencia parásita de la región P de b) es unas 50 veces inferior a la de a).

c) El diseño b) es más complicado de fabricar, ya que requiere disponer de una capa P poco dopada sobre un sustrato P⁺.

d) El diseño a) es preferible al b), ya que tiene menos capacidad parásita.

3. En la figura adjunta se representa la composición en planta de un circuito integrado y la correspondiente sección en profundidad. ¿Cuál de las siguientes afirmaciones es falsa?



a) Se trata de dos diodos en paralelo de entradas 2 y 3, en serie con una resistencia con salida al terminal 5.

b) Conectando 5 a 0 V se obtiene una puerta OR de entradas 2 y 3 y salida 4.

c) Para aislar los dispositivos el uno del otro es preciso aplicar 0 V al terminal 1 y 5 V al terminal 6.

d) Para conseguir un correcto funcionamiento del circuito es necesario conectar el terminal 4 a la tensión más negativa disponible.

3.1.3 La obtención del material semiconductor

La obtención de semiconductores es una rama especializada de la industria metalúrgica, cuyos clientes son los fabricantes de dispositivos semiconductores y circuitos integrados. La producción del silicio monocristalino comprende tres grupos de etapas: la purificación química del silicio, el crecimiento del cristal y la presentación en forma de oblea. Habitualmente, el material inicial es el cuarzo (SiO₂), tal y como se halla en la naturaleza. La reducción en horno por carbono proporciona Si de calidad metalúrgica, muy impuro. Para poder purificarlo por destilación se transforma en un compuesto líquido, habitualmente el triclorosilano (SiHCl₃). Finalmente, dicha sustancia se descompone y se obtiene Si muy puro, pero sin estructura cristalina.

La cristalización se puede conseguir por distintos procedimientos que tienen un principio común: la fusión del silicio seguida de una solidificación. Esa solidificación se realiza de manera que comience en un punto de contacto con una muestra de material monocristalino, denominada *germen* (en inglés, *seed*). De este modo, el material que se incorpora al sólido prolonga la ordenación cristalina de los átomos del germen. Existen dos variantes para provocar el crecimiento del cristal:

- El método Czochralski (CZ): el Si líquido se halla en un recipiente (un crisol) con el germen suspendido en la superficie. Un desplazamiento lento del germen hacia arriba, acompañado de una rotación, arrastra el Si fundido, que se enfría y solidifica. De este modo se genera un lingote cilíndrico de material monocristalino. La figura 3.5.a representa esquemáticamente este procedimiento.

- El método de zona flotante (FZ): el material no cristalino en forma de lingote está en contacto con un germen en uno de sus extremos. Se provoca la fusión local del lingote —fusión de zona— mediante una bobina de inducción envolvente, comenzando por el extremo que está en contacto con el germen. Cuando la bobina se aleja de ese extremo, el material fundido solidifica alrededor del germen, y adopta de ese modo la estructura cristalina, al mismo tiempo que se funde material más alejado de ese extremo. En la figura 3.5.b se esquematiza el método.

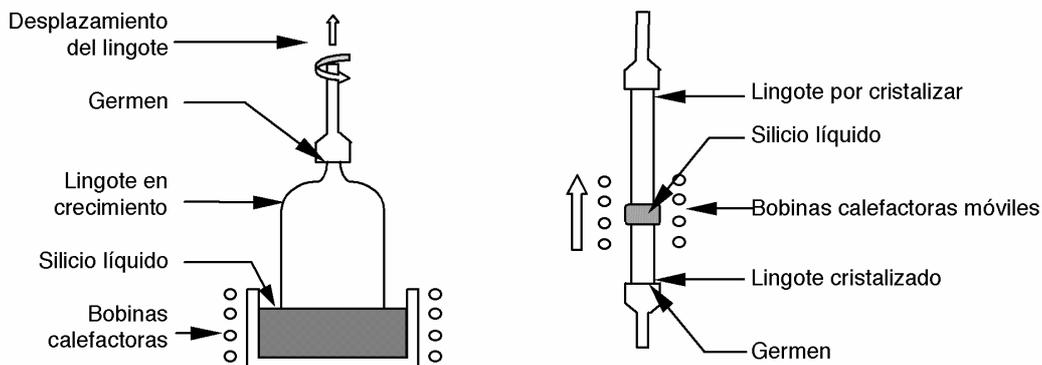


Fig. 3.5 Técnicas para el crecimiento del cristal de silicio. a) Método de Czochralski.
b) Método de zona flotante.

El método FZ proporciona un material más puro porque el material no está en contacto con el crisol, y además la fusión local arrastra las impurezas. El método CZ permite obtener lingotes de mayor diámetro. Este último es el más ampliamente utilizado.

En ambas técnicas se puede incorporar impurezas al silicio de partida, para obtener material dopado. Conviene saber que la mayoría de los dopantes son más solubles en el silicio líquido que en el sólido. Así, en el método CZ, a medida que el lingote crece aumenta la concentración del dopante en la masa fundida, y ello causa una variación de la concentración de impurezas a lo largo del lingote. En el caso del método FZ, es la zona fundida la que al desplazarse arrastra impurezas; este hecho, que es bueno para eliminar impurezas no deseadas (la fusión de zona es también una técnica de purificación), afecta a las impurezas utilizadas como dopantes, y provoca una variación de la concentración a lo largo del lingote. En todos los casos se halla una dispersión apreciable de los valores de conductividad del material, que el fabricante del silicio especifica.

EJEMPLO 3.4

La tabla 3.1 muestra datos de catálogo de un fabricante de silicio (TOPSIL), relativos a las tolerancias para la resistividad de obleas de silicio FZ:

Tabla 3.1 Datos de obleas de silicio comerciales

Resistividad ($\Omega\cdot\text{cm}$)	Tolerancia a lo largo del lingote		Variación radial típica (%)	Tiempo de vida (μs)
	Estándar (%)	Radial (%)		
8 a 100	± 15	± 12	20	> 500
100 a 1000	± 20	± 15	25	> 1000

Ejercicio 3.3

Estímense las concentraciones de impurezas en silicio de $20 \Omega\cdot\text{cm}$ de resistividad. Tómese como valor de movilidad de los portadores $1000 \text{ cm}^2/(\text{Vs})$ en el caso de los electrones, y $400 \text{ cm}^2/(\text{Vs})$ en el caso de los huecos.

Aproximadamente $3.1 \times 10^{14} \text{ donadores/cm}^3$ en material N, y $7.8 \times 10^{14} \text{ aceptores/cm}^3$ en material P.

El semiconductor se presenta en forma de discos, denominados *obleas* (en inglés, *wafers*). Para conseguirlos existen una serie de procesos mecánicos de corte y pulimento de la superficie. Las operaciones que tienen lugar en la fabricación de dispositivos exigen que la superficie tenga un pulimento óptico (aspecto especular). Los discos tienen unos pequeños cortes en la periferia (en inglés, *flats*) que indican la orientación de los planos cristalinos y el tipo de dopaje (v. fig. 3.6). El diámetro más frecuente de las obleas utilizadas industrialmente es de 200 mm, con tendencia hacia formatos mayores. Las obleas más pequeñas son útiles en el laboratorio.

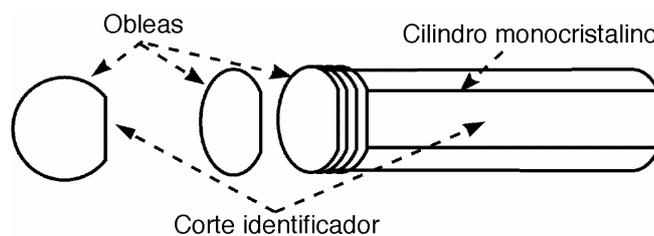


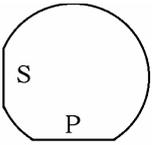
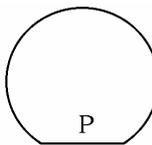
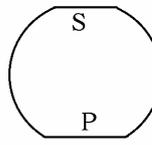
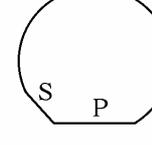
Fig. 3.6 Lingote de silicio y producción de obleas

Las técnicas de obtención de semiconductores III-V son algo distintas, pero los principios son los mismos. La principal dificultad proviene del hecho de que para fundir un compuesto se debe controlar muy bien la atmósfera que lo rodea, si se quiere evitar que el material se empobrezca en el componente más volátil. Las obleas de semiconductores III-V son generalmente de tamaño menor que las de silicio.

EJEMPLO 3.5

La tabla 3.2 presenta las posiciones de los cortes de la oblea más utilizadas. En cada uno de los cuatro casos se indica la posición del corte mayor, denominado *primario* (P), y del menor o *secundario* (S).

Tabla 3.2 Posiciones de los cortes en las obleas de silicio

	Orientación 100	Orientación 111
Silicio P		
Silicio N		

CUESTIONARIO 3.1.b

1. Supóngase un lingote de silicio obtenido por el método de Czochralski. Para doparlo se ha añadido fósforo a la masa de material líquido. Sabiendo que la impureza es más soluble en la fase líquida que en la sólida, ¿qué extremo del lingote queda más dopado?

- El extremo superior.
- El extremo inferior.
- El dopaje es uniforme.
- Depende del tipo de impureza.

2. La fusión local del lingote (fusión de zona) es una técnica empleada no solo para transformar material no cristalino en un monocristal, sino también para eliminar impurezas no deseadas. Indíquese qué afirmación no es correcta.

- La zona fundida tiene mayor densidad de impurezas que la cristalizada.
- La zona fundida tiene igual densidad de impurezas que la recién cristalizada.
- El extremo superior tiene más impurezas que el extremo inferior.
- La zona fundida en el extremo superior tiene la mayor densidad de impurezas del lingote.

3 El proveedor de silicio no especifica la concentración de impurezas que ha introducido en sus obleas, sino su resistividad. Si la precisión de los datos es de $\pm 10\%$, ¿con cuántas cifras significativas se expresa la concentración de impurezas?

- Una cifra.
- Dos cifras.
- Tres cifras.
- Cuatro cifras.

4 Dispositivos optoelectrónicos

El objetivo de este capítulo es el estudio de los dispositivos receptores de luz que se pueden construir a partir de un semiconductor —los fotoconductores— o a partir de un diodo —los fotodiodos y las células fotovoltaicas—, así como los dispositivos emisores de luz —diodos electroluminiscentes y el diodo láser—. Es requisito previo el conocimiento de los fenómenos de interacción entre la radiación electromagnética y los semiconductores, y por ello se resumen brevemente sus aspectos principales antes de pasar a presentar los dispositivos en cuestión. Algunos dispositivos, tales como los fototransistores, se analizan más adelante, una vez expuestos los conceptos necesarios para su estudio.

4.1 Radiación electromagnética y semiconductores

4.1.1 Radiación electromagnética

La radiación luminosa se manifiesta como una onda continua —radiación electromagnética clásica— en algunos fenómenos, y como un haz de corpúsculos —cuantos de energía— en otros. En tanto que onda, presenta fenómenos de interferencia, difracción, etcétera, y sus parámetros característicos son la frecuencia (f) y la longitud de onda (λ). En el vacío, la relación entre esas dos cantidades viene dada por la ecuación $\lambda = c/f$, donde c es una constante universal: la velocidad de la luz en el vacío (2.998×10^8 m/s aproximadamente). La relación entre c y las constantes electromagnéticas del vacío —permitividad eléctrica, ϵ_0 , y permeabilidad magnética, μ_0 — es

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \quad (4.1)$$

En un medio material, la velocidad de la luz, v , queda determinada por el índice de refracción del material, n , definido como

$$n \equiv \frac{c}{v} \quad (4.2)$$

El índice de refracción es una función de la longitud de onda, $n(\lambda)$. En un material caracterizado por una constante dieléctrica relativa ϵ_r y por una permeabilidad magnética relativa μ_r se cumple que

$$n = \sqrt{\epsilon_r \mu_r} \quad (4.3)$$

EJEMPLO 4.1

Algunos índices de refracción de interés para la optoelectrónica en los extremos del espectro visible:

Material	$\lambda = 0.4 \mu\text{m}$	$\lambda = 0.7 \mu\text{m}$	Material	$\lambda = 0.4 \mu\text{m}$	$\lambda = 0.7 \mu\text{m}$
GaAs	4.373	3.755	SiO ₂	1.470	1.455
Si	5.570	3.787	Si ₃ N ₄	2.072	2.013

Ejercicio 4.1

Calcúlense las velocidades de las radiaciones de los dos extremos del espectro visible en el dióxido de silicio.

2.060×10^8 m/s en el rojo, 2.039×10^8 m/s en el violeta.

Ejercicio 4.2

Considérese una fibra óptica fabricada con dióxido de silicio. Calcúlese la diferencia entre el tiempo que necesita una radiación de $0.4 \mu\text{m}$ de longitud de onda para recorrer 1 km de fibra, y el que necesita una radiación de $0.7 \mu\text{m}$.

50 ns.

Cuando un rayo de luz incide sobre la superficie que separa dos medios con distintos índices de refracción formando un ángulo θ_i —ángulo de incidencia— con respecto a la perpendicular al plano de separación, en el caso más general una parte de la luz se transmite al otro medio —constituye el denominado *rayo refractado*—, mientras que otra parte se reenvía al mismo medio —constituye el denominado *rayo reflejado*—; ello se ilustra en la figura 4.1.

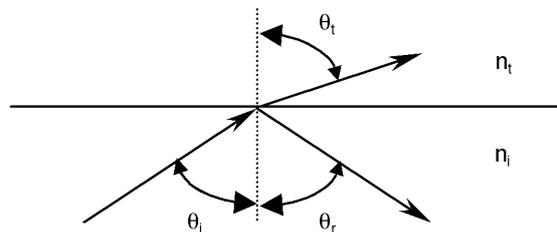


Fig. 4.1 Reflexión y refracción de un rayo de luz

Las relaciones entre las direcciones de esos rayos obedecen a las siguientes leyes:

- Ley de la reflexión: el ángulo de incidencia es igual al de reflexión, $\theta_i = \theta_r$; el rayo incidente, el reflejado y la normal se encuentran en el mismo plano.
- Ley de la refracción (o de Snell): entre el ángulo de incidencia y el de refracción existe la relación

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_t}{n_i} \quad (4.4)$$

donde n_i y n_t son, respectivamente, el índice de refracción del medio desde donde incide el rayo y el del medio al que se transmite.

Si se da el caso que $(n_i/n_t)\sin\theta_i > 1$, entonces no existe ningún ángulo θ_t que cumpla la ecuación 4.4, y no hay refracción del rayo, sino que toda la luz se refleja. Es el fenómeno conocido como reflexión total, y es particularmente importante para comprender el funcionamiento de las fibras ópticas, como se ve más adelante. El ángulo de incidencia máximo que permite la transmisión de la luz al segundo medio es

$$\theta_i = \arcsin \frac{n_t}{n_i} \quad (4.5)$$

En el fenómeno de la reflexión y refracción, una parte de la potencia incidente se transmite al segundo medio y el resto se refleja al primero. La fracción R de la potencia reflejada es:

$$R = \left(\frac{n_i - n_t}{n_i + n_t} \right)^2 \quad (4.6)$$

El flujo P de energía que se propaga se mide en $\text{J}/\text{cm}^2\text{s} = \text{W}/\text{cm}^2$. En el caso más general, la radiación es la superposición en proporciones variables de radiaciones de distintas longitudes de onda. La contribución de cada una de ellas a la potencia del flujo radiante viene dada por la función distribución espectral $dP/d\lambda$, expresada en $\text{W}/(\text{cm}^2 \cdot \mu\text{m})$.

Fig. 4.2 Curva normal de sensibilidad del ojo humano

Cuando el interés se centra en el efecto de la luz sobre la visión humana, entonces esa distribución se pondera según la sensibilidad del ojo humano, de acuerdo con una curva normal estándar que se representa en la figura 4.2. La unidad de potencia luminosa —la iluminación— es el *lumen*. Un vatio de luz de 555 nm de longitud de onda —máximo de la curva estándar— equivale a 680 lúmenes. Para otras longitudes de onda esta cantidad se debe multiplicar por el valor indicado en la curva normal. Un lumen por metro cuadrado proporciona una densidad de iluminación —denominada *luminancia*— conocida como *lux*.

EJEMPLO 4.2

Valores típicos de luminancia :

Fuente	Luminancia (lux)
A pleno sol	10^5
Día nublado	10^3
Estancia iluminada	10^2
Luna llena	1

Tabla 4.1 El espectro electromagnético

Nombre	Longitud de onda (μm)	Energía del fotón (eV)
Rayos cósmicos	$< 3 \times 10^{-7}$	$> 4.1 \times 10^6$
Rayos gamma	10^{-8} a 8×10^{-3}	1.24×10^8 a 155
Rayos X	2×10^{-6} a 0.2	6.2×10^{-5} a 6.2
Ultravioleta	Extremo	0.01 a 0.2
	Lejano	0.2 a 0.3
	Próximo	0.3 a 0.39
Visible	Violeta	0.39 a 0.455
	Azul	0.455 a 0.492
	Verde	0.492 a 0.577
	Amarillo	0.577 a 0.597
	Anaranjado	0.597 a 0.622
	Rojo	0.622 a 0.77
Infrarrojo	Próximo	0.77 a 1.5
	Medio	1.5 a 6
	Lejano	6 a 40
	Extremo	40 a 10^3
Ondas milimétricas	10^3 a 10^4	1.24×10^{-3} a 1.24×10^{-4}
Microondas	10^4 a 3×10^6	1.24×10^{-4} a 4.13×10^{-7}
Ondas de radio	3×10^6 a 2×10^{11}	4.13×10^{-7} a 6.2×10^{-12}
Oscilaciones eléctricas	$> 2 \times 10^{11}$	$< 6.2 \times 10^{-12}$

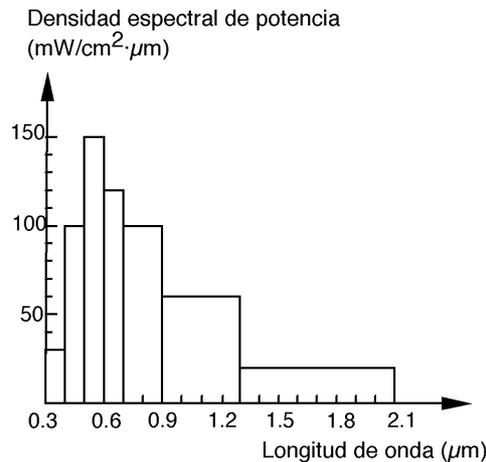
Ejercicio 4.3

La curva normal de respuesta del ojo humano da los valores del 30 % para una luz de 500 nm de longitud de onda, y del 60 % para una de 600 nm. Calcúlense los vatios de luz necesarios para iluminar una superficie de 100 m^2 con 100 lux.

49 W y 24.5 W, respectivamente.

Ejercicio 4.4

La curva adjunta representa una aproximación por tramos a la distribución espectral de la luz solar que llega a la superficie de la tierra en incidencia perpendicular, después de cruzar la atmósfera en condiciones óptimas; se conoce como *espectro estándar AM1*. Calcúlese la potencia de la radiación por unidad de superficie.



100 mW/cm².

El carácter corpuscular se manifiesta en los fenómenos de emisión y de absorción de la luz por la materia. La radiación se compone de unidades indivisibles de energía —los cuantos—, denominados *fonones*. El valor de la energía de un fotón de una radiación de frecuencia f es

$$E = hf = h \frac{c}{\lambda} = \frac{1.24 \text{ eV} \cdot \mu\text{m}}{\lambda} \quad (4.7)$$

donde h es la constante de Planck. En tanto que partícula, el fotón también tiene cantidad de movimiento (o momento). Su valor es h/λ . La tabla 4.1 presenta las regiones del espectro electromagnético.

EJEMPLO 4.3

Los fotones de la radiación para la cual la curva normal de sensibilidad presenta un máximo tienen una energía $1.24/0.555 = 2.23 \text{ eV}$.

Ejercicio 4.5

Calcúlese la energía de los fotones de los extremos del espectro visible (v. tabla 4.1).

1.61 eV en el extremo rojo, y 3.18 eV en el extremo azul.

CUESTIONARIO 4.1.a

1. Los valores límite de longitud de onda del espectro visible son $0.4 \mu\text{m}$ y $0.7 \mu\text{m}$. Determinense los límites de la frecuencia. Datos: velocidad de la luz en el vacío: $c = 3 \times 10^{10} \text{ cm/s}$.

- a) 7.5×10^{14} , 4.3×10^{14} b) 7.5×10^{12} , 4.3×10^{12}
 c) 7.5×10^{12} , 4.3×10^{14} d) 4.3×10^{12} , 7.5×10^{14}

2. ¿Cuántos fotones por centímetro cuadrado y por segundo inciden sobre una superficie que recibe una radiación de 1 mW/cm^2 de luz de $0.65 \mu\text{m}$ de longitud de onda? Datos: $hc = 1.24 \text{ eV} \cdot \mu\text{m}$.

- a) 3.3×10^{18} b) 3.3×10^{12} c) 3.3×10^{15} d) 3.3×10^{21}

3. Determinense la velocidad de la luz en el óxido de silicio. Datos: la constante dieléctrica relativa del óxido de silicio es $\epsilon_r = 3.9$, la permitividad del vacío es $\epsilon_0 = 8.85 \times 10^{-14} \text{ F/cm}$, la permeabilidad magnética relativa es $\mu_r = 0.55$ y la permeabilidad magnética del vacío $\mu_0 = 4\pi \times 10^{-7} \text{ Vs}^2\text{C}^{-1}$.

- a) $c/1.46$ b) $c \cdot 1.46$ c) $c/2.13$ d) $c \cdot 2.13$

4.1.2 Interacción entre la radiación electromagnética y los semiconductores**La absorción de la luz**

El proceso más importante de absorción de luz en un semiconductor es la creación de pares electrón-hueco. Cada fotón absorbido provoca una transición desde la banda de valencia a la de conducción. Para que un semiconductor absorba un fotón es preciso que la energía de éste sea mayor que la de la banda prohibida del material:

$$E > E_g \Leftrightarrow \lambda \leq \lambda_{\text{máx}} = \frac{1.24 \text{ eV} \cdot \mu\text{m}}{E_g} \quad (4.8)$$

Es necesario escoger en cada aplicación el semiconductor con la banda prohibida más adaptada a la radiación del problema. La tabla 4.2 presenta una lista de semiconductores y su margen de aplicación en la detección de la luz.

EJEMPLO 4.4

Las longitudes de onda de $1.3 \mu\text{m}$ y $1.5 \mu\text{m}$ se utilizan en comunicaciones por fibra óptica. Los semiconductores empleados para detectar las señales deben tener anchuras de banda prohibida no superiores a 0.95 eV y 0.83 eV , respectivamente.

Ejercicio 4.6

La fibra óptica también transmite bien la radiación de $0.8 \mu\text{m}$ de longitud de onda. Empleando un dispositivo de GaAs, ¿se pueden detectar señales transmitidas con esa luz?

Sí.

5 El transistor bipolar

5.1 Introducción

El transistor bipolar, conocido en inglés por el acrónimo BJT (*Bipolar Junction Transistor*, transistor bipolar de unión) es uno de los dispositivos más utilizados en los circuitos electrónicos. El nombre de *transistor* procede de la condensación de dos vocablos ingleses, *transfer* y *resistor*, y hace referencia al hecho de que la corriente que circula entre dos terminales está controlada por una señal aplicada al tercer terminal; el término *bipolar* es debido al hecho de que la corriente es transportada por portadores de ambas polaridades: electrones y huecos. La electrónica moderna, basada en circuitos integrados (CI) se inició de hecho con el descubrimiento de este dispositivo. Actualmente sigue siendo el dispositivo amplificador por excelencia, y el que más se emplea en los CI analógicos. En este capítulo se hace una introducción al transistor bipolar y se presenta la teoría de su funcionamiento, su proceso de fabricación, el comportamiento en continua y en señal, y sus modelos circuitales.

El transistor bipolar fue descubierto casualmente en diciembre de 1947 por Bardeen, Brattain y Shockley en los Laboratorios Bell, cuando intentaban realizar un «amplificador de estado sólido» basado en lo que más adelante se denominaría *transistor MOS*. Ese descubrimiento fue seguido casi inmediatamente de la teoría que explicaba su funcionamiento, y condujo a una revolución tecnológica que significó la desaparición, en pocos años, de la tecnología de las válvulas de vacío, que hasta entonces había proporcionado soporte físico a los circuitos electrónicos.

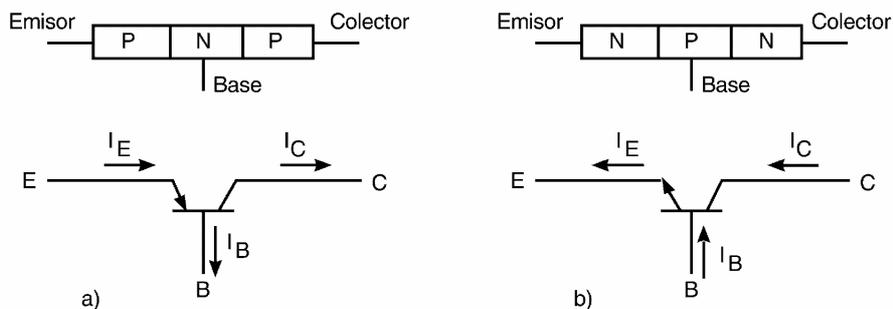


Fig. 5.1 Estructura, símbolo y sentido positivo de las corrientes del transistor bipolar.
a) PNP. b) NPN.

El transistor bipolar es un dispositivo de tres terminales denominados *emisor*, *base* y *colector*. Hay dos tipos de transistores bipolares: los NPN y los PNP, cuyas denominaciones responden a su estructura básica, la cual se esquematiza en la figura 5.1. El transistor bipolar tiene dos uniones PN:

una entre el emisor y la base —la unión *emisora*— y otra entre la base y el colector —la unión *colectora*—. En el símbolo de estos transistores se incluye una flecha en el terminal del emisor, que apunta siempre en el sentido de P a N, y sirve para identificar el tipo de transistor. Los sentidos de las corrientes indicadas en la figura 5.1. se consideran positivos. Estos sentidos se basan en la asignación al emisor del sentido de la corriente de su símbolo —de P a N—, y a la base y al colector de los derivados de la suposición de que la corriente de emisor es la suma de la corriente de base más la corriente de colector.

El transistor más utilizado es el NPN, y es el que se estudia en este texto. Según se ve más adelante, el comportamiento del PNP es dual con respecto al del NPN, lo cual significa que si se cambian electrones por huecos, junto con los sentidos de las corrientes y las polaridades de las tensiones, el comportamiento resulta idéntico.

5.1.1 Principio de funcionamiento

Cuando el transistor se emplea como amplificador, se polariza el diodo del emisor en directa y el del colector en inversa. Esa polarización se denomina *activa* y es la que se considera aquí. Conviene recordar de la teoría de la unión PN que con una polarización directa cada región inyecta sus portadores mayoritarios a la región adyacente, de forma que en el inicio de esa segunda región, justo en la frontera con la zona de carga espacial —la ZCE—, la concentración de los minoritarios inyectados es $m_0 \exp(V_D/V_T)$, siendo m_0 la concentración de minoritarios en esa región en condiciones de equilibrio térmico, y V_D la tensión de polarización de la unión. Cuando una unión se polariza en inversa, el campo eléctrico de la ZCE aumenta y domina el transporte de minoritarios de una región a la otra. Si la polarización es suficientemente negativa, la concentración de éstos en la frontera con la ZCE se anula.

Se analizan ahora las corrientes que circulan por el transistor. Al estar la unión emisora en directa, el emisor N inyecta electrones a la base P, y ésta inyecta huecos al emisor N. La teoría de la unión PN muestra que si el dopaje de la región N es mucho mayor que el de la P, la corriente de electrones a través de la ZCE de la unión es muy superior a la de huecos. Esa relación entre corrientes es la que la figura 5.2 representa. Por otro lado, al estar la unión colectora polarizada en inversa, la concentración de electrones en el punto I_B de la región P —frontera con la ZCE del colector— es cero. Por lo tanto, en la región neutra de la base P del transistor hay una diferencia de concentraciones de electrones entre los puntos 0_B —la frontera con la ZCE del emisor— y I_B , diferencia que origina un flujo de electrones por difusión desde la parte del emisor hacia la parte del colector. Cuando esos electrones llegan a la ZCE del colector, el campo eléctrico presente en esa región los arrastra desde la base hacia el colector. La corriente de huecos en esa segunda unión es prácticamente nula, porque está polarizada inversamente y casi no hay huecos en la parte N del colector.

Una parte de los electrones que se trasladan por difusión desde el emisor hacia el colector a través de la base se recombinan. Es la corriente denominada I_r en la figura 5.2. En régimen estacionario debe entrar por el terminal de la base una corriente de huecos igual a la de electrones que se recombinan, ya que de otro modo los huecos de la base terminarían por agotarse. Por la misma razón, por el terminal de la base deben entrar los huecos que la base inyecta al emisor, y deben salir los pocos huecos que el colector inyecta a la base.

Partiendo de esas corrientes elementales se puede escribir:

$$I_E = I_{En} + I_{Ep} \quad I_C = I_{Cn} + I_{Cp} \quad I_B = I_r + I_{Ep} - I_{Cp} \quad (5.1)$$

donde el subíndice E significa *emisor*, el C *colector* y el B *base*; el otro subíndice indica n para corriente de electrones y p para corriente de huecos. Si se tiene en cuenta que en un transistor NPN normal la corriente I_{En} es muy superior a I_{Ep} y a I_r , resulta que

$$I_E \cong I_{En} \cong I_{Es} (e^{V_{BE}/V_t}) \quad I_C \cong I_{Cn} \cong I_E \cong I_{Es} e^{V_{BE}/V_t} \quad (5.2)$$

que indica que la corriente de colector se controla con la tensión de polarización de la unión emisora y es independiente de la polarización de la unión colectora, siempre que el transistor esté polarizado en la región activa. Esa propiedad se denomina *efecto transistor*. Nótese que si se sustituye el transistor por dos diodos en oposición con un tercer terminal entre los dos ánodos —a efectos de simular las dos uniones del transistor NPN— no se produce el efecto transistor, ya que es necesario que las dos uniones compartan la región central —la base— y que esa región sea suficientemente delgada para permitir que los minoritarios inyectados por una unión alcancen la otra.

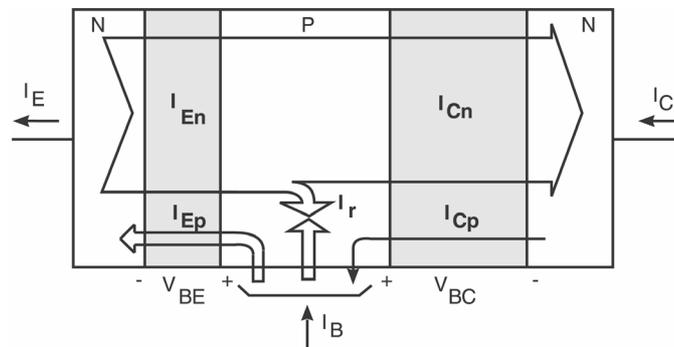


Fig. 5.2 Corrientes en un transistor NPN. Las flechas interiores del rectángulo señalan los flujos de portadores, y las externas los sentidos de las corrientes eléctricas. Se debe tener en cuenta que el sentido de la corriente eléctrica de los electrones es contrario al sentido de su flujo, es decir, I_{En} tiene el mismo sentido que I_E , e I_{Cn} el mismo que I_C .

Ejercicio 5.1

Un transistor NPN tiene una $I_{Es} = 10^{-16}$ A. ¿Cuál es la corriente de colector si $V_{BE} = 0.7$ V y la unión colectora está en inversa?

Aplicando la expresión 5.2, $I_C = 10^{-16} \exp(0.7/0.025) = 0.14$ mA.

Ejercicio 5.2

¿Cuál es la tensión V_{BE} en el transistor anterior cuando $I_C = 1$ mA?

$V_{BE} = 0.748$ V.

La capacidad del transistor como amplificador de señales está ligada a ese efecto. Considérese el circuito de la figura 5.3, en el que el transistor NPN tiene la unión emisora polarizada directamente

por la tensión $[V_{EE} - \Delta V_i(t)]$, siendo $\Delta V_i(t)$ una señal de pequeña amplitud que se desea amplificar. La unión colector está polarizada inversamente por V_{CC} . Aproximando la caída de tensión en el diodo del emisor con la tensión de codo $V_\gamma (\cong 0.7 \text{ V en el silicio})$ resulta

$$V_{EE} - \Delta V_i(t) = I_E R_E + V_\gamma \Rightarrow I_E = \frac{V_{EE} - V_\gamma - \Delta V_i(t)}{R_E} = I_{EQ} - \Delta I_E(t) \quad (5.3)$$

Suponiendo que I_C se pueda aproximar con I_E , y siempre que la polarización del diodo del colector sea inversa ($V_{CB} > 0$), resulta que

$$V_o = V_{CC} - I_C R_C = [V_{CC} - I_{EQ} R_C] + \frac{R_C}{R_E} \Delta V_i(t) = V_{oQ} + \Delta V_o(t); \quad \Delta V_o(t) = \frac{R_C}{R_E} \Delta V_i(t) \quad (5.4)$$

Como se puede ver, en la salida aparece una señal $\Delta V_o(t)$ que es proporcional a la señal de entrada. El factor R_C/R_E que multiplica a $\Delta V_i(t)$ en la última expresión se denomina *ganancia de tensión del amplificador*, la cual se puede controlar fijando la relación entre resistencias, con la única restricción de que V_o debe siempre ser positiva para asegurar que el transistor esté polarizado en activa.

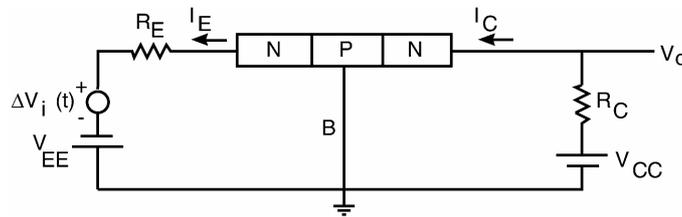


Fig. 5.3 Ejemplo de circuito amplificador

Ejercicio 5.3

En el amplificador de la figura anterior, $R_C = 5 \text{ k}\Omega$ y $R_E = 100 \Omega$. ¿Cuál es, aproximadamente, la ganancia de tensión del circuito?

La ganancia aproximada es $R_C/R_E = 50$.

Ejercicio 5.4

¿Qué valor debe tener R_C en el circuito de la cuestión anterior para conseguir una ganancia de 100?

$R_C = 10 \text{ k}\Omega$

El circuito de la figura 5.3 tiene el terminal de la base común a las mallas del emisor y del colector. Por ese motivo se dice que la configuración del circuito es *en base común*. Existen también las configuraciones *de emisor común* y *de colector común*, que presentan el terminal del emisor o el del colector comunes a las mallas de entrada y salida. La configuración de emisor común es la más utilizada en amplificación, mientras que la de colector común —también denominada *seguidor por emisor*— suele emplearse como etapa de salida.

Los transistores bipolares, además de emplearse como amplificadores, se usan en circuitos digitales, en los que los transistores trabajan en dos estados. Por este motivo esos dispositivos no se utilizan solo en polarización activa, sino también en los modos denominados *corte* y *saturación*. En la tabla 5.1 se presentan los cuatro modos posibles de polarización de un transistor. El modo inverso no se suele emplear en la práctica, dado que en ese modo las prestaciones del transistor son similares a las del modo activo, pero inferiores.

Tabla 5.1 Modos de funcionamiento del transistor

		Unión emisora	
		Directa	Inversa
Unión colectora	Directa	Saturación	Inverso
	Inversa	Activo	Corte

5.1.2 La estructura física del transistor bipolar. Su proceso de fabricación.

En la figura 5.4 se presenta la estructura física del transistor bipolar que se emplea en circuitos integrados. Esa estructura responde a la necesidad de disponer los tres terminales del transistor sobre la superficie de la oblea y de aislar unos de otros los dispositivos del CI que comparten un sustrato común. Tal y como se señala en el capítulo 3, dedicado a la tecnología de los semiconductores, ese aislamiento se consigue creando cada dispositivo dentro de un bloque con dopaje de tipo contrario al del sustrato, y polarizando inversamente la unión creada. Con esa polarización, la unión equivale a un circuito abierto y las corrientes que circulan por el interior del bloque incrustado no pueden acceder al sustrato, por lo que no llegan a los demás bloques. La comunicación de esas corrientes interiores a cada bloque se permite solo a través de las pistas metálicas realizadas sobre la superficie de la oblea.

La figura 5.4 muestra la estructura de un transistor NPN. El dispositivo se ha realizado en un bloque N creado dentro del sustrato P. La unión PN entre el sustrato y ese bloque se polariza inversamente, según se justifica en el párrafo anterior. Por otro lado, se accede a las regiones del emisor, la base y el colector desde la superficie. La corriente que circula entre el emisor y el colector se desplaza en vertical, tal y como representa la citada figura. Para acceder a la parte central de la base se crea una región P ancha, con la que se hace contacto desde la superficie (terminal B). Debe notarse que la corriente de base debe recorrer un trayecto largo antes de alcanzar la parte central de la base. Ese trayecto introduce una cierta resistencia, que se incluye en el circuito equivalente del transistor.

Algo parecido sucede con el terminal del colector. La corriente que sale por ese terminal debe recorrer un trayecto muy largo por el interior del silicio, lo cual introduce una resistencia considerable. Para disminuir la resistencia de ese trayecto se crea una región N^+ muy dopada, que presenta una resistencia reducida. Esa región se denomina *capa enterrada*. El trayecto «vertical» que sigue la corriente del transistor desde el emisor hasta el inicio de la capa enterrada permite que el comportamiento del dispositivo se pueda aproximar empleando un modelo unidimensional. Debe notarse, no obstante, que los terminales del emisor y del colector no son intercambiables, como podría suponerse partiendo de una estructura meramente unidimensional.

La figura 5.5 representa el proceso de fabricación de un transistor bipolar NPN. Partiendo de una oblea P se crea una capa muy dopada N^+ en la superficie, que se convierte finalmente en la capa enterrada. Para crear esa región N^+ se utiliza una primera máscara. El paso siguiente consiste en hacer

crecer una capa epitaxial de unos 10 a 20 micrómetros de espesor en la superficie de toda la oblea. Esa capa se hace crecer con el dopaje N que va a tener el colector del transistor. Con esa acción, la capa N^+ realizada en la anterior etapa queda «enterrada» en el fondo de la capa epitaxial N. Seguidamente se procede a dividir esa capa en porciones aisladas una de otra, en cada una de las cuales se fabrica un dispositivo. Para ello se realiza una difusión P de aislamiento desde la superficie, empleando una segunda máscara. La difusión se lleva a cabo de manera que la nueva región P atraviese completamente la capa epitaxial y alcance el sustrato P.

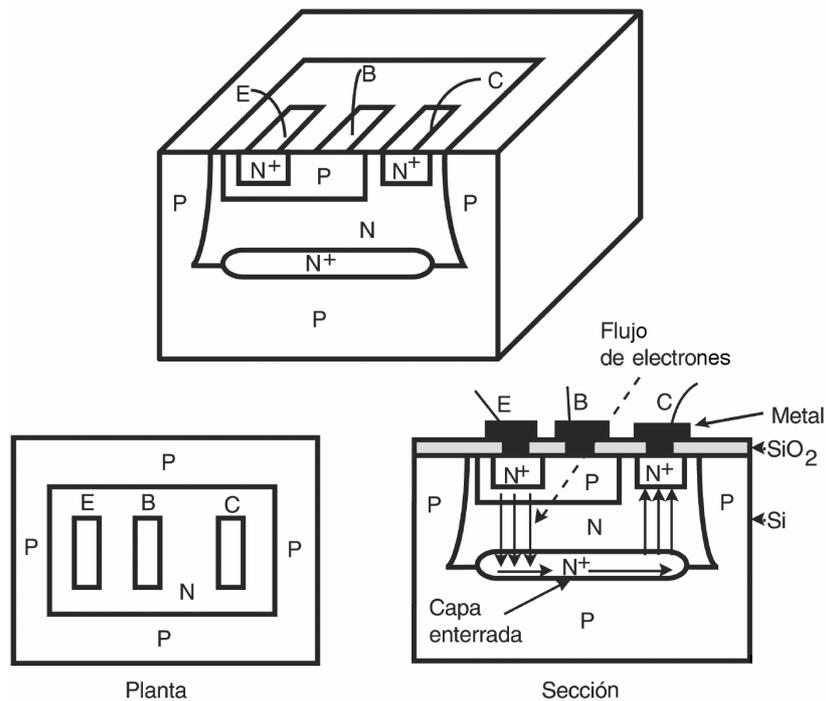


Fig. 5.4 Estructura física de un transistor bipolar NPN

Construidas las porciones o islas N, se procede a fabricar los transistores bipolares propiamente dichos. La cuarta etapa consiste en realizar la base mediante una difusión P, para lo que se emplea una tercera máscara. Después de oxidar de nuevo toda la oblea, se abren nuevas ventanas con la máscara correspondiente, y se realizan las difusiones N^+ del emisor y el colector. Se elimina el óxido y, tras otra oxidación, se abren las ventanas que van a permitir que el metal que se deposita seguidamente haga contacto con las regiones del emisor, la base y el colector del transistor. Finalmente, se elimina parte del metal depositado sobre toda la oblea, de modo que solo queden las pistas que interconectan los terminales entre sí.

El proceso que se ha descrito requiere seis máscaras. Mientras en unas islas se crean transistores bipolares, en otras se pueden utilizar los procesos descritos para realizar otros tipos de dispositivos (resistencias, diodos...) y crear así un circuito integrado. Este proceso tecnológico basado en la fabricación de un transistor bipolar se denomina *tecnología bipolar*.

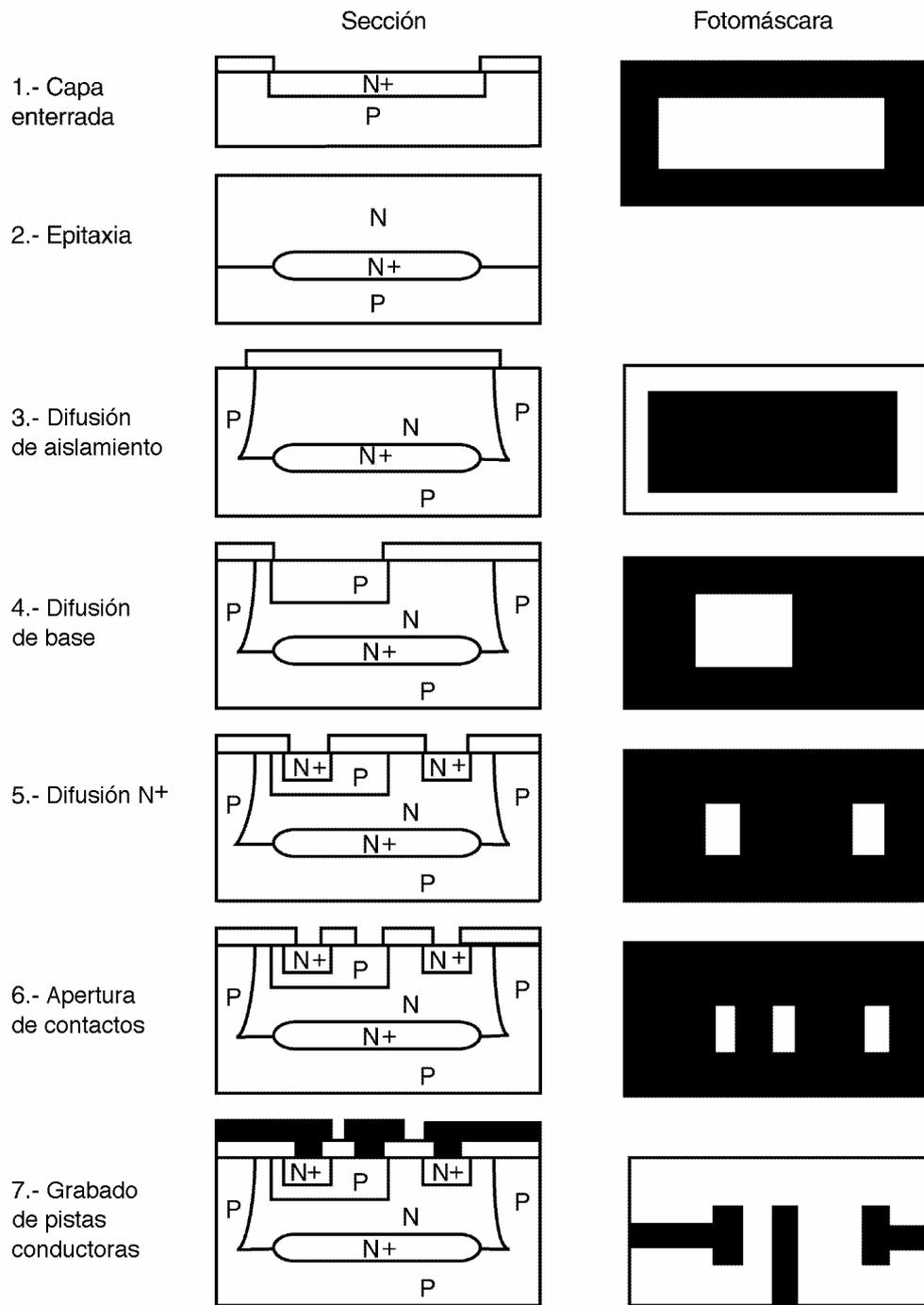
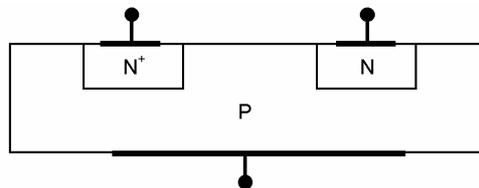


Fig. 5.5 Tecnología bipolar: proceso de fabricación del transistor bipolar

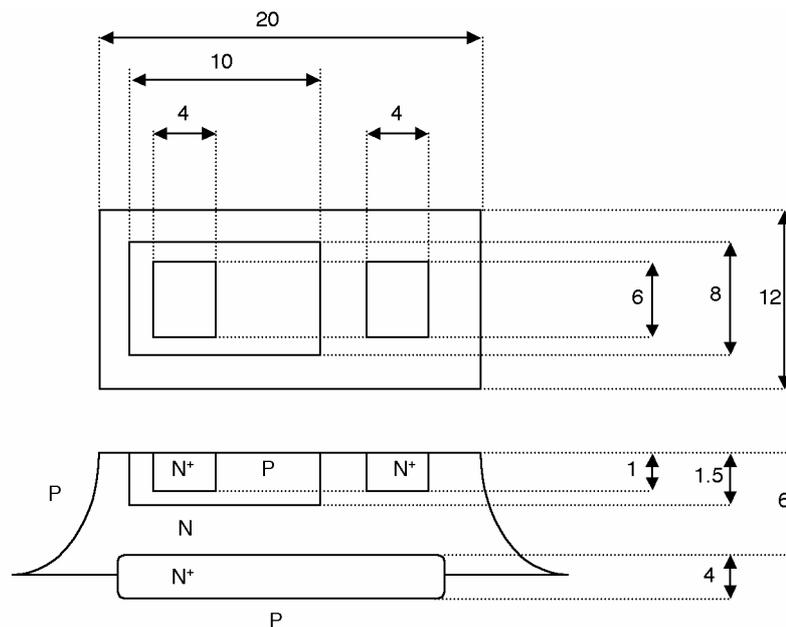
CUESTIONARIO 5.1

1. La estructura representada en la figura siguiente puede corresponder a un transistor bipolar o a dos diodos con los ánodos conectados entre sí.



¿Cuál de las siguientes condiciones se debe cumplir para que se trate de un transistor?

- La separación entre las regiones N^+ y N es mucho más pequeña que la longitud de difusión de los huecos en la región N^+ .
- La separación entre las regiones N^+ y N es mucho más pequeña que la longitud de difusión de los electrones en la región P .
- La separación entre las regiones N^+ y N es mucho más pequeña que la longitud de difusión de los huecos en la región N .
- La separación entre el contacto metálico de la región P y las regiones del tipo N es mucho más pequeña que la longitud de difusión de los electrones en la región N^+ .



2. La figura adjunta representa la vista en planta de un transistor bipolar integrado y una sección del mismo dispositivo. Las dimensiones indicadas se expresan en micrómetros (nótese que no se respeta ninguna escala).

6 Transistores de efecto de campo

La aparición del transistor fue el resultado de la búsqueda de componentes de estado sólido para sustituir a los tubos de vacío, que eran los dispositivos activos básicos de la electrónica de la primera mitad del siglo XX y que se preveían más pequeños y con un menor consumo de potencia. En el triodo, que era el tubo de vacío básico, la corriente que circulaba entre el ánodo y el cátodo se modulaba por la tensión aplicada a un tercer terminal, la rejilla. De esta forma se lograba amplificar la señal aplicada a este último terminal. Tubos más sofisticados eran mejoras de esta estructura básica. La idea inicial del transistor consistía en hacer circular corriente entre dos terminales a través de un semiconductor y modular esta corriente utilizando un tercer electrodo. Aunque se pensaba en la tensión como variable de control, las limitaciones tecnológicas de aquella época, -la segunda mitad de los años 40-, condujeron a un dispositivo en el que la variable de control era una corriente. Nació así el transistor bipolar en 1947. Una década más tarde, cuando la tecnología de los semiconductores se desarrolló considerablemente, se obtuvo el primer transistor de efecto de campo en el que la variable de control era una tensión. El desarrollo que ha tenido lugar desde entonces ha permitido crear una amplia gama de dispositivos que han ido desplazando al transistor bipolar en un gran número de aplicaciones.

El MOSFET, el JFET, el MESFET y el HEMT son familias de transistores de especial relevancia en la tecnología actual. El transistor metal-óxido-semiconductor (MOSFET, metal-oxide-semiconductor field effect transistor, o más brevemente MOST) es propio de la tecnología de silicio. En este dispositivo, la corriente circula entre dos terminales denominados *drenador* y *surtidor* (o fuente) a través de una región semiconductor denominada *canal*. Su intensidad se modula por la tensión aplicada a un electrodo denominado *puerta* (originariamente de metal), aislada del semiconductor por una capa de dieléctrico (óxido del propio silicio), de forma que no circula corriente por este terminal. El nombre de este dispositivo deriva de esta estructura. Históricamente, el MOSFET también ha sido denominado transistor de puerta aislada (IGFET, insulated gate field effect transistor). Actualmente es el dispositivo más utilizado en los circuitos integrados de silicio. Los MOST discretos se utilizan casi exclusivamente en los circuitos de electrónica de potencia

En la elaboración de este capítulo también han participado los profesores del Departamento de Ingeniería Electrónica de la UPC Ramón Alcubilla González y Ángel Rodríguez Martínez.

En el transistor de efecto de campo de unión (JFET, junction field effect transistor), la puerta no está aislada del canal por un dieléctrico, sino por una unión PN polarizada inversamente. Se suele fabricar con silicio y habitualmente se presenta en forma de componente discreto. Para aplicaciones de alta frecuencia se han desarrollado transistores rápidos, fabricados con semiconductores del grupo III-V. Entre estos destacan el MESFET (metal-semiconductor field effect transistor), en el que entre puerta y canal hay un contacto rectificador metal semiconductor polarizado en inversa. En los últimos años ha ganado importancia en aplicaciones de alta frecuencia una familia de transistores de electrones de alta movilidad (HEMT, High electron mobility transistor), basados en heterouniones. La mayor parte de este capítulo se dedicará al estudio del MOSFET (apartados 6.1 a 6.6), dada la extraordinaria importancia de este dispositivo en la electrónica actual. La descripción del MESFET y del JFET se hará de forma más breve y de manera conjunta, ya que tienen muchos aspectos comunes.

6.1 Electrostática del sistema metal-óxido-semiconductor

6.1.1 Fundamentos del MOSFET

Considérese la estructura representada en la figura 6.1. En un cristal de silicio tipo P, denominado *sustrato* (B, de *bulk*) se han creado dos regiones de tipo N, denominadas *drenador* (D, de *drain*) y *surtidor o fuente* (S, de *source*). La superficie del semiconductor comprendida entre estas dos regiones, que se denominará región de canal, está recubierta de material dieléctrico, el óxido de silicio (SiO_2). Encima de este material está el metal o electrodo de *puerta* (G, de *gate*).

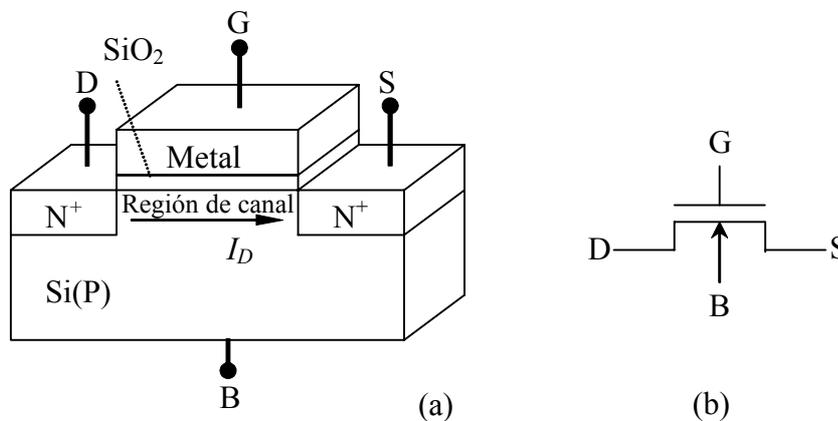


Fig. 6.1 Transistor MOS de canal N a) Estructura física b) Símbolo circuital. Los símbolos B, D, S y G designan indistintamente las regiones del dispositivo y los respectivos terminales.

Considérese una tensión V_{DS} aplicada entre los terminales de drenador y surtidor y se examinará la corriente I_D , denominada corriente de drenador, que puede circular entre ellos a través de la región de canal. Esta corriente deberá atravesar dos regiones PN conectadas en oposición: la del drenador con el

sustrato y la del sustrato con el surtidor. Por tanto, el paso de corriente está bloqueado. La situación cambia si se aplica una tensión positiva al terminal de puerta respecto al sustrato, ya que en estas condiciones los electrones, portadores minoritarios del sustrato, serán atraídos hacia el metal. Como el óxido impide que estos electrones puedan pasar al metal, quedan acumulados en la región superficial de canal. Si la concentración de estos electrones es suficiente, las regiones de drenador y surtidor quedan conectadas por una capa superficial de electrones, que se denominará *canal*, que permite el paso de corriente. El valor de la tensión de puerta determina el número de electrones presentes en el canal y, de esta forma, modula el valor de la corriente de drenador. Similares consideraciones se habrían podido hacer en un dispositivo construido en un cristal tipo N con regiones de drenador y surtidor de tipo P. En este caso, la tensión que debería aplicarse a la puerta sería negativa respecto al sustrato, y se diría que se trata de un transistor de canal P, mientras que el de la figura 6.1 sería un transistor de canal N.

Se iniciará el estudio del dispositivo acabado de presentar analizando el sistema metal (puerta) - óxido (dieléctrico) - semiconductor (sustrato), para determinar la relación entre la tensión de puerta y la concentración de portadores en el canal. Con esta información, se determinará la relación entre las tensiones continuas aplicadas a los terminales y la corriente de drenador. El estudio se completará analizando el comportamiento dinámico de estos transistores.

6.1.2 El sistema metal-óxido-semiconductor en equilibrio

El análisis de la formación de la capa conductora en la región de canal es más fácil de realizar en una estructura formada únicamente por las regiones de puerta, óxido y sustrato (estructura MOS). Este sistema puede ser visto como un condensador, de forma que el estudio de la carga acumulada en la "armadura" del semiconductor permitirá entender la formación del canal. El diagrama de bandas de la estructura determina la distribución de carga. Inicialmente, se considerará silicio tipo P y un metal con una función trabajo menor que la del semiconductor $q\Phi_m < q\Phi_s$. Se asignará al óxido un diagrama de bandas con una banda prohibida muy grande. La figura 6.2a representa los diagramas de bandas de los tres materiales separados, sin contacto entre ellos. La figura 6.2b corresponde al sistema formado por los tres materiales en contacto, suponiendo condiciones de equilibrio térmico, es decir, que el nivel de Fermi es constante. El eje de abscisas es perpendicular a la superficie del dispositivo.

El procedimiento para construir este diagrama es similar al utilizado en el estudio de los contactos entre metal y semiconductor (apartado 2.7). La diferencia de potencial entre los extremos de la estructura es $\Phi_s - \Phi_m$, pero mientras en aquel caso toda la caída de potencial tenía lugar en el semiconductor, aquí se reparte entre el semiconductor, V_{s0} , y el óxido, V_{ox0} . El dieléctrico se comporta como el de un condensador y, suponiendo que no haya cargas localizadas en su interior, la distribución de potencial a través de esta capa es lineal. Por tanto:

$$\Phi_s - \Phi_m = V_{ox0} + V_{s0} \quad (6.1)$$

En la figura 6.2b la curvatura de las bandas en la región del semiconductor corresponde a una zona vacía de mayoritarios, como la ZCE de la región P de una unión PN (figura. 2.6) o de un contacto metal semiconductor (figura. 2.21c). La carga neta por unidad de sección, Q_{s0} , localizada en esta región se puede calcular aplicando las ecuaciones de electrostática, de forma similar a como se hizo en

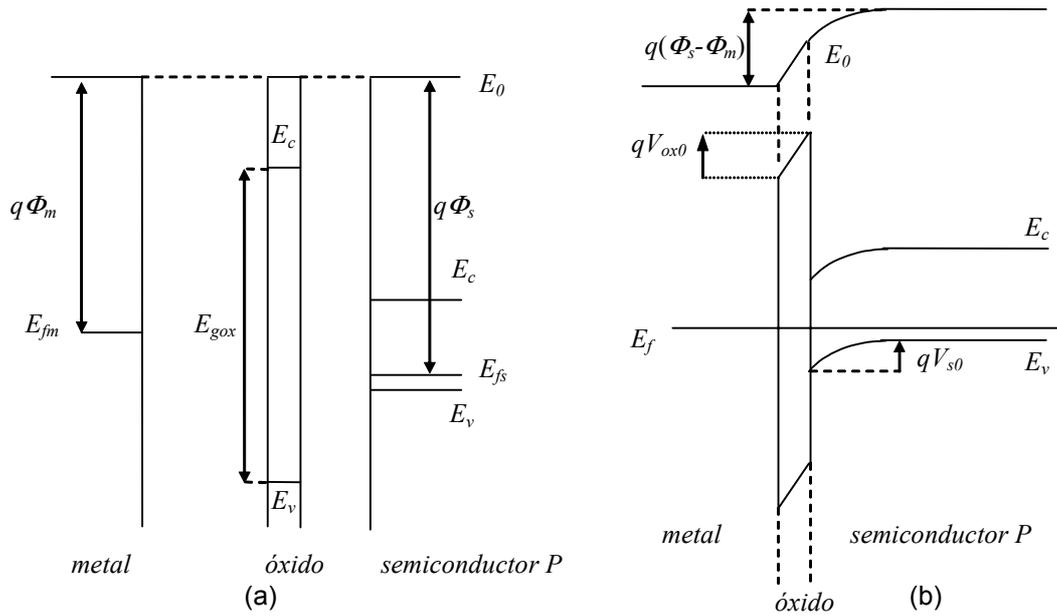


Fig. 6.2 Bandas de energía del sistema MOS en equilibrio: a) Diagramas de bandas de los materiales separados. b) Diagrama de bandas del sistema.

las expresiones 2.11 y 2.13. Teniendo en cuenta que ahora la diferencia de tensión entre los extremos de la zona con carga es V_{s0} , en lugar de $V_{bi} - V_D$:

$$Q_{s0} = -qN_A w_{d0} = -qN_A \sqrt{\frac{2\epsilon_s}{q} \frac{1}{N_A} V_{s0}} = -\sqrt{2q\epsilon_s N_A V_{s0}} \quad (6.2)$$

siendo N_A la concentración de impurezas aceptadoras en el silicio, ϵ_s su constante dieléctrica (el subíndice es necesario para no confundirla con la del óxido, ϵ_{ox}), y w_{d0} la anchura de la ZCE. La curvatura de las bandas en la figura 6.2 implica que la carga del semiconductor es negativa. Por tanto, en el metal de puerta habrá una carga positiva y del mismo valor absoluto, y tendrá una tensión positiva respecto al semiconductor. En la ecuación 6.1 los términos V_{s0} y V_{ox0} tienen signo positivo.

Por otra parte, como en todo condensador, la caída de potencial en el dieléctrico es:

$$V_{ox0} = \frac{-Q_{s0}}{C_{ox}} \quad (6.3)$$

siendo C_{ox} su capacidad, de valor (por unidad de área de la sección):

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad (6.4)$$

donde ϵ_{ox} es la constante dieléctrica del óxido, $\epsilon_{ox} = \epsilon_{rox}\epsilon_0$. La constante dieléctrica relativa del óxido de silicio vale 3.9. El grosor del óxido es t_{ox} y su valor, dependiendo de la tecnología usada, suele ser desde unas pocas decenas de nanómetros hasta algunas unidades. A partir de las expresiones 6.1 a 6.4 se puede determinar los valores de V_{ox0} , V_{s0} y Q_{s0} para un valor dado de $\Phi_s - \Phi_m$.

El diagrama de la figura 6.2b representa un caso de los cuatro posibles. Al igual que ocurría en los contactos metal semiconductor, los otros tres casos corresponden a silicio tipo P con $\Phi_m > \Phi_s$, a silicio tipo N con $\Phi_m < \Phi_s$ y a silicio tipo N con $\Phi_m > \Phi_s$. En este último caso, el semiconductor también presenta, en equilibrio, una zona vacía de mayoritarios. El procedimiento de análisis es paralelo al acabado de realizar, tal como se muestra en el ejercicio 6.1.

En los primeros tiempos de la tecnología MOS, el electrodo de puerta era de aluminio, metal utilizado habitualmente en las conexiones de circuitos integrados. Por razones tecnológicas, este material fue sustituido posteriormente por silicio policristalino (polisilicio) que, con un dopado elevado, presenta un comportamiento eléctrico casi metálico. El análisis de la estructura en este caso es similar al acabado de realizar. Sólo hay que sustituir en la figura 6.2a el diagrama de bandas del metal por el del silicio con un nivel de Fermi próximo a la banda de conducción, si se trata de material N, o a la banda de valencia, si es P. Cuando se construye el diagrama de la figura 6.2b, el nivel de Fermi del polisilicio no se desplaza, ya que se trata de un material con muchos portadores. En el ejercicio 6.2 se tratan estos conceptos.

Ejercicio 6.1

Calcular el diagrama de bandas de una estructura metal-óxido-semiconductor en equilibrio, siendo la puerta de aluminio, con una función trabajo de 4.1 eV, el grosor del óxido de silicio de 25 nm y el semiconductor de tipo P dopado con $N_A = 10^{16}$. La afinidad electrónica del silicio es de 4.05 eV. Datos: $\epsilon_{ox} = 3.45 \times 10^{-13}$ F/cm, $\Phi_s = 10^{-12}$ F/cm, $k_B T/q = 0.025$ eV, y $n_i = 1.5 \times 10^{10}$ cm⁻³.

$$\begin{aligned} \text{Función trabajo del silicio: } q\Phi_s &\approx q\chi_s + E_g/2 + (E_{fi} - E_v) = q\chi_s + E_g/2 + (k_B T/q) \ln(N_A/n_i) \\ &= 4.05 + 0.55 + 0.025 \cdot \ln(10^{16}/1.5 \times 10^{10}) = 4.935 \text{ eV. Por tanto: } q\Phi_s - q\Phi_m = 0.835 \text{ eV.} \end{aligned}$$

Por otra parte, $C_{ox} = \epsilon_{ox}/t_{ox} = 1.38 \times 10^{-7}$ F/cm². Substituyendo las expresiones 6.2 a 6.4 en 6.1 se tiene:

$$\Phi_s - \Phi_m = \frac{\sqrt{2q\epsilon_s N_A V_{s0}}}{C_{ox}} + V_{s0} \Rightarrow 0.835 \text{ V} = 0.41 \sqrt{V_{s0}} + V_{s0} \Rightarrow V_{s0} = 0.53 \text{ V}, V_{ox0} = 0.30 \text{ V}$$

La forma de las bandas es similar a la de la figura 6.2b. En la superficie del silicio la separación entre la banda de conducción, y el nivel de Fermi es:

$$\left[E_c - E_f \right]_s = \left[E_c - E_f \right]_{interior} - qV_{s0} = q\Phi_s - qV_{s0} - q\chi = 4.935 - 0.53 - 4.05 = 0.355 \text{ eV}$$

Apéndice B. Constantes, unidades y parámetros

Magnitud	Símbolo	Valor
Nº de Avogadro	N_{Av}	$6.022 \times 10^{23} \text{ mol}^{-1}$
Constante de Boltzmann	k_B	$1.38 \times 10^{-23} \text{ J/K}$
	k_B	$8.62 \times 10^{-5} \text{ eV/K}$
Carga del electrón	q	$1.602 \times 10^{-19} \text{ C}$
Electrón-voltio	eV	$1.602 \times 10^{-19} \text{ J}$
Masa del electrón en reposo	m_0	$0.911 \times 10^{-30} \text{ kg}$
Permeabilidad del vacío	μ_0	$1.262 \times 10^{-8} \text{ H/cm} \cdot (4\pi \times 10^{-9})$
Constante dieléctrica del vacío	ϵ_0	$8.854 \times 10^{-14} \text{ F/cm}$
Constante de Planck	h	$6.626 \times 10^{-34} \text{ J}\cdot\text{s}$
Velocidad de la luz	c	$2.998 \times 10^8 \text{ m/s}$
Tensión térmica a 300 K	$V_t(300\text{K}) = k_B T/q$	0.02586 V
	$k_B T(300\text{K})$	0.02586 eV
1 Ångstrom	1 Å	$1 \times 10^{-8} \text{ cm}$
1 micrómetro	1 μm	$1 \times 10^{-4} \text{ cm}$
1 pulgada (en inglés, <i>inch</i>)	1 in o 1"	2.54 cm

Magnitud	Si	Ge	GaAs	AlAs	GaP	InP	InGaAs
E_g (eV)	1.11	0.67	1.424	2.16	2.26	1.35	0.75
Tipo de BP	ind.	ind.	dir.	ind.	ind.	dir.	dir.
Afinidad $q\chi$ (eV)	4.05	4.13	4.07		4.3	4.4	4.6
Constante de red a (Å)	5.43	5.66	5.65	5.66	5.45	5.87	5.87
Densidad (g/cm^3)	2.33	5.32	5.31	3.60	4.13	4.79	
Temp. fusión ($^\circ\text{C}$)	1415	936	1238	1740	1467	1070	
ϵ_r	11.8	16	13.2	10.9	11.1	12.4	13.5
m_n/m_0	1.18	0.55	0.063			0.08	0.045
m_p/m_0	0.81	0.3	0.53			0.869	0.535
N_c (10^{18} cm^{-3})	3.22×10^{19}	1.04×10^{19}	4.7×10^{17}			5.68×10^{17}	2.8×10^7
N_v (cm^{-3})	1.83×10^{19}	6.0×10^{18}	7×10^{18}			6.35×10^{19}	6×10^{18}
n_i (cm^{-3})	1.02×10^{10}	2.33×10^{13}	2.1×10^6			1.0×10^7	6.5×10^{11}
μ_n (cm^2/Vs)	1450	3900	8500	180	300	4600	13800
μ_p (cm^2/Vs)	500	1900	500		150	150	100
Cond. térm. ($\text{W}/[\text{cm}\cdot^\circ\text{C}]$)	1.31	0.6	0.54	0.8	0.97	0.68	0.26

* Los parámetros del semiconductor compuesto $\text{Al}_x\text{Ga}_{1-x}\text{As}$ son función de x :

$$E_g = 1.424 + 1.247x \text{ (BP indirecta hasta } x = 0.4), q\chi = 4.07 - 1.1x, \epsilon_r = 13.1 - 3.12x$$

* En el semiconductor compuesto $\text{GaAs}_{1-x}\text{P}_x$, la banda prohibida varía linealmente entre $x = 0$ (GaAs, $E_g = 1.424 \text{ eV}$) y $x = 0.45$ ($E_g = 2 \text{ eV}$). En ese intervalo, la BP es del tipo directo.

* Los datos del semiconductor compuesto $\text{In}_x\text{Ga}_{1-x}\text{As}$ corresponden a un valor de $x = 0.53$.

Apéndice C. El cuestionario interactivo DELFOS*

El cuestionario interactivo DELFOS es un programa informático que contiene los enunciados y soluciones de los cuestionarios y problemas guiados contenidos en este libro. Está concebido como una herramienta de auto-aprendizaje que permite practicar de forma fácil y cómoda los conceptos básicos de dispositivos electrónicos y fotónicos expuestos en este texto.

DELFOs permite tres formas de trabajo: el modo estudio, el modo evaluación y el modo de resolución de los problemas guiados. El usuario debe elegir uno de los modos de trabajo (véase figura apéndice C-1), seleccionándolo en la página inicial.

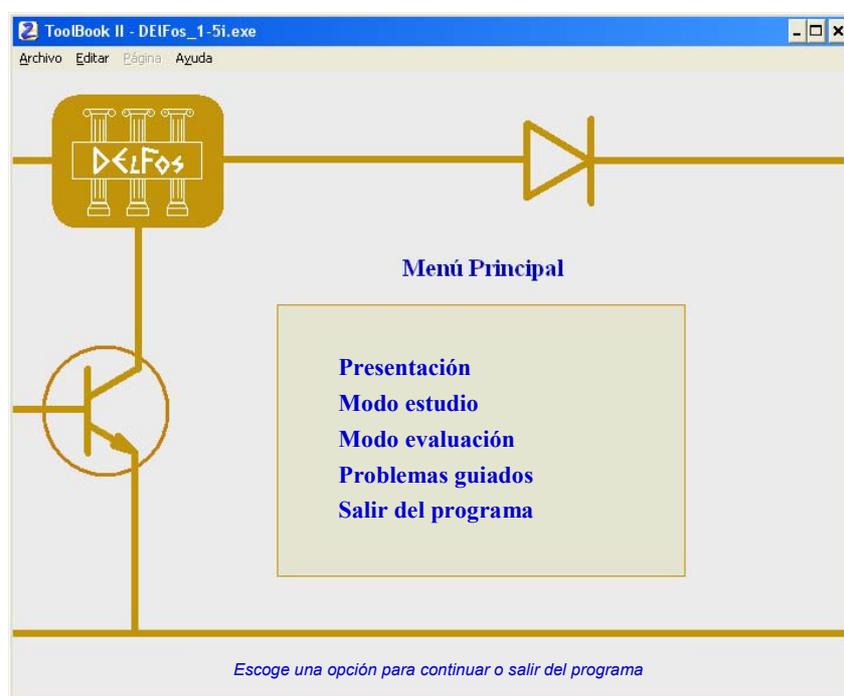


Fig. Apéndice C-1.- Modos de trabajo de DELFOS

* Los autores de DELFOS son los profesores del Departamento de Ingeniería Electrónica de la UPC: Lluís Prat Viñas, Josep Calderer Cardona, Vicente Jiménez Serres y Joan Pons Nin

Supongamos que se ha seleccionado el modo estudio. Al activarlo, aparece la pantalla de selección de cuestionario, que se muestra en la figura apéndice C-2. Debe seleccionarse un capítulo concreto y posteriormente el cuestionario que se desea trabajar. Una vez realizada la selección hay que pulsar el botón de navegación de la parte superior derecha que contiene una flecha apuntando a la derecha.

Aparece, entonces, una pantalla con la primera pregunta del cuestionario seleccionado. Con los dos botones de la izquierda, el usuario puede navegar por las diversas cuestiones. En cada cuestión (véase figura apéndice C-3), se presenta el enunciado y cuatro respuestas posibles, de las cuales sólo una es correcta. El usuario debe marcar la respuesta que considera correcta y después debe corregirla, pulsando el segundo botón de la parte superior derecha (en claro en la figura), lo que habilita la consulta de la solución correcta (botón inferior derecha).

En el modo de evaluación, el programa genera un examen de preguntas escogidas de los cuestionarios. El usuario puede elegir entre un examen de un capítulo o un examen global de todo el texto. Para cada cuestión del examen se ofrecen cuatro respuestas posibles (de las que sólo una es correcta), más la posibilidad de no contestar. El usuario puede navegar entre las cuestiones propuestas y acceder, en todo momento, una hoja resumen de las respuestas realizadas (véase figura apéndice C-4), pero no puede consultar las soluciones hasta después de haber corregido el examen. Cuando el usuario da por concluido el examen activa su corrección (pulsando el botón de navegación del extremo superior derecho de la hoja resumen). Entonces aparece en la hoja resumen los resultados del examen realizado: las respuestas correctas e incorrectas, el tiempo dedicado a realizar el examen, y la nota obtenida (que se calcula restando un tercio de punto por cada respuesta incorrecta). El usuario tiene la opción de obtener una copia impresa de esta página resumen (que se denomina *certificado de nota*) escribiendo su nombre y apellidos y pulsando la tecla OK en el recuadro correspondiente.

En el modo de resolución de los problemas guiados, el usuario debe empezar seleccionando uno de ellos. El programa presenta el enunciado del problema con diversos apartados, generalmente encadenados entre sí, junto a unos botones de navegación que permiten seleccionar la "pista" o la "solución" de cada apartado (véase figura apéndice C-5). La "pista" es una ayuda indicativa del procedimiento a seguir para resolver el apartado correspondiente.

La base de datos que contiene las cuestiones dispone de un procedimiento para cambiar de forma aleatoria el orden de las respuestas que aparecen en la pantalla, y también, cuando procede, los valores numéricos de los datos de la cuestión. Cada vez que se selecciona un cuestionario o un examen se realiza de forma automática esta inicialización aleatoria de datos numéricos y orden de las respuestas. En consecuencia, las cuestiones que se presentan al usuario, así como los exámenes que las contienen, son distintas cada vez que se genera un cuestionario o un examen.

El programa interactivo DELFOS puede obtenerse por Internet siguiendo las instrucciones de la página web: www.edicionsupc.es/poli131.