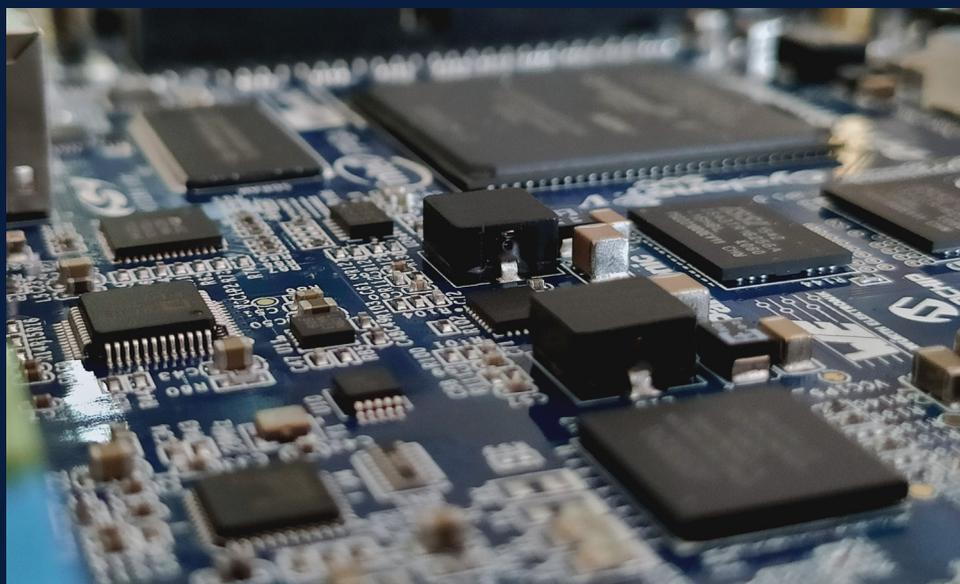


Fundamentos Físicos de la Informática



RAÚL RENGEL ESTÉVEZ

Ediciones Universidad
Salamanca

FUNDAMENTOS FÍSICOS
DE LA INFORMÁTICA

RAÚL RENGEL ESTÉVEZ

FUNDAMENTOS FÍSICOS
DE LA INFORMÁTICA



Ediciones Universidad
Salamanca

DOCUMENTOS DIDÁCTICOS, 167

© de esta edición:
Ediciones Universidad de Salamanca
y Raúl Rengel Estévez

Motivo de cubierta: © Raúl Rengel Estévez

1ª edición: octubre, 2020
ISBN: 978-84-1311-386-9 (PDF)
ISBN: 978-84-1311-387-6 (POD)
DOI: <https://doi.org/10.14201/0DD0167>

Ediciones Universidad de Salamanca
Plaza San Benito s/n
E-37002 Salamanca (España)
<http://www.eusal.es>
eus@usal.es

Realizado por:
Nueva Gráficas S.L.
Teléfono: 923 26 01 11
Salamanca (España)

Realizado en UE-Made in EU

*Todos los derechos reservados.
Ni la totalidad ni parte de este libro
puede reproducirse ni transmitirse sin permiso escrito de
Ediciones Universidad de Salamanca.*

Obra sometida a proceso de evaluación mediante sistema de doble ciego

Ediciones Universidad de Salamanca es miembro de la UNE
Unión de Editoriales Universitarias Españolas
www.une.es



Catalogación de editor en ONIX accesible en <https://www.dilve.es/> CEP

ÍNDICE GENERAL

I. Fundamentos de Electricidad y Magnetismo

1	Carga eléctrica	16
2	Campo eléctrico	19
3	Potencial	22
4	Corriente	24
5	Ley de Ohm	27
6	Potencia disipada	29
7	Almacenamiento de energía eléctrica: condensadores	30
8	Campo magnético	34
9	Inducción electromagnética	36
10	Bobinas	40
11	Campo electromagnético	41

II. Fundamentos de Teoría de Circuitos

1	Resistencias	47
2	Conexión en serie y paralelo	48
3	Definiciones útiles	53
4	Fuentes de alimentación	54
5	Leyes de Kirchhoff	55
6	Análisis de mallas	57
7	Teoremas de Thévenin y Norton	59
8	Circuitos de corriente alterna	64

III. Semiconductores

1	El átomo	73
2	El sólido cristalino	76
3	Bandas de energía	79

4	Conductores, aislantes y semiconductores	80
5	Semiconductores intrínsecos	86
6	Semiconductores extrínsecos	90
7	Ley de neutralidad de carga	92
8	Ley de acción de masas	93
9	Arrastre	94
10	Difusión	98
11	Generación y recombinación de portadores	100
12	Semiconductores más relevantes	103
IV. El diodo		
1	Principio físico	108
2	Polarización del diodo y curva característica	111
3	Aproximaciones	113
4	Rectificación	115
5	Diodos Zener	119
6	Otros tipos de diodos	121
V. El transistor bipolar		
1	Principio físico	127
2	Curvas características	131
3	Polarización del BJT	135
4	Circuito equivalente de pequeña señal	138
VI. El transistor MOSFET		
1	Principio físico	144
2	Curvas características	152
3	Polarización del MOSFET	155
4	Circuito equivalente de pequeña señal	157
VII. Dispositivos optoelectrónicos		
1	Diodos LED	163
2	Láseres de semiconductor	169
3	Fotodetectores	172
4	Células solares	175
5	Optoacopladores	177

VIII. Amplificación	
1 Concepto de amplificación	181
2 Un ejemplo: el amplificador de voltaje en fuente común	182
3 El amplificador operacional	187
IX. Bases de electrónica digital	
1 Álgebra de Boole	195
2 Representación física de señales digitales	199
3 Conversión de señales analógicas y digitales	201
X. Conmutación	
1 Conmutación de diodos	213
2 Conmutación de transistores	214
3 Un ejemplo: circuito inversor en fuente común	220
4 El inversor CMOS	227
5 Aplicaciones del MOSFET en conmutación	232
XI. Familias lógicas y diseño CMOS	
1 Principales familias lógicas	239
2 Funcionamiento de puertas lógicas CMOS básicas	244
3 Niveles lógicos y márgenes de ruido	251
4 Características generales y consumo de potencia	254
5 Diseño lógico CMOS	258
XII. Tecnología de fabricación de circuitos integrados	
1 Obleas de semiconductor	270
2 El entorno: la sala blanca	273
3 Litografía	275
4 Procesos tecnológicos	278
5 El escalado del transistor	281
Bibliografía	287
Índice alfabético	289

PRÓLOGO

Desde la invención del transistor en 1947 la electrónica ha conocido un desarrollo vertiginoso que ha hecho posible la aparición de la denominada *sociedad de la información*. Teléfonos inteligentes, ordenadores personales, cámaras digitales, videoconsolas o pulseras de actividad forman parte de la vida cotidiana de millones de personas en todo el mundo. Vivimos en una sociedad conectada a través de internet, con una plétora de aparatos electrónicos a nuestro alrededor, de cuya presencia no somos siempre conscientes. En las próximas décadas el internet de las cosas llevará al extremo esta situación. La biología, la agricultura, el diseño o incluso el arte emplean cada vez en mayor medida electrónica, sensores y ordenadores para extender las fronteras de estas disciplinas. Otras especialidades, como la medicina, se han visto también beneficiadas por este desarrollo tecnológico sin precedentes. Cada vez resulta más difícil imaginar un mundo sin electrónica.

La mayor parte de esos aparatos electrónicos necesitan ser programados para llevar a cabo sus tareas. La revolución electrónica ha traído consigo una evolución sin precedentes de la informática, que se ha beneficiado como ninguna otra disciplina de los avances conseguidos. La inmensa mayoría de aplicaciones software serían simplemente inimaginables sin el hardware que las hace posibles. En particular, la informática tal y como la conocemos hoy en día es una realidad gracias a la física de semiconductores, el transistor MOSFET de silicio y los circuitos integrados.

Este libro está orientado a alumnos de primer curso del grado universitario en ingeniería informática, aunque se ha intentado que los fundamentos aquí explicados puedan ser comprendidos, sin excesiva dificultad, por cualquier lector interesado con formación a nivel de secundaria en física y matemáticas. En el texto se pretende sentar las bases para la comprensión del funcionamiento físico de los circuitos digitales que constituyen los microprocesadores actuales. Comenzaremos con un repaso a los fundamentos básicos del electromagnetismo y de la teoría de circuitos, que nos permitan una primera aproximación a conceptos tales

como carga eléctrica, campo eléctrico, potencial, corriente, resistencias, fuentes de alimentación, etc. que serán esenciales para poder entender con posterioridad la operación de los circuitos digitales.

Seguidamente, se estudiarán los semiconductores, sus propiedades físicas, los fenómenos de transporte de corriente eléctrica en ellos y cómo se puede regular el número de cargas eléctricas en su interior, así como los procesos que afectan a estas cantidades. Todo ello es esencial para entender cómo funcionan los dispositivos electrónicos, como diodos y transistores, que serán objeto de estudio en los capítulos siguientes, junto con los principales dispositivos optoelectrónicos, de interés también en este contexto.

Las aplicaciones de los transistores ocuparán los posteriores capítulos. Si bien la amplificación de señales no es un proceso esencial para entender el funcionamiento de un circuito digital, se ha estimado conveniente incluir un capítulo dedicado a este tipo de aplicación por su importancia en la adquisición de señales y toma de datos de sensores, de gran importancia en la informática y la automática.

Posteriormente nos centraremos plenamente en los circuitos digitales, y para ello se comenzará con un resumen de los fundamentos del álgebra de Boole, la representación física de señales y la conversión de señales analógicas y digitales. Se proseguirá con el estudio de la conmutación de los dispositivos, principalmente del transistor MOSFET. Con ello estaremos en condiciones de abordar la realización física de puertas lógicas y el diseño de funciones lógicas en tecnología CMOS, lo que permitirá al lector abrir la puerta a la comprensión del funcionamiento físico interno de un microprocesador.

Se ha incluido un capítulo final dedicado a la fabricación de circuitos integrados, pues se ha considerado esencial para entender el contexto general de la tecnología de semiconductores actual, así como de la evolución de la arquitectura física del transistor MOSFET, el bloque constructor esencial de nuestro mundo digital actual.

En definitiva, el objetivo final de este libro es que el lector consiga dar respuesta a un pregunta fundamental: **¿cómo funciona realmente un ordenador?**

I.

FUNDAMENTOS DE
ELECTRICIDAD Y
MAGNETISMO

La humanidad ha tenido conocimiento de la existencia de fenómenos eléctricos y magnéticos desde tiempos inmemoriales. Los antiguos griegos ya constataron que el ámbar (una resina fosilizada que descubrieron hace casi 3000 años al abrirse las primeras rutas comerciales con el mar Báltico), al ser frotado con un trozo de lana o piel de animal, era capaz de atraer objetos ligeros, como cabellos, plumas o paja¹, en lo que constituía una de las primeras manifestaciones del fenómeno que hoy conocemos como electricidad estática. Por otra parte, en la región griega de Magnesia, en Tesalia, cerca del año 600 a.C. se observaron también los primeros fenómenos de naturaleza magnética. Algunas piedras minerales en aquella región eran capaces de atraer pequeños trozos de material que contuviesen vetas de hierro. Unos pocos siglos después, en la antigua China, se produjo la invención de la brújula, que tardaría más de mil años en llegar a Occidente, a principios del siglo XIII. Durante milenios los fenómenos eléctricos y magnéticos no pasaron de ser simples curiosidades (con una aplicación práctica útil para la navegación), cuyo origen y explicación era un misterio.

No es hasta el año 1600, con la llegada a Europa del Renacimiento, y en los albores del método científico, cuando aparece por fin un interés en el estudio riguroso de estos fenómenos. Es a William Gilbert² al que debemos el término *eléctrico* para referirse a los fenómenos de naturaleza electrostática³. Sus estudios sobre el magnetismo y la electricidad, si bien plagados de errores, fueron un paso fundamental para iniciar la investigación de estas materias. Junto con la inven-

¹Tales de Mileto (624 a.C - 546 a.C.) fue el primero en dejar constancia escrita de este fenómeno.

²William Gilbert (1544-1603), físico y médico inglés autor de *De magnete*, el primer libro relevante sobre el estudio del electromagnetismo.

³Término acuñado por Gilbert a partir de la palabra del griego antiguo *electrón* (ἤλεκτρον), ámbar.

ción en 1663 por parte de Otto von Guericke⁴ de un mecanismo capaz de generar carga eléctrica y almacenarla en una esfera metálica, abrieron la puerta a la experimentación sistemática de los fenómenos electrostáticos a lo largo del siglo XVIII.

En concreto, Benjamin Franklin⁵ estudió en profundidad la triboelectricidad⁶, y así, descubrió una serie de propiedades fundamentales de la carga eléctrica.



Figura I.1: Grabado en el que se representa el famoso experimento de la cometa. Benjamin Franklin fue uno de los pioneros en el estudio de los fenómenos eléctricos.

1 CARGA ELÉCTRICA

La **carga eléctrica** es una *propiedad física fundamental de la materia* (propia de ciertas partículas elementales) responsable en último término de las interacciones electromagnéticas. Así, decimos que un cuerpo está cargado eléctricamente cuando se observan en él fenómenos de naturaleza electrostática, como los descritos

⁴Otto von Guericke (1602-1686), científico e inventor alemán.

⁵Benjamin Franklin (1706-1790), científico, inventor, escritor y político norteamericano.

⁶La triboelectricidad es el fenómeno de aparición de una carga eléctrica al frotar un material.

por Benjamin Franklin. Además de su famoso experimento de la cometa⁷, Franklin realizó numerosos experimentos triboeléctricos sobre diferentes materiales. Encontró que si frotaba una barra de caucho con un trozo de piel, al acercarla a otra barra de caucho sometida al mismo tratamiento ambas se repelían. Lo mismo ocurría cuando se frotaban dos barras de vidrio con un trozo de seda; sin embargo, al acercar una barra de caucho electrificada a una de vidrio, tendían a atraerse. Realizando experimentos con más materiales encontró que los objetos susceptibles de ser electrificados se dividían en dos grupos: los que se comportaban como el vidrio y los que se comportaban como el caucho.

A esas cargas de los cuerpos electrificados las denominó **positivas y negativas**, asignándole al vidrio la carga positiva y al caucho la negativa⁸.

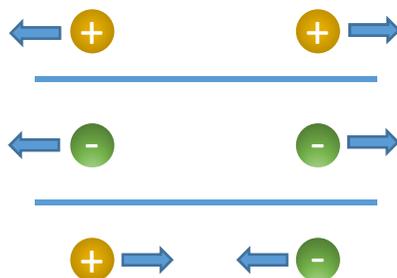


Figura I.2: Cargas de igual signo se repelen, mientras que cargas de signo opuesto se atraen.

De este modo, descubrió que la carga eléctrica tiene una naturaleza **dual**. Asumió que un cuerpo sin electrificar tiene una cierta cantidad de carga, y al ser frotado, ésta se transfiere desde o hacia el objeto, resultando en un exceso o defecto de carga.

Posteriores experimentos nos iban a dar más información acerca de la naturaleza de la carga eléctrica. Otra de sus propiedades relevantes es la **cuantización**: *la cantidad de carga que contiene cualquier cuerpo, sea positiva o negativa, es siempre múltiplo de la unidad elemental de carga, que corresponde a la carga que posee un electrón*⁹.

⁷Existen dudas sobre si el experimento se llevó a cabo en la realidad por parte de Franklin, o si se trató de un experimento conceptual. En todo caso, llevó a la invención del pararrayos y a la caracterización de los rayos como fenómenos de naturaleza esencialmente eléctrica.

⁸Franklin eligió los signos de la carga de manera arbitraria: sin embargo, su elección traería como consecuencia que el convenio que se elige para el signo de la corriente eléctrica corresponda a un movimiento de los electrones en sentido opuesto.

⁹Debe tenerse en cuenta que la mayor parte de los descubrimientos acerca de la naturaleza de

Otra propiedad interesante de la carga eléctrica es su **conservación**. La carga se puede transferir entre cuerpos, pero la suma de todas sus cargas antes y después de la transferencia debe permanecer constante. Es decir, *la carga no se crea o se destruye, pero sí puede transferirse*.

En su estado natural, los átomos presentan la propiedad de la **neutralidad**: la cantidad de carga positiva y negativa está equilibrada, de forma que la carga neta es nula. Los cuerpos son eléctricamente neutros hasta que por algún procedimiento se les transfiere o retira carga de algún tipo, momento en el que quedan electrificados.

La unidad de carga eléctrica es el **culombio**¹⁰, que representamos por la letra C. La unidad elemental de carga es la carga del electrón, que corresponde a $e = -1.6 \cdot 10^{-19}$ C.

Como acabamos de señalar, los materiales cargados eléctricamente poseen una cierta carga (múltiplo de la elemental) y se atraen o repelen en función de los signos de dicha carga. Consideremos esos materiales electrificados como cargas puntuales (es decir, que están concentradas en un punto del espacio). ¿Cuál es la fuerza con la que se atraen o repelen? La respuesta a esta pregunta nos la da la **ley de Coulomb**, que nos indica que la fuerza que ejerce una partícula con carga q_1 sobre otra partícula con carga q_2 se puede calcular como:

$$\vec{F}_{12} = K \frac{q_1 q_2}{r_{12}^2} \hat{r}_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2} \hat{r}_{12} \quad (\text{I.1})$$

donde \vec{r}_{12} es el vector que va desde q_1 a q_2 (véase la figura I.3), r_{12} su módulo y \hat{r}_{12} el vector unitario en la dirección de \vec{r}_{12} . Para el caso del vacío, la constante de proporcionalidad toma valor $K = 8.99 \cdot 10^9 \text{ Nm}^2\text{C}^{-2}$ (o bien se puede utilizar la permitividad $\epsilon_0 = 8.854 \cdot 10^{-12} \text{ N}^{-1}\text{m}^{-2}\text{C}^2$).

Esta fuerza es directamente proporcional al producto de las cargas e inversamente proporcional al cuadrado de la distancia. En este cálculo debe tenerse en cuenta de forma explícita el signo de las cargas: así, dos cargas de igual signo producirán una fuerza positiva a lo largo de la dirección que une las dos cargas (lo que indica que se repelen) mientras que si tienen signos distintos la fuerza será negativa (indicando que se atraen).

Es interesante notar la similitud formal entre esta ley y la ley de la gravedad

la carga eléctrica tuvieron lugar con anterioridad a que el inglés Joseph J. Thomson (1856-1940) demostrase la existencia de los electrones en 1897.

¹⁰En honor a Charles-Augustin de Coulomb (1736-1806), que realizó importantes descubrimientos en este campo.

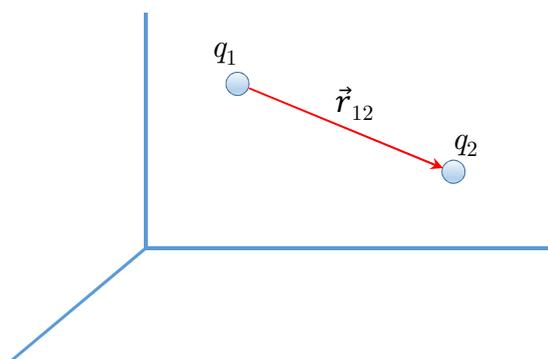


Figura I.3: La fuerza ejercida por una carga puntual sobre otra se dirige a lo largo de la línea que las une.

de Newton. Sin embargo, hay algunas diferencias significativas: en primer lugar, la gravedad siempre es una fuerza de naturaleza atractiva, mientras que la fuerza electrostática puede ser atractiva o repulsiva. Además, la fuerza electrostática es mucho más intensa que la gravitatoria: en el contexto del átomo de hidrógeno es aproximadamente 40 órdenes de magnitud mayor, por lo que la gravedad es despreciable frente a las fuerzas de naturaleza eléctrica. En objetos de escala macroscópica, sin embargo, la débil cantidad de carga neta o la neutralidad de los cuerpos hacen que esta fuerza sea por lo general despreciable comparada con la gravitatoria, y sólo se pueda apreciar en objetos muy ligeros, como hicieron los griegos cuando observaron por vez primera estos fenómenos.

Si en nuestro sistema existiesen varias cargas puntuales, podríamos calcular de forma sencilla la fuerza total que actúa sobre una de ellas sin más que sumar (vectorialmente) la fuerza que ejerce cada una de las otras cargas presentes.

2 CAMPO ELÉCTRICO

A medida que se avanzaba en el estudio y comprensión de los fenómenos electrostáticos, surgió una pregunta de difícil respuesta: ¿por medio de qué procesos se produce esta interacción a distancia? La fuerza electrostática no parecía transmitirse a través de ningún medio material, ya que tiene lugar incluso en el vacío. Tampoco estaba claro qué ocurría cuando una partícula cambiaba de posición, ni a través de qué mecanismos tenían lugar los cambios en la fuerza que experimentaba otra partícula separada de ella, y sin ningún contacto material entre ambas.

Fue Michael Faraday¹¹ quien introdujo el concepto de **campo eléctrico** para tratar de aliviar estos problemas conceptuales. Si bien el concepto de campo no responde de manera directa a la pregunta anteriormente planteada, sí permite tratar de forma relativamente sencilla y elegante el problema.

En el ámbito de la física, un campo (gravitatorio, eléctrico, etc.) es la representación de la variación espacial de una magnitud física. En el caso del campo eléctrico, lo que representamos es la *fuerza que ejercería la carga que crea el campo sobre una carga de valor 1 C en los diferentes puntos del espacio*. En definitiva, el concepto de campo eléctrico nos dice que una partícula cargada crea a su alrededor una región en la que si colocásemos otra partícula, ésta experimentaría una fuerza proporcional al valor del campo eléctrico: dicho de otra manera, el campo nos indica qué fuerza por unidad de carga crearía la carga original en cada punto del espacio.

Así, tendríamos que el campo eléctrico que crea q_1 se puede calcular como:

$$\vec{E} = \frac{\vec{F}_{12}}{q_2} = \frac{1}{4\pi\epsilon_0} \frac{q_1}{r_{12}^2} \hat{r}_{12} \quad (\text{I.2})$$

De este modo, en cada punto del espacio podemos calcular el valor del campo eléctrico asociado. Para representar el campo eléctrico, con frecuencia se utiliza el concepto de **líneas de campo**. Para una cierta carga dada, el campo eléctrico que crea se representa dibujando flechas que apuntan hacia fuera, si la carga es positiva, o hacia la carga, si la carga es negativa. Cuando las líneas de campo están más cercanas, significa que el campo eléctrico es más intenso en esa zona del espacio (véase por ejemplo la figura I.4).

Es interesante en este punto introducir el concepto de **dipolo eléctrico**. Un dipolo eléctrico es una *carga positiva separada una cierta distancia fija de una carga negativa de igual magnitud*. Cuando se da esta situación, el campo eléctrico asociado a ese dipolo se concentra principalmente en la línea que une ambas cargas, apuntando desde la carga positiva hacia la negativa (figura I.5).

Una aplicación práctica de las propiedades de los dipolos, que utilizamos de manera cotidiana, son los hornos microondas. Las moléculas de agua presentes en los alimentos forman de manera natural un dipolo. En el interior de un horno microondas se aplica un campo eléctrico oscilante. Las moléculas de agua tienden a alinearse con este campo, de modo que al hacerlo rotan y se calientan cuando colisionan con otros átomos y moléculas.

¹¹Michael Faraday (1791-1867) fue un físico y químico inglés que realizó algunas de las aportaciones más importantes en estos ámbitos científicos, en particular respecto al estudio de la inducción electromagnética y de los fenómenos electrolíticos.

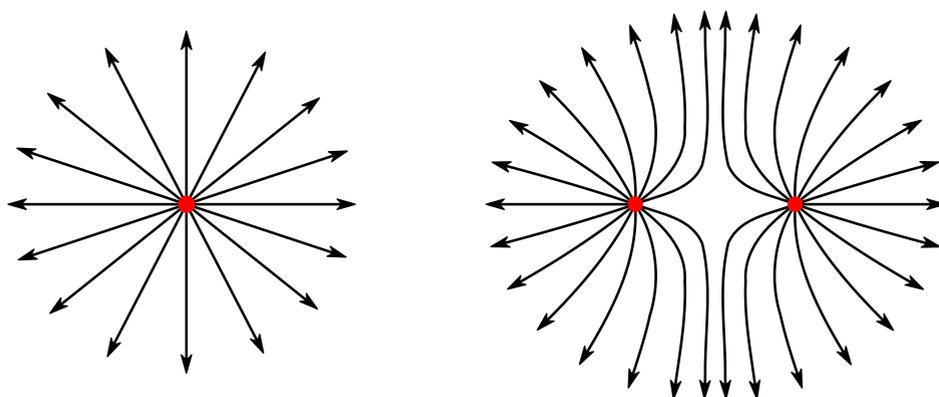


Figura I.4: Líneas de campo creadas por una carga positiva (izquierda) y por dos cargas positivas de igual valor (derecha).

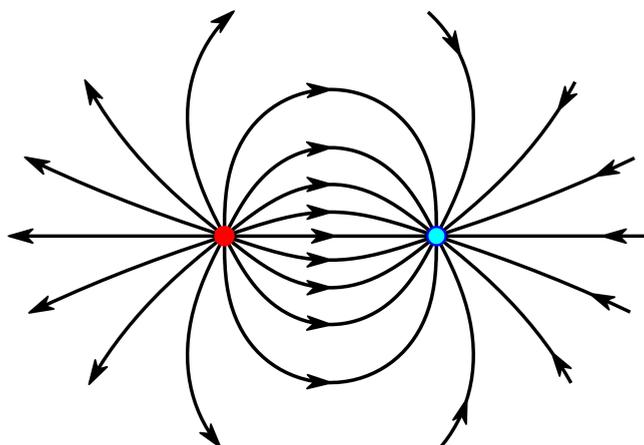


Figura I.5: Líneas de campo creadas por cargas de signo opuesto e igual módulo (positiva a la izquierda y negativa a la derecha), es decir, un dipolo.

En definitiva, si conocemos el valor del campo eléctrico en un punto dado del espacio, podemos conocer la fuerza que experimentaría una carga si se colocase en ese lugar. Esta fuerza es proporcional al campo eléctrico, y tiene su misma dirección, mientras que su sentido dependerá del signo de la carga: si la carga es positiva, la fuerza tiene el mismo sentido que el campo eléctrico, y si es negativa, tendrán sentidos opuestos:

$$\vec{F} = q\vec{E} \quad (\text{I.3})$$

donde tendremos que tener en cuenta que q puede ser negativa o positiva. Conforme a la segunda ley de Newton, esta fuerza genera una aceleración sobre la carga, de forma que:

$$\vec{F} = m\vec{a} = q\vec{E} \quad (\text{I.4})$$

Y por tanto:

$$\vec{a} = \frac{q}{m}\vec{E} \quad (\text{I.5})$$

En este fenómeno se basa el principio de funcionamiento de las placas deflectoras, utilizadas por ejemplo en los antiguos televisores de rayos catódicos, en los osciloscopios analógicos o en algunos tipos de impresoras, como las impresoras de inyección industriales basadas en un chorro continuo de tinta¹². En estas impresoras se bombea tinta líquida a través de un nebulizador microscópico, de forma que se crea un chorro de pequeñas partículas de tinta cargadas eléctricamente. Al pasar entre dos placas metálicas entre las que se ha creado un campo eléctrico, las gotas son deflectadas y se orientan de manera adecuada hacia el papel, formando progresivamente la imagen impresa.

3 POTENCIAL

El campo eléctrico es un campo de tipo conservativo. Un campo de fuerzas es conservativo si el trabajo total realizado por el campo sobre una partícula que realiza un desplazamiento en una trayectoria cerrada (como por ejemplo la órbita de un planeta, en el caso del campo gravitatorio) es nulo.

En este contexto, podemos intentar calcular cuál es el trabajo que realiza el campo eléctrico para desplazar una partícula desde un punto A hasta un punto B . Imaginemos que tenemos una partícula de carga positiva q en un punto A , bajo la acción de un campo eléctrico uniforme \vec{E} (figura I.6). Bajo la acción de este campo eléctrico, la partícula se desplaza una distancia l hasta el punto B .

El concepto físico de trabajo es esencialmente el producto escalar de la fuerza por la distancia recorrida. Así, si se recorre una distancia infinitesimal $d\vec{l}$ el trabajo realizado será:

$$dW = \vec{F}d\vec{l} \quad (\text{I.6})$$

¹²Las impresoras de tinta de uso doméstico no se basan en el chorro continuo de tinta, sino en la técnica de gota a demanda.

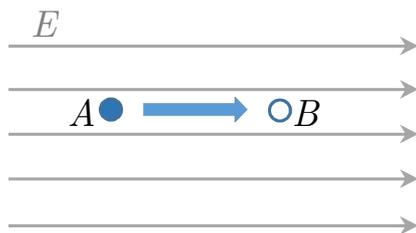


Figura I.6: Movimiento de una carga positiva bajo la acción del campo eléctrico, desde el punto A hasta el B .

El trabajo total realizado entre A y B será por tanto la integral (la suma de todos los trabajos infinitesimales) entre esos puntos:

$$W_A^B = \int_A^B \vec{F} d\vec{l} = \int_A^B q \vec{E} d\vec{l} = q \int_A^B \vec{E} d\vec{l} \quad (I.7)$$

El campo ha sido capaz de realizar este trabajo a costa de perder energía potencial, de manera totalmente análoga a como cuando dejamos caer un objeto y el campo gravitatorio realiza un trabajo (que se traduce en un incremento de energía cinética) a costa de perder energía potencial gravitatoria. Por tanto, podemos decir que el trabajo realizado es igual a la energía potencial perdida, es decir:

$$W_A^B = -\Delta U_A^B = -(U_B - U_A) = (U_A - U_B) \quad (I.8)$$

Podemos definir entonces el **potencial eléctrico** V como la energía potencial por unidad de carga eléctrica, es decir:

$$V = \frac{U}{q} \quad (I.9)$$

Así, tendremos entonces que:

$$W_A^B = (U_A - U_B) = q \int_A^B \vec{E} d\vec{l} \quad (I.10)$$

y por tanto:

$$\frac{(U_A - U_B)}{q} = \int_A^B \vec{E} d\vec{l} \quad (I.11)$$

Finalmente:

$$V_A - V_B = \int_A^B \vec{E} d\vec{l} \quad (\text{I.12})$$

Si el campo es homogéneo, tendremos que para este caso particular:

$$V_A - V_B = E \cdot l \quad (\text{I.13})$$

En definitiva, lo relevante no es tanto el potencial eléctrico en sí como la **diferencia de potencial**¹³ entre dos puntos del espacio (o dos puntos de un circuito). *La diferencia de potencial entre los puntos A y B nos indica cuánto trabajo debería realizar el campo eléctrico para desplazar una carga de valor 1 C desde el punto A hasta el B.*

El potencial eléctrico es una magnitud escalar (es decir, no tiene orientación, sino simplemente un valor en cada punto del espacio), y es continuo en todos los puntos del espacio. La unidad de potencial en el sistema internacional es el **voltio**, que corresponde a 1 J/C (julio dividido por culombio).¹⁴

4 CORRIENTE

Los fenómenos electrostáticos tienen un interés indudable, pero en la mayor parte de aplicaciones prácticas lo que nos interesa es el movimiento colectivo de partículas cargadas y cómo podemos controlarlo. La forma más sencilla de mover partículas cargadas en un objeto es inducir en él un campo eléctrico¹⁵ (o lo que es lo mismo, una diferencia de potencial entre sus extremos). El movimiento de carga que tiene lugar se traduce en una **corriente eléctrica**, que es la *cantidad de carga eléctrica que atraviesa una sección transversal por unidad de tiempo*¹⁶.

Por convenio, se entiende que una corriente positiva implica el movimiento de cargas positivas en la dirección de la corriente (o de cargas negativas en sentido

¹³A lo largo del texto usaremos diferencia de potencial y voltaje como sinónimos. Si bien en propiedad no son conceptos exactamente equivalentes, en el contexto de este libro pueden considerarse como tales.

¹⁴El voltio recibe su nombre en honor a Alessandro Volta (1745-1827), inventor de la pila eléctrica.

¹⁵Como veremos más adelante, existe otro tipo de corriente, denominada de difusión, que puede tener lugar sin aplicar un campo eléctrico externo.

¹⁶En capítulos sucesivos asumiremos que las corrientes son *filiformes*, es decir, que viajan por líneas geométricas y no por conductores con una cierta sección. Esta es una aproximación habitual en el estudio de teoría de circuitos.

opuesto al de la corriente)¹⁷. La unidad de medida de la corriente eléctrica es el **amperio**¹⁸, que equivale a C/s (culombio dividido por segundo).

Cuando en un material sólido disponemos de una cantidad de electrones libres¹⁹, en ausencia de campos eléctricos el movimiento de estas partículas es de naturaleza aleatoria. Cada electrón se mueve con una velocidad y una orientación distintas, sufriendo colisiones cada cierto tiempo con los átomos que forman el sólido y las imperfecciones e impurezas que existan en el mismo. Globalmente, la velocidad promedio de esos electrones libres es nula, dado que se mueven indistintamente en todas las direcciones.

Sin embargo, cuando aplicamos un campo eléctrico, dicho campo genera una aceleración sobre cada electrón de valor $\frac{-e\vec{E}}{m}$, por lo que tienden a ganar velocidad en la dirección del campo (pero con sentido opuesto al ser carga negativa). Sin embargo, no ganan velocidad de manera indefinida: cada cierto tiempo los electrones sufrirán colisiones con los átomos del medio en el que se están desplazando, lo que supone que serán sucesivamente frenados y acelerados.

Así, en un material en el que exista un determinado campo eléctrico acelerador, el conjunto de los electrones alcanzará una velocidad promedio constante que denominaremos **velocidad de deriva**. En el caso de los electrones, esta velocidad de deriva será (a campos no demasiado elevados) proporcional al campo aplicado y de signo contrario.

Vamos a determinar de manera general el valor de la corriente en función de la velocidad de deriva. Por simplicidad, supongamos que tenemos una corriente formada por partículas positivas, todas con idéntica carga q , que se desplazan con una velocidad de deriva promedio \vec{v}_d en un cable de sección S . En un tiempo Δt las cargas recorrerán una distancia $v_d\Delta t$, y por tanto, atravesarán la sección S las que ocupen el volumen $Sv_d\Delta t$. Por otra parte, el número de partículas cargadas que tendremos por unidad de volumen, es decir su **concentración** o **densidad**²⁰,

¹⁷En nuestro caso las cargas que físicamente van a estar en movimiento son, por lo general, electrones. En ese caso, la corriente tiene sentido opuesto al flujo de electrones, por tener estos carga negativa. Esta situación aparentemente contradictoria se debe a la elección arbitraria por parte de Franklin de los signos positivo y negativo de las cargas eléctricas, como ya hemos señalado.

¹⁸En honor a André-Marie Ampère (1775-1836), físico francés. La definición formal de amperio es la corriente que debe circular por dos hilos paralelos separados 1 metro, para que la fuerza que exista entre ellos sea igual a $2 \cdot 10^{-7}$ newtons.

¹⁹Un electrón libre es aquel que ha *escapado* del átomo en el que se encontraba originalmente, y puede por tanto desplazarse en el medio en que se encuentre.

²⁰La densidad o concentración de partículas se mide en número de partículas por unidad de

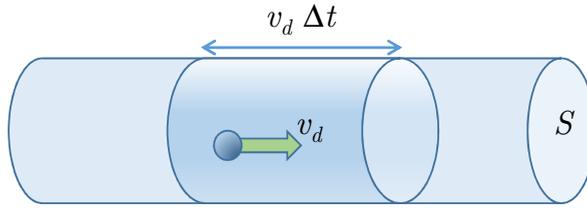


Figura I.7: Flujo de corriente en un cable.

que representaremos por la letra n . De este modo, la carga total que ocupa el volumen anteriormente mencionado será:

$$\Delta Q = qnSv_d\Delta t \quad (\text{I.14})$$

Y por tanto la corriente:

$$I = \frac{\Delta Q}{\Delta t} = qnSv_d \quad (\text{I.15})$$

La **densidad de corriente**, es decir, la corriente por unidad de área transversal, será entonces:

$$J = \frac{I}{S} = qnv_d \quad (\text{I.16})$$

Conviene detenerse un momento a analizar este resultado. Como vemos, la densidad de corriente va a depender, esencialmente, de dos factores:

- Por una parte, de la densidad de partículas cargadas que tengamos (por ejemplo, de cuántos electrones libres dispongamos en un material dado).
- Por otra parte, de la velocidad promedio, que dependerá a su vez fundamentalmente del campo aplicado y de la facilidad que tengan las partículas de carga para moverse en el interior del material.

Estos dos factores son esenciales para determinar la capacidad de conducir corriente de un material: si se dispone de muchos electrones pero no pueden desplazarse, no habrá corriente. Tampoco la habrá si existen muchas facilidades para

volumen. En electrónica se suele emplear el cm^3 para medir los volúmenes. Dado que el número de partículas no tiene unidades físicas, la unidad de concentración será cm^{-3} .

desplazarse dentro del material pero no hay un número relevante de electrones libres. Sobre estos aspectos volveremos a tratar en el capítulo III.

A modo de ejemplo, podemos decir que en el caso del cobre (uno de los materiales más empleados como conductor de la electricidad) la velocidad de deriva es relativamente baja (aproximadamente 80 cm/h). Sin embargo, la alta densidad de electrones libres en este metal (cerca de 10^{22} cm^{-3}) hace que se puedan alcanzar corrientes elevadas.

5 LEY DE OHM

Dependiendo de su facilidad para conducir corriente eléctrica podemos clasificar a los materiales en dos grandes categorías, **conductores** y **aislantes**²¹. Los materiales conductores son aquellos que, con mayor o menor dificultad, permiten el paso de una corriente eléctrica de valor apreciable; los aislantes serían aquellos que impiden el paso de corriente (o la corriente que puede circular por ellos es despreciable). Vamos a fijarnos primero en los materiales que permiten la conducción de corriente eléctrica.

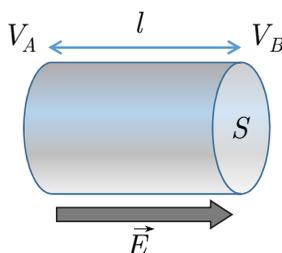


Figura I.8: Relación entre voltaje y campo eléctrico en un medio cilíndrico homogéneo.

Supongamos que tenemos un material conductor homogéneo con forma cilíndrica, con longitud l y sección S (figura I.8). Cuando apliquemos una diferencia de potencial entre sus extremos²², se creará un campo eléctrico constante, de

²¹Posteriormente veremos como se puede incluir una tercera categoría a caballo entre estas dos, los semiconductores.

²²Esto se puede realizar de manera sencilla, conectando, por ejemplo, los bornes de una batería entre extremos del material.

valor:

$$E = \frac{V_A - V_B}{l} \quad (\text{I.17})$$

Para hacer más sencillas las expresiones, vamos a simplificar la notación de la siguiente manera:

$$V_A - V_B = V_{AB} \quad (\text{I.18})$$

Usaremos de forma habitual esta notación a lo largo del libro. El campo eléctrico aplicado en el material arrastrará a la carga eléctrica presente en su interior y generará una corriente I . Como sabemos, dicha corriente va a depender de la densidad de partículas cargadas y de su velocidad de deriva, que a su vez depende de E , y en última instancia, de V_{AB} . Definimos por tanto la **resistencia** del conductor como *el cociente entre la diferencia de potencial en la dirección de la corriente y la intensidad de corriente que circula*, es decir:

$$R = \frac{V_{AB}}{I} \quad (\text{I.19})$$

Esta definición constituye lo que conocemos como **ley de Ohm**²³. La resistencia se mide en **ohmios** y se representa por la letra griega Ω ; un ohmio es equivalente a 1 V/A (voltio dividido por amperio).

En los metales, por lo general, el valor de resistencia R es constante a una temperatura dada e independiente de V_{AB} e I en un amplio rango. Además, el valor de resistencia tiende a aumentar con la temperatura.

Pero este valor de resistencia que hemos determinado está condicionado por el tamaño del conductor. La corriente depende de la sección transversal de manera directamente proporcional, y la diferencia de potencial producirá un campo eléctrico mayor o menor dependiendo de la longitud l . En consecuencia, para poder comparar la facilidad de conducción de materiales con tamaños o formas diferentes conviene definir una magnitud relacionada con la resistencia eléctrica que no dependa de esos factores geométricos.

Esa magnitud es la **resistividad**²⁴, que representamos por la letra griega ρ , y a partir de la cual podemos obtener la resistencia como:

$$R = \rho \frac{l}{S} \quad (\text{I.20})$$

²³En honor a Georg Simon Ohm (1789-1854), que fue quien la enunció.

²⁴En sistema internacional la resistividad se mide en $\Omega \cdot \text{m}$.

Alternativamente, se puede emplear la **conductividad eléctrica**²⁵ σ , que es el inverso de la resistividad:

$$\sigma = \frac{1}{\rho} \quad (\text{I.21})$$

De ese modo, la ley de Ohm se puede expresar de forma alternativa, teniendo en cuenta la densidad de corriente J y el campo eléctrico E :

$$J = \sigma E \quad (\text{I.22})$$

6 POTENCIA DISIPADA

Cuando una corriente eléctrica circula por un objeto, se produce un aumento de su temperatura por el denominado **efecto Joule**²⁶. Al chocar con los átomos del material, los electrones que forman la corriente eléctrica ceden energía cinética en forma de calor, produciéndose un aumento de la temperatura. Se produce por tanto una pérdida de energía, que implica una pérdida de potencia. La **potencia** se define como la *cantidad de trabajo*²⁷ *realizado por unidad de tiempo*, de modo que:

$$P = \frac{W_A^B}{\Delta t} = \frac{W_A^B}{\Delta Q} \frac{\Delta Q}{\Delta t} = V_{AB} \cdot I \quad (\text{I.23})$$

Siendo A y B los extremos del elemento por el que circula la corriente. Es decir, para cualquier componente que tengamos en un circuito podemos determinar la **potencia disipada** en él como el producto de la diferencia de potencial entre sus extremos multiplicado por la corriente que circula por él.

Para el caso de una resistencia, aplicando la ley de Ohm:

$$P = V_{AB} \cdot I = I^2 R = \frac{V_{AB}^2}{R} \quad (\text{I.24})$$

La potencia se mide en **vatios**²⁸ (w), siendo 1 vatio igual a 1 J/s (julio dividido por segundo), o 1 V·A (voltio multiplicado por amperio).

²⁵En sistema internacional la conductividad se mide en $(\Omega \cdot \text{m})^{-1}$

²⁶Estudiado por vez primera por James Prescott Joule (1818-1889).

²⁷El trabajo puede entenderse como una transferencia de energía.

²⁸De inglés *watt*, en honor a James Watt (1736-1819), inventor de la máquina de vapor.

7 ALMACENAMIENTO DE ENERGÍA ELÉCTRICA: CONDENSADORES

Un condensador es un dispositivo formado por dos conductores, denominados *placas* o *armadura*, separados por un material *dieléctrico*, es decir, un aislante. Debido a la presencia de este aislante, no puede haber un flujo continuo de corriente a través de un condensador²⁹. Un ejemplo son los condensadores **plano-paralelos**, cuya representación simplificada se muestra en la figura I.9.

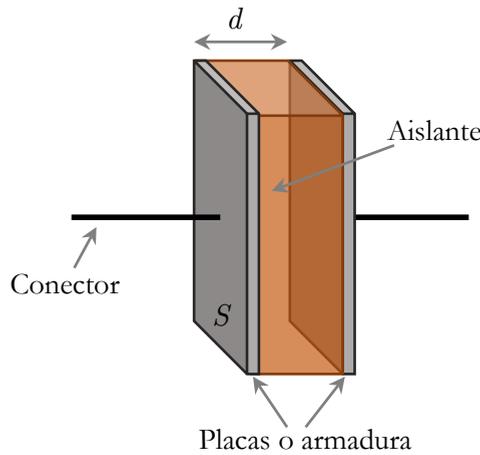


Figura I.9: Esquema básico de un condensador plano-paralelo.

Los condensadores nos permiten almacenar carga eléctrica: un condensador cargado tendrá, en cada una de sus placas conductoras, una cantidad de carga Q , con signo diferente en cada placa ($+Q$ en una y $-Q$ en la otra). El almacenamiento de carga implica un almacenamiento de energía eléctrica, de forma que un condensador cargado podrá proporcionar energía a un circuito durante un cierto tiempo, actuando brevemente como una batería.

En el proceso de carga de un condensador, inicialmente en las placas conductoras no habrá una carga neta (véase la figura I.10). Al conectar una pila o batería³⁰, los electrones empiezan a intentar fluir desde el polo negativo hacia el

²⁹En el caso de que el condensador forme parte de un circuito, ello implicará que en la rama en la que se encuentre la corriente estacionaria será nula, si bien pueden existir corrientes transitorias cuando se cambien las condiciones de alimentación del circuito. Los fundamentos de teoría de circuitos se estudiarán en el capítulo II.

³⁰Las fuentes de alimentación se estudiarán en el capítulo II.

positivo de la misma, pero sin poder completar el recorrido, dado que el dieléctrico del condensador lo impide. De este modo, se produce una separación de la carga: en un lado, hay un exceso de electrones (carga $-Q$) mientras que en el otro lado hay un defecto de electrones (carga $+Q$). A esas cargas separadas una cierta distancia les corresponde un campo eléctrico y una diferencia de potencial asociada, de modo que se alcanza la situación final cuando la diferencia de potencial entre placas del condensador se iguala con el voltaje de la pila externa V_S .

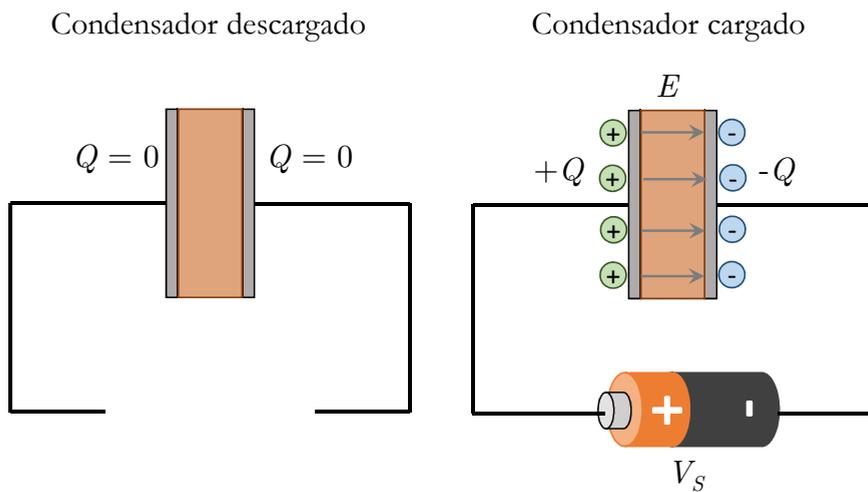


Figura I.10: Carga de un condensador.

Los condensadores se caracterizan por el valor de su **capacidad**, que es la *relación entre la carga que almacenan y la diferencia de potencial entre sus placas*, es decir:

$$C = \frac{Q}{\Delta V} \tag{I.25}$$

En el caso particular del condensador plano-paralelo, éste consta de dos placas metálicas paralelas, separadas entre sí una distancia d , con un material de permitividad³¹ ϵ entre ellas. En ese caso, si la superficie de los conductores es igual a S , la capacidad del condensador plano-paralelo viene dada por la expresión:

$$C = \epsilon \frac{S}{d} \tag{I.26}$$

³¹La permitividad es un parámetro que determina la respuesta electrostática de un material en presencia de un campo eléctrico externo.

La unidad de capacidad es el **faradio**³² (F), que corresponde a 1 C/V (culombio dividido por voltio). Las capacidades usuales de los condensadores empleados en la mayoría de circuitos electrónicos de interés se encuentra en el rango entre los nanofaradios y los microfaradios.

Los condensadores tienen múltiples aplicaciones en electrónica. En el ámbito de la informática, y a título de curiosidad, podemos mencionar los teclados capacitivos: cada tecla corresponde a una de las placas de un condensador, de modo que, según esté pulsada o no, la distancia de separación entre placas del condensador asociado cambia, variando por tanto su valor de capacidad y siendo posible detectar la pulsación (figura I.11).

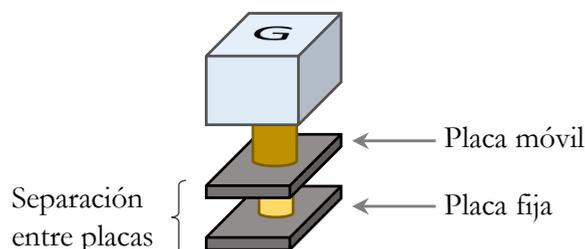


Figura I.11: Teclado capacitivo.

La **energía electrostática almacenada** en un condensador proviene del trabajo que ha sido necesario realizar para colocar la carga en sus placas, y viene dada por la expresión:

$$U = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} QV = \frac{1}{2} CV^2 \quad (\text{I.27})$$

El trabajo que tenemos que realizar es el doble de la energía almacenada (es decir, $W = 2U = QV$). La parte restante se pierde en la resistencia que, de una forma u otra, siempre está presente en el circuito (los cables en un circuito real poseen una cierta resistencia, aunque sea pequeña).

El proceso de carga de un condensador no tiene lugar de manera instantánea, sino que requiere una cierta cantidad de tiempo. Durante este tiempo, y de manera transitoria, podrán existir corrientes (movimiento de carga) en los cables que estén conectados al condensador, pero no a través de éste. Transcurrido este

³²En honor a Michael Faraday, que como ya sabemos fue el creador del concepto de líneas de campo.

tiempo y alcanzado el estado estacionario, la corriente será nula: es decir, en presencia de un condensador sólo podrá existir corriente en los cables a los que esté conectado durante el periodo transitorio, o, de manera general, mientras existan variaciones de voltaje entre sus extremos. En esas condiciones la corriente en los cables vendrá dada por:

$$I = \frac{dQ}{dt} = \frac{d(CV_C)}{dt} = C \frac{dV_C}{dt} \quad (\text{I.28})$$

La duración del tiempo transitorio depende fundamentalmente del valor de capacidad del transistor (a mayor capacidad más tiempo de carga) y de la resistencia de carga (cuanto mayor sea, más tardará en cargarse). El proceso de carga/descarga de un condensador lleva asociada una **constante de tiempo** $\tau = RC$. El voltaje entre extremos del condensador durante el transitorio de carga viene dado por la expresión:

$$V_C = V_S(1 - e^{-t/\tau}) \quad (\text{I.29})$$

Es decir, es un proceso de tipo exponencial (véase la figura I.12) en el que inicialmente hay un aumento rápido del voltaje, seguido de un transitorio más largo con una subida suave hasta alcanzar la diferencia de potencial de la pila externa V_S .

El régimen transitorio de carga y descarga de un condensador, el hecho de que no sea un proceso instantáneo sino que tome una cierta cantidad de tiempo, tiene una gran importancia en ciertos procesos que se estudiarán más adelante, como en el filtrado de señales rectificadas, y muy especialmente en el retardo de los circuitos digitales. La existencia de este retardo implica, por ejemplo, que existan límites a la velocidad a la que puede operar un ordenador³³.

Los condensadores tienen una serie de parámetros que deben ser tenidos en cuenta a la hora de trabajar con ellos en una aplicación real, como son por ejemplo:

- Voltaje máximo de trabajo: es la diferencia de potencial máxima que puede existir entre extremos de un condensador sin que se dañe. Depende fundamentalmente del tipo de dieléctrico y de la distancia de separación entre

³³Como estudiaremos más adelante, el proceso de conmutación de señales digitales en los transistores se asemeja en cierto modo a la carga y descarga de un condensador.

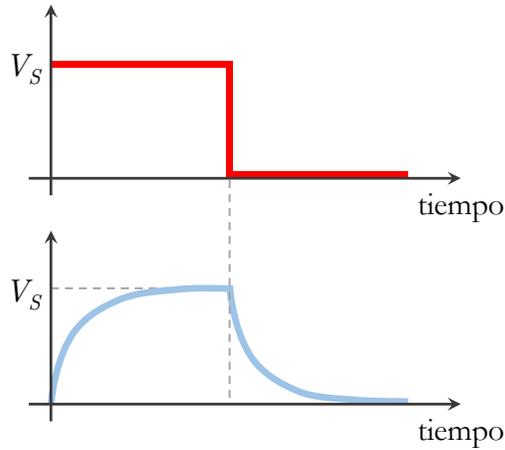


Figura I.12: Evolución temporal de la carga y descarga de un condensador. La figura superior muestra el potencial externo aplicado (según esté o no la pila conectada) y la inferior la evolución de la diferencia de potencial entre placas del condensador.

placas. Se debe siempre evitar el trabajar a un voltaje superior al máximo recomendado.

- Tolerancia: es el error máximo que puede existir entre el valor de capacidad nominal (el indicado por el fabricante) y el valor real de capacidad de un condensador.
- Polaridad: teóricamente debería ser indiferente la orientación con la que conectásemos un condensador. Sin embargo, algunos tipos de condensadores (en particular los condensadores electrolíticos), por el modo en el que están fabricados, tienen polaridad, es decir, deben colocarse con una orientación determinada, respetando los lados positivo y negativo indicados por el fabricante.

8 CAMPO MAGNÉTICO

Como ya indicamos al comienzo de este capítulo, el magnetismo es conocido por la humanidad desde hace casi tres mil años, cuando en la región griega de Magnesia los pastores de la zona observaron cómo algunos minerales eran capaces de atraer piedras que contuviesen vetas de hierro. Sin embargo, como en

el caso de la electricidad, el estudio científico detallado de los fenómenos magnéticos no se iba a producir prácticamente hasta el siglo XIX. Particularmente llamativos fueron los experimentos que evidenciaron una relación entre la electricidad y el magnetismo. En 1820, de manera casual, mientras realizaba una demostración en una de sus clases, Ørsted³⁴ descubrió que la aguja de una brújula cambiaba su orientación cuando se encontraba cerca de una corriente eléctrica. Este descubrimiento llamó la atención de Ampère y Faraday, que realizaron una serie de experimentos para analizar la relación entre los fenómenos eléctricos y magnéticos. Ampère, por ejemplo, estudió la fuerza de atracción magnética que se observaba entre dos hilos metálicos colocados en paralelo cuando por ellos circulaba corriente eléctrica. Faraday, por su parte, desarrolló los primeros motores basados en la acción de un imán sobre un conductor, y fue el responsable del descubrimiento de la inducción electromagnética.

De manera análoga a las cargas positivas y negativas, en el caso de los fenómenos magnéticos hablamos de **polo norte y polo sur** de un objeto imantado (que posee propiedades magnéticas). A diferencia del caso de las cargas eléctricas, que sí pueden separarse, los polos magnéticos aparecen siempre de forma conjunta. Si por ejemplo rompiésemos un imán por la mitad, cada una de las nuevas partes presentaría polo norte y polo sur. Polos iguales se repelen magnéticamente, mientras que polos distintos se atraen.

En nuestro caso, nos interesa conocer cómo afectan los campos magnéticos a cargas eléctricas en movimiento. Al igual que en el caso de los fenómenos eléctricos, la acción de la fuerza magnética puede caracterizarse por un **campo magnético**, que denotaremos con la letra \vec{B} ³⁵. Como para el campo eléctrico, pueden dibujarse unas líneas de campo magnético, que en este caso saldrán desde el polo norte y entrarán hacia el polo sur.

Si tenemos una carga eléctrica q que se desplaza con una cierta velocidad \vec{v} , en presencia de un campo magnético \vec{B} experimentará una fuerza denominada **fuerza de Lorentz**³⁶, que viene dada por:

$$\vec{F} = q\vec{v} \times \vec{B} \quad (\text{I.30})$$

El símbolo \times representa el producto vectorial³⁷, lo que en este caso se tra-

³⁴Hans Christian Ørsted (1777-1851), físico y químico danés.

³⁵En sentido estricto, esta magnitud se denomina *inducción magnética*.

³⁶Hendrik Lorentz (1853-1928) fue quien realizó la formulación moderna de la expresión de esta fuerza.

³⁷El producto vectorial de dos vectores es a su vez un vector, con dirección perpendicular al

duce en que la fuerza de Lorentz aparece en la dirección perpendicular al plano formado por \vec{v} y \vec{B} .

En consecuencia, si cerca de un cable por el que circule corriente eléctrica existe algún tipo de campo magnético externo, éste afectará al movimiento de las cargas. Este es un fenómeno que puede producir interferencias no deseadas en la transmisión de señales eléctricas (en definitiva, en la transmisión de información) y en ciertos casos puede ser un problema para los equipos informáticos si los campos magnéticos presentes son intensos.

9 INDUCCIÓN ELECTROMAGNÉTICA

La **inducción electromagnética** consiste en la generación de una diferencia de potencial (fuerza electromotriz)³⁸ V cuando un conductor se encuentra en presencia de un flujo variable de campo magnético. El flujo se define como:

$$\phi = \vec{B} \cdot \vec{S} = BS \cos(\theta) \quad (\text{I.31})$$

siendo \vec{S} un vector de magnitud igual al valor del área encerrada por el conductor y con dirección perpendicular a la misma, y θ el ángulo entre \vec{B} y \vec{S} (véase la figura I.13).

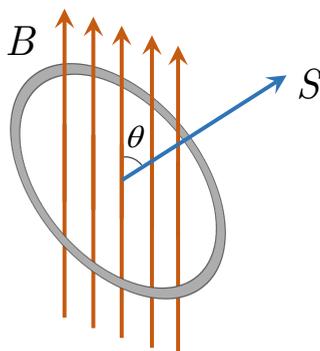


Figura I.13: Flujo magnético.

Como ya se ha mencionado, Faraday fue el descubridor del fenómeno de la inducción. En su experimento, Faraday acercaba un imán a un conductor con plano que conforman los dos vectores del producto.

³⁸El concepto de fuerza electromotriz puede considerarse equivalente al de la generación de una diferencia de potencial entre dos puntos.

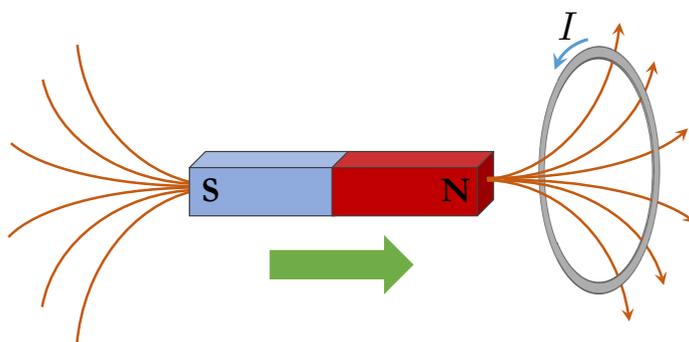


Figura I.14: Inducción electromagnética.

forma de espira. En dicha espira se medía una corriente únicamente cuando se acercaba o se alejaba el imán, pero no cuando éste permanecía quieto. Este fenómeno se conoce como la **ley de Faraday**, que en este contexto puede expresarse como³⁹:

$$V = -\frac{d\phi}{dt} \quad (\text{I.32})$$

Donde V es la diferencia de potencial o fuerza electromotriz inducida que produce la corriente I en la espira (véase la figura I.14). La ley de Faraday expresa que la fuerza electromotriz aparece *cuando el flujo cambia*: en el experimento mencionado, el flujo cambia al variar la intensidad del campo magnético que atraviesa la espira, es decir, \vec{B} . Cuando se acerca el imán aumenta, y cuando se aleja disminuye. Cuando está parado, al ser constante la intensidad de \vec{B} y no variar la orientación del imán y la espira, la variación de flujo magnético es nula y por tanto no aparece ninguna corriente (no se genera una fuerza electromotriz que mueva las cargas en el conductor).

Los fenómenos de inducción tienen multitud de aplicaciones, desde la generación de energía eléctrica hasta el calentamiento de alimentos. También puede tener consecuencias no deseadas, como mencionamos con anterioridad: si una línea de comunicación de datos se encuentra en un entorno donde puede haber campos magnéticos variables (por ejemplo, cerca de motores industriales) la inducción puede provocar corrientes en la línea que alteren la integridad de los

³⁹Esta formulación en concreto se conoce también como ley de Lenz, por Heinrich F. R. Lenz (1804-1865). En realidad, la ley de Lenz está contenida en la formulación más general que realizó Faraday para la inducción.

datos enviados. Una posible solución para paliar este fenómeno es el uso de cables de par trenzado, que minimizan el área efectiva y reducen el flujo magnético.

La **generación de energía eléctrica** es quizá una de las aplicaciones más extendidas de la inducción electromagnética. El principio se basa en el movimiento rotatorio de una espira (o un conjunto de espiras) en un campo magnético constante, como el que puede producir un imán estático. Si rotamos las espiras con frecuencia de giro constante, el ángulo entre el campo y la espira variará como $\theta = \omega t$, siendo ω la velocidad o frecuencia angular⁴⁰ y t el tiempo. En este caso, aparece una fuerza electromotriz⁴¹:

$$V = -\frac{d\phi}{dt} = -\frac{d(BS \cos \theta)}{dt} = A \sin(\omega t) \quad (\text{I.33})$$

donde A es la amplitud (el valor máximo que puede tomar V), igual en este caso a $BS\omega$.

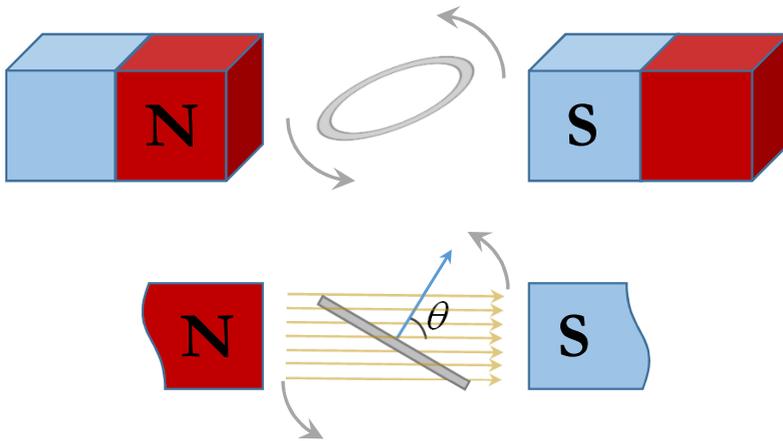


Figura I.15: Generación de una fuerza electromotriz senoidal a partir del fenómeno de la inducción.

La fuerza electromotriz obtenida tiene una dependencia con el tiempo según la función seno, por lo que decimos que es una señal *senoidal*. En una señal

⁴⁰La frecuencia angular se mide en rad/s (radianes dividido por segundo). Un radián equivale a $360/2\pi$ grados.

⁴¹Este ejemplo corresponde a una generación de tipo *monofásico*, que implica que en ciertos instantes de tiempo la potencia sería nula (cada vez que V pase por cero). Para evitar este problema la generación industrial es de tipo trifásico, es decir, consiste en tres señales desfasadas 120° para mantener un valor constante de potencia.

senoidal, nos interesan parámetros tales como el periodo T (el tiempo que se tarda en hacer un ciclo completo), la frecuencia lineal f (el inverso del periodo)⁴², la frecuencia angular (ω , correspondiente a $2\pi f$), y, en el caso de campos electromagnéticos, como veremos posteriormente, la longitud de onda λ (igual a $cT = c/f$, siendo c la velocidad de la luz, es decir, 300.000 km/s). En caso de que la señal no oscile en torno a cero voltios, decimos que presenta un *offset* o desplazamiento, que corresponderá al valor medio en voltaje de la señal.

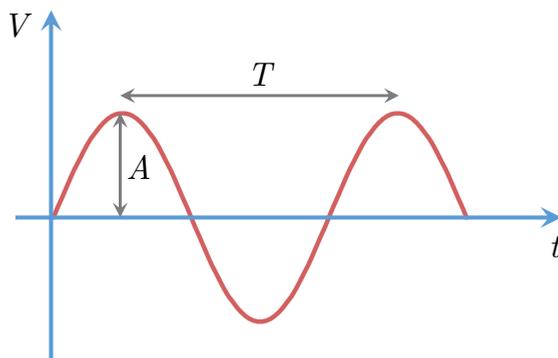


Figura I.16: Señal senoidal: amplitud y periodo.

La existencia de señales de voltaje oscilantes abre la puerta a los llamados **circuitos de corriente alterna (AC)**⁴³, por contraposición a los de **corriente continua (DC)**⁴⁴, en los que voltajes y corrientes no varían en el tiempo. A finales del siglo XIX, hubo una enorme disputa sobre cuál era la mejor forma de transportar la corriente eléctrica, dando lugar a la denominada **guerra de los corrientes**. Por un lado, Thomas Alva Edison (1847-1931) defendía las ventajas de la corriente continua, mientras que la compañía Westinghouse, basándose en los trabajos y patentes de Nikola Tesla (1856-1943) abogaba por el transporte a través de líneas de corriente alterna. Tras varios años de una agria controversia, se acabó imponiendo la evidencia de las ventajas de transportar la energía eléctrica mediante líneas alternas de alto voltaje, que minimizaban las pérdidas y posibi-

⁴²La frecuencia lineal se mide en *hercios*, en honor a Heinrich Hertz (1857-1894). $1 \text{ Hz} = 1 \text{ s}^{-1}$.

⁴³AC se refiere a las siglas inglesas *Alternating Current*, corriente alterna, indicando que la corriente cambia su sentido de flujo y magnitud de forma periódica durante la operación.

⁴⁴DC se refiere a las siglas inglesas *Direct Current*, corriente directa, en referencia a que la corriente no cambia de sentido en la operación del circuito.

litaban que las centrales productoras estuvieran a decenas o incluso cientos de kilómetros de distancia del consumidor final.

Más del 99% de la energía eléctrica que se produce hoy en día proviene de generadores de corriente alterna. La existencia de transformadores que permiten aumentar el voltaje de forma sencilla y eficiente permite que esta energía pueda ser transportada a un coste reducido en las líneas de alto voltaje. En el contexto de estas señales AC, que oscilan entre valores positivos y negativos, es importante el denominado **valor eficaz**, que viene dado por la expresión:

$$V_{eficaz} = \sqrt{\frac{1}{T} \int_0^T V^2(t) dt} \quad (\text{I.34})$$

El valor eficaz de un voltaje AC corresponde al *valor de un voltaje DC que proporciona la misma potencia*.

10 BOBINAS

Una **bobina de autoinducción** L es un dispositivo pasivo de dos terminales que permite almacenar energía en un campo magnético. Suelen consistir en una espira de hilo conductor enrollada alrededor de aire o de un núcleo de ferrita. Se emplean con frecuencia en circuitos AC dado que pueden servir, por ejemplo, para bloquear corrientes alternas de una cierta frecuencia. Cuando circula una corriente por una espira, se genera un campo magnético perpendicular al plano de la espira. Si esa corriente varía, el campo magnético también variará y aparecerá una fuerza electromotriz debido al fenómeno de inducción. Dicha fuerza electromotriz tiende a oponerse al aumento de corriente y es igual a la diferencia de potencial entre extremos de la bobina, que viene dada por la expresión:

$$V_L = L \frac{dI}{dt} \quad (\text{I.35})$$

El valor de autoinducción L se mide en Henrios⁴⁵ (H), siendo 1 Henrio = 1 Ω/s (ohm dividido por segundo). Para una bobina con N espiras de área A y con longitud l , tenemos que:

$$L = \mu \frac{N^2 A}{l} \quad (\text{I.36})$$

⁴⁵En honor a Joseph Henry, físico estadounidense (1797-1878).

siendo μ la permeabilidad magnética del material que haya en el interior de la espira.

Bobinas, resistencias y condensadores permiten crear circuitos que actúan como filtros, circuitos resonantes, osciladores, etc. que son particularmente importantes en el ámbito de las telecomunicaciones.

11 CAMPO ELECTROMAGNÉTICO

A lo largo del siglo XIX, la experimentación sobre los fenómenos eléctricos y magnéticos fue dejando cada vez más patente la relación entre ambos. Fue Maxwell⁴⁶ quien proporcionó el soporte teórico para la descripción conjunta de ambos fenómenos, a través de las que hoy son conocidas como **ecuaciones de Maxwell**. Si bien en este texto no las emplearemos, las enunciamos aquí por su importancia e interés en el estudio del electromagnetismo:

- Ley de Gauss:

$$\oint_S \vec{E} d\vec{S} = \frac{1}{\epsilon_0} Q_{interior} \quad (I.37)$$

- Ley de Faraday

$$\oint_C \vec{E} \cdot d\vec{l} = -\frac{d}{dt} \int_S \vec{B} d\vec{S} = -\int_S \frac{\partial \vec{B}}{\partial t} d\vec{S} \quad (I.38)$$

- Ley de Gauss del magnetismo

$$\oint_S \vec{B} d\vec{S} = 0 \quad (I.39)$$

- Ley de Ampère

$$\oint_C \vec{B} \cdot d\vec{l} = \mu_0 I + \mu_0 \epsilon_0 \int_S \frac{\partial \vec{E}}{\partial t} d\vec{S} \quad (I.40)$$

⁴⁶James Clerk Maxwell (1831-1879), científico escocés que realizó la unificación teórica de los fenómenos de la electricidad y el magnetismo.

Estas cuatro leyes unificaron la descripción de los campos eléctricos y magnético, demostrando que se trata de dos caras del mismo fenómeno. En consecuencia, hablamos de **campo electromagnético**. El campo electromagnético se propaga en forma de ondas, a la velocidad de la luz. Dichas **ondas electromagnéticas** pueden tener distinta frecuencia o **longitud de onda**⁴⁷, de modo que dependiendo de ellas tendrán distintas propiedades. Los rayos X, los rayos gamma, la luz visible, las microondas, las ondas de radio comercial... todas ellas son ondas electromagnéticas que se diferencian por su longitud de onda, conformando el **espectro electromagnético**.

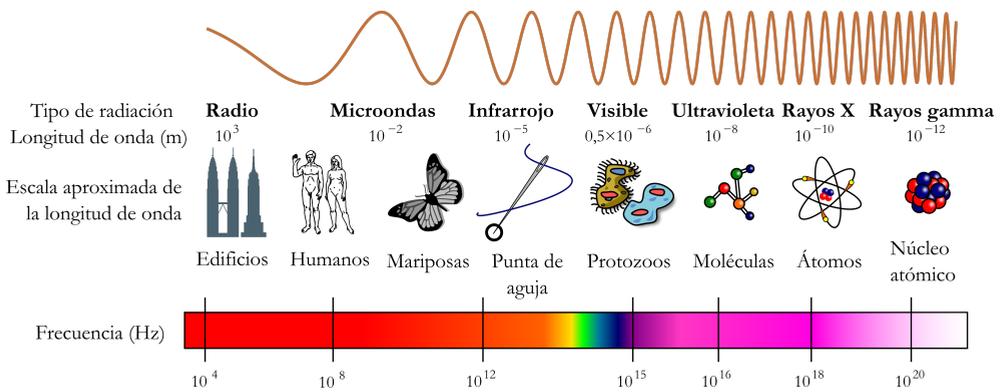


Figura I.17: Espectro electromagnético.

Las señales **wifi**, por ejemplo, trabajan en muchos casos a una frecuencia de 2.4 GHz (longitud de onda de 12.5 cm). En los 60 y 70, cuando se definió el estándar para el funcionamiento de los hornos microondas, se eligió la frecuencia de 2.4 GHz a partir de ciertos factores relacionados con la eficiencia en el calentamiento de distintos tipos de alimentos, construcción de los hornos, etc. Este rango de frecuencias quedó sin utilizar mediante licencias gubernamentales, para evitar interferencias, y quedó libre para otros tipos de usos. No es casual por tanto que un horno microondas en funcionamiento pueda interferir con la señal wifi.

Además, esta frecuencia tiene la ventaja de corresponder a radiación electromagnética no ionizante, y por tanto, no daña para el ADN humano (al contrario que la radiación muy energética, como pueden ser los rayos X, los rayos gamma

⁴⁷La longitud de onda, que suele representarse por la letra griega λ , es la distancia que recorre una onda durante el tiempo correspondiente a un periodo.

o los ultravioletas). Tampoco tiene un alcance excesivo entre muros, lo que es bueno para aplicaciones de interior, en el hogar, la oficina, etc. En los usos comerciales habituales no se emplean potencias elevadas (máximo 100 mw o 20 dBm -decibelio-milivatio-⁴⁸ para 2.4 GHz, 200 mw o 23 dBm para 5 GHz), por lo que no son perjudiciales para la salud humana. Como ejemplo sirva decir que la potencia de un horno microondas es de unos 800 w, es decir, 8000 veces mayor a la máxima legal que puede emitir un router wifi.

⁴⁸El dBm o decibelio-milivatio es una unidad de medida alternativa de potencia empleada en comunicaciones por radio y fibra óptica, basada en logaritmos en base 10. Así, 0 dBm equivalen a 1 mw, 10 dBm a 10 mw, 20 dBm a 100 mw, etc.

II.

FUNDAMENTOS DE
TEORÍA DE
CIRCUITOS

Los ordenadores son, en esencia, circuitos electrónicos extraordinariamente complejos, constituidos por una serie de elementos como transistores, resistencias, condensadores, diodos, etc. Antes de abordar el estudio de los dispositivos basados en semiconductores, como transistores o diodos, vamos a conocer los fundamentos de la teoría de circuitos eléctricos. Nos centraremos principalmente en los circuitos de corriente continua, que tienen mayor interés para el caso de los circuitos digitales, en los que los dispositivos van a trabajar en uno de dos posibles estados o niveles.

Un **circuito eléctrico** no es más que un *conjunto de elementos interconectados mediante cables conductores*. Por estos cables y elementos usualmente circulará una corriente, que podrá ser continua o alterna (oscilante). La idea de *circulación* de corriente ya nos indica que en el circuito deberá haber algún camino cerrado que puedan recorrer de forma repetitiva las partículas cargadas eléctricamente. Asimismo tendrá que existir algún elemento que proporcione energía a esas partículas para que puedan recorrer el circuito, en definitiva, algún tipo de fuente de alimentación.

1 RESISTENCIAS

El elemento más sencillo que podemos encontrar en un circuito son las **resistencias**. Se trata de dispositivos de dos terminales (es decir, dos conectores) que se oponen en mayor o menor grado al paso de corriente según su valor de R (concepto que ya vimos en el capítulo I). Las resistencias se caracterizan por ser dispositivos pasivos (no proporcionan ganancia¹) y además lineales (hay una re-

¹La ganancia es la relación entre la amplitud de la señal eléctrica a la salida de un determinado elemento y la amplitud de la señal en la entrada, concepto que veremos con mayor detalle en el

lación de proporcionalidad lineal entre la diferencia de voltaje entre sus extremos y la corriente que circula por ellas, determinada precisamente por el valor de R).

Para representar una resistencia usaremos el símbolo mostrado en la figura II.1:



Figura II.1: Símbolo correspondiente a una resistencia eléctrica.

Cuando utilizamos resistencias como *componentes discretos*², podemos conocer su valor midiéndolo con un **multímetro**³, pero también mediante el denominado **código de colores**, una serie de líneas de color serigrafiadas en su encapsulado que nos indican el valor nominal de la resistencia.

En la figura II.2 se muestra el código de colores de cuatro líneas⁴. Las dos primeras líneas corresponden a cifras significativas según su color. En el ejemplo mostrado, la primera línea sería amarilla (4) y la segunda violeta (7), conformando el número 47. La tercera cifra es el multiplicador: en este ejemplo, de color rojo, correspondería a un factor $10^2 = 100$, que es el número por el que tenemos que multiplicar el obtenido anteriormente, resultando $47 \times 100 = 4700$. Finalmente, la última línea indica la **tolerancia**, es decir, el margen de error entre el valor nominal y el valor real. En este caso, al ser de color plata, correspondería a un margen de error del 10%.

2 CONEXIÓN EN SERIE Y PARALELO

En un circuito, podremos encontrarnos resistencias (o cualquier otro elemento) colocadas principalmente de dos maneras, en **serie** o en **paralelo**.

La **conexión en serie** consiste en el encadenamiento de un elemento a continuación de otro, sin otros elementos entre medias.

La **conexión en paralelo** consiste en conectar dos elementos del circuito de manera que tengan en común su punto inicial y su punto final.

capítulo VIII.

²Un componente discreto es aquel objeto que sólo contiene un dispositivo electrónico, en este caso una resistencia, por contraposición a los circuitos integrados, donde en el mismo objeto (el *microchip*) están incluidos múltiples dispositivos.

³Un multímetro es un dispositivo que nos permite medir varias magnitudes eléctricas, como

Color	1ª Cifra	2ª Cifra	Multiplicador	Tolerancia
Negro	0	0	10^0	
Marrón	1	1	10^1	
Rojo	2	2	10^2	
Naranja	3	3	10^3	
Amarillo	4	4	10^4	
Verde	5	5	10^5	
Azul	6	6	10^6	
Violeta	7	7	10^7	
Gris	8	8	10^8	
Blanco	9	9	10^9	
Dorado				5%
Plata				10%

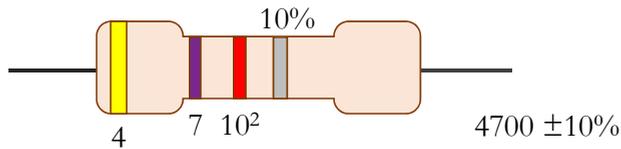


Figura II.2: Código de colores de resistencias. Ejemplo para una resistencia de 4700Ω con tolerancia de $\pm 10\%$.

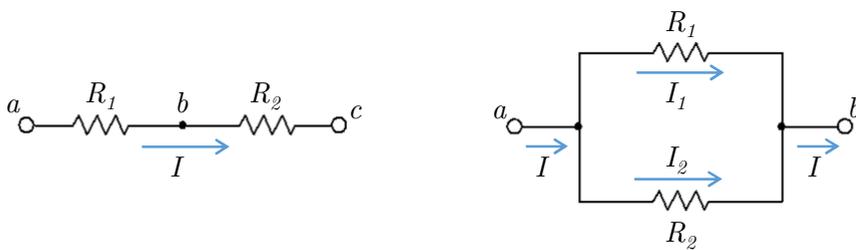


Figura II.3: Conexiones en serie (izquierda) y paralelo (derecha).

Cuando se colocan dos resistencias en serie, por ellas circula *la misma corriente eléctrica*. Sin embargo, la diferencia de potencial en cada una de ellas (o, si se

voltaje, corriente, resistencia, etc.

⁴Existen otras variantes de código, de tres, cinco o seis líneas, por ejemplo.

prefiere, el voltaje que cae en cada una) será diferente si las resistencias tienen distinto valor. En todo caso, la suma de diferencias de potencial deberá ser igual a la diferencia de potencial total entre el comienzo de la primera resistencia y el final de la segunda.

De este modo, podemos definir una **resistencia equivalente en serie** que agrupe el efecto de las dos resistencias, del siguiente modo:

$$V_{ac} = V_{ab} + V_{bc} = IR_1 + IR_2 = I(R_1 + R_2) = IR_{serie} \quad (\text{II.1})$$

De forma general, dadas n resistencias colocadas en serie, la resistencia equivalente en serie total será la suma de todas las resistencias:

$$R_{serie} = \sum_{i=1}^n R_i \quad (\text{II.2})$$

En cambio, cuando se colocan dos resistencias en paralelo, es la diferencia de potencial la que es igual para las dos resistencias, mientras que la corriente será distinta si las resistencias no tienen el mismo valor. La suma de las corrientes que circulan por cada resistencia será igual a la corriente total, por lo que tendremos:

$$I = I_1 + I_2 = \frac{V_{ab}}{R_1} + \frac{V_{ab}}{R_2} = V_{ab} \left(\frac{1}{R_1} + \frac{1}{R_2} \right) = \frac{V_{ab}}{R_{paralelo}} \quad (\text{II.3})$$

Luego, de manera general, podemos decir que dadas n resistencias en paralelo, el inverso de la resistencia equivalente es igual a la suma de los inversos de todas las resistencias:

$$\frac{1}{R_{paralelo}} = \sum_{i=1}^n \frac{1}{R_i} \quad (\text{II.4})$$

En algunas ocasiones no podremos afirmar con claridad si varias resistencias en un circuito están conectadas en serie o en paralelo. En esos casos, puede ser útil emplear las transformaciones **triángulo-estrella** definidas por el **teorema de Kennelly**⁵.

⁵Por Arthur E. Kennelly (1861-1939), ingeniero norteamericano de origen británico.

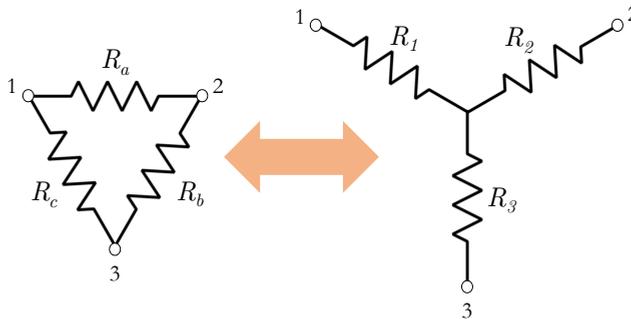


Figura II.4: Resistencias equivalentes en configuración de triángulo y estrella.

Paso de triángulo a estrella

Las transformaciones son:

$$R_1 = \frac{R_a R_c}{R_a + R_b + R_c} \tag{II.5}$$

$$R_2 = \frac{R_a R_b}{R_a + R_b + R_c} \tag{II.6}$$

$$R_3 = \frac{R_b R_c}{R_a + R_b + R_c} \tag{II.7}$$

Paso de estrella a triángulo

Las transformaciones son:

$$R_a = \frac{R_1 R_2 + R_1 R_3 + R_2 R_3}{R_3} \tag{II.8}$$

$$R_b = \frac{R_1 R_2 + R_1 R_3 + R_2 R_3}{R_1} \tag{II.9}$$

$$R_c = \frac{R_1 R_2 + R_1 R_3 + R_2 R_3}{R_2} \tag{II.10}$$

Al igual que en el caso de las resistencias, los condensadores pueden asociarse en serie o en paralelo, de modo que podremos calcular la capacidad equivalente de un conjunto de condensadores conforme a los tipos de conexión que tengan.

En el caso de la asociación de **condensadores en serie**, debe tenerse en cuenta que en ese caso la carga almacenada en cada uno de ellos es la misma, Q , mientras que el voltaje total es la suma de los voltajes de cada condensador, por lo que tenemos que:

$$V_{ac} = V_{ab} + V_{bc} = \frac{Q}{C_1} + \frac{Q}{C_2} = \frac{Q}{C_{serie}} \quad (\text{II.11})$$

Por tanto, de manera general la capacidad equivalente en serie será:

$$\frac{1}{C_{serie}} = \sum_{i=1}^n \frac{1}{C_i} \quad (\text{II.12})$$

Por otra parte, la asociación de **condensadores en paralelo** implica que la diferencia de potencial en ellos será la misma, pero la carga almacenada total será la suma de las cargas almacenadas por cada condensador, por lo que tendremos que:

$$Q_{paralelo} = Q_1 + Q_2 = C_1 V_{ab} + C_2 V_{ab} = C_{paralelo} V_{ab} \quad (\text{II.13})$$

En definitiva, la capacidad equivalente en paralelo será:

$$C_{paralelo} = \sum_{i=1}^n C_i \quad (\text{II.14})$$

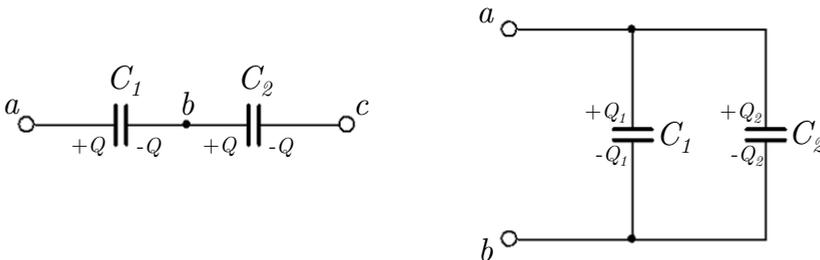


Figura II.5: Asociación de condensadores en serie y en paralelo.

En el caso de las **autoinducciones (bobinas)**, la asociación en serie y paralelo se realiza de igual forma que en el caso de las resistencias.

3 DEFINICIONES ÚTILES

Como ya se ha comentado anteriormente, los circuitos eléctricos pueden clasificarse en circuitos de corriente continua (DC), cuando trabajan con señales eléctricas, corrientes o voltajes, que son constantes en el tiempo y siempre tienen el mismo signo, o de corriente alterna (AC), cuando trabajan con señales eléctricas oscilantes.

En este capítulo vamos a ocuparnos principalmente de los circuitos de corriente continua compuestos de fuentes de alimentación y resistencias, es decir, circuitos lineales⁶. Además, los componentes empleados (fuentes o resistencias) tendrán solamente dos conexiones o terminales. Estos componentes se conectarán empleando conductores (cables) que, idealmente, supondremos que tienen resistencia nula. Al realizarse las conexiones, podremos identificar en un circuito varios elementos de interés:

- **Nodo:** un nodo es un punto del circuito donde se unen 3 o más componentes. Es necesario aclarar que en este caso no nos referimos exactamente a un punto geométrico, sino de manera general, a una conexión eléctrica.
- **Rama:** es cualquier trayecto entre dos nodos.
- **Malla:** es cualquier camino cerrado en un circuito (es decir, que empieza y termina en el mismo punto).

Además, en ocasiones pueden darse las siguientes situaciones:

- **Cortocircuito:** consiste en unir dos puntos con un cable. En consecuencia, la diferencia de potencial entre esos dos puntos es nula, ya que asumimos que el cable no tiene resistencia. Si por ejemplo cortocircuitamos una resistencia (unimos sus extremos con un cable), a efectos prácticos es como si la hubiéramos eliminado del circuito (no pasaría corriente por ella ya que toda la corriente se desviaría por el cortocircuito). Unir dos nodos con un cortocircuito hace que estos nodos pasen a ser eléctricamente el mismo nodo.

⁶En los circuitos lineales se aplica el principio de superposición: si tenemos varias fuentes, es posible resolver el sistema obteniendo la solución (por ejemplo, el cálculo de una corriente determinada) como la suma de las soluciones considerando cada fuente por separado.

- **Circuito abierto:** dos puntos están en circuito abierto cuando no hay ningún componente o cable que los una, es decir, es como si existiese una resistencia infinita entre ellos. Por una rama en la que haya un circuito abierto no puede circular corriente.

4 FUENTES DE ALIMENTACIÓN

En los circuitos deberán existir **fuentes** de voltaje o corriente que proporcionen la energía necesaria para el funcionamiento del circuito. Un ejemplo típico de fuente de voltaje lo tenemos en una batería o pila, que transforma energía química en energía eléctrica. Sobre los electrones que salen por el lado negativo de la pila se ha realizado un trabajo que, por unidad de carga, corresponde al **voltaje** de la batería.

Una **fuente de voltaje ideal** es *aquella capaz de mantener una diferencia de potencial constante entre sus extremos*, independientemente de los elementos que le acoplemos⁷ (o, si se prefiere, de la corriente que circule a su través). Para ello debe cumplirse que la resistencia interna de la batería sea nula. Esto en la realidad nunca se cumple: toda fuente de voltaje tiene asociada una resistencia interna r no nula (normalmente de valor muy pequeño) que conlleva que el voltaje que se cede a los elementos acoplados (carga externa) nunca sea exactamente el nominal, sino ligeramente inferior. En los circuitos que estudiaremos, sin embargo, usualmente consideraremos las fuentes como ideales, y en todo caso se tratará el efecto de la resistencia interna r tomando una fuente de voltaje ideal en serie con dicha resistencia (figura II.6).

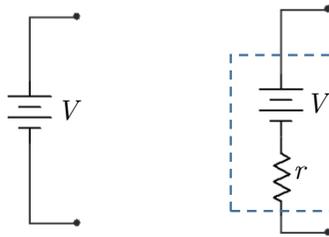


Figura II.6: Fuente de voltaje ideal (izquierda) y real (derecha).

Por otra parte, una **fuente de corriente ideal** es *aquella que proporciona una*

⁷De manera genérica nos referiremos a esos elementos como la *carga* acoplada al circuito, que no debe confundirse con el concepto de carga eléctrica.

corriente constante en la rama en la que se encuentre colocada, independientemente de la carga acoplada (o si se prefiere, de la diferencia de potencial entre sus extremos). Al igual que en el caso de las fuentes de voltaje, en las fuentes de corriente reales existe una resistencia interna r que debe ser tomada en cuenta. En nuestro caso consideraremos las fuentes de corriente reales como fuentes de corriente ideales con una resistencia en paralelo de valor igual a la resistencia interna r (que en este caso, debe ser de un valor muy grande, idealmente infinito), como se muestra en la figura II.7.

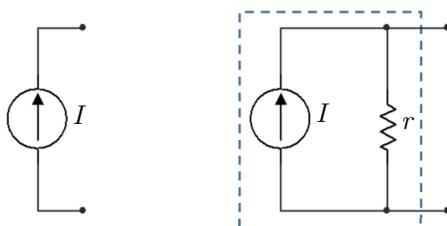


Figura II.7: Fuente de corriente ideal (izquierda) y real (derecha).

5 LEYES DE KIRCHHOFF

Las leyes de Kirchhoff⁸ son dos herramientas muy útiles para la resolución de circuitos. Son dos: la ley de corrientes de Kirchhoff (LCK) y la ley de voltajes de Kirchhoff (LVK).

La ley de corrientes de Kirchhoff (LCK) es consecuencia del principio de conservación de carga (la suma algebraica de las cargas en un sistema debe permanecer constante), y a efectos prácticos, de ella se desprende que *la suma de las corrientes que entran a un nodo es igual a la suma de las corrientes que salen*.

$$\Sigma I_{in} = \Sigma I_{out} \quad (\text{II.15})$$

La ley de voltajes de Kirchhoff (LVK) es consecuencia del principio de conservación de la energía: a efectos prácticos, esta ley implica que *la suma del potencial ganado es igual a la suma de las pérdidas o caídas de potencial*, o dicho de otra manera, que *cuando recorremos una malla de manera completa, empezando y terminando en el mismo punto, la ganancia o pérdida neta de voltaje es nula*.

⁸Llamadas así por Gustav Kirchhoff (1824-1887), que las enunció en 1847.

Para utilizar esta ley a la hora de resolver circuitos, se puede realizar de varias formas. En nuestro caso separaremos las ganancias de potencial en las fuentes de voltaje (baterías) de las pérdidas de potencial en los demás elementos (por ejemplo, resistencias). Así, tendremos que:

$$\Sigma V_{baterias} = \Sigma V_{resistencias} \quad (\text{II.16})$$

Las leyes de Kirchhoff constituyen el principal arsenal teórico que necesitamos para resolver circuitos eléctricos sencillos.

Criterios para determinar las variaciones de voltaje

El criterio que vamos a seguir para determinar las variaciones de voltaje o potencial cuando evaluemos la LVK va a ser el que describimos a continuación. Consideraremos dos casos, el de las baterías y el de las resistencias. A la hora de evaluar estas variaciones de potencial, **partiremos de un punto del circuito y recorreremos una malla hasta volver al mismo punto, siguiendo un determinado sentido en el recorrido** (por ejemplo, sentido de las agujas del reloj, o sentido contrario a las agujas del reloj, como prefiramos).

En el caso de las fuentes de voltaje (baterías), para determinar el valor de $\Sigma V_{baterias}$ seguiremos el siguiente criterio: si una fuente de voltaje V tiene la misma orientación que el sentido en el que estamos recorriendo la malla, se suma, y si tiene orientación opuesta, se resta (véase la figura II.8).

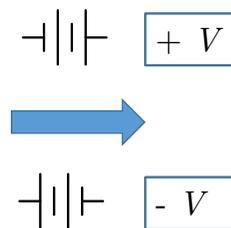


Figura II.8: Criterio para evaluar la variación de potencial en las baterías. La flecha indica el sentido en el que estamos evaluando la malla.

En el caso de las resistencias, vamos a evaluar las caídas de voltaje (pérdidas) mediante la ley de Ohm. Para ello tendremos que tener en cuenta cuál es el sentido en el que estamos recorriendo la malla, y cuál es la orientación de la corriente en cada rama. Tendremos por tanto dos posibilidades. Si la corriente tiene la misma orientación que el sentido en el que estamos recorriendo la malla, consideramos

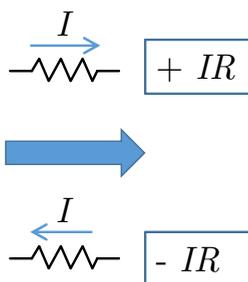


Figura II.9: Criterio para evaluar las caídas de potencial en las resistencias. La flecha indica el sentido en el que estamos evaluando la malla.

que la caída de potencial será $+I \cdot R$. En cambio, si la corriente tiene orientación opuesta al sentido en el que estamos evaluando las ganancias o pérdidas de voltaje, la caída de potencial será $-I \cdot R$. Ambos casos están ilustrados en la figura II.9.

6 ANÁLISIS DE MALLAS

El **análisis de mallas** consiste en obtener las corrientes en cada rama de un circuito aplicando las leyes de Kirchhoff. Para ello primero hay que indicar mediante flechas el sentido (elegido arbitrariamente) de la corriente en cada rama. Posteriormente, hay que aplicar la LCK en cada nodo y la LVK en cada malla hasta obtener tantas ecuaciones como incógnitas (corrientes) tengamos, para finalmente resolver el sistema de ecuaciones obtenido⁹.

Apliquémoslo a un ejemplo. Sea el circuito de la figura II.10, compuesto de dos fuentes de voltaje reales (con resistencias internas r_1 y r_2) y dos resistencias R_1 y R_L conectadas como se indica en el diagrama.

El primer paso es dibujar las corrientes en cada rama del circuito (en nuestro caso tres ramas, y por tanto, tres corrientes que serán las incógnitas a resolver), e identificar los nodos y mallas. En la figura II.11 se muestran para nuestro caso.

⁹Habitualmente se considera este método de análisis empleando corrientes asignadas a cada malla. Ello permite reducir el número de incógnitas en el sistema asumiendo de manera no explícita la LCK en las ramas comunes a dos mallas, pero puede dar lugar a una cierta confusión conceptual respecto a la corriente que fluye en cada rama del circuito. Para evitarlo, en este texto definiremos las corrientes por rama, y no por malla, y consideraremos la LCK explícitamente, por lo que nuestro método (más general y conceptualmente sencillo) puede entenderse como una combinación de los métodos de análisis de mallas y análisis de nodos que se suelen describir en otros textos.

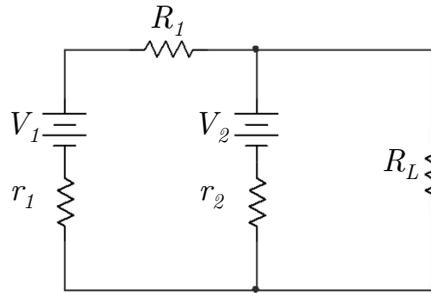


Figura II.10: Circuito de ejemplo compuesto por dos fuentes de voltaje reales y dos resistencias adicionales. Los valores de los elementos son los siguientes: $V_1 = 6 \text{ V}$, $r_1 = 2 \Omega$, $V_2 = 4 \text{ V}$, $r_2 = 1 \Omega$, $R_1 = 4 \Omega$ y $R_L = 2 \Omega$.

Los sentidos de las corrientes se han elegido en principio arbitrariamente. Ello no supone un problema, pues en caso de fluir en sentido contrario al supuesto inicialmente, simplemente obtendríamos un valor negativo para la corriente. Como nuestras incógnitas son tres, serán necesarias tres ecuaciones independientes para resolver el sistema.

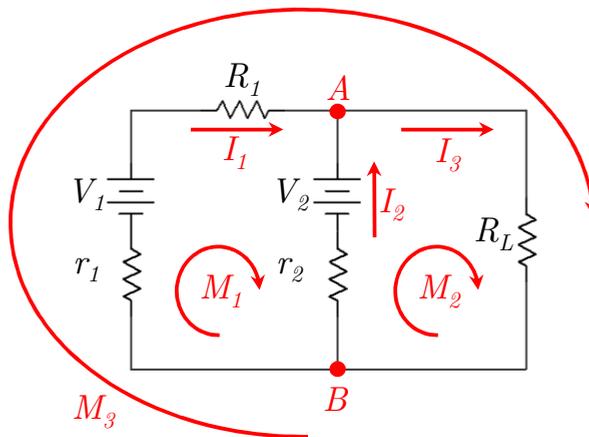


Figura II.11: Corrientes, nodos y mallas en el circuito.

Respecto a los nodos, tenemos dos, A y B . Si aplicamos la LCK en ellos (la suma de las corrientes que entran a un nodo debe ser igual a la suma de las corrientes que salen), encontraremos las siguientes ecuaciones:

- Nodo A : $I_1 + I_2 = I_3$
- Nodo B : $I_3 = I_1 + I_2$

Como vemos, se trata de la misma ecuación, por lo que de la LCK en este caso únicamente podemos obtener una ecuación, es decir, bastaba con que la hubiésemos aplicado a uno de los dos nodos. Las otras dos ecuaciones podemos obtenerlas aplicando, por ejemplo, la LK a las mallas M_1 y M_2 , recorriéndolas en el sentido de las agujas del reloj, por ejemplo:

- Malla M_1 : $V_1 - V_2 = I_1 r_1 + I_1 R_1 - I_2 r_2$
- Malla M_2 : $V_2 = I_2 r_2 + I_3 R_L$

Podríamos haber obtenido una ecuación adicional considerando la malla M_3 ; sin embargo, la ecuación resultante sería combinación lineal de las obtenidas a partir de M_1 y M_2 , por lo que no sería una ecuación independiente. En todo caso, ya tenemos un número suficiente de ecuaciones para resolver el problema, con los valores indicados para las resistencias y fuentes de voltaje empleadas sería:

$$I_1 + I_2 - I_3 = 0$$

$$6I_1 - I_2 = 2$$

$$I_2 + 2I_3 = 4$$

Que como resultado nos da $I_1 = 0.5$ A, $I_2 = 1$ A, $I_3 = 1.5$ A.

7 TEOREMAS DE THÉVENIN Y NORTON

Los teoremas de Thévenin y Norton¹⁰ son muy útiles a la hora de simplificar los cálculos sobre un circuito. Un ejemplo típico es cuando se nos pide determinar cuánto vale la corriente a través de una determinada resistencia y cuánto valdría la corriente si esa misma resistencia es reemplazada por un cortocircuito.

¹⁰Existen además otros teoremas que pueden ser útiles en la resolución de problemas de circuitos, por ejemplo el teorema de superposición, o el teorema de Millman, pero nosotros nos centraremos en los dos mencionados por su particular utilidad práctica.

En vez de resolver dos veces el circuito, podemos simplificar considerablemente los cálculos gracias a estos circuitos equivalentes sencillos. La idea general es convertir un circuito relativamente complejo (que puede tener varias resistencias, fuentes de voltaje o corriente, ramas, etc.) en uno sencillo, compuesto solamente por una fuente (de voltaje en el caso de Thévenin y de corriente en el caso de Norton) y una resistencia (en serie o en paralelo para Thévenin y Norton respectivamente), al cual le acoplamos el elemento que puede variar (en nuestro ejemplo la resistencia que era reemplazada por un cortocircuito).

Teorema de Thévenin

El teorema de Thévenin¹¹ dice que cualquier circuito lineal activo (con fuentes), con terminales de salida A y B, **puede sustituirse por una sola fuente de voltaje ideal, en serie con una resistencia equivalente**, de modo que:

1. El voltaje o fuerza electromotriz de la fuente equivalente (**voltaje Thévenin**) es igual a la diferencia de potencial entre los terminales A y B en circuito abierto.
2. La resistencia equivalente del circuito (**resistencia Thévenin**) es la resistencia entre los puntos A y B obtenida reemplazando cada fuente por su resistencia interna, si la tuviera.

A efectos prácticos, para determinar el equivalente Thévenin de un circuito entre dos puntos A y B se sigue el siguiente **procedimiento**:

- Determinar el voltaje Thévenin: se trata de calcular cuánto vale $V_A - V_B$ utilizando el análisis de mallas. Para ello debemos tener en cuenta que la diferencia de potencial entre dos puntos se calcula como:

$$V_A - V_B = \Sigma V_{baterias} - \Sigma V_{resistencias} \text{ (en el camino desde B hasta A)}$$

- Determinar la resistencia Thévenin: las fuentes de voltaje en el circuito se cortocircuitan (dejando únicamente su resistencia interna si la tuvieran) y las fuentes de corriente se eliminan (dejando únicamente su resistencia interna si la tuvieran). Una vez hecho esto, se calcula cuánto vale la resistencia entre A y B a partir del análisis de resistencias en serie y paralelo.

¹¹Debido a Léon Charles Thévenin (1857–1926).

Teorema de Norton

El teorema de Norton¹² dice que cualquier circuito lineal activo (con fuentes), con terminales de salida A y B, **puede sustituirse por una sola fuente de corriente ideal, en paralelo con una resistencia equivalente**, de modo que:

1. La corriente de la fuente equivalente (**corriente Norton**) es igual a la corriente que circularía entre los terminales A y B si estuvieran cortocircuitados.
2. La resistencia equivalente del circuito (**resistencia Norton**) es la resistencia entre los puntos A y B obtenida reemplazando cada fuente por su resistencia interna, si la tuviera. Es decir, **se calcula igual que la resistencia Thévenin** (de hecho, ambas resistencias son iguales).

A efectos prácticos, para determinar el equivalente Norton de un circuito entre dos puntos A y B se sigue el siguiente **procedimiento**:

- Determinar la corriente Norton: se trata de calcular cuánto vale la corriente entre A y B cuando están cortocircuitadas (unidas por un cable). Para ello empleamos el análisis de mallas.
- Determinar la resistencia Norton: se calcula exactamente igual que en el caso de la resistencia Thévenin.

Volvamos al ejemplo considerado con anterioridad en la figura II.10. Imaginemos que ahora nos piden determinar la corriente que circula por R_L si en vez de tomar valor de $2\ \Omega$ toma valor de $3\ \Omega$. Podríamos recurrir al análisis de mallas de forma análoga a como hicimos anteriormente, obteniendo un nuevo sistema de ecuaciones a resolver. El procedimiento sería un tanto tedioso cada vez que cambiásemos el valor de R_L , por lo que conviene recurrir a uno de los teoremas que acabamos de presentar, por ejemplo, el teorema de Thévenin. El primer paso por tanto es obtener un circuito equivalente sencillo para todo el circuito salvo la resistencia R_L , es decir, la parte correspondiente a la figura II.12.

En la figura II.12 también se han indicado las mallas (en este caso una única malla) y la única corriente que habría en ese caso. Se ha respetado la notación A y B para indicar los dos puntos entre los que calcularemos el equivalente Thévenin,

¹²Debido a Edward Lawry Norton (1898-1983).

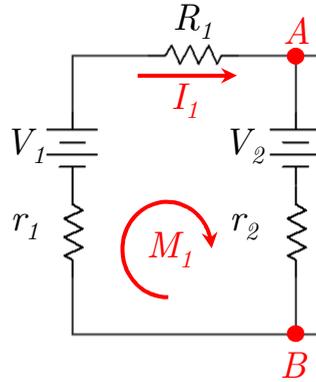


Figura II.12: Parte del circuito para la cual calcularemos su equivalente Thévenin.

aunque debe tenerse en cuenta que en este caso particular ahora no se trata de nodos, por lo que en ellos no se aplicará la LCK.

El siguiente paso es calcular, en primer lugar, el voltaje Thévenin. Para ello tenemos en cuenta que la LVK en la malla M_1 quedará como:

$$V_1 - V_2 = I_1 r_1 + I_1 R_1 + I_1 r_2$$

Con los valores considerados, se reduce a

$$2 = 7I_1$$

Por lo que tendremos que $I_1 = \frac{2}{7}$ A. Con este valor de corriente podremos calcular el voltaje Thévenin. Para ello, debemos recorrer el camino desde B hasta A , evaluando las ganancias y pérdidas en el mismo.

Así, si vamos de B hasta A por V_2 y r_2 , tendremos que:

$$V_{TH} = V_A - V_B = (+V_2) - (-I_1 r_2) = V_2 + I_1 r_2 = 4 + \frac{2}{7} \cdot 1 = \frac{30}{7}$$

Es decir, $V_{TH} = \frac{30}{7} = 4.29$ V.

Para calcular la resistencia equivalente Thévenin reemplazamos todas las fuentes de alimentación por sus resistencias internas. De este modo, tendremos que entre A y B las resistencias que nos quedarán serán las mostradas en la figura II.13.

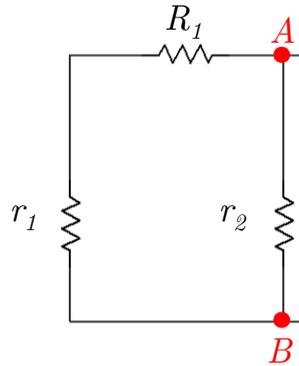


Figura II.13: Cálculo de la resistencia equivalente Thévenin.

Como puede comprobarse, tenemos por una parte a las resistencias r_1 y R_1 en serie, y el conjunto de estas dos, en paralelo con r_2 . En este caso particular la resistencia Thévenin valdrá:

$$R_{TH} = \left(\frac{1}{r_1 + R_1} + \frac{1}{r_2} \right)^{-1}$$

Que con los valores de nuestro problema será $R_{TH} = 0.86 \Omega$.

Por lo tanto, el equivalente Thévenin completo será el mostrado en la figura II.14.

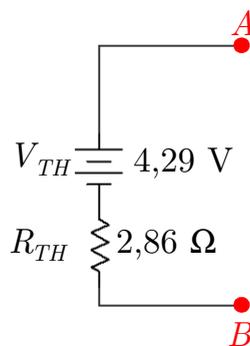


Figura II.14: Equivalente Thévenin del subcircuito considerado entre A y B .

Si a este circuito le acoplamos entre A y B una resistencia R_L es fácil comprobar que la corriente por dicha resistencia valdría:

$$I = \frac{V_{TH}}{R_{TH} + R_L}$$

De este modo, es sencillo calcular dicha corriente si cambia el valor de R_L o, por ejemplo, si la reemplazamos por un cortocircuito (que equivaldría a tener $R_L = 0$). El problema se podría haber resuelto de forma completamente análoga empleando el equivalente Norton, sin más que tener en cuenta el método para calcular la fuente de corriente equivalente en ese caso.

8 CIRCUITOS DE CORRIENTE ALTERNA

Hasta ahora, nos hemos centrado en circuitos de corriente continua o DC, en los que las fuentes de alimentación proporcionaban valores de voltaje o corrientes constantes en el tiempo. Este es el caso de mayor interés para el problema que nos ocupa, el funcionamiento de un ordenador. Si bien es verdad que en sentido estricto los ordenadores funcionan controlados por señales variables (el reloj interno de un ordenador no es más que una señal de tipo cuadrado, que varía entre dos valores de voltaje con una cierta frecuencia), para comprender el tratamiento de las señales digitales que realizan los microprocesadores nos basta con estudiar qué ocurre a cada elemento del circuito en uno de los dos posibles estados de operación, en definitiva, como si operase en condiciones de corriente continua en uno de ambos estados.

Sin embargo, por su importancia especialmente en el campo de las telecomunicaciones y la adquisición de señales, vamos a introducir brevemente algunos conceptos relacionados con los circuitos de corriente alterna.

Como ya hemos indicado anteriormente, los circuitos de corriente alterna o AC son circuitos que trabajan con señales eléctricas que varían de forma periódica en el tiempo, típicamente con señales de tipo senoidal, que ya presentamos en el capítulo I cuando discutimos el fenómeno de la inducción.

En este tipo de señales, si aplicamos por ejemplo un voltaje oscilante¹³ (por ejemplo de la forma $V = A \cos(\omega t)$) a los extremos de una resistencia, la respuesta en corriente también será de tipo oscilante, en fase con la señal aplicada¹⁴.

¹³La fuente de voltaje en alterna se representa por un círculo en cuyo interior hay dibujada una onda.

¹⁴Dos señales están en fase si, teniendo la misma frecuencia, sus máximos y sus mínimos coinciden en los mismos instantes temporales.

Es decir, a partir de la ley de Ohm la corriente que circulará por la resistencia será simplemente:

$$I = \frac{V}{R} = \frac{A}{R} \cos(\omega t)$$

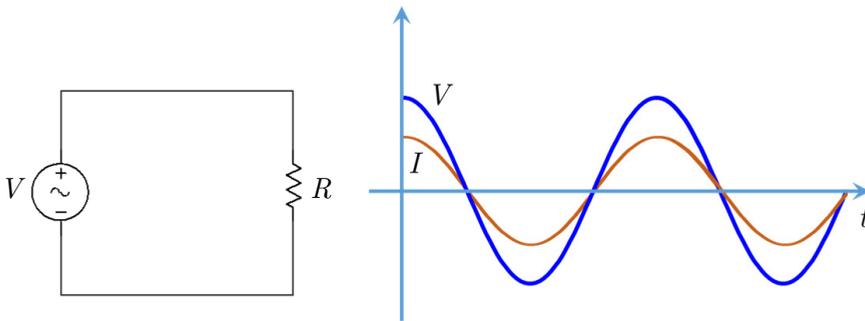


Figura II.15: Circuito AC compuesto por una fuente de voltaje en alterna y una resistencia. A la derecha se muestra la evolución temporal del voltaje aplicado y de la corriente en el circuito.

En la figura II.15 se muestra el circuito y la respuesta temporal del voltaje aplicado y la corriente que circularía por la resistencia. Ambas señales están en fase, aunque lógicamente su amplitud sea distinta por tratarse de magnitudes diferentes.

Pero si en vez de colocar una resistencia en el circuito colocásemos un condensador (figura II.16), la situación variaría notablemente. En un circuito de corriente continua, en situación estacionaria no podría haber corriente en el circuito, dado que el aislante del condensador impide que haya movimiento de carga eléctrica entre sus placas. En un circuito AC la situación es distinta. La señal externa oscilante provoca una fluctuación de las cargas en los cables que se traduce en variaciones de la carga almacenada en el condensador, de modo que si bien sigue sin ser posible el paso de electrones de una placa a otra del condensador por la presencia del aislante, sí existe una corriente en alterna en el circuito.

Es posible calcular la corriente teniendo en cuenta la definición de corriente y la de capacidad del condensador, así como el circuito en particular que estamos considerando. La corriente, recordemos, es la variación de carga por unidad de tiempo. En su definición más estricta, sería:

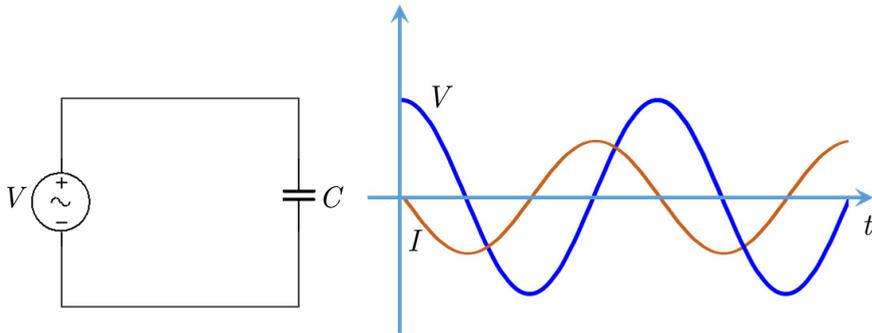


Figura II.16: Circuito AC compuesto por una fuente de voltaje en alterna y un condensador. A la derecha se muestra la evolución temporal del voltaje aplicado y de la corriente en el circuito.

$$I = \frac{dQ}{dt}$$

Por otra parte, la capacidad del condensador relaciona la carga almacenada en el mismo con la diferencia de potencial entre sus extremos V_C , es decir:

$$C = \frac{Q}{V_C}$$

En nuestro caso, dado que la fuente de voltaje AC está conectada directamente a los extremos del condensador, tendremos que $V_C = V = A \cos(\omega t)$. Por tanto, tendremos que:

$$Q = CV_C = CA \cos(\omega t)$$

En consecuencia, la corriente será:

$$I = \frac{dQ}{dt} = CA \frac{d(\cos(\omega t))}{dt} = -\omega CA \sin(\omega t) = \omega CA \cos(\omega t + \pi/2)$$

Es decir, aparte de la lógica diferencia en amplitud (que en este caso, a diferencia del caso de la resistencia, no sólo depende de la amplitud de la señal aplicada y del valor de C , sino también de la frecuencia angular de la señal ω), comprobamos cómo la corriente está **desfasada** respecto al voltaje de la fuente, en concreto está adelantada un cuarto de ciclo ($\pi/2$ radianes).

Si en vez de un condensador en el circuito hubiese una bobina (figura II.17), de nuevo la situación es diferente. En ese caso, el voltaje entre extremos de la bobina, V_L , debe coincidir con el aplicado por la fuente, V , y como ya vimos además deberá cumplirse que:

$$V_L = L \frac{dI}{dt}$$

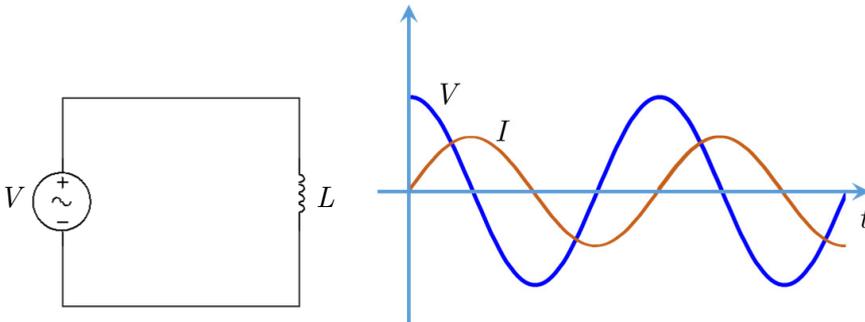


Figura II.17: Circuito AC compuesto por una fuente de voltaje en alterna y una autoinducción o bobina. A la derecha se muestra la evolución temporal del voltaje aplicado y de la corriente en el circuito.

En consecuencia la corriente en el circuito será:

$$I = \int_0^t \frac{V_L}{L} dt = \int_0^t \frac{A \cos(\omega t)}{L} dt = \frac{A}{\omega L} \sin(\omega t) = \frac{A}{\omega L} \cos(\omega t - \pi/2)$$

De nuevo podemos extraer dos conclusiones interesantes: por una parte, la amplitud de la corriente vendrá determinada por la de la fuente de voltaje, el valor del componente (L) y además, por la frecuencia de la señal ω . Por otra parte, la señal está de nuevo desfasada, en este caso con un retraso de cuarto de ciclo ($-\pi/2$ radianes).

En realidad, estos tres circuitos se pueden resolver de idéntica manera introduciendo un nuevo concepto, la **impedancia** Z . Este concepto generaliza el de resistencia en el sentido de oposición al paso de corriente, pero permite incluir también a otros elementos como condensadores o bobinas, sin más que definir adecuadamente su valor de Z . Así, todos los circuitos que acabamos de ver podrían resolverse con la ley de Ohm para los circuitos AC definida como:

$$Z = \frac{V}{I}$$

En los circuitos de corriente alterna, con frecuencia en vez de trabajar con funciones seno o coseno se trabaja con **notación fasorial**, es decir, las señales se representan como magnitudes correspondientes a números complejos (con parte real y parte imaginaria), por ejemplo de la forma:

$$V = Ae^{j(\omega t + \varphi)} = Ae^{j\omega t} e^{j\varphi} = \mathbf{A}e^{j\omega t} = \mathbf{A}(\cos(\omega t) + j \sin(\omega t))$$

Donde \mathbf{A} es un fasor, un número complejo de la forma $\mathbf{A} = Ae^{j\varphi}$. Así, por ejemplo, podríamos considerar que la señal aplicada $V = A \cos(\omega t)$ en nuestros ejemplos es la parte real de $\mathbf{A}e^{j\omega t}$, considerando que \mathbf{A} tiene un $\varphi=0$.

Empleando esta notación fasorial, resulta sencillo definir la impedancia de cada elemento. **La impedancia de una resistencia** coincide simplemente con su valor de resistencia, es decir:

$$Z_R = R$$

Conforme a la ley de Ohm, en el circuito de la figura II.15 tendríamos el resultado ya conocido:

$$I = \frac{V}{Z} = \frac{\mathbf{A}e^{j\omega t}}{R}$$

Quedándonos con la parte real del resultado, tendremos:

$$I = \frac{A}{R} \cos(\omega t)$$

En el caso de un condensador, su impedancia viene dada por:

$$Z_C = \frac{1}{j\omega C}$$

Aplicando la ley de Ohm para señales AC en el circuito de la figura II.16, tendremos:

$$I = \frac{V}{Z} = \frac{V}{\frac{1}{j\omega C}} = j\omega C \mathbf{A}e^{j\omega t} = \omega C \mathbf{A}(j \cos(\omega t) - \sin(\omega t))$$

Luego la parte real será el resultado que ya conocíamos:

$$I = -\omega CA \sin(\omega t) = \omega CA \cos(\omega t + \pi/2)$$

Finalmente, la **impedancia de una bobina** vendrá dada por:

$$Z_L = j\omega L$$

Respecto al circuito de la figura II.17, de nuevo si aplicamos la ley de Ohm para las impedancias:

$$I = \frac{V}{Z} = \frac{V}{j\omega L} = -j \frac{V}{\omega L} = -j \frac{\mathbf{A}}{\omega L} e^{j\omega t} = \frac{\mathbf{A}}{\omega L} (-j \cos(\omega t) + \sin(\omega t))$$

Cuya parte real se reduce, como ya sabíamos, a:

$$I = \frac{A}{\omega L} \sin(\omega t) = \frac{A}{\omega L} \cos(\omega t - \pi/2)$$

En definitiva, la notación fasorial y el concepto de impedancia nos permite tratar de forma relativamente sencilla los circuitos de corriente AC. Consideremos por ejemplo un circuito que contuviese los tres elementos estudiados, una resistencia, un condensador y una bobina conectados en serie a una fuente de voltaje en alterna, como se muestra en la figura II.18.

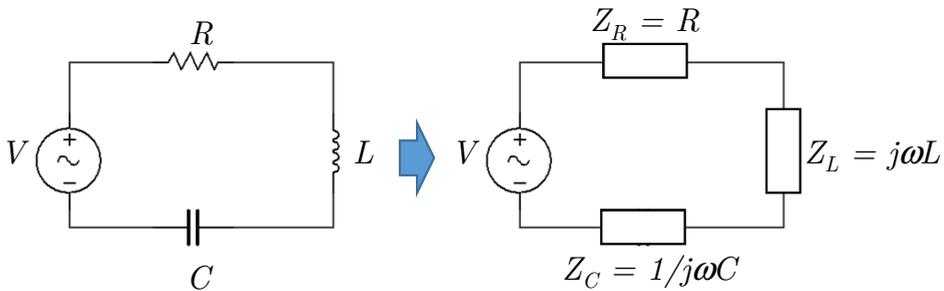


Figura II.18: Circuito AC compuesto por una fuente de voltaje en alterna, una resistencia, un condensador y una bobina conectados en serie. A la derecha se muestra el mismo circuito considerando sus impedancias.

Si reemplazamos los componentes pasivos por sus impedancias, tenemos un circuito muy simple con tres impedancias en serie, de modo que la impedancia total será:

$$Z_T = Z_R + Z_L + Z_C = R + j\omega L + \frac{1}{j\omega C} = R + j\omega L - j\frac{1}{\omega C} = R + j\left(\omega L - \frac{1}{\omega C}\right)$$

La corriente en el circuito será, por tanto:

$$I = \frac{V}{Z_T} = \frac{V}{R + j\left(\omega L - \frac{1}{\omega C}\right)}$$

Esta corriente tomará el máximo valor posible cuando el denominador (que depende de la frecuencia angular ω) tome su valor mínimo, lo cual ocurrirá cuando $\omega L - \frac{1}{\omega C} = 0$, o lo que es lo mismo, cuando se cumpla que:

$$\omega = \frac{1}{\sqrt{LC}}$$

Este es un circuito **resonante**, que ofrece la corriente máxima para una señal V oscilante con frecuencia lineal $f = \frac{\omega}{2\pi} = \frac{1}{2\pi\sqrt{LC}}$.

Los circuitos de corriente alterna tienen infinidad de aplicaciones, y su análisis puede realizarse con gran parte de las herramientas básicas que hemos presentado en este apartado. En el resto del texto, como ya hemos comentado, trabajaremos principalmente con circuitos de corriente continua.

III.

SEMICONDUCTORES

La electrónica nació a principios del siglo XX con los tubos de vacío. Durante la primera mitad de ese siglo, los aparatos electrónicos (incluidos los primeros ordenadores de propósito general) se fabricaron empleando este tipo de dispositivos, que eran voluminosos, consumían mucha energía eléctrica y eran muy frágiles. Fueron los tiempos de la **electrónica de vacío**.

El escenario cambió radicalmente tras la segunda guerra mundial a partir de la invención del transistor, basado en un tipo de materiales llamados **semiconductores**, que trataremos en este tema. La electrónica pasó a ser **electrónica de estado sólido**: los dispositivos ya no se fabricaban con una campana de vidrio en la que se había hecho el vacío, sino en un trozo de semiconductor sólido. Ello supuso una verdadera revolución que ha acabado transformando nuestra sociedad: todo ello ha ocurrido gracias a los semiconductores, que hacen posible los ordenadores, tablets y teléfonos móviles actuales.

1 EL ÁTOMO

Para entender qué es un semiconductor, en primer lugar debemos recordar algunos conceptos fundamentales respecto a la configuración interna de los átomos. Un **átomo**¹ es el constituyente más básico de la materia. En su interior contiene diversos tipos de partículas, repartidas entre el **núcleo** (constituido por **protones** y **neutrones**) y los **electrones** orbitando alrededor de éste. Los elec-

¹El concepto de átomo lo debemos a los antiguos griegos, en concreto a Demócrito. Demócrito especuló con que la materia estaba constituida por pequeñas partículas de tamaño minúsculo a las que denominó *átomos*. No sería hasta los experimentos de Thomson, a finales del siglo XIX, cuando se retomaría este concepto y se establecería definitivamente el camino para la comprensión de la naturaleza última de la materia.

trones son partículas con una masa muy pequeña², con carga negativa, mientras que los protones, mucho más masivos, tienen carga positiva, de igual valor en módulo a la carga del electrón (la carga elemental). Dado que en un átomo existe el mismo número de protones y electrones, globalmente el átomo posee una carga *neutra* (cargas positivas y negativas se compensan). Los neutrones son partículas con masa pero sin carga eléctrica. El número de protones (o electrones) que tenga un determinado átomo es lo que determina de qué tipo de elemento se trata: por ejemplo, un átomo de hidrógeno siempre contendrá un único protón (y un único electrón), mientras que un átomo de carbono se caracteriza por tener seis protones (y seis electrones). El número de protones o electrones que posee un átomo se denomina **número atómico**, *Z*. Dentro de cada tipo de elemento, sin embargo, pueden existir diferencias en cuanto al número de neutrones que contiene en su núcleo. A las diferentes variedades para un mismo elemento se las denomina **isótopos**³.

Sin embargo, desde el punto de vista de las propiedades electrónicas de los materiales lo que nos interesa es el comportamiento de los electrones moviéndose alrededor del núcleo. A principios del siglo XX se pensaba que los electrones orbitaban alrededor del núcleo de manera no muy distinta a como lo hacen los planetas alrededor del Sol⁴. Posteriormente, el desarrollo de la mecánica cuántica demostró que los electrones no orbitan alrededor del núcleo en trayectorias predecibles, sino que se encuentran en constante movimiento en determinadas regiones del espacio alrededor del núcleo, dependiendo de su energía: estas regiones reciben el nombre de **orbitales atómicos**.

Existen diferentes tipos de orbitales atómicos (véase la figura III.1), dependiendo de los niveles energéticos en los que se pueden encontrar los electrones ligados al núcleo atómico. Los niveles energéticos se numeran empezando por el 1, siendo el nivel más bajo de energía el correspondiente a los electrones más cercanos al núcleo, y se continúa en orden (niveles 2, 3, 4, etc.). En el nivel 1 hay

²La masa del electrón es igual a $9.1 \cdot 10^{-31}$ kg.

³Quizá el ejemplo más conocido de isótopo es el *carbono 14*. La mayor parte de átomos de carbono tienen 12 partículas en su núcleo, 6 protones y 6 neutrones, pero existe un isótopo del carbono que tiene 8 neutrones, y por tanto, 14 partículas en su núcleo. Este isótopo no es estable y tiende a desintegrarse con el tiempo, por lo que midiendo el porcentaje de carbono 14 que posee una muestra orgánica puede determinarse su antigüedad.

⁴A pesar de tratarse de una hipótesis incorrecta, la idea de que los electrones giraban en órbitas circulares con ciertas restricciones, debida a Niels Bohr (1885-1962), fue extraordinariamente útil para explicar determinadas observaciones experimentales y para permitir el posterior desarrollo de la teoría mecánico-cuántica.

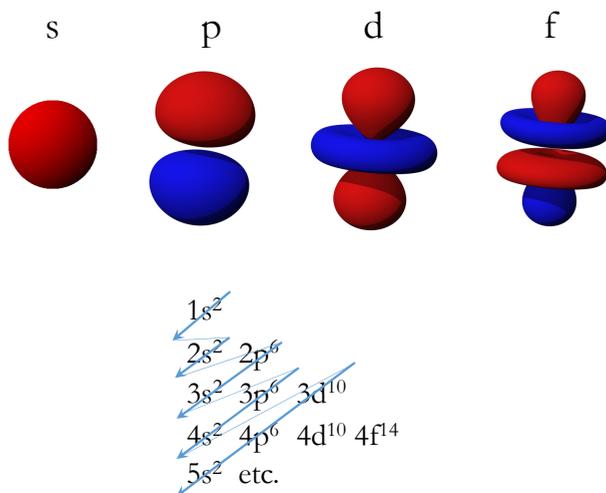


Figura III.1: Tipos de orbitales y configuración electrónica.

un único tipo de orbital, denominado orbital s, que tiene forma esférica. En el nivel 2, sin embargo, hay dos tipos de orbitales: un orbital s, con forma esférica⁵, y un orbital denominado p, con una forma similar a lóbulos unidos. En el nivel 3 hay un orbital tipo s, otro tipo p, y un tercer tipo d, y así sucesivamente. Los electrones van a ocupar estos orbitales disponibles pero con una condición: el número de electrones que caben en cada orbital está limitado. En los orbitales tipo s únicamente caben dos electrones, en los tipo p sólo caben seis, en los tipo d diez, etc. En este contexto, los electrones que contenga el átomo van a ocupar los diferentes niveles **empezando siempre por los niveles menos energéticos**.

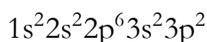
En este punto debemos imaginar los orbitales atómicos como apartamentos en un edificio sin ascensor. Los electrones tienden a ocupar primero los «apartamentos» del primer piso, pues evidentemente suponen un escenario energéticamente más favorable que subir a pisos más altos. De este modo, los apartamentos se van ocupando conforme al número de habitaciones que tienen y el piso en el que están, hasta que todos los electrones han encontrado su lugar⁶. Esta colocación ordenada de los electrones en los orbitales atómicos recibe el nombre de **configuración electrónica**: el orden de llenado es el indicado en la figura III.1,

⁵Podemos imaginar esta esfera como una esfera concéntrica, de radio mayor que la del nivel 1, y conteniendo a ésta.

⁶En este imaginario y particular «edificio», a medida que subiéramos de piso habría más apartamentos en cada planta, y el número de pisos sería idealmente infinito.

es decir, 1s-2s-2p-3s-3p-4s-3d-4p-5s...

Para determinar la configuración electrónica de un determinado elemento, debemos seguir el siguiente procedimiento. En primer lugar, debemos conocer el número atómico del elemento, que determina cuántos electrones hay que colocar en los orbitales siguiendo la secuencia que acabamos de describir. Imaginemos por ejemplo el caso del silicio, el material más comúnmente utilizado en la fabricación de microchips. El número atómico del silicio es 14, es decir, posee 14 electrones. Siguiendo el orden de llenado de orbitales, y considerando el número de electrones que caben en cada orbital, llegamos a que la configuración electrónica del silicio es:



Como vemos, en el último nivel hay 4 electrones (2 en el nivel 3s y 2 en el nivel 3p).

El número de electrones que se encuentren en el último nivel energético (los llamados **electrones de valencia**)⁷ es crítico a la hora de determinar las propiedades de los elementos. En la tabla periódica los elementos se agrupan conforme a su configuración electrónica de valencia (es decir, de su última capa o nivel energético). Así, por ejemplo, los elementos del grupo 13 (antiguamente llamado grupo III) tienen en común poseer tres electrones de valencia en su último nivel energético. O, por ejemplo, el carbono (C) y el silicio (Si) tienen en común el poseer cuatro electrones en su última capa, y por ello ambos pertenecen al grupo 14 (antiguo IV) de la tabla periódica (ver figura III.2).

2 EL SÓLIDO CRISTALINO

Cuando hablamos de las propiedades electrónicas de un determinado elemento normalmente no nos referimos a sus átomos aislados, en forma gaseosa. Como ya se ha comentado, para construir dispositivos electrónicos y microchips nos interesan las propiedades electrónicas de los materiales en su estado sólido: tan importante como la configuración electrónica de un determinado elemento es la disposición de sus átomos en el material solidificado. De manera general, podemos clasificar los sólidos en tres tipos: amorfos, policristalinos y monocristalinos (véase la figura III.3).

Los sólidos amorfos se caracterizan porque no existe una disposición regular y periódica de sus átomos en la estructura del material solidificado. Es decir,

⁷El concepto de *valencia atómica* proviene de la química, y originalmente designaba la capacidad o poder un elemento para combinarse con átomos de otro elemento o de su mismo tipo.

Grupo →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
↓ Período																			
1	1 H																		2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne	
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar	
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr	
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe	
6	55 Cs	56 Ba	57 La	* 72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn	
7	87 Fr	88 Ra	89 Ac	* 104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og	
				* 58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu		
				* 90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr		

Figura III.2: Tabla periódica de los elementos.

si examinásemos con un microscopio electrónico el interior del sólido, llegando al nivel atómico, no podríamos definir un patrón regular que se repitiese en regiones amplias del material. Tomando de nuevo como ejemplo el caso del silicio y sus usos en electrónica, en su estado sólido amorfo las aplicaciones son más bien escasas: se utiliza principalmente en las pequeñas células solares de las calculadoras, así como en ciertos tipos de transistores empleados en las pantallas planas.

Los sólidos policristalinos son aquellos en los que en el interior del material existen determinadas regiones dentro de las cuales hay patrones de ordenación de los átomos que se repiten en zonas de extensión reducida (típicamente de varias micras⁸). Sus propiedades electrónicas son mucho mejores que las del material amorfo; por ejemplo, en el caso del silicio, su versión policristalina⁹ es la más empleada hoy en día para la fabricación de células solares, por su excelente relación rendimiento/precio.

Finalmente, denominamos **sólido monocristalino** a aquel en cuyo interior la disposición de los átomos es perfectamente ordenada, siguiendo un patrón regular a lo largo y ancho de todo el material, conformando lo que se conoce

⁸La micra o micrómetro es la millonésima parte de un metro.

⁹En realidad la denominación del silicio empleado en la mayoría de las células solares comerciales es *multicristalino*, ya que suele tener regiones de mayor tamaño y con ciertas orientaciones privilegiadas por la forma en que se fabrican estos dispositivos.

como **red cristalina**. Las más sencillas son las mostradas en la figura III.4: la red cúbica simple, la cúbica centrada en el cuerpo y la cúbica centrada en las caras. Los sólidos monocristalinos son los que presentan mejores propiedades electrónicas: la totalidad de microprocesadores y memorias están fabricados a partir de obleas de silicio monocristalino, por citar un ejemplo.

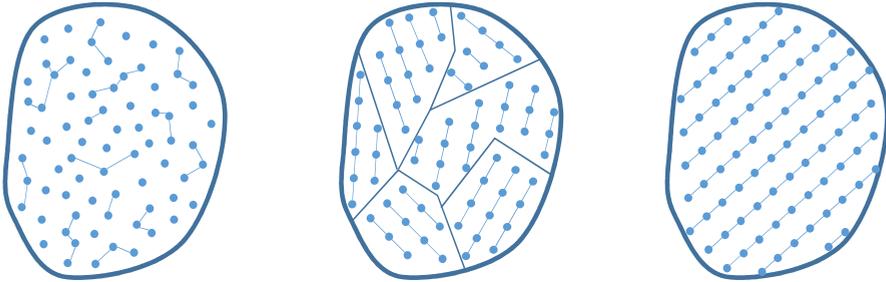


Figura III.3: Los materiales sólidos pueden clasificarse como amorfos (izquierda), policristalinos (centro) o monocristalinos (derecha).

Existen catorce tipos de estructuras cristalinas básicas, llamadas **redes de Bravais**¹⁰. Los materiales de mayor interés en electrónica (semiconductores como el silicio, el germanio o el arseniuro de galio) cristalizan en estructuras derivadas de la cúbica¹¹.

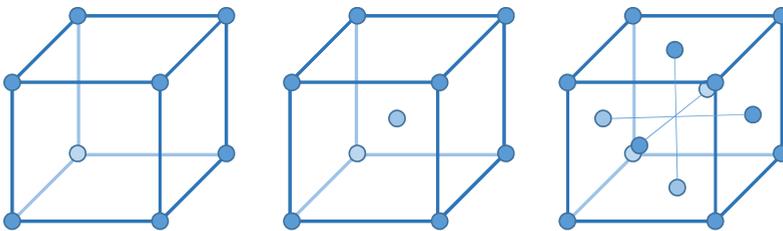


Figura III.4: Posibles redes cristalinas cúbicas: cúbica simple (izquierda), cúbica centrada en el cuerpo (centro) y cúbica centrada en las caras (derecha).

¹⁰Auguste Bravais (1811-1863), físico francés que realizó aportaciones clave en el campo de la cristalografía.

¹¹En concreto, es una estructura correspondiente a dos redes cúbicas centradas en las caras e interpenetradas, de modo que cada átomo está rodeado por cuatro vecinos más cercanos, a igual distancia.

3 BANDAS DE ENERGÍA

Retomemos cómo era la distribución de los niveles energéticos de, por ejemplo, un átomo de silicio. Como hemos visto anteriormente, en su última capa o nivel de valencia existían 4 electrones repartidos en los orbitales 3s y 3p. Pero, ¿qué ocurre cuando un átomo, en vez de estar aislado, se encuentra rodeado de otros átomos? En ese caso, los orbitales se ven alterados y ocurre un fenómeno muy interesante.

Imaginemos que dos átomos de silicio se encuentran próximos entre sí. En ese caso forman un sistema combinado, en el que existirán 8 electrones de valencia (4 por cada átomo). Esos 8 electrones inicialmente estaban en los niveles 3s y 3p de cada átomo, pero cuando un átomo está cerca de otro ocurre entra en juego un efecto de naturaleza cuántica, el **principio de exclusión de Pauli**¹², que impone que en un determinado sistema no puede haber más de dos electrones en cada nivel energético¹³. Por tanto, cuando los dos átomos forman un sistema combinado, los niveles energéticos, los orbitales, se ven alterados para respetar este principio. Se produce un **desdoblamiento de niveles**: en vez de mantenerse en el nivel energético inicial, los niveles 3s y 3p de cada átomo cambian ligeramente su energía para respetar el principio de exclusión.

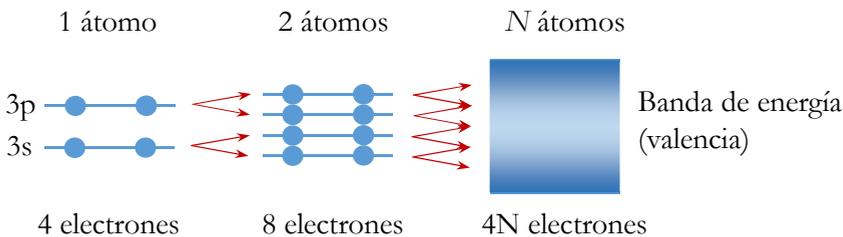


Figura III.5: Desdoblamiento de niveles para los electrones de valencia del silicio. Inicialmente, con un único átomo, tenemos dos niveles, 3s y 3p, siendo ligeramente más energético este último. Estos niveles se desdoblamiento hasta formar una banda de energía cuando hay un número de átomos elevado en el sistema.

Así, por ejemplo, lo que inicialmente era un nivel 3s por separado en cada átomo, al acercarse los átomos se convierten en dos niveles muy cercanos entre sí. Si en vez de tener 2 átomos tuviéramos N átomos, obtendríamos N niveles

¹²Wolfgang Ernst Pauli (1900-1958), uno de los pioneros de la mecánica cuántica y ganador del premio Nobel por este principio.

¹³Esos dos electrones, además, deben tener una propiedad denominada *spin* diferente entre sí.

desdoblados, con una separación muy pequeña entre ellos. Ese conjunto de niveles discretos (es decir, que no forman un continuo, sino que hay una pequeñísima separación energética entre ellos) conforman una **banda de energía**.

En la capa de valencia del silicio existían 4 electrones, y hasta 8 niveles disponibles (los 2 del orbital s y 6 de los orbitales p)¹⁴. Si pasamos a tener N átomos de silicio formando un sistema sólido cristalino, tendremos $4N$ electrones y $8N$ niveles correspondientes a las capas energéticas más externas. Para la distancia de separación entre átomos en silicio¹⁵ a_0 los niveles energéticos del conjunto se reparten entre los $4N$ más energéticos por un lado, formando una banda de energía, y los $4N$ menos energéticos por otro, formando otra banda (ver figura III.6). Entre ellos existe una zona energética en la que no existen niveles disponibles (es decir, que no pueden existir electrones con esas energías). ¿Dónde se situarán los electrones, qué posiciones energéticas ocuparán? Como siempre, la tendencia natural de los electrones es ocupar los estados de menor energía, por lo que los $4N$ electrones de valencia estarán en su totalidad en los $4N$ niveles más bajos. Los $4N$ niveles más altos conforman la **banda de conducción**, y los $4N$ niveles más bajos la **banda de valencia**. La zona de energías no permitidas entre las dos bandas, en las que no hay estados disponibles, recibe el nombre de banda prohibida, brecha o **gap**. Inicialmente la banda de conducción estará completamente vacía (son estados energéticos disponibles, susceptibles de ser ocupados, pero en principio vacíos) mientras que la banda de valencia estará completamente llena (totalmente ocupada por todos los electrones de valencia del sistema).

Este fenómeno (la formación de bandas de energía en un sólido cristalino) ocurre también en otros materiales: la diferencia respecto al caso del silicio será la separación entre las bandas de conducción y de valencia, es decir, el gap, que será diferente en cada material.

4 CONDUCTORES, AISLANTES Y SEMICONDUCTORES

El tamaño (en energía) del gap es un parámetro fundamental a la hora de determinar la capacidad de un material para conducir la corriente eléctrica. Recordemos que el gap es la zona de energías prohibidas para los electrones entre las bandas de conducción y de valencia. La energía del gap es de un valor muy

¹⁴Los orbitales p están a su vez compuestos por tres suborbitales, p_x p_y y p_z , cada uno con capacidad para dos electrones.

¹⁵La distancia de separación entre átomos de silicio en su forma sólida cristalina es de $2,34 \text{ \AA}$, es decir, $2,34 \cdot 10^{-10} \text{ m}$.

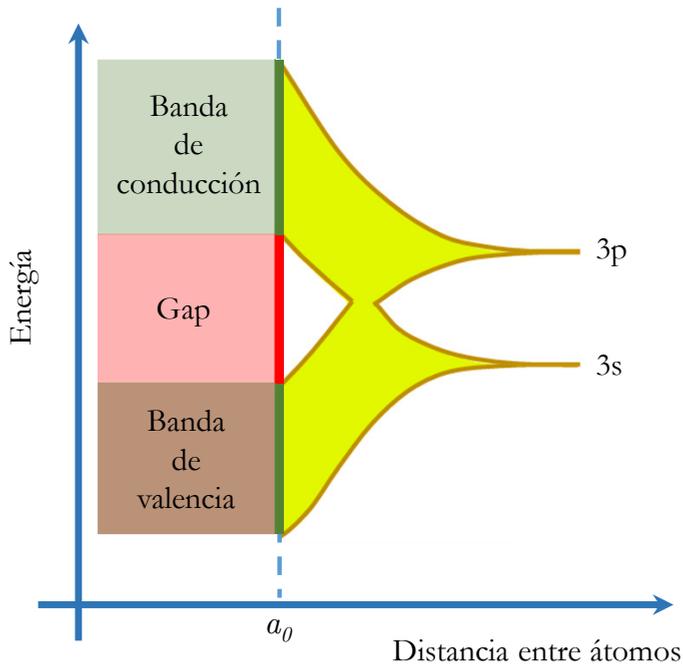


Figura III.6: Formación de las bandas de conducción y valencia en silicio. La distancia a_0 corresponde a la separación entre átomos del material sólido.

pequeño, y se suele medir en unidades de **electrón-voltio, eV**¹⁶. En el caso del silicio, su gap tiene un valor de aproximadamente 1,11 eV a temperatura ambiente¹⁷, esto es, a unos 300 grados Kelvin (300 K) aproximadamente.

Pero no todos los materiales tienen gaps de este valor. Podemos hacer una clasificación de los materiales atendiendo al valor de su gap, y así, tenemos:

Conductores Son aquellos materiales cuyas bandas de conducción y de valencia están **solapadas**. Es decir, su gap tiene valor nulo. En esta categoría entran por ejemplo los metales.

Aislantes En estos materiales el valor del gap es superior a unos 3 eV. La conducción de corriente eléctrica es prácticamente despreciable. En esta categoría entran ciertos óxidos empleados en electrónica. En ocasiones nos referiremos a ellos también como dieléctricos.

¹⁶Un electrón-voltio son $1,6 \cdot 10^{-19}$ julios.

¹⁷Esta energía corresponde a aproximadamente $1,778 \cdot 10^{-19}$ julios; el valor del gap puede cambiar ligeramente con la temperatura.

Semiconductores Son aquellos materiales cuyo valor del gap suele estar entre 0,5 y 2 eV. A temperaturas muy bajas se comportan como aislantes, pero a temperatura ambiente tienen una cierta capacidad de conducir la corriente eléctrica.

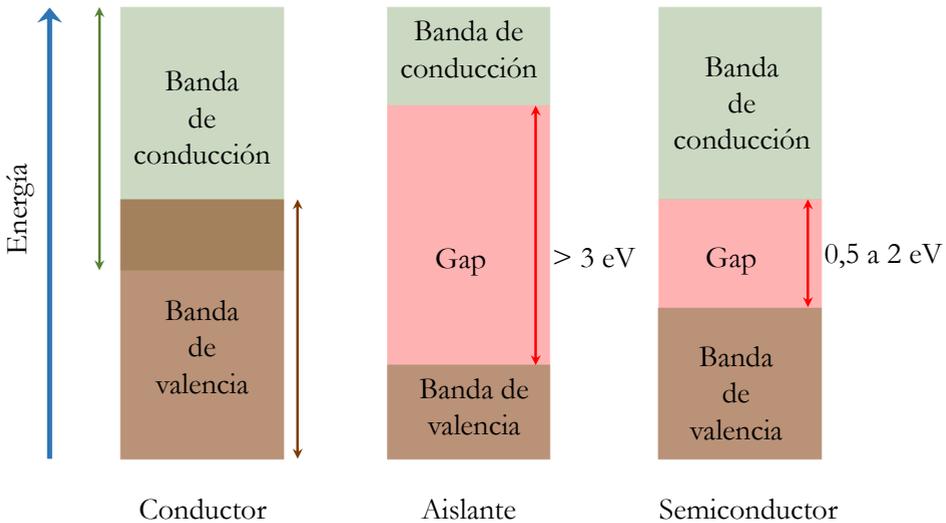


Figura III.7: Tipos de materiales según el valor del gap.

El valor del gap es un parámetro fundamental a la hora de hacer que un material pueda transportar corriente eléctrica. Pero, ¿qué es lo que posibilita realmente que se pueda producir ese proceso? Para conducir corriente eléctrica necesitamos esencialmente dos cosas: tener partículas con carga (electrones) que se puedan desplazar en el interior del material, es decir, disponer de **electrones libres**, y deben existir además **estados energéticos disponibles** que puedan ocupar esos electrones.

Volvamos ahora a la situación del silicio y su estructura de bandas. Para el caso de un sólido con N átomos de silicio cristalino se formaba una banda de valencia con $4N$ estados energéticos y $4N$ electrones: es decir, se trata de una banda completamente ocupada. Bajo esa premisa, los electrones en la banda de valencia no pueden contribuir al transporte de corriente eléctrica. Al aplicar un campo eléctrico al material (por ejemplo, aplicando una diferencia de potencial o voltaje entre sus extremos con una batería) los electrones no podrían variar su energía de manera neta, ya que no existen estados energéticos susceptibles de ser ocupados en su banda.

¿Y qué ocurre con la banda de conducción? En ese caso la banda de conducción estaba formada por $4N$ estados en los que no hay ningún electrón (todos estaban en la banda de valencia). Es por tanto una banda completamente vacía, por lo que al no existir carga en ella no hay posibilidad de transportar corriente.

En definitiva, el material se comportaría como un aislante, ya que al aplicarle un campo eléctrico no existiría corriente en el mismo¹⁸. Volvemos por tanto a la pregunta: *¿cómo se produce el transporte de carga eléctrica (corriente) en un material?*

La pregunta puede hacerse de manera análoga (salvando lógicamente las distancias) a si nos preguntásemos *¿qué es necesario para que una persona se mueva en una sala?* Si se trata de una sala completamente abarrotada, en la que hay tanta gente que no existe posibilidad de moverse (coloquialmente diríamos que estamos como sardinas en lata), resultaría evidente que nadie puede desplazarse, porque no hay espacios disponibles. Si por contra, la sala está totalmente vacía, ninguna persona puede desplazarse porque simplemente no hay nadie.

Imaginemos ahora el semiconductor como una sala dividida en dos partes, separadas por un muro bajo. En una parte de la sala hay una cierta cantidad de personas que abarrotan completamente su zona, sin dejar ningún espacio vacío. La otra parte de la sala, separada por el muro, está totalmente vacía. En esa situación seguiríamos estando en el caso en el que ninguna persona se desplaza. Pero, ¿y si en un momento dado una persona de la parte abarrotada es capaz de dar un salto, superar el muro, y llegar a la parte vacía de la sala? En ese caso podría tener lugar evidentemente el desplazamiento, ya que esa persona que ha saltado el muro estaría en una situación muy confortable, con una parte de la sala para sí sola. Y, ¿qué ha ocurrido con la parte de la sala que estaba abarrotada? Pues que, al haberse marchado una persona, el espacio que ocupaba queda disponible para que alguien pueda ocuparlo. La persona que ocupe ese espacio dejará tras de sí otro espacio, que podrá a su vez ser ocupado por otra persona, etc. Es decir, existe **un hueco**, un espacio que permite una cierta movilidad de las personas en esa parte de la sala.

Teniendo en cuenta las diferencias conceptuales con la analogía (en un semiconductor hablamos de estados energéticos y no de posiciones en el espacio), podemos ahora entender por qué tiene lugar el transporte de corriente en un semiconductor. A temperaturas cercanas al cero absoluto (0 K) ningún electrón tiene capacidad para *dar el salto* desde la banda de valencia hasta la banda de con-

¹⁸En los materiales aislantes, si aplicásemos un campo eléctrico muy elevado (dependiendo de cada tipo de material) se produciría la denominada *ruptura* del material, fenómeno por lo general no deseado.

ducción. A temperaturas por encima del cero absoluto, por ejemplo a temperatura ambiente (300 K), tiene lugar el fenómeno de **agitación térmica**, por el cual cualquier partícula sufre un movimiento errático a escala atómica, con mayor intensidad cuanto mayor sea la temperatura. Así, los electrones poseerán una cierta energía, denominada **energía térmica**, de valor:

$$\varepsilon_{th} = \frac{3}{2} K_B T \quad (\text{III.1})$$

donde ε_{th} es la energía térmica, K_B es la **constante de Boltzmann**¹⁹ y T es la temperatura expresada en grados Kelvin. Esta es una energía promedio, es decir, podrá haber electrones con energía superior o inferior. Unos pocos tendrán energía suficiente como para superar la barrera energética del gap y saltar de la banda de valencia a la de conducción, obteniéndose como resultado electrones en la banda de conducción y espacios energéticos libres en la banda de valencia. Por tanto, estaremos en condiciones de conducir corriente eléctrica si aplicásemos una diferencia de potencial en el material.

Es por ello que los materiales con un valor del gap intermedio reciben el nombre de semiconductores, ya que dependiendo de las condiciones (por ejemplo de la temperatura) se pueden comportar como un aislante o como un conductor (un mal conductor, para ser más precisos, pero conductor al fin y al cabo)²⁰.

La idea de un electrón saltando de la banda de valencia a la de conducción también puede interpretarse como un electrón que escape del sistema de enlaces, dejando tras de sí una vacante, como veremos posteriormente. El gap se puede interpretar por tanto como la energía necesaria para liberar a un electrón de valencia del sistema de enlaces del sólido.

Además del valor del gap, podemos catalogar a los materiales entre conductores, semiconductores y aislantes en función de su conductividad eléctrica, concepto que ya introdujimos en el capítulo I.

Los conductores poseen una conductividad muy elevada, típicamente superior a 10^6 $(\Omega \text{ m})^{-1}$ a temperatura ambiente. Los aislantes, inferior a 10^{-10} $(\Omega \text{ m})^{-1}$. Los semiconductores estarán entre ambas categorías, pero con una característica muy especial: su conductividad puede variarse dentro de un determinado

¹⁹Ludwig Eduard Boltzmann (1844-1906), físico austríaco. El valor de K_B es de $1,38 \cdot 10^{-23}$ J/K, o si se prefiere, $8,625 \cdot 10^{-5}$ eV/K.

²⁰Pensemos que, por ejemplo en el caso del silicio, a temperatura ambiente únicamente uno de cada catorce billones, aproximadamente, de electrones en la banda de valencia es capaz de dar el salto hasta la banda de conducción.

rango de valores mediante ciertos procesos tecnológicos²¹. Las conductividades típicas de los semiconductores suelen situarse entre 10^{-8} y 10^3 $(\Omega \text{ m})^{-1}$ a temperatura ambiente.

La conductividad eléctrica es un parámetro que varía con la temperatura, aunque lo hace de manera diferente en metales y en semiconductores. En el caso de los **metales**, al aumentar la temperatura la conductividad tiende a reducirse. En cambio, en el caso de los **semiconductores**, cuando la temperatura aumenta también lo hace la conductividad (véase la figura III.8).

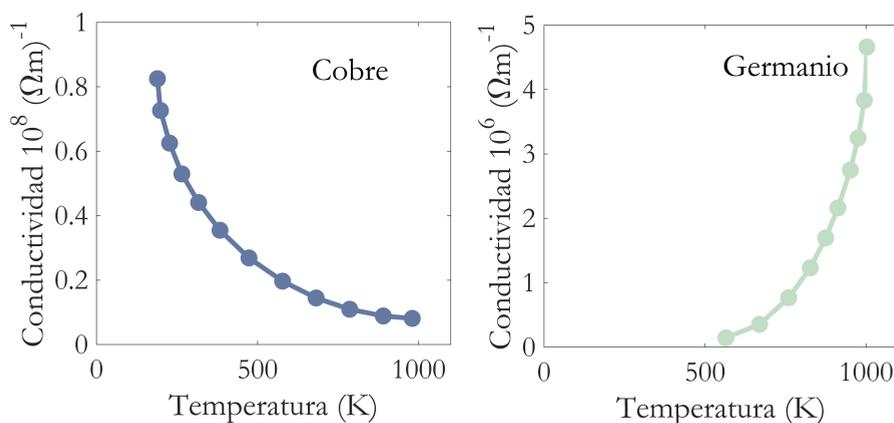


Figura III.8: Dependencia con la temperatura de la conductividad de un metal (cobre) y un semiconductor (germanio).

Para entender este comportamiento, debemos tener en cuenta los factores que afectan a la conductividad. Por una parte, es fundamental la cantidad de electrones libres de que dispongamos: cuantos más electrones, más carga, y por tanto, más corriente. Por otra parte, no debemos olvidar que la definición de corriente se refiere a la cantidad de carga que atraviesa una superficie por unidad de tiempo. Por tanto, no importa sólo el número de electrones, sino también lo rápido que puedan moverse en el interior del sólido.

Fijémonos, en primer lugar, en el caso de un metal. En un metal, se dispone de un número muy elevado de electrones capaces de desplazarse libremente en el interior del sólido. Al aumentar la temperatura, este número de electrones no aumenta. En cambio, debido al efecto de agitación térmica, los núcleos atómicos que conforman la red cristalina vibran con mayor intensidad cuanto mayor es

²¹Esta característica es la que permitirá fabricar los distintos tipos de dispositivos electrónicos que estudiaremos más adelante.

la temperatura, por lo que obstaculizan en mayor medida el movimiento de los electrones. Ello hace que disminuya la capacidad de conducir la corriente.

En un semiconductor también se produce este efecto y en consecuencia los electrones tienen mayor dificultad para desplazarse cuanto mayor es la temperatura. Pero al aumentar la temperatura, también aumenta el número de electrones libres, ya que un mayor número de ellos podrán saltar de la banda de valencia a la de conducción. Este incremento del número de electrones libres tiene mayor peso que el aumento de la dificultad para moverse, por lo que globalmente se observa un aumento de la conductividad.

5 SEMICONDUCTORES INTRÍNSECOS

Volvamos ahora al caso del silicio, que tomaremos como ejemplo en las siguientes páginas. Como ya se ha comentado, sus electrones de valencia son 4, teniendo 8 estados energéticos disponibles en ese último nivel. Cuando forma un sólido cristalino, cada átomo de silicio se rodea de otros cuatro átomos, compartiendo un electrón con cada uno de ellos y formando enlaces covalentes. En cierto modo podríamos pensar que tiende a completar su último estado energético de valencia, buscando una configuración parecida a la de un gas noble²².

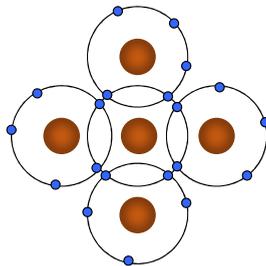


Figura III.9: Cada átomo de silicio se rodea de otros cuatro para formar la estructura de enlaces en el sólido.

Cuando nos encontramos ante un material semiconductor completamente puro decimos que se trata de un semiconductor **intrínseco**. En ese caso, ya sabemos que a temperatura cercana a 0 K (-273°C) se comporta como un material aislante porque todos sus electrones de valencia ocupan completamente esta banda, y no hay ningún electrón en la banda de conducción. Visto de otro modo,

²²Debe tenerse en cuenta que esta analogía no es del todo exacta, ya que la configuración electrónica en un sólido no es la de un átomo aislado, como ya hemos visto.

todos los electrones de valencia están formando parte de los enlaces covalentes del sólido y no hay electrones libres. Sin embargo, a temperatura ambiente, algunos electrones ganan suficiente energía de agitación térmica como para saltar a la banda de conducción: pensando en el sistema de enlaces, tendrían energía suficiente para liberarse de un enlace.

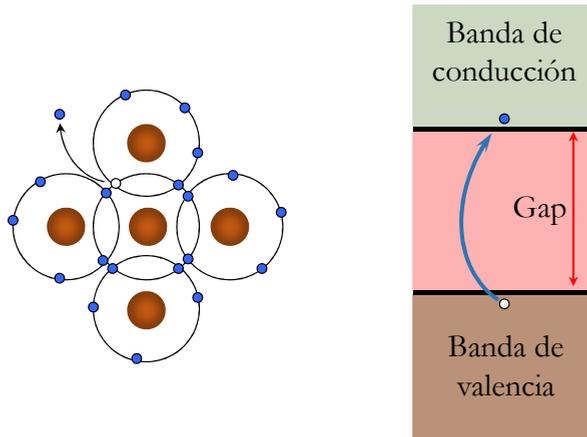


Figura III.10: Aparición de un hueco al liberarse un electrón de la banda de valencia.

Al escapar del enlace (saltar a la banda de conducción), el electrón deja tras de sí un espacio energético vacante en el sistema de enlaces (o si se quiere, una vacante en la banda de valencia), que vamos a denominar **hueco** (figura III.10). El resultado por tanto es que tendremos un nuevo electrón libre en la banda de conducción y un hueco en la banda de valencia: ambos podrán contribuir al transporte del corriente. El electrón, por motivos obvios, ya que al estar liberado del enlace podrá desplazarse por el interior del material cuando se aplique un campo eléctrico al mismo, moviéndose en sentido contrario al del campo eléctrico. Pero el hueco también contribuye al transporte de corriente. Al aplicar un campo externo, un electrón que esté en un enlace podrá saltar a ese hueco, dejando tras de sí otro hueco, que podrá ser ocupado por un nuevo electrón, que a su vez dejará otro hueco... y así sucesivamente. Es decir, habrá un movimiento de electrones en sentido contrario al campo, o, si se quiere, un movimiento del hueco en la misma dirección que el campo.

Por tanto, el comportamiento del hueco va a ser similar al de una partícula

parecida a un electrón **pero con carga positiva**²³. A la hora de describir el movimiento de cargas en los espacios vacantes de la banda de valencia es más sencillo considerar un hueco que se desplaza que múltiples electrones que se mueven sucesivamente (véase la figura III.11). Por tanto, trataremos al hueco como una partícula ficticia con carga $+q$.

Electrones en banda de conducción, huecos en banda de valencia. De manera general, hablaremos de **portadores de carga**, o simplemente **portadores**, para referirnos tanto a electrones como a huecos, indistintamente.

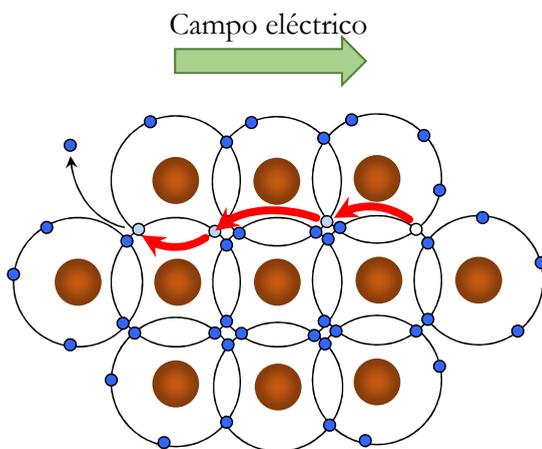


Figura III.11: La existencia de huecos también permite el transporte de carga eléctrica, por ocupación sucesiva de los mismos. Al liberarse un electrón (a la izquierda) y dejar una vacante, esta puede ir ocupándose sucesivamente por electrones que estaban en otros enlaces (movimientos indicados por las flechas), lo que también puede interpretarse como movimiento del hueco en el sentido opuesto al de las flechas, es decir, en el sentido del campo eléctrico aplicado.

En un semiconductor intrínseco (puro), existe el mismo número de electrones libres que de huecos: cada electrón de la banda de conducción ha dejado tras de sí un hueco en la banda de valencia, por lo que deben coincidir en número. A la densidad o concentración (cantidad de partículas por unidad de volumen) de portadores en esas condiciones la llamamos **concentración intrínseca**, n_i . En

²³No debe confundirse un hueco con un protón. El protón es una partícula real que forma parte del núcleo atómico, y que no va a desplazarse. El hueco es la ausencia de un electrón, que tratamos como una partícula ficticia.

electrónica típicamente se usa el cm^3 como unidad de volumen, por lo que las unidades de concentración serán cm^{-3} (partículas por centímetro cúbico).

La concentración intrínseca de portadores viene dada por la siguiente expresión:

$$n_i = A_0 T^{3/2} e^{\left(-\frac{\varepsilon_{gap}}{2K_B T}\right)} \quad (\text{III.2})$$

donde A_0 es una constante que depende del semiconductor en cuestión, T es la temperatura (en grados Kelvin), ε_{gap} es la energía del gap o banda prohibida, y K_B es la constante de Boltzmann.

Como vemos en la anterior expresión, la concentración intrínseca crece exponencialmente con la temperatura, y variará notablemente de unos semiconductores a otros, siendo clave el valor del gap del material, dado que cuanto menor sea, mayor será la concentración intrínseca. Este hecho hace que los semiconductores con un valor del gap relativamente pequeño, como el germanio, no sean demasiado adecuados para fabricar dispositivos, ya que presentarán de manera natural una conductividad demasiado alta y además funcionarán peor a temperaturas elevadas.

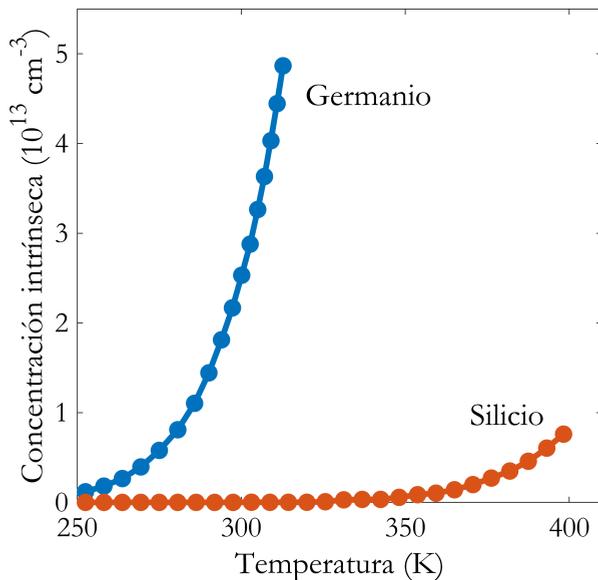


Figura III.12: La concentración intrínseca de portadores crece exponencialmente con la temperatura. El germanio posee un gap inferior al del silicio, por lo que el incremento de n_i con la temperatura es mayor en su caso.

La concentración intrínseca de portadores, que, recordemos, se debe a los electrones que por efecto de la agitación térmica pueden liberarse del sistema de enlaces, es muy pequeña en comparación con la concentración total de electrones de valencia disponible en el material. Para hacernos una idea, en silicio, a temperatura ambiente, únicamente uno de cada 13.8 billones de electrones es capaz de liberarse del sistema de enlaces y pasar a la banda de conducción. Por ello los semiconductores en su estado puro no son especialmente interesantes para fabricar dispositivos, ya que conducen muy mal la electricidad por el escaso número de portadores que poseen.

6 SEMICONDUCTORES EXTRÍNSECOS

El interés de los semiconductores no está en su versión pura o intrínseca. De ser así, los semiconductores no pasarían de ser materiales muy malos conductores de la electricidad, con escasas aplicaciones prácticas. El verdadero potencial de los semiconductores se desata cuando realizamos una serie de procesos tecnológicos en ellos para modificar sus propiedades conductoras. Así, decimos que tenemos un **semiconductor extrínseco, dopado o impurificado** cuando le *be-mos incorporado, de manera controlada, cierta cantidad de átomos de otros elementos*.

Dependiendo del tipo de elemento que incorporemos, vamos a tener dos posibilidades, los semiconductores extrínsecos **tipo n** y los semiconductores extrínsecos **tipo p**.

Semiconductor tipo n

Consideremos una muestra de silicio puro. Supongamos ahora que somos capaces de reemplazar un átomo de silicio por un átomo de un elemento del grupo 15 (por ejemplo, fósforo, arsénico o antimonio), que llamaremos **átomo de impureza** o simplemente, **impureza**²⁴. Este átomo tendrá cinco electrones de valencia. Al colocarse en la red cristalina, cuatro de esos electrones formarán enlaces covalentes con los cuatro átomos de silicio que rodean al átomo de impureza. El electrón en exceso quedará débilmente ligado al átomo de impureza²⁵, de modo que, a temperatura ambiente, la energía de agitación térmica le permitirá liberarse y pasar a ser un electrón de la banda de conducción. Así, a temperatura ambiente, cada impureza contribuye aportando un electrón libre, por lo que este

²⁴Es importante que la impureza reemplace a un átomo de silicio en su posición de la red cristalina. En ese caso diremos que la impureza está activada.

²⁵La energía de ligadura ε_D suele ser del orden de unas pocas decenas de meV.

tipo de impurezas reciben el nombre de **impurezas donadoras**. De este modo podemos regular la cantidad de electrones libres en el material dentro de un cierto rango de valores.²⁶

Cuando la impureza cede (dona) su electrón extra, la impureza queda **ionizada**, dado que pierde un electrón y por tanto queda como un ión cargado positivamente que no puede contribuir al transporte de corriente, porque forma parte de la red cristalina. Sin embargo, sí tiene una influencia importante en el transporte de carga, dado que al ser un ión cargado eléctricamente, actúa como un centro coulombiano que puede desviar la trayectoria de los portadores. La concentración de impurezas donadoras la denotaremos por N_D , y la de impurezas donadoras ionizadas por N_D^+ . A temperatura ambiente, ambas coincidirán.

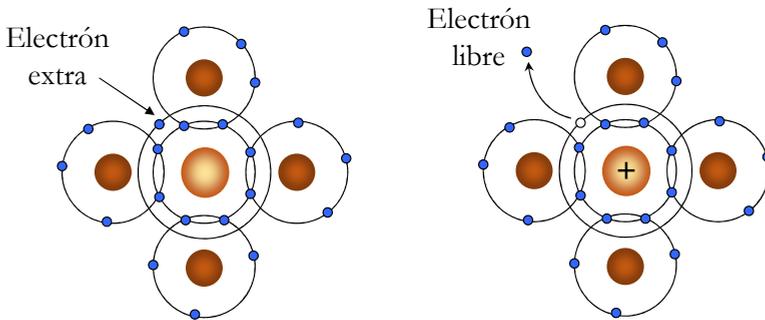


Figura III.13: Semiconductor extrínseco tipo n. La impureza donadora (en el centro) permite disponer de un electrón libre extra. Dicha impureza queda cargada positivamente (impureza ionizada, a la derecha de la imagen) al liberarse el electrón extra.

Semiconductor tipo p

Otra posibilidad diferente es incorporar un átomo de un elemento del grupo 13 (típicamente boro) que reemplace a un átomo de silicio. Al tener únicamente tres electrones de valencia, existe un defecto en uno de los enlaces, en el que hay un electrón ausente. Este defecto es un estado energético muy atractivo²⁷ que, a temperatura ambiente, será ocupado por otro electrón del sistema de enlaces,

²⁶Típicamente las concentraciones de impurezas varían entre 10^{15} y 10^{20} cm^{-3} ; si se empleasen concentraciones mayores el material pasaría a ser una aleación, que podría perder sus propiedades semiconductoras.

²⁷De nuevo, existe una energía de ligadura asociada, ε_A , de pocas decenas de meV.

de modo que tras de sí generará un **hueco**. La impureza en ese caso recibe el nombre de **impureza aceptadora**: cuando capture un electrón (generando un hueco adicional) la impureza quedará **ionizada**, cargada negativamente en este caso. Al igual que para las impurezas donadoras, la impureza aceptadora ionizada no contribuye al transporte de corriente por ocupar una posición fija en la red cristalina (sí lo hace en cambio el hueco adicional que ha aparecido), pero es un centro de dispersión muy eficiente respecto al movimiento de los portadores. La concentración de impurezas aceptadoras la denotaremos por N_A , y la de impurezas aceptadoras ionizadas por N_A^- . De nuevo ambas coincidirán a temperatura ambiente.

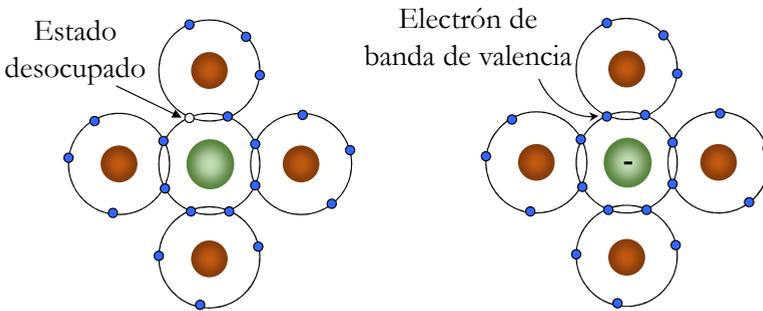


Figura III.14: Semiconductor extrínseco tipo p. La impureza aceptadora (en el centro) permite disponer de un hueco extra al capturar un electrón de la banda de valencia. Dicha impureza queda cargada negativamente (impureza ionizada, a la derecha de la imagen) al realizarse esta captura.

7 LEY DE NEUTRALIDAD DE CARGA

En el material semiconductor, bien sea intrínseco o extrínseco, necesariamente debe cumplirse que el número total de cargas positivas sea igual al número total de cargas negativas, es decir, debe respetarse la condición de neutralidad eléctrica. Supongamos el caso más general posible, en el que tuviésemos tanto impurezas donadoras como aceptadoras.²⁸ En ese caso, las cargas negativas en el sistema serán, por una parte, los electrones libres (en concentración n) y las impurezas aceptadoras ionizadas (que quedaron cargadas negativamente al captar un elec-

²⁸Se trataría del caso de un semiconductor **compensado**, aunque no es lo más habitual. Usualmente tendremos impurezas de un tipo o del otro.

trón del sistema de enlaces, con concentración N_A^-). Por otra parte, las cargas positivas en el material serán los huecos de la banda de valencia (con concentración p) y las impurezas donadoras ionizadas (que quedaron cargadas positivamente al ceder su electrón extra, con concentración N_D^+). Como en conjunto el cristal semiconductor debe ser neutro, tendremos que:

$$n + N_A^- = p + N_D^+ \quad (\text{III.3})$$

Que es lo que conocemos como ley o condición de **neutralidad de carga**.

8 LEY DE ACCIÓN DE MASAS

Por otra parte, en cualquier semiconductor, bien sea intrínseco o extrínseco, se da también una relación entre la cantidad de electrones y huecos que tenemos, a través de la denominada **ley de acción de masas**:

$$n \cdot p = n_i^2 \quad (\text{III.4})$$

En el caso de un semiconductor intrínseco, esta relación es directa dado que $n = p = n_i$. En un semiconductor extrínseco debemos tener en cuenta que el número de electrones es diferente del número de huecos, siendo mayores unos u otros dependiendo del tipo de dopaje.

Para un semiconductor tipo n, las impurezas donadoras han aportado electrones extra, y por tanto los portadores **mayoritarios** son los electrones, y los **minoritarios** los huecos, por lo que siempre se cumplirá que:

$$n > n_i > p \quad (\text{III.5})$$

En el semiconductor tipo p la concentración de huecos se reduce respecto a la intrínseca porque para los electrones en la banda de valencia se hace más complicado pasar a la banda de conducción, al estar ocupados en esta última más estados (por los electrones extra aportados por las impurezas).

En cambio, en un semiconductor tipo p, la situación es la contraria. Las impurezas aceptadoras han generado huecos extra, de modo que ahora los portadores mayoritarios son los huecos y los minoritarios los electrones, cumpliéndose que:

$$p > n_i > n \quad (\text{III.6})$$

En este caso la concentración de electrones se reduce respecto a la intrínseca porque las impurezas aceptadoras reducen enormemente las probabilidades de

que un electrón promocione a la banda de conducción (es mucho más probable que sean capturados por las impurezas).

A temperatura ambiente (300 K aproximadamente), en un semiconductor extrínseco la concentración de mayoritarios es prácticamente igual a la de impurezas incorporadas al material. A otras temperaturas esto puede no ser cierto. A baja temperatura puede que no todas las impurezas estén ionizadas, y a temperaturas muy elevadas la concentración intrínseca, que aumenta exponencialmente con T , puede tener un valor comparable a la densidad de impurezas y su contribución debe ser tomada en cuenta.

9 ARRASTRE

El número de portadores en un semiconductor es un factor importante, pero no debemos perder de vista que nuestro objetivo es construir dispositivos en los que estas cargas se muevan y den lugar a corrientes. Por tanto, los procesos de transporte de carga eléctrica serán fundamentales.

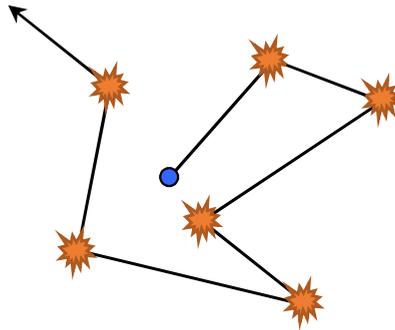


Figura III.15: En ausencia de campo aplicado, el movimiento de los portadores es aleatorio, y consiste en una sucesión de colisiones cada cierto tiempo, que reorientan a la partícula.

Imaginemos que disponemos de una muestra semiconductor homogénea y en equilibrio termodinámico, con temperatura uniforme a lo largo de toda ella. En ausencia de un campo eléctrico externo, los portadores se mueven debido a su energía de agitación térmica, que se traduce en una cierta velocidad para cada partícula, con módulo y orientación diferentes. En consecuencia, las partículas se desplazan pero de manera aleatoria, sufriendo cada cierto tiempo colisiones con los átomos que forman el sólido (figura III.15). El promedio de velocidad

de todo el conjunto de portadores es nulo: no existe un movimiento colectivo en el interior del material, como es lógico en ausencia de campo externo y con una muestra totalmente homogénea y a temperatura constante.

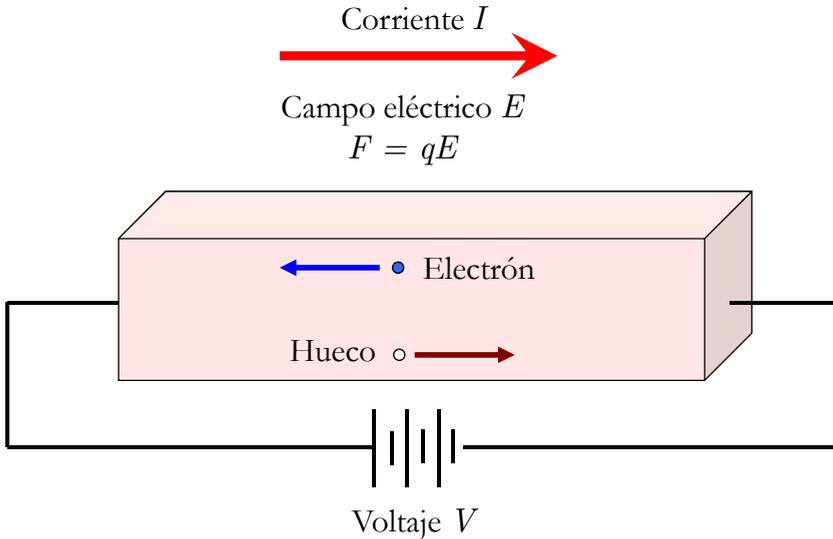


Figura III.16: Respuesta de los portadores libres ante la aplicación de un campo eléctrico: los electrones (carga negativa) sufrirán una fuerza que los desplazará en sentido contrario al del campo, mientras que los huecos (carga positiva) se desplazarán en el mismo sentido del campo.

Sin embargo, si aplicamos un campo eléctrico externo (por ejemplo, conectando una batería entre extremos del semiconductor), este campo ejercerá una fuerza sobre los portadores y producirá un movimiento colectivo. En el caso de los electrones, estos se desplazarán en sentido contrario al campo eléctrico, mientras que los huecos, al comportarse como si tuvieran carga positiva, se moverán en el mismo sentido que el campo aplicado. Aparecerá por tanto una corriente I en el semiconductor (figura III.16), con el mismo sentido que el campo eléctrico aplicado.

Microscópicamente, el movimiento de los portadores tiene lugar en una sucesión de vuelos libres y colisiones (figura III.17). El campo eléctrico provoca una aceleración de los electrones y los huecos, que ganan progresivamente velocidad y energía, hasta que sufren una colisión (bien sea con los átomos del semiconductor, con átomos de impurezas o con defectos en el material) que los frena. En consecuencia, para un cierto campo eléctrico dado los portadores alcanzan una

velocidad promedio, la **velocidad de deriva**, que podrá ser distinta para electrones y para huecos, y que es directamente proporcional al campo aplicado²⁹. La constante de proporcionalidad se denomina **movilidad**, y es la magnitud que mide la *facilidad que tienen los portadores para desplazarse en el interior del sólido cristalino*.

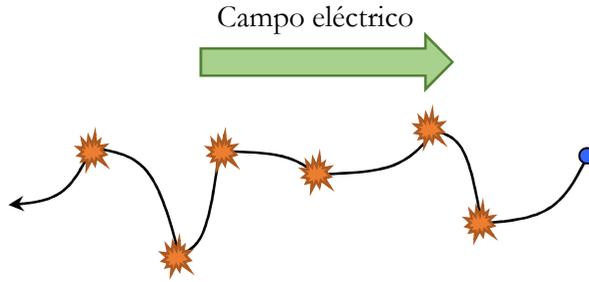


Figura III.17: Bajo la acción de un campo eléctrico, un electrón (o, análogamente, un hueco) sufre una sucesión de vuelos libres acelerados por el campo y colisiones (principalmente con la red cristalina).

Tendremos por tanto que las velocidades de deriva de electrones (v_n) y huecos (v_p) para un cierto campo aplicado (no excesivamente elevado) serán:

$$v_n = -\mu_n E \quad (\text{III.7})$$

$$v_p = \mu_p E \quad (\text{III.8})$$

El signo $-$ en la expresión de v_n da cuenta de que el movimiento de los electrones es en sentido opuesto al del campo eléctrico.

La movilidad de electrones y huecos va a depender esencialmente de dos factores. Por una parte, el tiempo promedio entre colisiones (**tiempo de vuelo libre**), τ_n o τ_p . Por otra parte, de la **masa efectiva** de los portadores. En un semiconductor, podemos describir el movimiento de los portadores durante los vuelos libres mediante la llamada **aproximación semiclásica**, que consiste en que la trayectoria del portador se puede calcular utilizando la segunda ley de Newton, pero asumiendo que la partícula se comporta con una masa diferente a la del

²⁹Esto es cierto únicamente si el campo no es demasiado elevado: a campos muy altos la velocidad en un semiconductor tiende en general a alcanzar un valor límite denominado **velocidad de saturación**.

vacío, precisamente la masa efectiva.³⁰

De este modo, las movilidades de electrones y huecos se pueden determinar mediante las siguientes expresiones:

$$\mu_n = \frac{q\tau_n}{m_n^*} \quad (\text{III.9})$$

$$\mu_p = \frac{q\tau_p}{m_p^*} \quad (\text{III.10})$$

Por regla general las movilidades de electrones y huecos son diferentes, lo cual tiene implicaciones importantes a la hora de fabricar dispositivos y circuitos electrónicos, en particular en el caso de la electrónica digital, donde se deben emplear transistores con transporte debido tanto a electrones como a huecos.

Como ya vimos en el capítulo I, si las partículas cargadas (electrones o huecos en nuestro caso) se mueven con una cierta velocidad promedio, generan una corriente que depende de la densidad de carga, la velocidad y la sección transversal del conductor. En nuestro caso esta corriente se debe a un campo eléctrico externo, y diremos que se trata de un proceso de **arrastre**, ya que los electrones son *arrastrados* por el campo. Calculando esta corriente de arrastre tanto para electrones como para huecos tendremos:

$$I_{an} = -qn v_n S = qn \mu_n E S \quad (\text{III.11})$$

$$I_{ap} = qp v_p S = qp \mu_p E S \quad (\text{III.12})$$

En definitiva, la corriente de arrastre total será:

$$I_a = I_{an} + I_{ap} = qn \mu_n E S + qp \mu_p E S = (qn \mu_n + qp \mu_p) E S \quad (\text{III.13})$$

La densidad de corriente por tanto será:

$$J_a = \frac{I_a}{S} = (qn \mu_n + qp \mu_p) E \quad (\text{III.14})$$

Por otra parte, en el capítulo I ya vimos que la ley de Ohm se podía expresar como

$$J = \sigma E \quad (\text{III.15})$$

³⁰En el caso del silicio, por ejemplo, la masa efectiva promedio para los electrones es $m_n^* = 0.25 m_0$, siendo m_0 la masa del electrón en reposo.

Por lo que a partir de aquí podemos obtener la expresión para la conductividad σ como:

$$\sigma = qn\mu_n + qp\mu_p = \sigma_n + \sigma_p \quad (\text{III.16})$$

Es decir, la conductividad total es la contribución de la conductividad debida a electrones más la contribución debida a huecos. Así, dependiendo del tipo de semiconductor tendremos diferentes situaciones.

En un **semiconductor intrínseco**, la cantidad de electrones y huecos coincide, de modo que $n = p = n_i$. Por tanto, la conductividad será:

$$\sigma = qn\mu_n + qp\mu_p = qn_i(\mu_n + \mu_p) \quad (\text{III.17})$$

En un **semiconductor extrínseco tipo n** la concentración de electrones es mayor que la de huecos. A temperatura ambiente³¹ la práctica totalidad de impurezas está ionizada, la concentración de minoritarios es despreciable frente a la de mayoritarios y se cumple que:

$$\sigma = qn\mu_n + qp\mu_p \simeq qn\mu_n \simeq qN_D\mu_n \quad (\text{III.18})$$

En un **semiconductor extrínseco de tipo p**, de forma análoga tendremos que a temperatura ambiente:

$$\sigma = qn\mu_n + qp\mu_p \simeq qp\mu_p \simeq qN_A\mu_p \quad (\text{III.19})$$

La principal característica de los semiconductores es, por tanto, que su conductividad es controlable (hasta un cierto punto) mediante el proceso tecnológico del dopado³².

10 DIFUSIÓN

Supongamos ahora que disponemos de una muestra de material semiconductor en la que la concentración de portadores no fuera homogénea, es decir,

³¹A temperaturas más elevadas debe tenerse en cuenta que la concentración intrínseca aumenta exponencialmente y no debe ocurrir necesariamente que la concentración de mayoritarios coincida con la de impurezas.

³²El dopado del semiconductor siempre mejora su conductividad, fundamentalmente debido al aumento de electrones o huecos según el tipo de dopaje. La movilidad se ve afectada por el dopaje, al aumentar el número de impurezas ionizadas, y por tanto se reduce. Sin embargo, es mayor la importancia del aumento de la concentración de portadores, por lo que globalmente la conductividad aumenta al dopar.

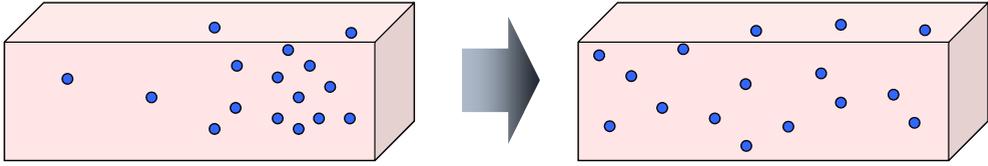


Figura III.18: Difusión de portadores en presencia de un gradiente de concentración. Inicialmente (izquierda) existe una mayor concentración de electrones en un lado de la muestra. Los portadores tienden de manera natural a repartirse homogéneamente, por lo que se produce una redistribución (derecha).

que existiesen regiones del material en las que concentración fuese más elevada que en otras. En esa situación, los portadores libres, de manera natural, tienden a distribuirse de manera uniforme, a repartirse de forma homogénea. Ello implica el movimiento de carga eléctrica y, en consecuencia, la aparición de una corriente debida a este fenómeno.

La **difusión** consiste en el *movimiento de redistribución de los portadores como consecuencia de la existencia de gradientes de concentración o de temperatura en la muestra*; es decir, si en una muestra semiconductora tenemos variaciones de estos parámetros, se produce el movimiento de electrones o huecos anteriormente mencionado. Junto con el arrastre, la difusión es el otro proceso de relevancia en el transporte de carga, y debe ser tenido también en cuenta a la hora de computar la corriente total.

Imaginemos una muestra semiconductora en la que la concentración de electrones variase con la posición, es decir, existiese un gradiente de n ³³. En ese caso, la corriente debida a la difusión de electrones viene dada por la llamada **ley de Fick**³⁴:

$$I_{dn} = qD_n \frac{dn}{dx} S \quad (\text{III.20})$$

donde D_n es el **coeficiente de difusión de electrones**. Al tener los electrones carga negativa la corriente fluye en sentido contrario al movimiento de difusión de los electrones, que tenderá a contrarrestar la existencia del gradiente. En consecuencia, la corriente tiene el mismo signo que $\frac{dn}{dx}$.

³³ Asumiendo que la variación se produce en una dimensión, podemos expresar el gradiente como $\frac{dn}{dx}$.

³⁴ En honor a Adolf Fick (1829-1901), fisiólogo alemán que estudió la difusión de gases a través de membranas.

En el caso de que hubiera un gradiente (variación) de la concentración de huecos p , tendríamos que:

$$I_{dp} = -qD_p \frac{dp}{dx} S \quad (\text{III.21})$$

donde D_p es el **coeficiente de difusión de huecos**. En este caso la corriente tiene el mismo sentido que el movimiento de los huecos, es decir, contrario al gradiente de concentración (los huecos se moverán hacia las zonas de menor concentración p para intentar repartirse uniformemente). Por tanto, mientras que en el caso del arrastre tanto electrones como huecos contribuían a la corriente en el mismo sentido, en el caso de la difusión la corriente tendrá un signo u otro dependiendo del tipo de portador involucrado.

Los coeficientes de difusión están relacionados con la movilidad de los portadores y la temperatura, a través de la denominada **relación de Einstein**³⁵:

$$D_n = \frac{K_B T}{q} \mu_n \quad (\text{III.22})$$

$$D_p = \frac{K_B T}{q} \mu_p \quad (\text{III.23})$$

Por tanto, tener una movilidad elevada o trabajar a una temperatura alta son factores que favorecen la difusión.

La corriente total por procesos de transporte se podrá calcular entonces, para electrones y para huecos, como:

$$I_n = I_{an} + I_{dn} = qn\mu_n ES + qD_n \frac{dn}{dx} S \quad (\text{III.24})$$

$$I_p = I_{ap} + I_{dp} = qp\mu_p ES - qD_p \frac{dp}{dx} S \quad (\text{III.25})$$

La corriente total en el material será simplemente la suma de ambas, $I = I_n + I_p$.

11 GENERACIÓN Y RECOMBINACIÓN DE PORTADORES

Además de los procesos de transporte (arrastre y difusión), debemos tener en cuenta otros tipos de procesos que pueden afectar a la cantidad de portadores

³⁵Debida a Albert Einstein (1879-1955).

libres que podemos tener en un material semiconductor, es decir, los procesos que afectan al **número** de electrones y huecos.

El proceso de **generación** se refiere al *paso de un electrón de banda de valencia a banda de conducción*, de forma que se genera un **par electrón-hueco**, es decir, que pasamos a tener dos nuevos portadores libres, un electrón en banda de conducción y un hueco en banda de valencia.

Por su parte, el proceso de **recombinación** se refiere a la *caída de un electrón desde la banda de conducción a la banda de valencia*³⁶, ocupando un hueco allí existente, es decir, que se produce la **aniquilación de un par electrón-hueco**, y perdemos dos portadores libres.

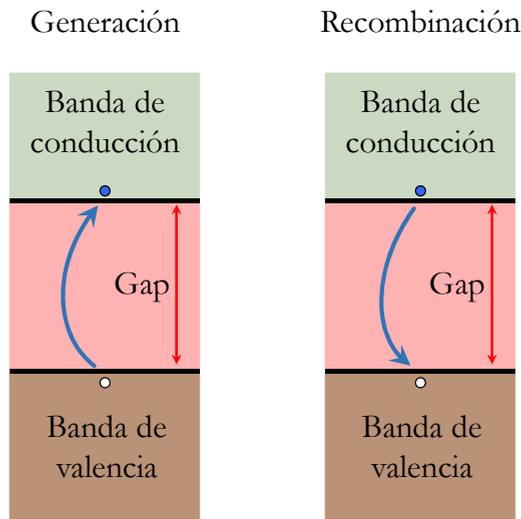


Figura III.19: Generación y recombinación de portadores.

Ambos fenómenos puede ser descritos por una tasa de generación (G) y una tasa de recombinación (R) respectivamente, que se refieren a la cantidad de portadores que se crea o se destruye en cada caso por unidad de volumen y de tiempo.

En un semiconductor en equilibrio, ambos procesos están compensados, es decir, se debe cumplir que $G = R$. Constantemente hay pares de portadores que se crean, permanecen un tiempo en su banda correspondiente y se aniquilan, siendo reemplazados por otros para que las poblaciones en banda de conducción y de valencia permanezcan constantes. El tiempo promedio que permanecen los

³⁶Podríamos decir también que este proceso ocurre cuando un electrón libre es recapturado por el sistema de enlaces.

portadores en la banda correspondiente se denomina **tiempo de vida media** (τ) y depende de factores como la calidad del cristal semiconductor y la cantidad de defectos que haya en el mismo.

Los procesos de generación y recombinación, además de ocurrir de manera natural por efectos térmicos (por el simple hecho de encontrarse a una temperatura superior al cero absoluto), también pueden tener lugar en combinación con la absorción o emisión de luz.

Como ya señalamos en el capítulo I, la luz no es más que una onda electromagnética con una longitud de onda determinada (el espectro visible por el ojo humano, por ejemplo, va desde longitudes de onda de 700 nm en el caso del color rojo a los 400 nm en el caso del color violeta). En su interacción con la materia, las ondas electromagnéticas se comportan como partículas sin masa, **fonones**, que tienen una energía (ε_{ph}) que depende de su longitud de onda de la forma:

$$\varepsilon_{ph} = h \frac{c}{\lambda} \quad (\text{III.26})$$

donde h es una constante, la **constante de Planck**,³⁷ c es la velocidad de la luz³⁸ y λ es la longitud de onda.

Así, por ejemplo, la luz de color violeta es más energética que la luz de color rojo. Cuando un fotón de luz de un determinado color³⁹ incide sobre un semiconductor, en ciertos casos puede ceder su energía a un electrón de la banda de valencia. Si la energía es suficiente para superar el valor del gap, el electrón podrá liberarse del sistema de enlaces y pasar a la banda de conducción, apareciendo simultáneamente un hueco en banda de valencia. Es decir, se produce un **proceso de generación estimulado por la luz**.

Resultado de este proceso es que se incrementa el número de portadores, y por tanto, aumentará también la conductividad del semiconductor, lo cual podrá ser de utilidad en ciertos dispositivos, como las células solares o los fotodetectores.

Pero en la interacción entre los semiconductores y la luz también puede darse el proceso inverso. En el proceso de recombinación de un electrón en banda de conducción que cae a banda de valencia, aniquilando un par electrón-hueco, la energía que pierde el electrón puede producir la **emisión de un fotón**, que tendrá

³⁷Determinada por Max Planck (1958-1947). La constante de Planck toma un valor $h = 6.62607 \cdot 10^{-34}$ J·s.

³⁸299792458 m/s, es decir, aproximadamente $3 \cdot 10^8$ m/s.

³⁹Puede tratarse también de *colores* no visibles para el ojo humano, como el infrarrojo o el ultravioleta.

una energía de al menos el valor del gap: es decir, el color de la luz emitida estará fijado por el valor de este parámetro, que depende del tipo de material semiconductor con el que estemos tratando. La aplicación más habitual de este fenómeno es en los diodos de tipo LED, muy empleados hoy en día en iluminación, y que serán tratados en el capítulo VII.

12 SEMICONDUCTORES MÁS RELEVANTES

En la actualidad el **silicio** es el material básico para la fabricación de dispositivos electrónicos. Sus propiedades electrónicas, así como la existencia de un óxido propio, el SiO_2 , de extraordinaria calidad, lo han convertido en el material clave de la tecnología actual.

El silicio fue descubierto en 1823 por Berzelius⁴⁰, siendo uno de los elementos más abundantes en nuestro planeta. En concreto, cerca del 28% de la corteza terrestre está compuesta por silicio, en forma de sílice (dióxido de silicio, SiO_2). Se trata por tanto de un material fácilmente disponible, pero tal y como se encuentra en la naturaleza no se puede emplear de manera directa para fabricar circuitos electrónicos. Para ello, es necesario disponer de **silicio puro y monocristalino**. Por tanto, deben realizarse dos procesos previos a la fabricación de los microchips, el refinado del material hasta niveles de pureza muy elevados, y la cristalización adecuada. Los procesos básicos para estos fines se describirán en el capítulo XII, donde se explicarán también brevemente los métodos de fabricación de microchips.

Otro semiconductor de interés es el **germanio**, que logró aislarse por primera vez en 1886. El germanio es un elemento mucho más escaso que el silicio (0.00014% de la corteza terrestre), pero en los inicios de la electrónica del estado sólido fue muy empleado⁴¹, porque no requiere de unos niveles de pureza tan elevados para funcionar de forma adecuada. Sin embargo, el valor de su gap es muy pequeño, de tan solo unos 0.67 eV a temperatura ambiente. Por ello, no es un material idóneo para fabricar dispositivos, porque su uso implica elevadas corrientes inversas en diodos y en estado de corte en transistores, así como problemas a temperaturas elevadas. Aún así, hoy en día aún se emplea en ciertas aplicaciones, como en transmisión de información en fibras ópticas o en optoelectrónica de infrarrojos, así como en ciertos tipos de transistores donde se

⁴⁰Jöns Jacob Berzelius (1779-1848), científico sueco considerado uno de los padres de la química moderna.

⁴¹De hecho el primer transistor de la historia fue fabricado con germanio.

emplea combinado con silicio formando aleaciones de silicio-germanio o SiGe.

El **arseniuro de galio (GaAs)** es un semiconductor compuesto, una aleación de galio y arsénico. El galio es un elemento del grupo 13 (antiguo III) de la tabla periódica, y el arsénico del grupo 15 (antiguo V), por lo que decimos que es un semiconductor III-V, en contraste con el silicio o el germanio que son semiconductores elementales. Al ser compuesto y realizado con materiales escasos como el galio y el arsénico, con características químicas (punto de fusión, presión de vapor, etc.) muy distintas, su coste es notablemente mayor, y su tecnología de fabricación no está tan desarrollada como la del silicio. Es útil en aplicaciones de alta velocidad y en optoelectrónica, junto con otros semiconductores de aleación III-V.

IV.

EL DIODO

El **diodo** es el dispositivo electrónico más simple. Se trata de un dispositivo pasivo¹ no lineal² de dos terminales. Existen diferentes tipos de diodos, pero en este capítulo vamos a centrarnos principalmente en el denominado **diodo de unión**.

Un diodo de unión está formado por un semiconductor con dos zonas de dopaje opuesto, una de tipo p y otra de tipo n. Por ello en ocasiones nos referimos a este tipo de estructura como **unión p-n**. Cada una de las zonas tiene además un contacto metálico adyacente al semiconductor para facilitar la circulación de corriente por los cables o pistas metálicas que unan al diodo con el resto del circuito. En su forma más esquemática³ representamos esta unión como se muestra en la figura IV.1.



Figura IV.1: Esquema simple de un diodo.

¹Los dispositivos activos son aquellos capaces de controlar el flujo de corriente eléctrica entre dos terminales mediante la aplicación de una señal en un tercer terminal. Por contra, los dispositivos pasivos no pueden realizar esta función de control.

²Un dispositivo no lineal es aquel en el que no existe una relación de proporcionalidad constante entre voltaje y corriente.

³En el caso de un diodo en un circuito integrado, la configuración es ligeramente diferente, debido a que debe realizarse con tecnología planar, es decir, ambos contactos deberán encontrarse en la parte superior del semiconductor. Sin embargo, para entender el funcionamiento del diodo podemos asumir que tiene una configuración como la mostrada aquí.

Los dos terminales del diodo reciben el nombre de **ánodo y cátodo**. El ánodo es el contacto adyacente a la zona p y el cátodo el adyacente a la zona n. El símbolo circuital del diodo es el mostrado en la figura IV.2.



Figura IV.2: Símbolo circuital de un diodo.

El lado p (ánodo) corresponde a la base del triángulo, y el n (cátodo) a la línea vertical. Podemos interpretar este símbolo como una flecha que nos indica en qué sentido va a poder fluir la corriente en el diodo. Los diodos son dispositivos con **polaridad**, es decir, que debe respetarse su orientación. Si trabajamos con componentes discretos, el lado n del diodo se puede identificar por una línea de diferente color en uno de los lados del dispositivo, como se muestra en la figura IV.3. Es importante, por tanto, colocar adecuadamente un diodo en su posición en el circuito, ya que en caso contrario no funcionará correctamente.



Figura IV.3: Imagen de un diodo discreto. La franja gris indica el lado n.

1 PRINCIPIO FÍSICO

Para entender el comportamiento de un diodo debemos recordar cuáles son los procesos de transporte de carga en un semiconductor, es decir, el arrastre y la difusión. En un diodo de unión, tenemos dos zonas de dopajes opuestos. En el lado n del diodo, los portadores mayoritarios son los electrones, y los minoritarios los huecos. En cambio, en el lado p, los portadores mayoritarios son los huecos y los minoritarios los electrones. Existe en consecuencia una inhomogeneidad en el material en cuando a la distribución de la concentración de portadores. Aparecerá por tanto el fenómeno de la **difusión**, es decir, el movimiento natural de cargas desde donde la concentración es elevada hacia donde es reducida. En nuestro caso, existirá un movimiento de electrones desde el lado n (donde son

numerosos) hacia el p (donde son escasos), mientras que los huecos seguirán el movimiento opuesto, tenderán a ir desde el lado p hacia el n (figura IV.4).

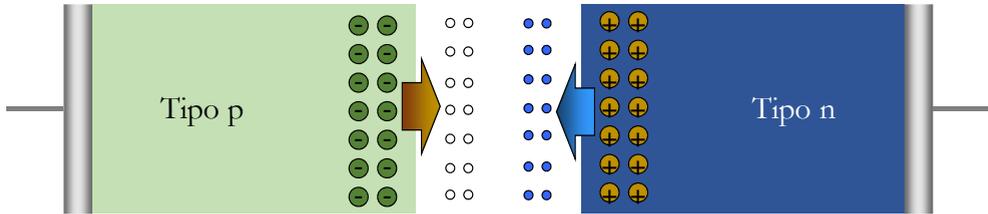


Figura IV.4: Procesos de difusión en la zona de la unión de un diodo: electrones (círculos azules) desde el lado n al p y huecos (círculos blancos) desde el lado p al n.

Sin embargo, debemos tener en cuenta que en cada uno de los lados del diodo además de los portadores libres existen iones de **impurezas ionizadas** que forman parte de la red cristalina, y que por tanto no pueden desplazarse. En la zona de la unión entre las dos regiones del diodo tendrá lugar el siguiente fenómeno. En el lado n, los electrones que se encontraban en la frontera al desplazarse dejarán al descubierto las cargas positivas de las impurezas donadoras ionizadas. En el lado p, los huecos que se difunden dejarán tras de sí las cargas negativas de las impurezas aceptadoras ionizadas. En consecuencia, en la zona de la frontera quedan al descubierto una serie de cargas fijas (las impurezas ionizadas no pueden moverse de las posiciones que ocupan en la red cristalina) que generan un campo eléctrico interno, que apunta desde el lado n hacia el p (figura IV.5).

Este campo eléctrico tiende a oponerse al movimiento por difusión de los

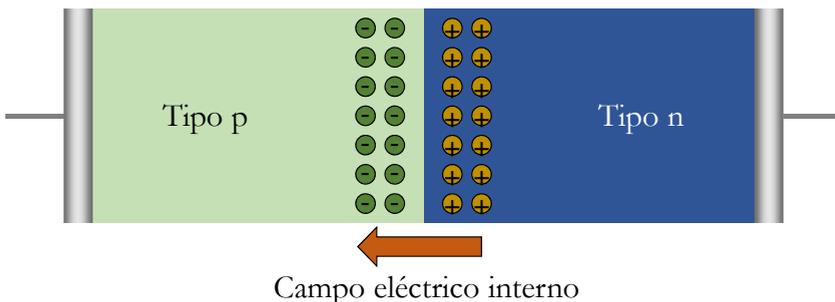


Figura IV.5: Aparición de un campo eléctrico interno debido a las impurezas ionizadas en la zona de la frontera entre las dos regiones impurificadas del diodo.

portadores libres (intenta devolver a los portadores a la región donde son mayoritarios), y provoca que en la frontera entre las dos regiones de distinto dopaje aparezca una **zona de vaciamiento** sin electrones ni huecos. Así, un electrón que quiera pasar del lado n al p por difusión deberá tener una energía suficiente como para vencer la oposición del campo eléctrico de la unión y llegar al otro lado. De manera análoga, los huecos del lado p deberán tener al menos una cierta energía para pasar al lado n por difusión.

Podemos hablar entonces de la existencia de una **barrera de energía**, que correspondería a la *energía mínima que debe tener un portador para pasar por difusión de la región donde es mayoritario a la región donde es minoritario*. Vamos a representar este fenómeno mediante un diagrama de bandas de energía, como el mostrado en la figura IV.6.

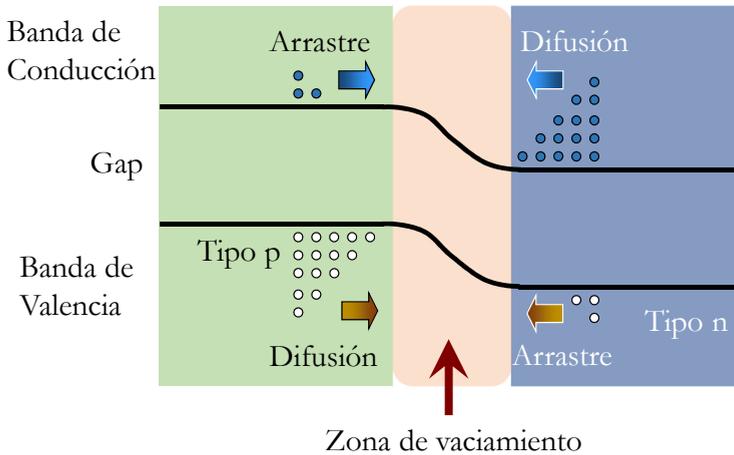


Figura IV.6: Bandas de energía en una unión p-n. En equilibrio las cuatro corrientes indicadas por las flechas se encuentran compensadas, dando lugar a una corriente neta nula.

En el diagrama, las líneas negras representan el fondo de la banda de conducción (el mínimo valor de energía de los electrones en esa banda) y el máximo de la banda de valencia (mayor valor de energía posible para los electrones en esa banda, o, si se quiere, mínima energía para los huecos⁴). El fenómeno del va-

⁴En el diagrama de bandas de energía debemos tener en cuenta que la tendencia natural de los electrones es situarse en los estados menos energéticos. En consecuencia, los huecos ocuparán los estados de energía electrónica más altos. Es sencillo interpretar el comportamiento de ambos tipos de portadores en el diagrama imaginando a los electrones como piedras en un estanque, y a

ciamiento en la frontera se representa mediante la transición curva entre las dos regiones de dopaje, que conforma la barrera de energía antes mencionada. Los electrones deberán remontar esa barrera para pasar del lado n al p, y los huecos, por su parte, verán una barrera análoga para pasar del p al n⁵. Los movimientos en sentido inverso (para los portadores minoritarios en cada lado) en cambio son muy sencillos, porque el campo eléctrico interno de la unión favorece su paso al lado opuesto mediante el proceso de arrastre. Existirá por tanto una pequeña corriente asociada a estos portadores, que necesariamente se compensará con la corriente debida a la difusión de mayoritarios con energía suficiente para remontar la barrera. Por tanto, la corriente total en el dispositivo (suma de la corriente de arrastre de electrones y huecos minoritarios y de difusión de electrones y huecos mayoritarios) será, como no puede ser de otra manera en ausencia de un campo externo aplicado, nula.

Esta situación evidentemente cambiará cuando apliquemos una diferencia de potencial entre terminales del diodo, que generará un campo eléctrico externo. Tendremos dos posibles situaciones dependiendo de que la diferencia de potencial entre los lados p y n del diodo (V_{pn}) sea positiva o negativa.

2 POLARIZACIÓN DEL DIODO Y CURVA CARACTERÍSTICA

La primera situación que consideraremos es el caso de la **polarización directa**. En este caso, el potencial en el lado p es mayor que en el lado n, es decir, $V_{pn} > 0$. La aplicación de una diferencia de potencial provoca la aparición de un campo eléctrico asociado, que tenderá a concentrarse en la zona más resistiva del dispositivo, es decir, en la zona de vaciamiento⁶. En el caso de polarización directa, la orientación del campo eléctrico externo es opuesta a la del campo interno de la unión: en consecuencia, el campo en la zona de vaciamiento se debilita, lo que implica que los portadores mayoritarios afrontan una barrera de energía menor, y además se produce un estrechamiento de la zona de vaciamiento (véase la figura IV.7).

los huecos como aire en una botella: los electrones siempre tenderán a ir al fondo, y los huecos a la parte más alta, cada uno en su banda correspondiente, conducción y valencia respectivamente.

⁵Podemos verlo fácilmente para el caso de los huecos si miramos el diagrama boca abajo.

⁶Recuérdese que en esta zona los portadores de carga, tanto electrones como huecos, no puede permanecer, por lo que al ser prácticamente despreciable su concentración la conductividad local de esa zona es bajísima, o lo que es lo mismo, su resistencia es muy grande. Ello supone que la mayor parte del potencial aplicado externamente cae en esa región de la frontera.

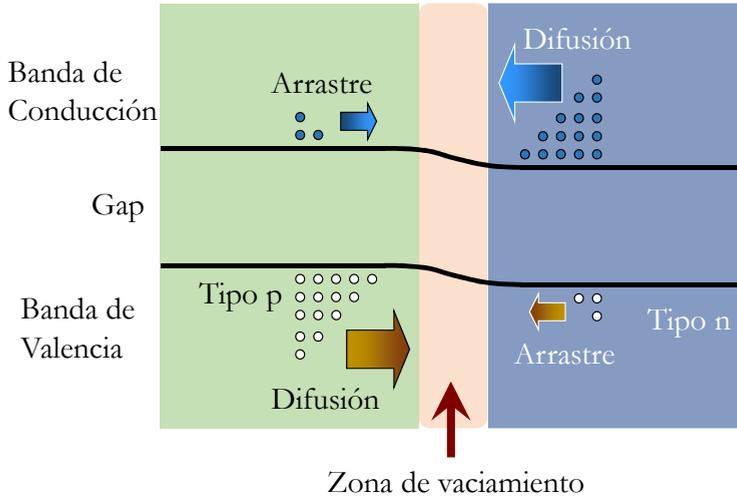


Figura IV.7: Bandas de energía en una unión p-n en directa. La corriente se produce por difusión al disminuir las barreras de energía.

Al reducirse la barrera de energía para los portadores mayoritarios, se favorece enormemente la difusión, mientras que la corriente de arrastre de minoritarios permanece prácticamente invariante, con un valor mucho más pequeño. Aparece por tanto una corriente neta, con sentido del lado p hacia el n, y que crece exponencialmente al aumentar la diferencia de potencial aplicada. Bajo condiciones de polarización directa, el diodo conduce. El aumento de corriente se produce de manera particularmente intensa a partir de un cierto voltaje, denominado **voltaje umbral del diodo**, que representaremos por V_γ .

La otra posibilidad es que apliquemos una diferencia de potencial de modo que $V_{pn} < 0$, lo que conocemos como **polarización inversa**. En este caso, el campo eléctrico externo tendrá la misma orientación que el interno y lo reforzará, de modo que se produce un aumento del campo efectivo en la zona de vaciamiento, que será más ancha, y además la barrera de energía será mayor. Por tanto, los portadores mayoritarios van a tener mayores dificultades para atravesar la unión por difusión, mientras que los minoritarios verán reforzado el efecto de arrastre. Las corrientes de arrastre y difusión ya no estarán compensadas como ocurría en equilibrio, sino que dominará la corriente de arrastre de minoritarios. En consecuencia, la corriente en el diodo será extraordinariamente pequeña⁷; a

⁷La corriente en este régimen de polarización inversa se denomina **corriente inversa de saturación**, I_S .

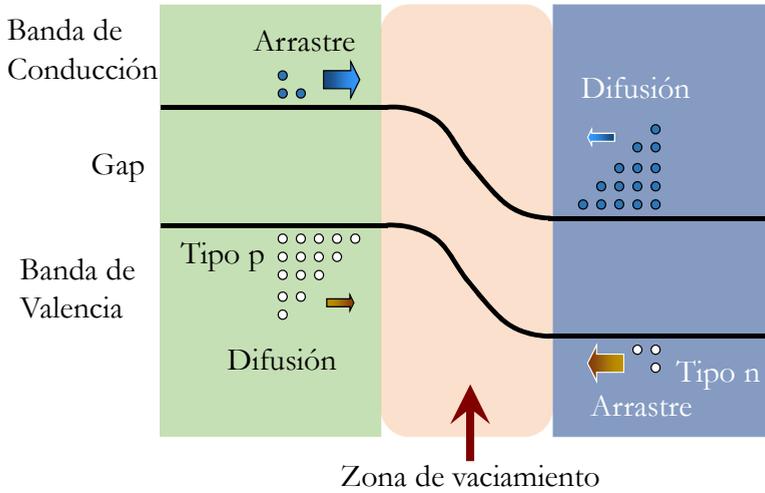


Figura IV.8: Bandas de energía en una unión p-n en inversa. La corriente es prácticamente despreciable.

efectos prácticos, despreciable. La situación respecto al diagrama de bandas puede verse en la figura IV.8.

En definitiva, el diodo es un dispositivo que, en polarización directa, y especialmente a partir del voltaje umbral V_γ , conduce muy bien, mientras que en polarización inversa no deja prácticamente pasar la corriente. Por tanto, como habíamos señalado al comienzo de esta sección, es un dispositivo no lineal, cuya característica corriente-voltaje tiene la forma mostrada en la figura IV.9.

3 APROXIMACIONES

La corriente en el diodo puede expresarse mediante la ecuación de Shockley⁸:

$$I = I_S \left(e^{\frac{qV_{pn}}{\eta k_B T}} - 1 \right) \quad (\text{IV.1})$$

donde η es un parámetro de idealidad o calidad, que depende de cómo esté fabricado el diodo, y que usualmente toma un valor entre 1 y 2.

Sin embargo, para efectuar un análisis rápido del comportamiento de un circuito que contenga diodos, en ingeniería normalmente se emplean una serie de

⁸Por William Bradford Shockley (1910-1989), coinventor del transistor y premio Nobel de física en 1956.

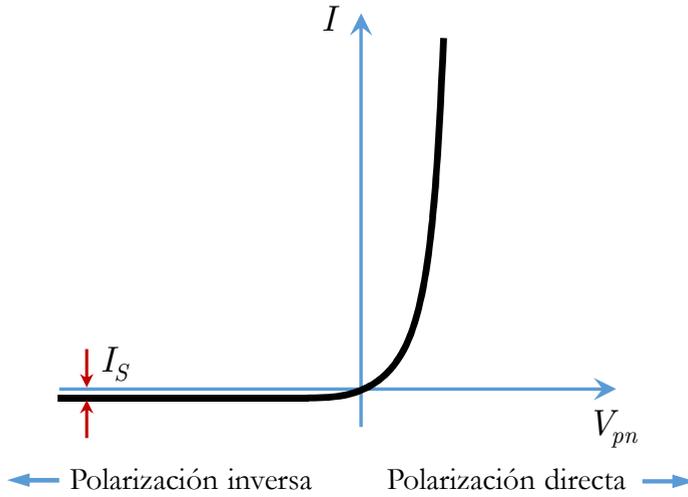


Figura IV.9: Característica corriente-voltaje de un diodo.

aproximaciones para describir el funcionamiento de este dispositivo.

La primera aproximación consiste en suponer que el diodo se comporta como un interruptor, es decir, que para $V_{pn} > 0$ es como un cortocircuito y para

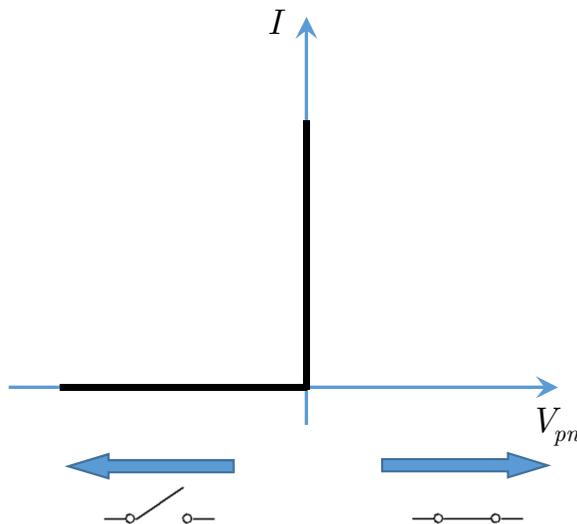


Figura IV.10: Primera aproximación de un diodo.

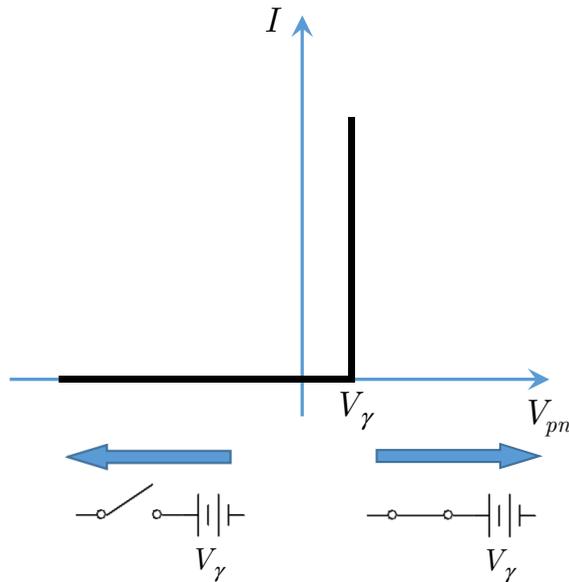


Figura IV.11: Segunda aproximación de un diodo.

$V_{pn} < 0$ es como un circuito abierto (véase la figura IV.10). Esta aproximación no es demasiado adecuada dado que no tiene en cuenta el voltaje umbral del diodo⁹, pero puede ser útil para tener una primera visión de la operación del circuito.

La segunda aproximación es similar a la primera, pero se incluye una fuente de voltaje de valor V_{γ} para dar cuenta del efecto del voltaje umbral (figura IV.11). Es la más utilizada para resolver circuitos de manera rápida, ya que da una idea razonable de cómo se comporta el diodo. Incorpora también un interruptor, que estará cerrado para voltajes iguales o superiores a V_{γ} y abierto para los inferiores.

Finalmente, la **tercera aproximación** se basa en la segunda, pero en vez de considerar que el diodo conduce idealmente para diferencias de potencial aplicadas superiores al voltaje umbral, presenta una pequeña resistencia de valor R_D .

4 RECTIFICACIÓN

Una de las aplicaciones más habituales de los diodos es la rectificación de señales en alterna. Decimos que una señal eléctrica es **bidireccional** cuando es variable en el tiempo y oscila entre valores positivos y negativos, como puede ser

⁹En el caso de un diodo de silicio el voltaje umbral suele valer en torno a 0.6 o 0.7 V, por lo cual no es despreciable.

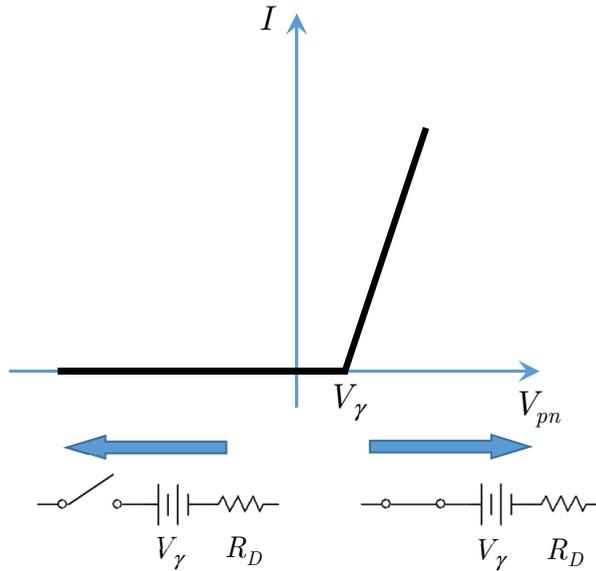


Figura IV.12: Tercera aproximación de un diodo.

el caso de una señal de voltaje senoidal que oscile en torno a 0 V. En cambio, una señal es **unidireccional** cuando siempre tiene el mismo signo, bien sea positivo o negativo.

La **rectificación** es el *proceso de convertir una señal bidireccional en unidireccional*. Se trata de una etapa fundamental en ciertas aplicaciones, como por ejemplo en los transformadores de alimentación alterna a continua¹⁰. El circuito rectificador más sencillo es el **rectificador de media onda**, que se muestra en la figura IV.13. Debemos señalar que en la circuito hemos incorporado la conexión de **tierra**¹¹, representada por las tres líneas paralelas de longitud decreciente, que es el punto que tomaremos como referencia (valor 0 V) para los voltajes.

Consideremos que la señal de entrada V_{sin} es una señal senoidal AC con una cierta amplitud, y sin offset, es decir, que oscila en torno a 0 V. Cuando el valor

¹⁰Los circuitos presentes en el interior de los dispositivos electrónicos trabajan con alimentación en continua. Dado que la red industrial proporciona una señal alterna (en Europa de valor eficaz de 220 V y 50 Hz de frecuencia), es necesario transformar la señal AC tomada en un enchufe en un valor continuo de voltaje adaptado al circuito o dispositivo en cuestión (usualmente de 5 o 12 V).

¹¹También nos referiremos a ella como masa, referencia o GND (*ground*). En los sistemas eléctricos se incluye una conexión física a la tierra (en los edificios una barra metálica larga que se clava en el subsuelo) por razones de seguridad.

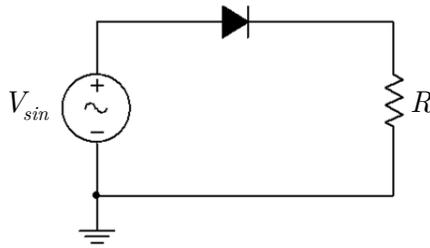


Figura IV.13: Circuito rectificador de media onda.

instantáneo de V_{sin} es inferior al voltaje umbral del diodo, V_γ , no se ha alcanzado la condición mínima para la conducción, por lo que la corriente en el diodo (y en consecuencia, en la resistencia R que está en serie con él) es nula. De este modo, la caída de potencial en la resistencia, conforme a la ley de Ohm, es cero, ya que no circula corriente por ella.

Sin embargo, cuando el voltaje de la fuente es superior al voltaje umbral, el

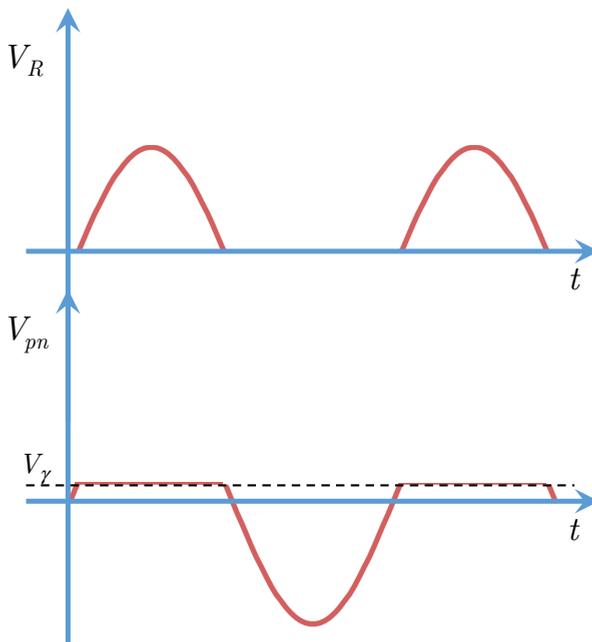


Figura IV.14: Caídas de potencial en la resistencia (arriba) y en el diodo (abajo).

diodo estará polarizado en directa y en condiciones de conducir corriente. Conforme a la segunda aproximación, podemos considerar que se comporta de manera ideal teniendo en cuenta una caída de potencial igual a V_γ entre sus extremos. Por tanto, el resto del voltaje aplicado (es decir, $V_{sin} - V_\gamma$) tendrá que caer en la resistencia R .

Es decir, en la resistencia hemos obtenido una señal *rectificada*, que sólo toma valores positivos, y que es prácticamente igual (salvo por el voltaje V_γ perdido en el diodo) a los semiciclos positivos de la señal de partida (figura IV.14).

Como vemos, este es un circuito muy sencillo que emplea únicamente un diodo, pero a cambio tiene el problema de que estamos desaprovechando la mitad de la onda de partida. Una posible solución es emplear más diodos para realizar un **rectificador de onda completa**, que aproveche también los semiciclos negativos de la señal de entrada. El esquema de este circuito se muestra en la figura IV.15.

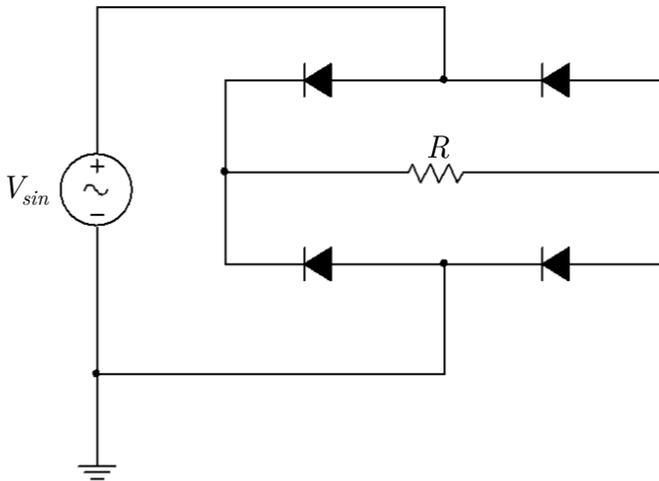


Figura IV.15: Circuito rectificador de onda completa.

Con este circuito se consigue una onda en la resistencia R que siempre toma valor positivo, como se observa en la figura IV.16, reproduciendo los ciclos de la señal de entrada (salvo las pérdidas por el voltaje umbral de los diodos).

La señal rectificada en todo caso dista aún mucho de ser una señal continua DC. Para poder lograr esto, es necesario realizar un proceso de **filtrado** de la señal rectificada. El filtrado elimina la oscilación y proporciona una señal prácticamente constante. Para ello puede ser suficiente con acoplar un condensador

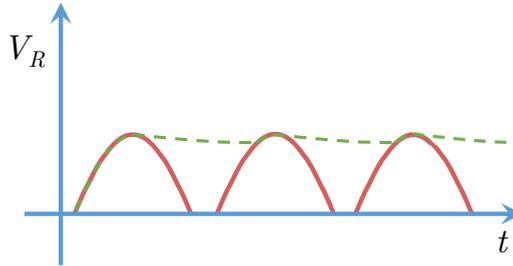


Figura IV.16: Señal rectificada en la resistencia, en el circuito rectificador de onda completa. La línea discontinua indica la forma de la señal filtrada.

(de capacidad relativamente elevada) en paralelo con la resistencia R . Al acoplar el condensador, éste se carga rápidamente durante la fase de subida del voltaje de entrada, pues lo hace gracias a la corriente proporcionada por el diodo (que posee muy baja resistencia interna). Tras alcanzar el máximo, el condensador colocado en paralelo con la resistencia debería reducir el voltaje entre sus extremos de manera acorde con la caída de voltaje de la señal. Ello implica que el condensador debería descargarse, proceso que se debe realizar a través de la resistencia R y que lleva un determinado tiempo, por lo que la bajada de voltaje observada en la resistencia sería más lenta que la variación de la señal de entrada. Con un condensador de capacidad suficientemente grande como para tener una constante de tiempo elevada, el efecto que se observa es el de una señal de voltaje en la resistencia prácticamente constante (ver figura IV.16).

Rectificación y filtrado son las etapas fundamentales del proceso de transformación de una señal AC en una señal DC, aunque no son las únicas. También es necesario un bloque transformador AC, que adapte inicialmente el voltaje de la señal de la red al valor de voltaje final que nos interesa, y un bloque final regulador, que asegure que la señal DC de salida permanezca estable ante variaciones de temperatura o picos de voltaje en la entrada.

5 DIODOS ZENER

En un diodo ideal, en polarización inversa un diodo no debería conducir la corriente eléctrica en ninguna circunstancia. Sin embargo, en la realidad, a tensiones negativas elevadas se puede producir una conducción eléctrica no deseada a partir de un cierto valor de V_{pn} en inversa, llamado voltaje de ruptura del diodo.

Esta conducción puede producirse por varios tipos de fenómenos. Por ejem-

plo, puede ocurrir que al aplicar una diferencia de potencial muy negativa, el elevado campo eléctrico provoque que un portador minoritario que está siendo arrastrado gane mucha energía y colisione con un electrón del sistema de enlaces, liberándolo. Aumentará por tanto el número de portadores, que a su vez serán arrastrados por el campo, pudiendo llegar a producirse el **fenómeno de avalancha**, que produce una multiplicación de la corriente.

Otro proceso por el que puede tener lugar una conducción de corriente en inversa es el llamado **efecto túnel**. La curvatura de las bandas de energía puede ser tan fuerte que, por ejemplo, electrones en el lado p del diodo, que estaban formando parte del sistema de enlaces, *vean* estados libres en la banda de conducción del lado n y pasen a ellos mediante un proceso de naturaleza cuántica, el efecto túnel, por el cual los electrones son capaces de atravesar una barrera de energía.

Existe un tipo de diodos que son fabricados *ex profeso* para aprovechar un voltaje inverso negativo de ruptura (figura IV.17). Se trata de los **diodos Zener**¹². Los diodos Zener pueden ser utilizados, por ejemplo, para proteger partes de un circuito frente a picos de voltaje, pues colocados adecuadamente en paralelo con el circuito que se desea proteger, derivan a través de ellos la corriente en el momento en que el voltaje externo aplicado supere el voltaje Zener V_Z .

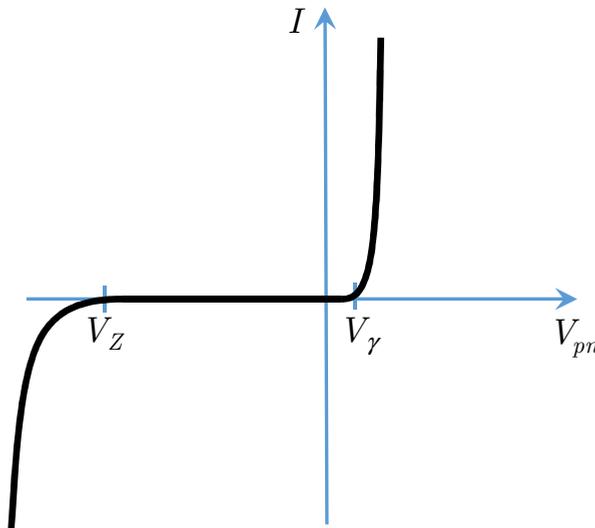


Figura IV.17: Curva característica de un diodo Zener.

¹²En honor a su inventor, Clarence Melvin Zener (1905-1993).

El símbolo del diodo Zener es ligeramente diferente al del diodo normal, como se muestra en la figura IV.18.



Figura IV.18: Símbolo circuital del diodo Zener.

6 OTROS TIPOS DE DIODOS

Además del diodo normal y del diodo Zener, existen otros tipos de diodos con múltiples usos en electrónica. Mencionaremos a continuación los más importantes, algunos de los cuales (LEDs y fotodiodos) serán tratados con más detalle en el capítulo VII.

El **diodo Schottky**¹³ se fabrica mediante la unión de un metal y un semiconductor dopado tipo n. Es, por tanto, una unión metal-semiconductor. Este tipo de uniones pueden formar una barrera energética natural por las diferencias entre las características físicas del metal y el semiconductor, de modo que en inversa no es posible el flujo de electrones, mientras que en directa pueden atravesar la barrera por emisión termoiónica (electrones con energía suficiente para superarla) y en menor medida por efecto túnel. Los diodos Schottky se caracterizan por tener un tiempo de respuesta muy rápido (por lo que son muy adecuados para aplicaciones de alta frecuencia, donde puede cambiar su polarización muy deprisa) y un voltaje umbral bajo (en torno a 0.25 V). Su símbolo se muestra en la figura IV.19.



Figura IV.19: Símbolo circuital del diodo Schottky.

Un tipo de diodo muy popular en la actualidad es el **diodo emisor de luz o diodo LED**. Son diodos capaces de emitir radiación luminosa (usualmente en el rango del visible o del infrarrojo) cuando circula corriente por ellos. La operación

¹³Por Walter Hans Schottky (1886-1976), físico alemán que sentó las bases teóricas de este dispositivo.

inversa la realizan los **fotodiodos**, que al recibir luz de una cierta longitud de onda generan una corriente eléctrica (de forma análoga a una célula solar, que en esencia es también un diodo). Ambos tipos de dispositivos los estudiaremos con más detalle en el capítulo VII.

Por último, el **varactor** es un tipo de diodo que se comporta como un condensador con valor de capacidad variable. La zona de vaciamiento del diodo, en la que no pueden permanecer las cargas libres, actúa de forma similar a la zona del dieléctrico de un condensador. Según sea mayor o menor la polarización aplicada, la anchura de la zona de vaciamiento varía: es decir, sería análogo a una variación de la distancia entre placas de un condensador, lo que modifica su capacidad. Su símbolo se muestra en la figura IV.20.



Figura IV.20: Símbolo circuital del diodo varactor.

V.

EL TRANSISTOR
BIPOLAR

En sus comienzos, la electrónica se basaba en tubos de vacío para el control de la corriente eléctrica. Los primeros ordenadores, aparecidos tras la Segunda Guerra Mundial, estaban fabricados con estos dispositivos. Ello implicaba que ocupaban una gran superficie, consumían mucha potencia eléctrica y tenían unas capacidades operativas muy limitadas. Además, los tubos de vacío son dispositivos muy frágiles, que con frecuencia se estropeaban o degradaban. La invención del **transistor**¹, un dispositivo de estado sólido realizado con semiconductores, abrió la puerta a una nueva forma de entender la electrónica. Nació así la **electrónica de estado sólido**, en la cual los dispositivos se fabrican con materiales semiconductores. El posterior desarrollo de los **circuitos integrados**² marcó el inicio de la revolución electrónica y de la sociedad de la información actual. En definitiva, es imposible entender la informática actual sin la aparición del transistor, posiblemente el invento más trascendental del siglo XX, y que hoy sigue teniendo plena vigencia.

Existen diferentes tipos de transistores, entre los que destacamos dos: el **transistor bipolar de unión** y el **transistor MOSFET**. Ambos se basan en principios físicos diferentes, pero su funcionalidad es similar. Un transistor es, en esencia, un dispositivo que controla el paso de corriente entre dos terminales mediante la aplicación de una señal eléctrica (corriente o voltaje, dependiendo del tipo de transistor) en un tercer terminal de control. En cierto modo, el terminal de control es responsable de modular la resistencia del semiconductor entre los otros

¹El transistor fue inventado en los laboratorios Bell en 1947 por John Bardeen (1908-1991) y Walter Brattain (1902-1987), que crearon inicialmente el transistor de punta de contacto. William Shockley (1910-1989) perfeccionó este diseño e ideó el transistor bipolar de unión. Los tres recibieron el premio Nobel de física en 1956.

²Los procesos de fabricación de circuitos integrados se tratarán en el capítulo XII.

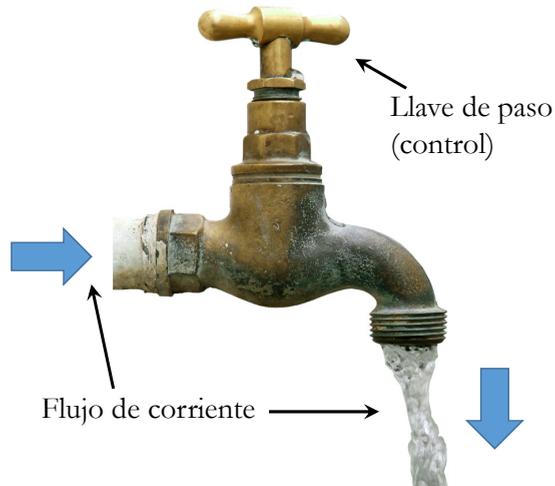


Figura V.1: El funcionamiento de un transistor es conceptualmente análogo al de un grifo: un elemento de control regula la corriente entre dos puntos.

dos contactos³.

Podemos entender el comportamiento básico de un transistor mediante una sencilla analogía. Pensemos en un grifo (figura V.1). El flujo de agua entre la cañería y la boca del grifo (la *corriente* de agua que mana del grifo) se regula mediante la maneta o llave. Girando ésta más o menos podremos tener mayor o menor corriente de agua, es decir, controlamos el flujo de agua entre dos puntos mediante un elemento de control. Habrá ocasiones en que deseemos interrumpir completamente el flujo de corriente, y en ese caso no queremos que el grifo gotee o pierda agua. En otras ocasiones, nos interesará que salga una cantidad de agua muy grande, y abriremos hasta el tope la llave para permitir el máximo flujo de agua posible. Otras veces nos interesará que salga una cantidad moderada de agua, y ajustaremos el giro de la maneta del grifo hasta la posición deseada. El transistor realiza una operación parecida, pero con un flujo de carga (corriente) eléctrica. Además, el control de esa corriente es por una señal eléctrica, y no por un mecanismo. Cómo conseguir ese control es lo que diferencia a los diferen-

³Tras su invención, los laboratorios Bell buscaron un nombre comercial para el nuevo dispositivo. El término *transistor* fue acuñado por el ingeniero John R. Pierce (1910-2002), jugando con el concepto de *transresistance* (que hacía referencia a la modulación de la resistencia del dispositivo por el terminal de control) y la terminación fonética de otros dispositivos como el *termistor* o el *varactor*.

tes tipos de transistores, pero lo importante es que todos pueden realizar esta función.

El transistor es el dispositivo clave de la electrónica moderna, y el bloque básico de construcción de la práctica totalidad de circuitos analógicos y digitales de interés hoy en día. La informática en la forma en que la conocemos es posible gracias al transistor. Los microprocesadores no son más que circuitos digitales extraordinariamente complejos, compuestos por miles de millones de transistores diminutos fabricados e interconectados entre sí en un pequeño trozo de semiconductor: en definitiva, un circuito integrado.

1 PRINCIPIO FÍSICO

El transistor bipolar de unión o **BJT** (*Bipolar Junction Transistor*) fue la evolución natural del primer transistor inventado en los laboratorios Bell en 1947, el transistor de punta de contacto. Los primeros dispositivos electrónicos de estado sólido se basaron en el BJT hasta finales de los años 60 del siglo XX, cuando fueron reemplazados de manera masiva por el transistor MOSFET en aplicaciones digitales. Hoy en día los transistores BJT siguen siendo muy utilizados, principalmente en electrónica analógica, amplificadores, control de potencia, etc.

La estructura de un BJT se basa en dos uniones p-n opuestas. Por tanto, existen dos tipos posibles de BJT, el **npn** y el **pnP**: en el primero la conducción será por electrones y en el segundo por huecos. El transistor bipolar dispone de tres terminales, que reciben el nombre de **emisor, base y colector**. Los portadores entrarán al dispositivo por el emisor y lo atravesarán hasta salir por el colector, siendo la base quien regula el flujo. La representación esquemática de ambos tipos de transistores⁴ se muestra en la figura V.2.

Se puede entender la estructura de dopaje del transistor bipolar como dos diodos adyacentes con orientaciones opuestas, con un contacto a cada lado, y otro contacto adicional en la zona central que va a ser el que regule la corriente entre emisor y colector (figura V.3). La señal eléctrica de control va a ser una corriente, la corriente de base, que usualmente tendrá un valor pequeño. Los símbolos circuitales para el BJT son los mostrados en la figura V.4.

Para entender el funcionamiento del transistor bipolar, en primer lugar de-

⁴Al igual que en el caso del diodo, en un circuito integrado el transistor bipolar debe realizarse en una configuración planar más compleja, con todos sus contactos en la superficie. Sin embargo, los esquemas simples aquí mostrados nos sirven para entender de forma sencilla el comportamiento del dispositivo.

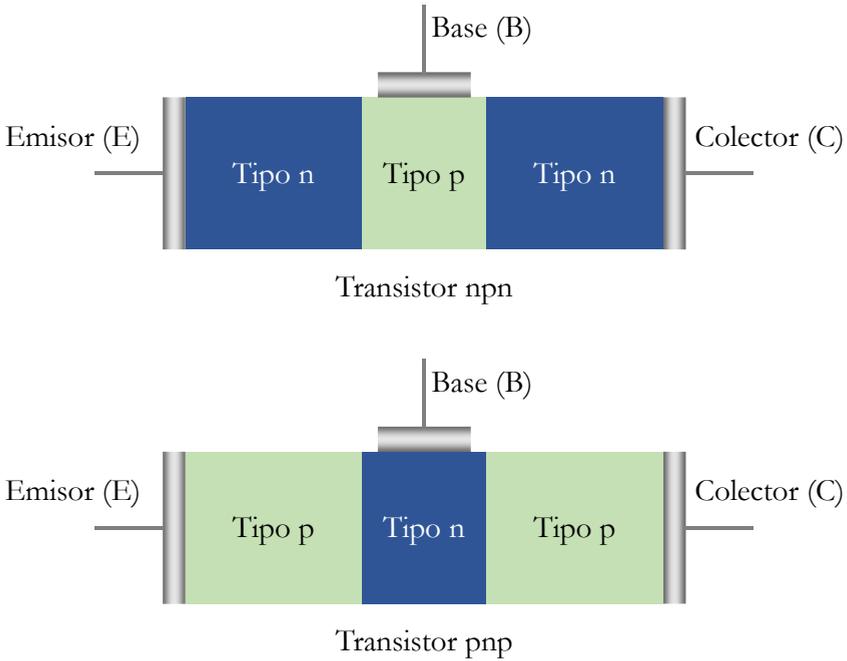


Figura V.2: Tipos de transistor bipolar.

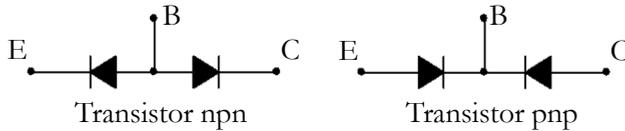


Figura V.3: La estructura de dopaje de un BJT corresponde a dos diodos adjuntos con orientaciones opuestas.

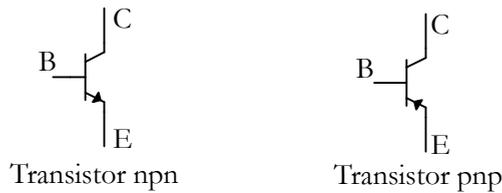


Figura V.4: Símbolos circuitales del transistor BJT npn y pnp, junto con la indicación de los terminales correspondientes.

bemos fijarnos en su estructura de bandas en equilibrio. Nos centraremos en el caso del transistor npn, pues el pnp se puede entender de forma totalmente análoga. En equilibrio, el diagrama de bandas corresponde al de dos uniones diodo opuestas, de manera que las corrientes de difusión de mayoritarios y las de arrastre de minoritarios están equilibradas, siendo la corriente total en el dispositivo nula, como cabe esperar en ausencia de fuentes de alimentación externas (figura V.5).

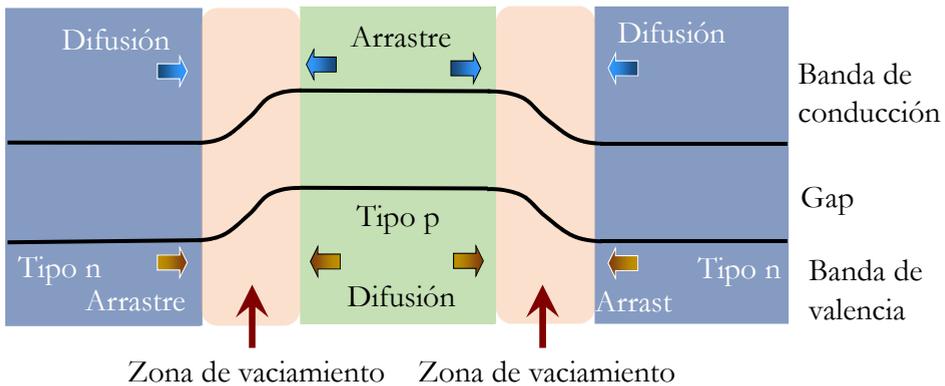


Figura V.5: Diagrama de bandas de un transistor bipolar en equilibrio.

Sin embargo, si existen fuentes de alimentación externas la situación cambia notablemente. En primer lugar, es fundamental el papel de la corriente que es inyectada o extraída en el terminal de base. Supondremos que la base es estrecha, de modo que los portadores que la alcanzan desde las regiones de dopaje n (electrones, que serán minoritarios en la base tipo p) no tienen tiempo de recombinarse con los huecos mayoritarios allí presentes. Si la corriente aplicada en el terminal de base, I_B , es inyectada hacia el interior del semiconductor, dicha corriente corresponderá a electrones que son extraídos⁵. El efecto de extraer electrones de la zona p de la base es reducir las bandas de energía en esa región. Si por el contrario se introdujesen electrones (es decir, la corriente de base saliera del dispositivo), las bandas en la zona p subirían en energía.

Por otra parte, el segundo factor a tener en cuenta es la diferencia de potencial aplicada entre los terminales de colector y emisor, es decir, V_{CE} . Para el caso del transistor npn, únicamente consideraremos potenciales positivos, que muevan a los electrones desde el emisor hasta el colector.

⁵Recuérdese que, por convenio, el sentido de la corriente es opuesto al del movimiento de electrones.

En la figura V.6 se muestra la situación del diagrama de bandas para un valor pequeño de V_{CE} cuando se está inyectando una corriente I_B en la base. En esa situación, la bajada de las bandas de energía en la zona p es mayor que en la zona n de colector. Ello se traduce en que ambas uniones p-n (la emisor-base y la base-colector) están polarizadas en directa y se favorece el paso de mayoritarios por difusión. Este régimen de operación en el BJT recibe el nombre de **región de saturación**, y en él el transistor deja fluir gran cantidad de corriente eléctrica I_C por el colector, comportándose de forma parecida a una resistencia de valor pequeño.

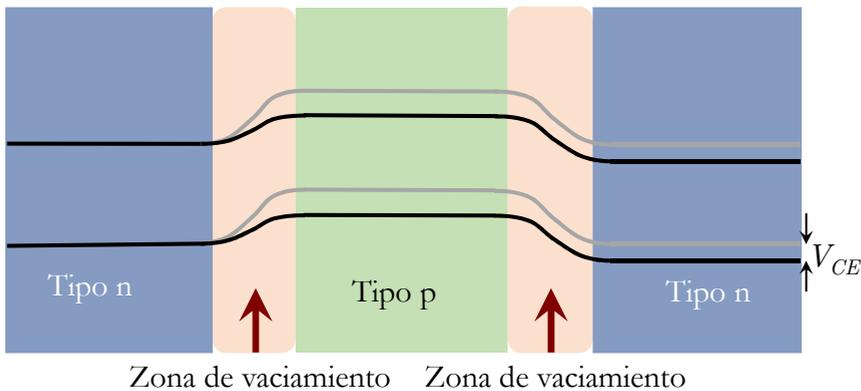


Figura V.6: Diagrama de bandas de un transistor bipolar en el régimen de saturación. Las líneas grises muestran la posición de las bandas en equilibrio.

Ahora bien, si incrementamos progresivamente el potencial aplicado V_{CE} , las bandas de energía siguen *cayendo* en el lado de colector y se produce un cambio de situación. La unión p-n emisor-base estará polarizada en directa, como antes, pero la unión p-n base-colector pasará a tener una situación de polarización en inversa, con una zona de vaciamiento más ancha que en equilibrio (véase la figura V.7). En esta unión base-colector, los electrones sólo podrán atravesarla por arrastre desde la base hacia el colector, estando impedida la difusión en sentido opuesto. Así, el flujo de electrones tendrá lugar de la siguiente manera: los electrones en el emisor alcanzarán la base por difusión (al ser una unión p-n en directa). Si la base no es excesivamente ancha, algunos saldrán de ella por el terminal de base (corriente I_B) y la mayor parte alcanzarán la unión base-colector, que atravesarán por arrastre hacia el colector. Un aumento del voltaje V_{CE} simplemente se traducirá en un mayor arrastre, pero no en un aumento significativo de la corriente de colector I_C , que en gran medida vendrá marcada por los procesos de

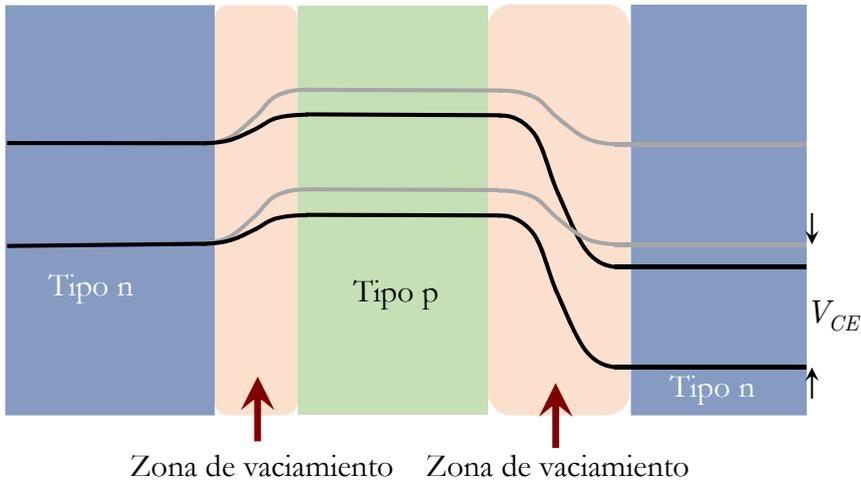


Figura V.7: Diagrama de bandas de un transistor bipolar en la región activa directa. Las líneas grises muestran la posición de las bandas en equilibrio.

difusión que ocurren en la unión emisor-base. Este régimen de funcionamiento, en el que la corriente prácticamente no aumenta al aumentar V_{CE} , se denomina **región activa directa**.

La última situación de interés en el BJT tiene lugar si en vez de inyectar una corriente I_B en la base del transistor la extraemos (es decir, inyectamos electrones en vez de sacarlos). En ese caso, las bandas de energía en la zona p suben, llevando a que la unión emisor-base se encuentre también en una condición de polarización inversa, y por tanto, no permitiendo que la atraviese una cantidad significativa de corriente. En esa situación, la corriente de colector I_C será prácticamente nula, y diremos que el transistor se encuentra en la **región de corte**. A pesar de que se aplica una diferencia de potencial V_{CE} , el transistor no conduce corriente. Esta situación corresponde al diagrama de la figura V.8.

2 CURVAS CARACTERÍSTICAS

Vamos a examinar con más detalle la respuesta eléctrica del transistor BJT frente a las condiciones de corriente en la base, I_B , y el voltaje entre colector y emisor V_{CE} .

En general, en el transistor bipolar nos van a interesar las siguientes magnitudes eléctricas:

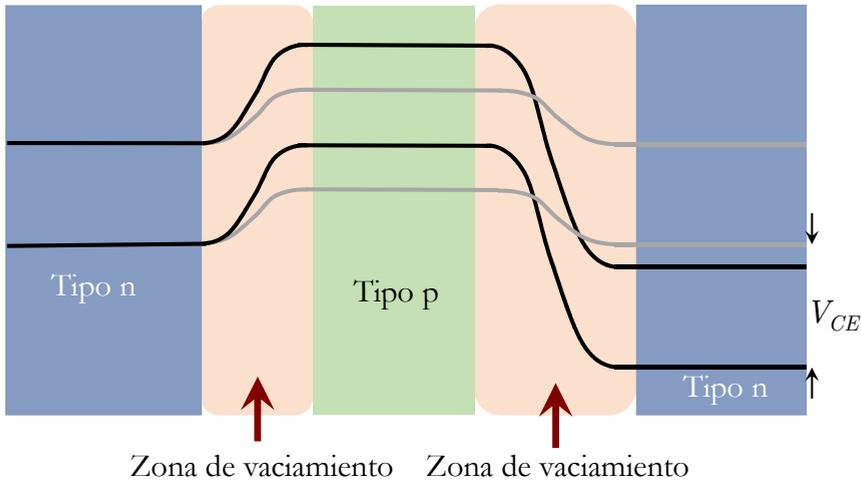


Figura V.8: Diagrama de bandas de un transistor bipolar en la región de corte. Las líneas grises muestran la posición de las bandas en equilibrio.

- La corriente de colector I_C : en el BJT npn se considera que entra al colector y en el pnp que sale del colector (en ambos casos, portadores que abandonan el dispositivo por el colector, electrones para el primero y huecos para el segundo).
- La corriente de base I_B : es la señal de control. Es de un valor pequeño y la responsable del valor final de I_C . Se considera que entra a la base en el npn y sale de la base en el pnp (de nuevo, correspondiendo al movimiento de portadores, electrones o huecos respectivamente, hacia fuera del dispositivo).
- La corriente de emisor I_E : debe complementar a las otras dos corrientes teniendo en cuenta que debe cumplirse el principio de conservación de carga (los tres terminales del dispositivo se comportarían como un nodo y tendrían que cumplir la ley de corrientes de Kirchhoff), es decir, $I_C + I_B = I_E$.
- Las diferencias de potencial entre terminales, y en particular V_{BE} (diferencia de potencial entre base y emisor, es decir, $V_B - V_E$) y V_{CE} (diferencia de potencial entre colector y emisor, o sea, $V_C - V_E$).

A la hora de conocer el comportamiento de un transistor, es importante co-

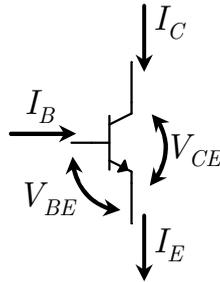


Figura V.9: Corrientes y diferencias de potencial relevantes en un BJT de tipo npn. En un BJT de tipo pnp los signos de las corrientes serían los opuestos.

nocer sus **curvas características**, es decir, el comportamiento de la corriente en el terminal de salida del transistor en función de otras magnitudes eléctricas relevantes (como por ejemplo el voltaje aplicado entre colector y emisor).

En el caso del BJT, nos interesan especialmente sus **características de salida**, que corresponden a los valores de la corriente I_C en función de la diferencia de potencial V_{CE} , usualmente para varios valores de la corriente de base I_B . En la figura V.10 se muestra la corriente I_C frente a V_{CE} para seis valores de I_B , desde 0 hasta $100 \mu\text{A}$, para un determinado BJT⁶.

En las curvas de salida resulta sencillo identificar las posibles regiones de operación del transistor, que resumimos de nuevo en el contexto de estas características:

- **Región de corte:** en ella la corriente de colector es nula, dado que no se ha alcanzado la corriente de base mínima para permitir la conducción. En este caso tanto la unión base-emisor como la unión base-colector están en inversa, y la corriente en el transistor es prácticamente nula.
- **Región activa directa:** es la región *plana* de las curvas. En ella la unión base-emisor está en directa y la unión base-colector en inversa, como ya vimos anteriormente, y la corriente de colector I_C permanece prácticamente constante al aumentar V_{CE} . Un aspecto interesante a destacar es que para incrementos constantes de I_B (por ejemplo, en la figura V.10 cada $20 \mu\text{A}$) se producen incrementos constantes de I_C ; es decir, existe una proporcionalidad entre I_C e I_B . Así, se cumple que $I_C = \beta_F I_B$, siendo β_F la

⁶Estas curvas serán diferentes para cada modelo de BJT, según sus características geométricas, dopajes, etc.

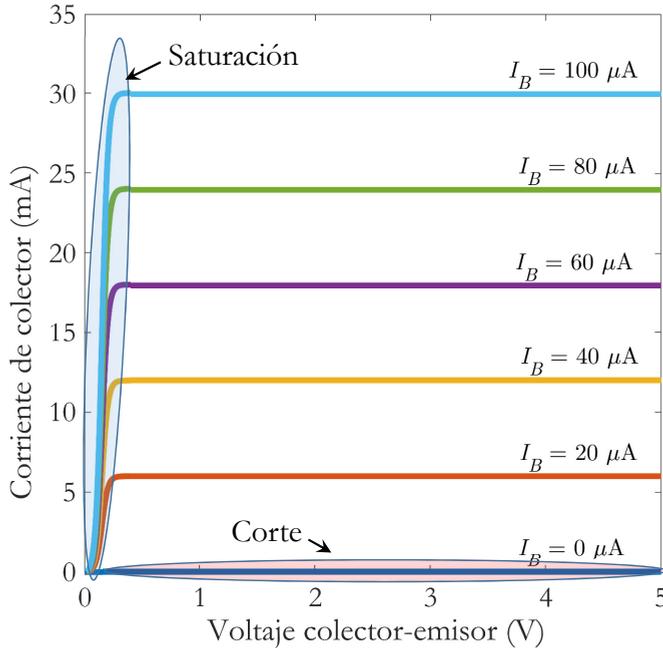


Figura V.10: Características de salida de un BJT.

ganancia en corriente⁷. Los valores habituales de ganancia en corriente de un BJT se sitúan entre 100 y 300, dependiendo del modelo.

- **Región de saturación:** región de aumento fuerte y casi lineal de I_C con V_{CE} . En ella tanto la unión base-emisor como la unión base-colector están en directa.

La región activa directa es de interés para aplicaciones de amplificación de señales, de las cuales hablaremos en el capítulo VIII, mientras que las regiones de saturación y corte son interesantes para aplicaciones digitales⁸.

⁷La ganancia en corriente no tiene unidades: se trata de un número que indica el factor de multiplicación de la corriente de colector en relación a la de base.

⁸Existiría una cuarta posibilidad de funcionamiento, la **región activa inversa**, en la que la unión base-emisor estaría polarizada en inversa y la unión base-colector en directa. Por los perfiles de dopaje empleados en la fabricación de los transistores BJT, esta región no es de interés práctico (el dopaje es asimétrico, estando el emisor más fuertemente dopado), salvo en un caso en particular, las familias lógicas TTL, que se estudiarán en el capítulo XI.

3 POLARIZACIÓN DEL BJT

Para conseguir que un transistor trabaje en alguna de las regiones de interés (corte, saturación o activa directa en el caso del BJT) debemos polarizarlo (aplicarle voltajes y/o corrientes) adecuadamente. La **red de polarización** es el conjunto de resistencias y fuentes que fuerzan al transistor a trabajar con una corriente de colector I_C y un potencial V_{CE} determinados.

Dicha red de polarización establece un **punto de operación estacionario o P.O.E.**⁹, que es el valor de corriente I_C y voltaje V_{CE} concretos que tiene el transistor BJT en situación estacionaria, es decir, cuando estamos en unas condiciones estables de polarización en continua.

Para calcular el valor del punto de operación estacionario, en primer lugar es conveniente determinar la **recta de carga**. En el caso del BJT, la recta de carga es una ecuación que nos indica el valor de I_C frente a los posibles valores de V_{CE} , es decir, $I_C = f(V_{CE})$. La recta de carga es particularmente importante porque nos indica cuáles son los posibles valores del P.O.E. una vez que se ha establecido un voltaje de alimentación en el circuito y se han colocado unas resistencias acompañando al transistor en su colector y emisor.

Para entender cómo se determina la recta de carga y se calcula el P.O.E. para un BJT, vamos a considerar un ejemplo. Imaginemos que disponemos de un circuito en **emisor común** (con el emisor conectado a tierra¹⁰) como el de la figura V.11, con el transistor operando en la región activa directa.

El circuito contiene un BJT de tipo npn (por ello se han dibujado las corrientes I_C e I_B entrando hacia el transistor), dos resistencias (siendo la de base muy elevada, para que la corriente I_B sea pequeña) y una fuente de voltaje ideal de valor V_{CC} .

Si aplicamos la ley de voltajes de Kirchhoff (LVK) a la malla que contiene a la fuente de alimentación y a la resistencia de colector, obtendremos:

$$V_{CC} = I_C R_C + V_{CE}$$

Es decir, el voltaje V_{CC} debe caer en la resistencia R_C y entre colector y emisor del transistor (V_{CE}). De aquí, despejando I_C podemos determinar la recta de carga (I_C en función de V_{CE}) para este circuito:

⁹En ocasiones también se le denomina **punto quiescente o punto Q**.

¹⁰Como ya vimos en el capítulo IV, se trata del punto que tomaremos como referencia de voltaje, es decir, el valor de 0 V.

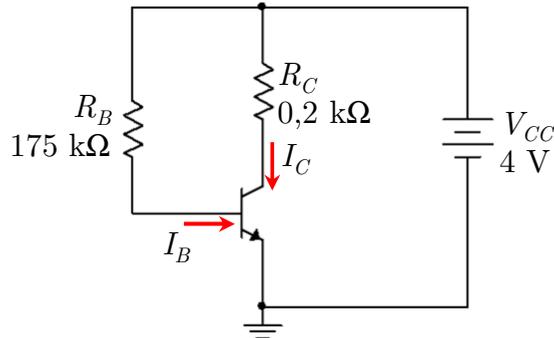


Figura V.11: Circuito con polarización en emisor común (el terminal de emisor del transistor está conectado a tierra).

$$I_C = \frac{V_{CC}}{R_C} - \frac{V_{CE}}{R_C}$$

Esta ecuación corresponde gráficamente a una recta con pendiente negativa, que cortaría al eje vertical en el punto $I_{Cmax} = \frac{V_{CC}}{R_C}$ y al eje horizontal en el punto $V_{CEmax} = V_{CC}$, que, con los valores de nuestro ejemplo, corresponden a 20 mA y 4 V respectivamente.

La recta de carga contiene, como ya se ha señalado, todos los puntos de operación estacionarios posibles de este circuito una vez que se ha fijado el valor de R_C y la V_{CC} conectada a ella. Para conocer el P.O.E. en concreto en el que estamos trabajando, es necesario conocer la corriente de base I_B . Para ello, podemos emplear la LVK en la malla que contiene a V_{CC} y R_B , obteniendo:

$$V_{CC} = I_B R_B + V_{BE}$$

Es decir, el voltaje V_{CC} cae en la resistencia R_B y entre la unión base-emisor (V_{BE}). Dado que estamos considerando que el transistor opera en la región activa directa, la unión base-emisor estará polarizada en directa. En ese caso, el voltaje V_{BE} tomará un valor prácticamente constante, correspondiente al voltaje umbral de dicha unión (aproximadamente 0,5 V). Con los valores del ejemplo, si despejamos la corriente de base de la anterior ecuación obtenemos que su valor es de 20 μA .

Para poder determinar entonces el P.O.E. será necesario tener algunos datos sobre el transistor concreto que estemos empleando en el circuito. O bien debere-

mos conocer su ganancia en corriente β_F o bien sus curvas de salida. Si tomamos en consideración un transistor con curvas como las mostradas en la figura V.10, sobre ellas podemos dibujar la recta de carga y determinar el P.O.E. como el punto de corte entre esta recta y la curva de salida para $I_B = 20 \mu\text{A}$, como se muestra en la figura V.12. Así, para este caso concreto el P.O.E. correspondería a $\{I_C = 6 \text{ mA}, V_{CE} = 2,8 \text{ V}\}$

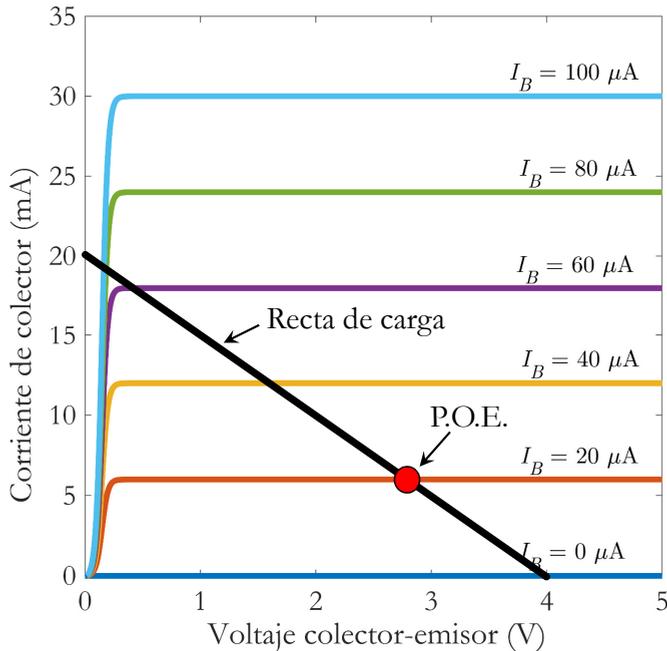


Figura V.12: Determinación gráfica del P.O.E. para el ejemplo considerado.

Si por el contrario no conociésemos las curvas de salida, tendríamos que recurrir a la ganancia en corriente (β_F). Para el transistor considerado su ganancia en corriente es igual a 300. Al estar operando en la región activa directa, sabemos que entre la corriente de colector y la de base existe la relación:

$$I_C = \beta_F I_B$$

En nuestro caso como obtuvimos que la corriente de base valía $20 \mu\text{A}$, ello nos da directamente la I_C de 6 mA. Despejando de la ecuación de la recta de carga, obtenemos la otra magnitud del P.O.E., V_{CE} , con su valor de 2.8 V.

4 CIRCUITO EQUIVALENTE DE PEQUEÑA SEÑAL

¿Por qué es importante saber calcular el valor del P.O.E.? El motivo es que dependiendo del tipo de aplicación que queramos construir con nuestro circuito, deberemos elegir un punto de operación estacionario en alguna de las regiones de operación del transistor. Por ejemplo, la región activa directa, como ya hemos comentado, es especialmente interesante para construir circuitos que trabajan con señales variables en el tiempo, amplificando el valor de su oscilación: en definitiva, el objetivo es aumentar la amplitud de señales AC, como veremos en el capítulo VIII. Se trabajará por tanto con señales variables en el tiempo, que variarán respecto a un cierto valor de continua.

Estas señales, en nuestro caso, van a ser las magnitudes correspondientes a los voltajes y corrientes característicos de un transistor. Así, en un instante dado, tendremos que para un transistor BJT su voltaje colector-emisor, su corriente de colector y su corriente de base serán:

$$V_{ce}(t) = V_{CE} + v_{ce}(t) \quad (\text{V.1})$$

$$I_c(t) = I_C + i_c(t) \quad (\text{V.2})$$

$$I_b(t) = I_B + i_b(t) \quad (\text{V.3})$$

Donde V_{CE} , I_C y I_B son los valores de continua correspondientes al punto de operación estacionario (que asumimos constantes en el tiempo), y $v_{ce}(t)$, $i_b(t)$ e $i_c(t)$ son las magnitudes en alterna, la parte variante de las señales con el tiempo¹¹.

Para el análisis de circuitos analógicos de pequeña señal (con magnitudes con parte variable de pequeña amplitud en torno a un punto de operación estacionario) es posible definir un circuito equivalente de pequeña señal que represente, mediante resistencias, fuentes y capacidades el comportamiento circuital del BJT. En la figura V.13 mostramos un circuito equivalente ideal para un transistor BJT operando en la región activa directa.

El circuito consta de una resistencia entre base y emisor (r_π), una fuente de corriente de valor $\beta_f i_b$ que fija la corriente de salida i_c (que será proporcional a la

¹¹La variación de la señal de entrada debe ser de pequeña amplitud (pequeña señal) para garantizar un proceso de amplificación lo más fiel posible, como veremos más adelante.

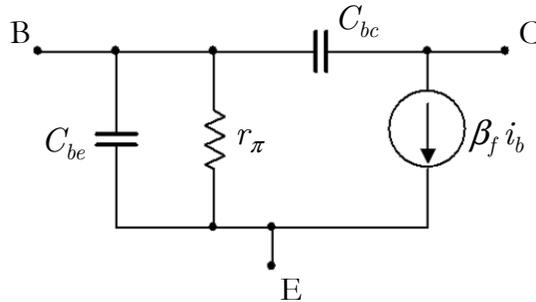


Figura V.13: Circuito equivalente ideal de pequeña señal de un BJT.

corriente de base), así como dos capacidades entre los terminales base-colector y base-emisor. El parámetro β_f es la ganancia en corriente en alterna, que vale:

$$\beta_f = \frac{i_c}{i_b} \quad (\text{V.4})$$

Su valor es similar a la de la ganancia en continua β_F . Si las señales AC son de baja frecuencia, podrá despreciarse el efecto de las capacidades, pues su impedancia será muy elevada, y eliminarlas del circuito equivalente, considerándolas como un circuito abierto. Asimismo, podrán incluirse otros elementos en el circuito equivalente para el tratamiento de fenómenos no ideales, como por ejemplo una resistencia entre el emisor y el colector, que idealmente debería ser nula pero que en la realidad tiene un pequeño valor, haciendo que las curvas de salida no sean completamente planas en la región activa directa.

VI.

EL TRANSISTOR
MOSFET

Las décadas iniciales de la electrónica del estado sólido estuvieron protagonizadas por el transistor bipolar. Sus extraordinarias prestaciones, especialmente su rapidez de respuesta, permitieron enormes avances en telecomunicaciones y amplificación de señales, siendo hoy en día aún muy empleado en estos ámbitos. Gracias a él, fueron posibles también las primeras aplicaciones digitales, pero en este caso pronto se iba a ver superado por un tipo diferente de transistor, capaz de realizar las mismas funciones básicas, pero con un principio de funcionamiento muy distinto. Un tipo diferente de transistor cuyas bases teóricas se habían establecido muchos años antes, a mediados de los años 20 del siglo pasado, pero que tardó más de treinta años en ser viable tecnológicamente, pues hasta 1959 no se pudo fabricar en la configuración que conocemos hoy en día. Se trata del transistor **MOSFET**. La palabra MOSFET es en realidad el acrónimo de *Metal-Oxide-Semiconductor Field Effect Transistor*, que podemos traducir por *Transistor de efecto de campo metal-óxido-semiconductor*. Dos conceptos fundamentales subyacen tras estas palabras:

- **Metal-Óxido-Semiconductor (MOS)**: hace referencia a la estructura física esencial en la que se basa este tipo de transistor, consistente en el apilamiento de tres materiales distintos. Por una parte, un metal (o un material altamente conductor) que actuará como contacto, y ejercerá el papel de terminal de control con el que regularemos la posibilidad de conducir corriente. Seguidamente, un óxido o dieléctrico, que actúa como separador entre el metal del contacto y el material semiconductor: este óxido es una barrera al paso de portadores, por lo que no podrá circular corriente por el terminal de control (esta es una característica importante del transistor MOSFET). Y finalmente, el semiconductor: la parte más cercana al óxido va a constituir el canal conductor, por donde circularán los electrones (o

huecos, en su caso).

- **Efecto de campo:** este término se refiere a que la modulación o regulación de la conductividad del canal conductor va a realizarse mediante un campo eléctrico perpendicular al mismo. En este caso, dicho campo va a generarse a través de la aplicación de un potencial en el terminal de control, siendo ésta otra característica importante del transistor MOSFET: la señal eléctrica responsable de regular el flujo de corriente es un voltaje, y no una corriente, como ocurría en el BJT.

El transistor MOSFET es la pieza fundamental de la electrónica digital actual, lo que es prácticamente equivalente a decir que lo es de la electrónica en general, y de la informática en su conjunto. Desde 1971 ha sido el bloque de construcción elemental con el que se han constituido la práctica totalidad de microprocesadores y memorias RAM. El transistor MOSFET de silicio es, en definitiva, el dispositivo que ha hecho posible la revolución electrónica e informática de nuestro tiempo.

1 PRINCIPIO FÍSICO

Para entender cómo funciona un transistor MOSFET debemos comenzar por la estructura metal-óxido-semiconductor, o simplemente MOS. Como ya hemos señalado, el primer elemento es un contacto¹. Este contacto recibe el nombre de terminal de **puerta** (*gate* en inglés²). El óxido de la siguiente capa suele ser SiO_2 , es decir, dióxido de silicio³. Su espesor es muy delgado, pero no tanto como para no actuar como barrera al paso de portadores a su través. Debajo del óxido, un semiconductor (usualmente silicio) que debe estar dopado. Así, tendremos dos posibilidades, que se trate de un dopaje de tipo n o de tipo p (ambos serán útiles - y necesarios- para la realización de circuitos digitales). El semiconductor recibe el nombre de **sustrato**. Vamos a considerar inicialmente el caso de tener un sustrato

¹Este contacto puede estar fabricado con un metal o con polisilicio de alta conductividad, que se comporta casi de forma metálica. A efectos prácticos y por simplicidad supondremos que se trata de un metal.

²De forma simplificada usaremos la letra G para referirnos a este contacto, por la inicial de su denominación en inglés.

³El hecho de que el silicio disponga de un óxido propio de excelente calidad es una de las causas del éxito de este semiconductor para su uso en electrónica. Sin embargo, en la fabricación de los MOSFETs desde 2010 se emplean también de forma habitual otros tipos de óxido, los dieléctricos de alta permitividad, que permiten un mejor rendimiento en dispositivos de tamaño ultrarreducido, como veremos en el capítulo XII.

de tipo p (figura VI.1), en el que los portadores mayoritarios serán los huecos y los minoritarios los electrones. En ese caso, consideraremos que el sustrato está conectado a tierra (0 V), mientras que en el terminal de puerta aplicaremos un voltaje V_G .

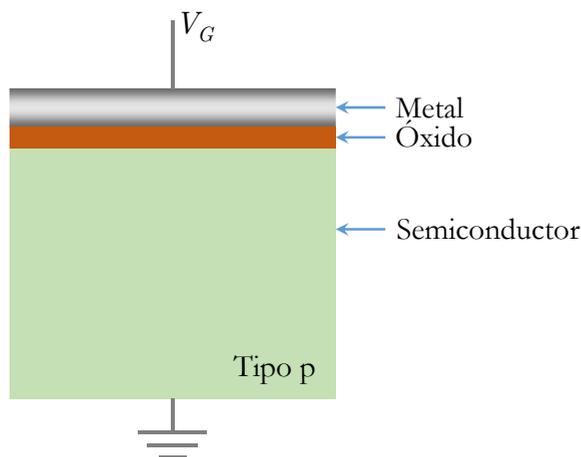


Figura VI.1: Estructura Metal-Óxido-Semiconductor con sustrato de tipo p.

Dependiendo del valor del potencial V_G tendremos diferentes situaciones respecto a la carga eléctrica libre en la zona cercana a la interfaz semiconductor-óxido. Esa zona próxima a la interfaz la denominaremos **canal conductor**. Las situaciones posibles son las siguientes (figura VI.2):

- **Acumulación:** ocurre por lo general si se aplica una diferencia de potencial elevada, negativa para semiconductor p y positiva para semiconductor n. Se produce una agrupación de portadores mayoritarios en la zona del canal conductor.
- **Vaciamiento:** en ese caso en el canal conductor (recordemos, la parte del semiconductor más cercana a la interfaz con el óxido) no existen portadores libres, ni mayoritarios ni minoritarios.
- **Inversión:** consiste en la formación de una capa delgada de portadores minoritarios (electrones en semiconductor tipo p y huecos en semiconductor tipo n) con una concentración muy elevada, superior a la de mayoritarios en equilibrio, que denominaremos **capa de inversión**.

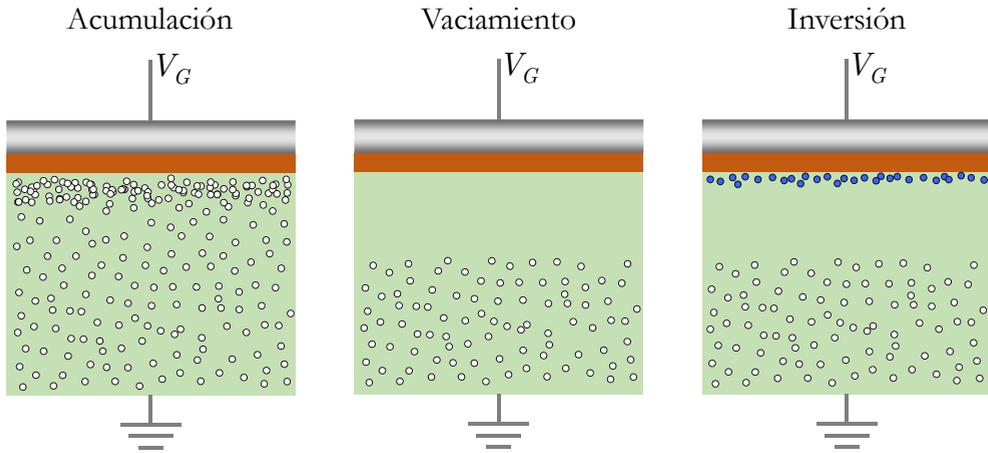


Figura VI.2: Acumulación, vaciamiento e inversión en una estructura MOS con sustrato de tipo p.

Desde el punto de vista práctico, el régimen de inversión es especialmente interesante porque mediante el potencial de puerta podremos regular de manera muy precisa la cantidad de carga en el canal (la concentración de minoritarios en esa zona), y por tanto su conductividad. Cuando queramos que el transistor no conduzca lo llevaremos a vaciamiento o incluso acumulación. El resto de la estructura del transistor será la responsable del comportamiento final del dispositivo, pero en definitiva, el papel de la puerta es clave para regular la existencia o no del canal conductor.

La estructura del transistor MOSFET se completa extendiendo a lo ancho el sustrato e incluyendo dos islas de dopaje opuesto al de aquel, las islas de **fuelle** y **drenador** (*source* y *drain* en inglés, respectivamente⁴). Son zonas de dopaje elevado (lo cual indicaremos con un símbolo $+$), y disponen además de contactos metálicos directamente adyacentes al semiconductor⁵. Dependiendo del tipo de dopaje del sustrato tendremos dos tipos de transistores: los MOSFET **canal n** (con sustrato tipo p, y en los que la conducción será por electrones que forman la capa de inversión) y los MOSFET **canal p** (sustrato tipo n, con conducción por huecos que forman la capa de inversión⁶).

La figura VI.3 nos da una idea más clara del funcionamiento de este tipo de

⁴Para referirnos a ellas de forma simplificada emplearemos las iniciales S y D. El sustrato (*bulk* en inglés) será referenciado con la letra B.

⁵Estos contactos son de tipo **ohmico**, es decir, permiten el paso de corriente en ambos sen-

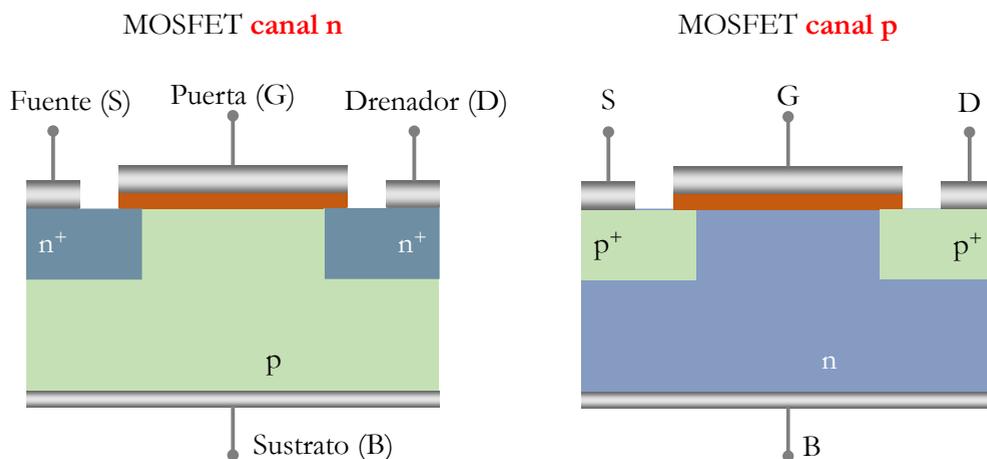


Figura VI.3: Tipos de transistor MOSFET: canal n (izquierda) y canal p (derecha). Usualmente el contacto de sustrato está cortocircuitado con la fuente, por lo que los tres terminales del transistor son fuente (S), puerta (G) y drenador (D).

transistor. Usualmente, los terminales de sustrato (B) y fuente (S) están cortocircuitados, es decir, unidos eléctricamente. Quien determinará por tanto el estado de acumulación, vaciamiento o inversión en el canal es la diferencia de potencial entre la puerta y la fuente, es decir, V_{GS} ⁷. Si tiene un valor suficiente como para formar una capa de inversión, ésta actuará como una especie de puente o enlace entre las dos islas de fuente y drenador. Cuando apliquemos una diferencia de potencial V_{DS} el MOSFET estará en condiciones de conducir corriente (figura VI.4) y por tanto mover los portadores⁸, que entrarán por el terminal de fuente, pasarán por el canal conductor y saldrán del dispositivo por el terminal de drenador. En el caso del MOSFET canal n, estos portadores serán electrones, por lo que la corriente tendrá sentido opuesto a su movimiento, es decir, consideraremos que la corriente entra por el drenador y sale por la fuente. Para el MOSFET

tidos, con una resistencia muy baja.

⁶La inmensa mayoría de transistores MOSFET trabajan en modo de inversión. Existe también la posibilidad de fabricar transistores que funcionen en modo de acumulación, pero su uso práctico es en general muy limitado, pues el control de la corriente a través del sustrato es problemático en este caso.

⁷Por simplicidad nos referiremos a esta magnitud como voltaje de puerta, aunque en propiedad corresponde a la diferencia de potencial entre puerta y fuente.

⁸Si no se aplica una diferencia de potencial entre drenador y fuente no habrá movimiento de portadores y la corriente será nula.

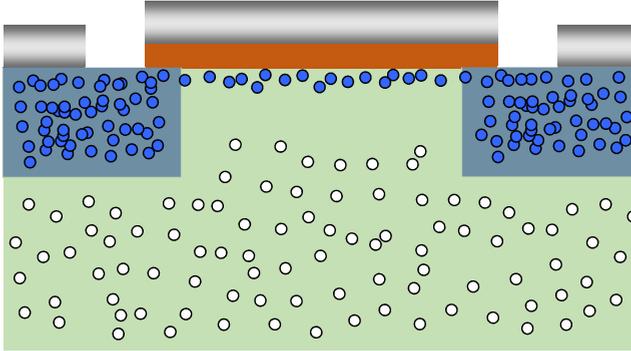


Figura VI.4: Situación en un MOSFET canal n con voltaje V_{GS} suficiente para permitir la conducción: los portadores pueden fluir por el canal conductor formado entre fuente y drenador.

canal p, con conducción por huecos, el sentido de la corriente será el mismo que el del movimiento de los huecos, es decir, saliendo por el drenador.

En caso de que el voltaje V_{GS} no sea suficiente como para trabajar en inversión, no existirá corriente aunque apliquemos una diferencia de potencial V_{DS} , dado que la puerta estará bloqueando el paso de electrones (o, respectivamente, huecos) entre las dos islas de fuente y drenador, y no existe un canal conductor como tal (figura VI.5).

Existe por tanto un valor mínimo de voltaje de puerta V_{GS} que se debe aplicar para que el transistor comience a conducir. Este voltaje recibe el nombre de **vol-**

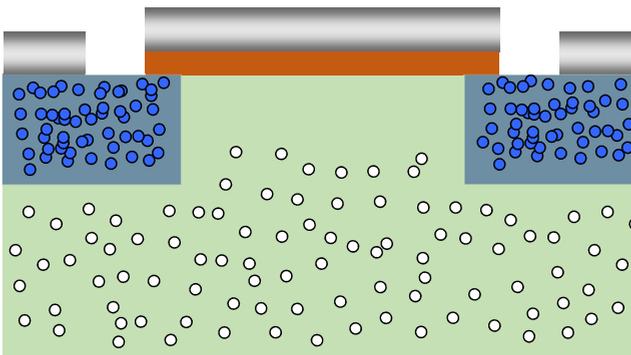


Figura VI.5: Situación en un MOSFET canal n con voltaje V_{GS} insuficiente: no existe un canal conductor que conecte fuente y drenador.

taje umbral (*threshold voltage* en inglés), y lo denotamos como V_T . En definitiva, el voltaje de puerta V_{GS} regula en primer lugar la posibilidad de que el transistor conduzca: si no se alcanza el voltaje umbral no existe un canal conductor (capa de inversión) que una las islas de fuente y drenador. Creado el canal conductor (cuando aplicamos un V_{GS} suficiente), la corriente se produce al aplicar una diferencia de potencial V_{DS} entre drenador y fuente, que pone en marcha a los portadores del canal.

El valor concreto del voltaje umbral depende fundamentalmente del tipo de metal que conforme el contacto de puerta, del espesor y tipo del óxido, y del sustrato semiconductor y su dopaje. En función de estos factores, V_T puede ser positivo o negativo tanto para transistores canal n como para canal p, y ello conformará diferentes subtipos de transistores MOSFET. Podremos tener por tanto dos situaciones (tanto para un MOSFET canal n como para un MOSFET canal p):

- Si con $V_{GS} = 0$ V no existe una capa de inversión (y por tanto no puede haber corriente), decimos que se trata de un MOSFET **de realce, de enriquecimiento o *normally off***.
- Si con $V_{GS} = 0$ V sí existe una capa de inversión (y por tanto hay corriente si se aplica un V_{DS}) decimos que se trata de un MOSFET **de vaciamiento, de empobrecimiento o *normally on***.

Cuando utilicemos un MOSFET en un circuito, lo representaremos mediante un símbolo, que será ligeramente diferente dependiendo de si es canal n o canal p y de si es de realce o de vaciamiento. Los símbolos que emplearemos en este texto se muestran en la figura VI.6: hemos considerado que sustrato y fuente se encuentran siempre cortocircuitados, dado que es la configuración más habitual⁹.

Los MOSFET canal n y canal p operan con voltajes de signo distinto, dado que en un caso son electrones y en el otro huecos los portadores involucrados. De manera general, por tanto, tendremos que:

- Un MOSFET canal n podrá conducir si $V_{GS} - V_T > 0$ V y $V_{DS} > 0$ V (es decir, el voltaje de puerta debe ser más positivo que el umbral y el voltaje de drenador debe ser positivo).

⁹Existen otras variantes de símbolos ligeramente diferentes que también se emplean, dependiendo de si se tiene en cuenta o no el contacto de sustrato, por ejemplo. En el caso del diseño lógico CMOS emplearemos una versión simplificada de los símbolos del MOSFET para hacer más claros y sencillos los diagramas, como veremos más adelante.

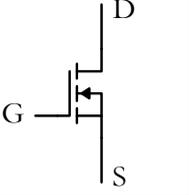
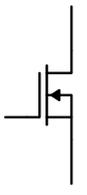
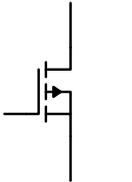
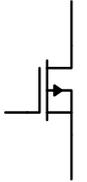
	MOSFET de realce (<i>normally off</i>)	MOSFET de vaciamiento (<i>normally on</i>)
MOSFET Canal n		
MOSFET Canal p		

Figura VI.6: Símbolos circuitales de los diferentes tipos de transistor MOSFET.

- Un MOSFET canal p podrá conducir si $V_{GS} - V_T < 0$ V y $V_{DS} < 0$ V (es decir, el voltaje de puerta debe ser más negativo que el umbral y el voltaje de drenador debe ser negativo).

Dependiendo de los voltajes aplicados el transistor MOSFET se encontrará en diferentes regiones de operación, como ocurría en el caso del BJT. La primera de ellas ya la conocemos: se trata de la **región de corte** o estado **off**, que ocurre cuando el voltaje de puerta no supera al umbral (es decir, $V_{GS} - V_T < 0$ V para un MOSFET canal n y $V_{GS} - V_T > 0$ V para un MOSFET canal p). Independientemente de la diferencia de potencial entre drenador y fuente, V_{DS} , no habrá corriente en el transistor, debido a la ausencia de un canal conductor.

En caso de que el voltaje de puerta sea suficiente para crear un canal conductor ($V_{GS} - V_T > 0$ V para un MOSFET canal n y $V_{GS} - V_T < 0$ V para un MOSFET canal p) estaremos en un estado de conducción (**on**), pero el régimen de funcionamiento concreto dependerá del valor de V_{DS} . Consideremos el caso de un MOSFET canal n con voltaje de puerta superior al umbral, y supongamos que incrementamos progresivamente el valor de V_{DS} , empezando desde un valor igual a 0 V. Mientras el voltaje de drenador V_{DS} sea pequeño, el canal conductor será relativamente homogéneo y se comportará desde el punto de vista eléctrico como una resistencia de valor bajo, es decir, dejará fluir la corriente

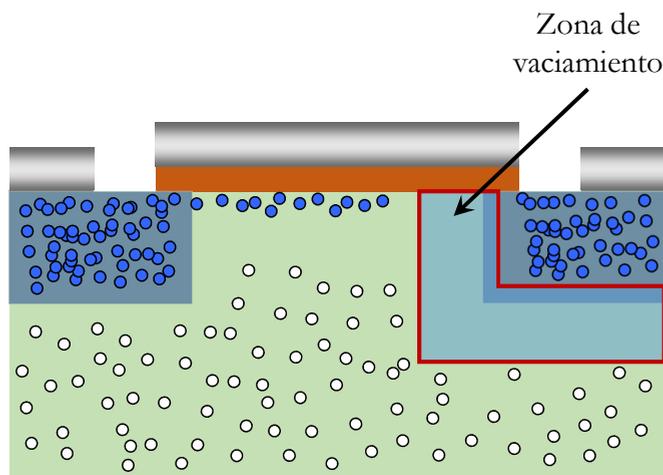


Figura VI.7: Aumento de la zona de vaciamiento en la interfaz entre la isla de drenador y el sustrato, y aparición del pinzamiento del canal.

con facilidad¹⁰. En este caso decimos que el transistor se encuentra en la **región triodo**¹¹.

Pero si continuamos incrementando el valor de V_{DS} , comienza a ocurrir un fenómeno interesante. La isla de drenador y el sustrato forman una unión de tipo diodo, en la que el voltaje más positivo (para nuestro caso de transistor canal n) se aplica en el lado n de la unión, es decir, en la isla de drenador. Ello implica que esta unión p-n está polarizada en inversa¹². En consecuencia, debe crecer la zona de vaciamiento en la interfaz entre los dos tipos de semiconductor a medida que aumentamos V_{DS} . Esta zona de vaciamiento, como ya vimos en el capítulo IV, se caracteriza por la ausencia de portadores libres en ella. Por tanto, en el extremo del canal más cercano a la zona de drenador se produce la competencia entre dos fenómenos: por una parte, la puerta tiende a atraer portadores minoritarios (electrones en nuestro caso de MOSFET canal n, que forman la capa de inversión), y la unión p-n polarizada en inversa que constituyen la isla de drenador y el sustrato tiende a expulsar cualquier tipo de portador, incluidos esos electro-

¹⁰El valor de la resistencia equivalente dependerá de la cantidad de carga que exista en el canal conductor, lo que viene controlado por el voltaje de puerta: a mayor voltaje de puerta, mayor carga en el canal y menor resistencia.

¹¹También podemos referirnos a ella como **región óhmica** o **región lineal**.

¹²Para el caso de un transistor canal p, el voltaje V_{DS} sería negativo, y estaría aplicado en una isla de drenador de tipo p, por lo que la unión también estaría polarizada en inversa.

nes. El canal conductor por tanto deja de ser homogéneo y ya no podemos decir que se comporta como una resistencia, pues al aumentar V_{DS} la corriente tenderá a crecer menos de lo que lo hacía en la región triodo. Cuando la zona de vaciamiento se impone finalmente frente al voltaje de puerta en el extremo de drenador del canal conductor tiene lugar el **estrangulamiento** o **pinzamiento del canal** (figura VI.7), en el que la capa de inversión desaparece en esa zona del canal. A partir de ese momento la corriente permanece constante¹³ a medida que aumentamos V_{DS} . Dicha región de funcionamiento se denomina, en el caso del MOSFET, **región de saturación**.¹⁴

2 CURVAS CARACTERÍSTICAS

Al igual que hicimos con el transistor bipolar en el capítulo V, vamos a analizar la respuesta eléctrica del transistor MOSFET cuando variamos las condiciones, en este caso, de V_{GS} y de V_{DS} . Como ejemplo consideraremos un transistor MOSFET canal n de realce: para el resto de posibilidades existirán curvas y regiones análogas, con la salvedad de los diferentes signos de los voltajes involucrados.

En primer lugar, vamos a ver la **característica de transferencia**, en la cual representamos el valor de la corriente de drenador (I_D) frente al voltaje de puerta V_{GS} , para un valor de V_{DS} constante (figura VI.8).

Si V_{GS} es inferior al voltaje umbral no existe corriente aunque se aplique V_{DS} (no se ha creado la capa de inversión y por tanto no se ha formado el canal conductor). Esa situación corresponde a la **región de corte** o **estado off**. Para voltajes de puerta superiores al umbral, sin embargo, sí existe corriente, y esta crece con el valor de V_{GS} . Nos encontraríamos por tanto en un estado de conducción o estado *on*.

Veamos a continuación qué ocurre si tenemos un voltaje de puerta constante y suficientemente grande como para crear la capa de inversión. Si representamos la corriente I_D en función de V_{DS} para V_{GS} constante obtendremos la **característica de salida**. En la figura VI.9 mostramos las curvas de salida para varios valores de voltaje de puerta V_{GS} .

En primer lugar, si el voltaje de puerta es inferior al voltaje umbral (como

¹³En dispositivos con longitud de puerta muy pequeña existe un ligero incremento de corriente, debida a los efectos de canal corto.

¹⁴Téngase en cuenta la diferencia de nomenclatura en las regiones de operación del BJT y el MOSFET: la región triodo del MOSFET es equivalente a la de saturación del BJT, mientras que la región de saturación del MOSFET es equivalente a la región activa directa del transistor bipolar.

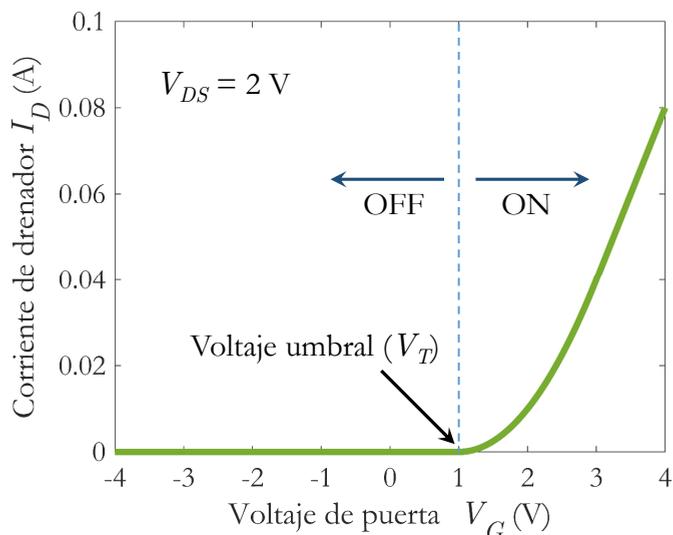


Figura VI.8: Característica de transferencia de un transistor MOSFET canal n de realce, para un determinado valor de V_{DS} . Se indica también el voltaje umbral (1 V en este caso), que marca la separación entre la zona de corte (*off*) y la de conducción (*on*).

ocurre en la figura para el caso de $V_{GS} = 0$ V) la corriente es nula para todos los posibles valores de V_{DS} . El transistor está, por tanto, en la región de corte.

Si el voltaje de puerta es superior al umbral, entonces inicialmente hay una región de crecimiento casi lineal de la corriente al aumentar V_{DS} : se trata, como ya conocemos, de la región triodo (este régimen prácticamente lineal se puede apreciar mejor en la figura VI.9 para el valor de V_{GS} más elevado), donde el transistor se comporta de forma parecida a una resistencia. También podemos observar cómo la pendiente de las curvas es mayor cuanto mayor es V_{GS} . Ello corresponde al hecho de que el aumento del voltaje de puerta supone un incremento de la carga en el canal conductor, aumentando su conductividad (o, equivalentemente, reduciendo la resistencia efectiva del canal).

Al aumentar el valor de V_{DS} , como ya se ha explicado comienza a aumentar la zona de vaciamiento de la unión drenador-sustrato (por estar polarizada en inversa). La pendiente de las curvas decae progresivamente hasta que aparece el fenómeno del pinzamiento (marcado en la figura VI.9 por la línea discontinua), momento a partir del cual la corriente no aumenta aunque se aumente V_{DS} . Dicha corriente constante es la corriente de saturación del transistor, que como vemos

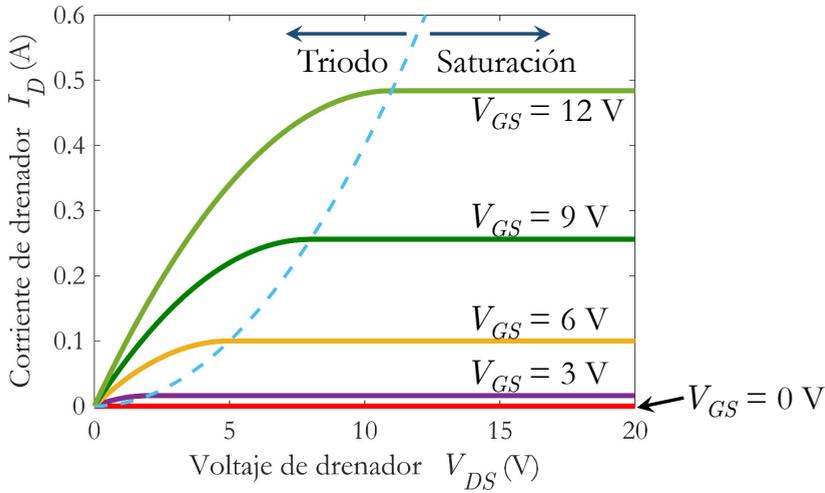


Figura VI.9: Características de salida para varios valores de V_{GS} para un transistor MOSFET determinado, con $K = 4 \text{ mA/V}^2$ y $V_T = 1 \text{ V}$. La línea discontinua marca la transición entre la región triodo y la región de saturación.

depende del valor de V_{GS} considerado.

En el caso del transistor MOSFET, nos va a interesar particularmente poder determinar con precisión la corriente I_D . Para ello podemos emplear en cada una de las regiones de operación una serie de ecuaciones analíticas sencillas¹⁵, que resumimos en la siguiente tabla:

	I_D	Condición Canal n	Condición Canal p
Corte	0	$V_{GS} - V_T < 0$	$V_{GS} - V_T > 0$
Triodo	$K [2(V_{GS} - V_T)V_{DS} - V_{DS}^2]$	$V_{GS} - V_T > 0$ y $V_{DS} < V_{GS} - V_T$	$V_{GS} - V_T < 0$ y $V_{DS} > V_{GS} - V_T$
Saturación	$K(V_{GS} - V_T)^2$	$V_{GS} - V_T > 0$ y $V_{DS} > V_{GS} - V_T$	$V_{GS} - V_T < 0$ y $V_{DS} < V_{GS} - V_T$

K se denomina **coeficiente de transconductancia**, y depende de ciertas características físicas del transistor, como el espesor y tipo del óxido (que deter-

¹⁵Estas ecuaciones son válidas para el caso de un MOSFET de canal relativamente largo (superior a $1 \mu\text{m}$). Para transistores con canales de tamaño más reducido, como los actuales, es necesario introducir factores de corrección para tener en cuenta los efectos de canal corto.

minan su capacidad equivalente por unidad de área $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$, la movilidad de los portadores μ en el canal, y la anchura (W) y longitud (L) de la puerta:

$$K = \frac{\mu C_{ox} W}{2 L} \quad (\text{VI.1})$$

3 POLARIZACIÓN DEL MOSFET

Al igual que en el caso del transistor bipolar, es necesario polarizar adecuadamente el transistor MOSFET para trabajar en un determinado punto de operación estacionario. Para ello nos valdremos de una red de polarización con fuentes de alimentación y resistencias. La elección del punto de operación estacionario es crítica para establecer la región de trabajo del transistor, que dependerá del tipo de aplicación (digital o analógica).

Las magnitudes que definen el punto de operación estacionario del transistor MOSFET son la corriente de drenador I_D y la diferencia de potencial drenador-fuente V_{DS} . Ambas se relacionarán a través de la recta de carga correspondiente, que en este caso será de la forma $I_D = f(V_{DS})$. Como ya sabemos, la recta de carga define todos los posibles P.O.E. del transistor dependiendo de las resistencias y fuentes que le acompañen: el valor de voltaje de puerta aplicado será el que determine el punto de operación estacionario concreto con el que trabajaremos.

Vamos a tomar como ejemplo un transistor MOSFET polarizado en configuración de **frente común** (es decir, con la fuente del transistor directamente conectada a un voltaje fijo, en este caso tierra, como en la figura VI.10).

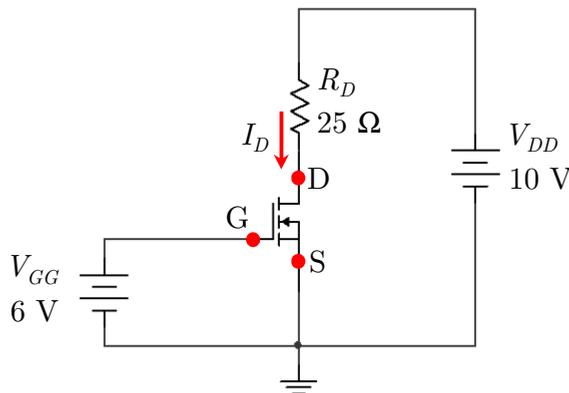


Figura VI.10: Circuito de polarización del transistor MOSFET en fuente común.

En primer lugar, para determinar la recta de carga aplicamos la ley de voltajes de Kirchhoff (LVK) en la malla que contiene al drenador y fuente del transistor y a la fuente de alimentación V_{DD} . De este modo, obtendremos que:

$$V_{DD} = I_D R_D + V_{DS}$$

Es decir, en este circuito el voltaje de la fuente de alimentación cae en la resistencia R_D y entre el drenador y fuente del transistor. De aquí, despejando I_D obtendremos la ecuación:

$$I_D = \frac{V_{DD}}{R_D} - \frac{V_{DS}}{R_D}$$

que es la recta de carga para este ejemplo en particular. Los valores concretos de voltaje de la fuente V_{DD} y de la resistencia R_D establecen, para este caso, cuáles son los posibles valores que puede tomar I_D en función de V_{DS} . Si representamos gráficamente esta ecuación, corresponde a una línea con pendiente negativa, como se observa en la figura VI.11. Para este ejemplo en concreto esta recta corta al eje vertical de la gráfica en el punto $I_{Dmax} = \frac{V_{DD}}{R_D}$, y al eje horizontal en el punto $V_{DSmax} = V_{DD}$, que en nuestro caso valen 0.4 A y 10 V, respectivamente.

Finalmente el P.O.E. estará determinado por el punto de corte entre la recta de carga y la curva de salida correspondiente al valor de V_{GS} de nuestro problema,

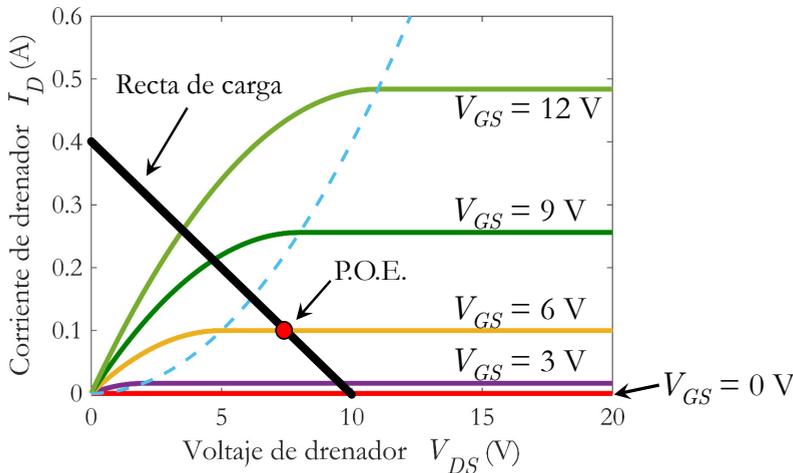


Figura VI.11: Recta de carga y punto de operación estacionario para el ejemplo considerado.

en este caso, 6 V. Las coordenadas de dicho punto (los valores de I_D y V_{DS} en definitiva) serán el punto de operación estacionario (en el ejemplo, 0.1 A y 7.5 V).

En caso de que no dispusiésemos de las características de salida, sería necesario tener datos sobre el transistor, como su factor de transconductancia o su voltaje umbral. En el ejemplo que nos ocupa, $K = 4 \text{ mA/V}^2$ y $V_T = 1 \text{ V}$. Dado que en este ejemplo el voltaje aplicado en la puerta es superior al voltaje umbral, el transistor está en estado *on*, o bien en saturación o bien en triodo. En cualquier caso tendremos dos ecuaciones (la recta de carga y la ecuación correspondiente de I_D) y dos incógnitas (I_D y V_{DS} , las variables que determinan el punto de operación estacionario). Resolviendo el sistema de ecuaciones obtendríamos el P.O.E.

4 CIRCUITO EQUIVALENTE DE PEQUEÑA SEÑAL

Como ya sabemos, fijar un punto de operación estacionario es el primer paso para construir una aplicación práctica con un transistor. Dependiendo del régimen de operación en el que se encuentre dicho P.O.E., tendremos diferentes tipos de aplicaciones. En el caso del transistor MOSFET, por ejemplo, las **aplicaciones digitales** necesitarán que trabaje entre dos estados, uno de conducción nula y otro de alta conductividad: es decir, se trabaja en la región de corte o en la región triodo.

En cambio, las **operaciones sobre señales analógicas** harán trabajar al transistor MOSFET en la región de saturación, donde existe la posibilidad de amplificar señales, como veremos en el capítulo VIII. En este último caso trabajaremos con señales variables en el tiempo, que oscilarán en torno al punto de operación estacionario en continua. Para un transistor MOSFET las magnitudes relevantes serán su voltaje drenador-fuente, su corriente y su voltaje de puerta:

$$V_{ds}(t) = V_{DS} + v_{ds}(t) \quad (\text{VI.2})$$

$$I_d(t) = I_D + i_d(t) \quad (\text{VI.3})$$

$$V_{gs}(t) = V_{GS} + v_{gs}(t) \quad (\text{VI.4})$$

Donde V_{DS} , I_D y V_{GS} son los valores de continua del punto de operación estacionario (que asumimos constantes en el tiempo), y $v_{ds}(t)$, $i_d(t)$ y $v_{gs}(t)$ son

las magnitudes en alterna, la parte variante de las señales con el tiempo y que, como ya comentamos en el caso del transistor bipolar, deben ser de pequeña amplitud (pequeña señal) para que la señal amplificada sea fiel a la señal original, como veremos más adelante.

Para el caso del transistor MOSFET podemos definir un **circuito equivalente de pequeña señal**, que facilitará realizar los cálculos en circuitos amplificadores. Si consideramos un MOSFET ideal, un posible circuito equivalente es el mostrado en la figura VI.12.

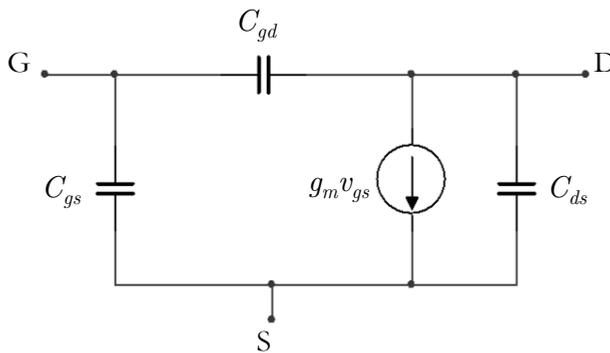


Figura VI.12: Circuito equivalente de pequeña señal del transistor MOSFET ideal.

Este circuito consta de varias capacidades entre los tres terminales (dos asociadas al contacto de puerta y una entre drenador y fuente) y una fuente de corriente de valor $i_d = g_m v_{gs}$ entre drenador y fuente. Es decir, la corriente de pequeña señal i_d es directamente proporcional al voltaje de puerta de pequeña señal v_{gs} .

El factor g_m se denomina **transconductancia**, y se calcula (para señales en alterna de pequeña amplitud) como:

$$g_m = \frac{i_d}{v_{gs}} \quad (\text{VI.5})$$

Una definición más general de la transconductancia se puede realizar como:

$$g_m = \left. \frac{\Delta I_D}{\Delta V_{GS}} \right|_{V_{DS}=cte} \quad (\text{VI.6})$$

La transconductancia se miden en **siemens**¹⁶. Para realizar un análisis de un

¹⁶El siemens es el inverso del ohmio, por lo que sus unidades son Ω^{-1} .

circuito en pequeña señal se reemplaza el transistor por su circuito equivalente y se realizan los análisis correspondientes teniendo en cuenta las leyes de Kirchhoff y los fundamentos de teoría de circuitos.

En el caso de trabajar con señales de baja frecuencia, podríamos prescindir de las capacidades del circuito equivalente de pequeña señal, quedándonos únicamente la fuente de corriente, y encontrándose el terminal de puerta en circuito abierto con la fuente.

Al circuito equivalente de pequeña señal podemos añadirle otros elementos para dar cuenta de situaciones no ideales en el transistor¹⁷.

¹⁷Por ejemplo, es muy común incorporar una resistencia de valor pequeño entre drenador y fuente para dar cuenta del pequeño incremento de corriente en saturación con V_{DS} que se observa en los dispositivos reales; el inverso de esta resistencia se denomina conductancia de salida.

VII.

DISPOSITIVOS
OPTOELECTRÓNICOS

Además de los dispositivos que hemos estudiado hasta el momento, que se centran principalmente en el control del flujo de corriente eléctrica, existe otro tipo de dispositivos que aprovechan las propiedades especiales de la interacción entre la luz y los semiconductores. La **optoelectrónica** es la *rama de la electrónica que trata de los dispositivos que emiten, modulan, transmiten y/o detectan la luz* (normalmente luz visible o cercana en el espectro, como puede ser la luz infrarroja). Existen principalmente dos posibilidades respecto a estos dispositivos:

- Dispositivos que emiten luz (por ejemplo, el LED, el láser, etc.).
- Dispositivos que generan corriente absorbiendo luz (por ejemplo, células solares, fotodetectores, etc.).

Algunos dispositivos combinan ambos efectos en un único chip, como por ejemplo los optoacopladores. En este capítulo vamos a conocer cuáles son los dispositivos optoelectrónicos de uso más común y sus principios de operación y características físicas básicas.

1 DIODOS LED

LED es el acrónimo de *Light Emitting Diode*, o **diodo emisor de luz**. Como su propio nombre indica, se basa en una unión p-n (diodo) de dos terminales, pero que presenta unas características especiales respecto a un diodo normal. Al polarizar este diodo LED en directa, y por encima del voltaje umbral, comienza a conducir corriente eléctrica, emitiendo luz de un color que estará determinado por el material semiconductor con el que esté fabricado el dispositivo.

Para permitir el paso de la luz, el encapsulado del diodo LED no puede ser opaco, sino que debe ser transparente o al menos translúcido. Además, el área

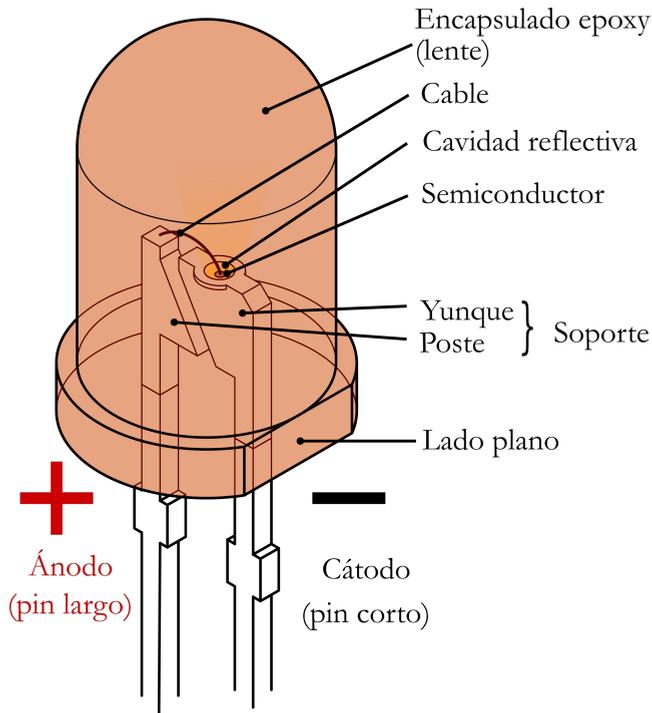


Figura VII.1: Partes internas de un diodo LED.

de la unión p-n debe ser relativamente grande, para que se emita una cantidad apreciable de luz.

En la figura VII.1 se muestran las partes internas de un diodo LED. En primer lugar, el diodo está recubierto por un encapsulado de resina epoxy transparente o translúcido¹, que corresponde a la zona sombreada del dibujo. El diodo tiene dos pines o terminales metálicos, que corresponderán al ánodo (terminal positivo) y cátodo (terminal negativo) de un diodo clásico. En la práctica, en la configuración más habitual de los LEDs discretos estos terminales se diferencian porque el pin correspondiente al ánodo se fabrica con una longitud mayor que el correspondiente al cátodo, por lo que son distinguibles rápidamente a simple vista. Ello nos facilitará su colocación con la orientación adecuada en un circuito. Además, se puede identificar el cátodo por ser el terminal más cercano al corte plano que tiene la base redonda del encapsulado.

¹En el caso de los diodos de color rojo, por ejemplo, el encapsulado está tintado con este mismo color para reforzar su tonalidad cromática.

El semiconductor se sitúa en una parte muy pequeña en la cima del cátodo, con un pequeño cable que une el lado dopado tipo p con el ánodo. Además, está colocado en una cavidad reflectiva que permite una salida más direccional de la luz emitida.

El principio de funcionamiento del LED se basa en el fenómeno de la recombinación, que ya se estudió en el capítulo III. Para desencadenar este fenómeno es necesario polarizar el diodo LED en directa, es decir, hay que forzar que exista una fuerte difusión de electrones del lado n al p y de huecos del p al n. El LED se diseña para que esta fuerte difusión se traduzca en un elevado número de recombinaciones.

En el proceso de recombinación un electrón presente en la banda de conducción *cae* a la banda de valencia y ocupa un hueco allí presente. En ese proceso, el electrón pierde energía (de un valor comparable al del gap del semiconductor). Si esa energía se emite en forma de fotón, se produce la generación de luz. Este es el fenómeno conocido como **electroluminiscencia**, que consiste en *la emisión de fotones (luz) como consecuencia de la recombinación del exceso de portadores creado por la aplicación de un campo eléctrico (o diferencia de potencial)*.

El símbolo de un diodo LED es el mostrado en la figura VII.2. Como vemos, es similar al del diodo convencional, con la única salvedad de las flechas que indican la emisión de luz. Al igual que un diodo normal es un dispositivo con polaridad, por lo que debe conectarse en el circuito con la orientación correcta.



Figura VII.2: Símbolo de un diodo LED.

Como hemos mencionado, el color de la luz emitida está relacionado con el gap del material semiconductor. De manera aproximada, como ya vimos en el capítulo III se cumple que:

$$\varepsilon_{gap} = \varepsilon_{ph} = h \frac{c}{\lambda} \quad (\text{VII.1})$$

donde ε_{gap} es la energía del gap del semiconductor con el que está hecho el LED, ε_{ph} es la energía de los fotones emitidos, h es la constante de Planck, c la velocidad de la luz y λ la longitud de onda, como ya sabemos.

En consecuencia, un diodo LED no podrá emitir radiación electromagnética (luz) de cualquier longitud de onda, sino sólo de aquella que esté relacionada con el material semiconductor con el que esté fabricado. La mayoría de semiconductores poseen valores del gap que corresponden a longitudes de onda entre el infrarrojo (longitudes de onda superiores a 700 nanómetros² o nm, es decir, luz no visible para el ojo humano) y el visible (longitudes de onda que varían entre 700 nm para el color rojo y 400 nm para el color violeta).

En la figura VII.3 se muestra con más detalle estos rangos de longitudes de onda y los valores del gap de algunos semiconductores significativos.

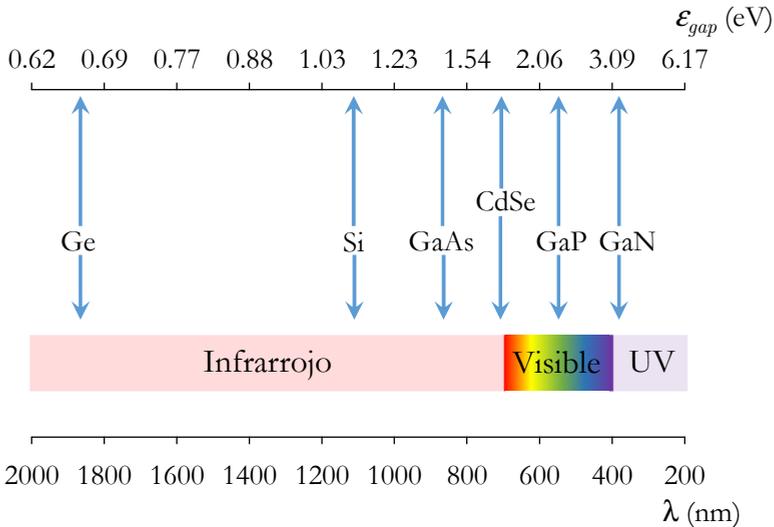


Figura VII.3: Valor del gap de algunos semiconductores y su relación con los rangos infrarrojo, visible y ultravioleta del espectro electromagnético.

El silicio (el semiconductor más ampliamente utilizado) tiene un valor del gap que corresponde al infrarrojo, por lo que no se pueden fabricar LEDs que emitan luz visible con este material. Además, el silicio tiene una característica adicional que hace que no sea un material especialmente eficiente para aplicaciones optoelectrónicas: posee un **gap indirecto**. En los materiales de gap indirecto el proceso de recombinación (y correspondientemente, el de generación) puede ser más complejo que el simple paso de un electrón desde banda de conducción hacia banda de valencia con la siguiente emisión de un fotón. En el silicio, por ejemplo,

²Un nanómetro es la milmillonésima parte de un metro.

el electrón necesita una pequeña ayuda (si se prefiere, un pequeño empujón) para completar la transición, por lo que decimos que se trata de un proceso asistido. Esa ayuda se produce a través de una interacción adicional con los átomos de la red cristalina³, lo que hace que el proceso sea menos eficiente.

En cambio, otros semiconductores poseen **gap directo**: la recombinación puede producirse sin la necesidad de estar asistida por una interacción adicional con la red cristalina, por lo que puede tener lugar con mayor facilidad. Este es el caso por ejemplo del GaAs (arseniuro de galio), lo que lo convierte en un material semiconductor especialmente interesante para aplicaciones optoelectrónicas. En este caso concreto su gap recae también en el infrarrojo.

Para fabricar LEDs que emitan en colores correspondientes al visible, es necesario recurrir a aleaciones semiconductoras, en las que se puede variar el contenido proporcional de cada tipo de elemento para ajustar con precisión el valor del gap. Ello se conoce como **ingeniería del gap**. Los avances en ciencia de materiales han permitido controlar con gran precisión la composición de las aleaciones semiconductoras y disponer de LEDs de todo tipo de colores⁴.

La emisión de luz en un LED no es monocromática (es decir, la intensidad luminosa no se concentra en una única longitud de onda), sino que tiene una cierta amplitud espectral. Así, por ejemplo, un LED de color naranja tiene su máxima intensidad en aproximadamente 600 nm, pero también emite cierta cantidad de luz en longitudes de onda cercanas. Por tanto, la emisión de un LED no es una luz de un color completamente puro, sino una combinación de colores cercanos a una longitud de onda principal.

Al ser un LED un diodo, la corriente que circula por el mismo aumenta muy rápidamente con el voltaje aplicado una vez superado el umbral. Por tanto, se debe tener cuidado al emplearlos en un circuito eléctrico, dado que una diferencia de potencial excesiva entre sus terminales puede producir una corriente demasiado elevada, que acabe quemando o dañando al LED. Para evitar este problema, se

³A temperaturas por encima del cero absoluto los átomos del sólido no están completamente quietos, sino que vibran en torno a sus posiciones en la red cristalina. Estas vibraciones se pueden caracterizar mediante unas cuasipartículas -modos vibracionales- llamadas *fonones*, de cuyo estudio se ocupa la física del estado sólido. Los fonones son los responsables de *ayudar* al electrón a recombinarse, en este caso.

⁴Los primeros LEDs capaces de emitir luz visible fueron desarrollados en 1962: emitían en color rojo. En años posteriores se fueron sucediendo los LEDs de color naranja, amarillo, verde, etc. El LED azul de alta eficiencia se resistió hasta 1993, cuando Shuji Nakamura (n. 1954) logró fabricarlo partiendo del GaN (nitruro de galio). Nakamura recibió el premio Nobel de Física en 2014 por este descubrimiento.

suelen conectar en serie con una resistencia para limitar el valor de corriente eléctrica que circule por el LED. La elección del valor de resistencia debe hacerse teniendo en cuenta la potencia que se disipará en el diodo, cuyo valor máximo suele estar indicado por el fabricante en la **hoja de especificaciones**⁵ del modelo concreto con el que se esté trabajando.

Las **aplicaciones de los diodos LED** son múltiples en la actualidad. Una de las más habituales en la vida cotidiana es la iluminación, reemplazando a las bombillas tradicionales (incandescentes o halógenas). La luz blanca no corresponde a una única longitud de onda, sino a la mezcla de los colores del espectro visible al ojo humano. Por ello, no es posible fabricar un LED que emita directamente luz blanca, sino que para ello debe recurrirse a otras soluciones.

Un LED de luz blanca se puede fabricar de dos formas:

- Empleando tres diodos LED que emitan colores básicos (por ejemplo rojo, verde y azul, LED RGB). Regulando la intensidad de corriente en cada uno de estos tres diodos puede conseguirse cualquier color, lo que hace de estos diodos dispositivos muy versátiles. Sin embargo, son caros y poco eficientes, por lo que no se emplean habitualmente en iluminación.
- Empleando un LED azul que tenga un encapsulado con un recubrimiento fluorescente, de modo que al pasar parte de la luz azul por este recubrimiento se produzca la emisión de colores complementarios, consiguiendo el aspecto de una luz blanca. Esta es la solución habitual en las bombillas LED comerciales.

La luz blanca producida por los diodos LED se caracteriza por una alta eficiencia energética y durabilidad en comparación con las bombillas tradicionales, ya que las pérdidas por calentamiento son notablemente reducidas y los materiales semiconductores son muy robustos siempre que los diodos estén sometidos a una polarización en continua regular y estable⁶. Además de su uso en iluminación doméstica y urbana, los LEDs son fundamentales en las pantallas planas, tabletas y teléfonos móviles, como fuente de retroiluminación de alta eficiencia

⁵La hoja de especificaciones o *datasheet* es un documento en el que el fabricante de un dispositivo electrónico incluye toda la información necesaria para trabajar de forma práctica con él, desde las dimensiones y tamaños hasta las características eléctricas, valores máximos admitidos y rangos de operación.

⁶Una de las causas más comunes por las que las bombillas LED pueden estropearse es por una fuente de alimentación de baja calidad, no por el semiconductor.

y bajo consumo. También son muy empleados como indicadores (en semáforos, *displays*, etc.).

Otra aplicación de uso cotidiano es la transmisión de información. El ejemplo más claro es el mando a distancia. En este caso, la información está codificada en forma de un tren de pulsos de luz infrarroja, que normalmente emplea una señal portadora de 38 KHz. En la mayor parte de mandos a distancia se puede ver el diodo LED en la cara superior del mando a distancia. Al pulsar una tecla, se emite una ráfaga característica que un sensor en el dispositivo controlado (por ejemplo el televisor) es capaz de identificar, actuando en consecuencia (cambiar de canal, subir el volumen, etc.). Esta ráfaga de luz no es visible al ojo humano, pero puede observarse con, por ejemplo, la cámara de un teléfono móvil, cuyo sensor sí se ve alterado por la luz infrarroja.

Los LED infrarrojos también se emplean en transmisión de información a través de fibra óptica, aprovechando la baja atenuación de ciertas longitudes de onda en este medio. Asimismo, se emplean en la parte emisiva de optoacopladores (que estudiaremos posteriormente), sensores de apertura automática de puertas en ascensores, edificios, etc. y otros múltiples usos.

2 LÁSERES DE SEMICONDUCTOR

La palabra *láser* es el acrónimo de *Light Amplification by Stimulated Emission of Radiation* (amplificación de luz por emisión estimulada de radiación). Da nombre a un tipo especial de luz (y por extensión a los dispositivos que la producen) que posee una serie de características especiales. Se trata de una luz altamente energética, coherente (la orientación y fase de las ondas electromagnéticas que la forman son constantes) y altamente monocromática (está muy centrada en una longitud de onda concreta).

El funcionamiento del láser se basa en la **emisión estimulada** de luz (figura VII.4): en el caso de un láser semiconductor, un fotón interactúa con un electrón de la banda de conducción, provoca su recombinación y la aparición de un segundo fotón (con idénticas características que el primero).

Este fenómeno se consigue mediante una estructura de tipo diodo con una serie de características geométricas especiales, que forman una cavidad resonante haciendo que la luz emitida en cierto modo *rebote* en las caras reflectantes y sea amplificada, de forma análoga a como el cuerpo de una guitarra actúa como caja de resonancia y amplifica el sonido. La luz es emitida por una cara que es parcialmente reflectante (figura VII.5), de modo que algunos fotones escapan,

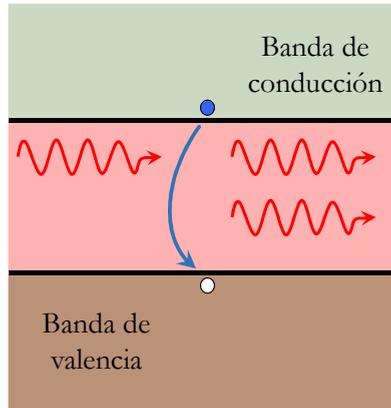


Figura VII.4: Emisión estimulada de luz en un láser semiconductor.

mientras que otros rebotan hacia el interior de nuevo para sostener el proceso de emisión estimulada. En la fabricación de láseres de semiconductor se emplean **heteroestructuras semiconductoras**⁷, acumulaciones de capas de materiales semiconductores con diferentes tamaños de red cristalina y valor del gap.

Las aplicaciones de los láseres en la actualidad son múltiples y abarcan numerosos ámbitos, siendo imposible detallarlas todas. Citaremos aquí simplemente algunas de las más comunes:

- **Industriales:** sistemas de alineamiento o localización, medida de distancias, corte de piezas a medida, etc.
- **Científicas:** contadores de partículas, física atómica, análisis químicos, etc.
- **Médicas:** instrumentos de corte y cauterización, estética, fotocoagulación, etc.
- **Artes gráficas:** impresoras láser, fotocopadoras.
- **Comunicaciones:** fibras ópticas.
- **Audiovisuales:** lectores de CD, DVD, Blu-ray, etc.

⁷Zhores Alferov (1930-2019) y Herbert Kroemer (n. 1928) recibieron el premio Nobel de Física en el año 2000 por sus descubrimientos de heteroestructuras semiconductoras, que hicieron posible en 1970 la aparición del láser semiconductor de diodo capaz de funcionar a temperatura ambiente.

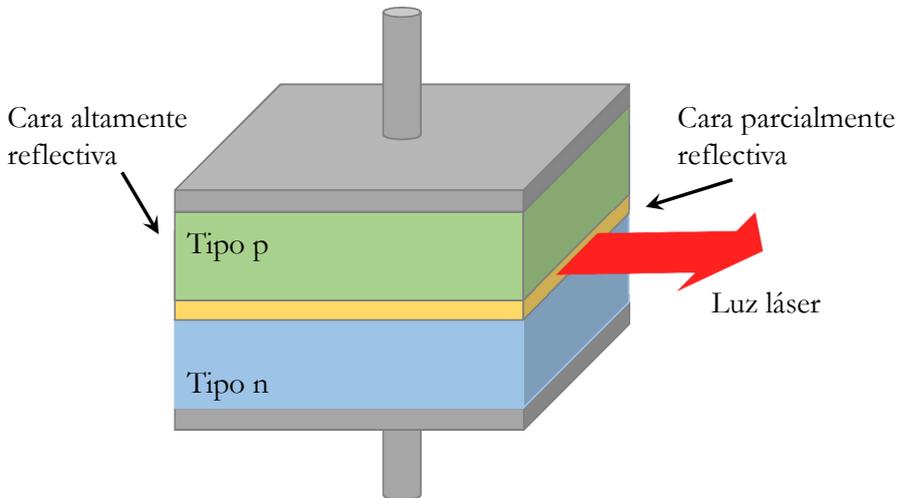


Figura VII.5: Estructura de un láser semiconductor.

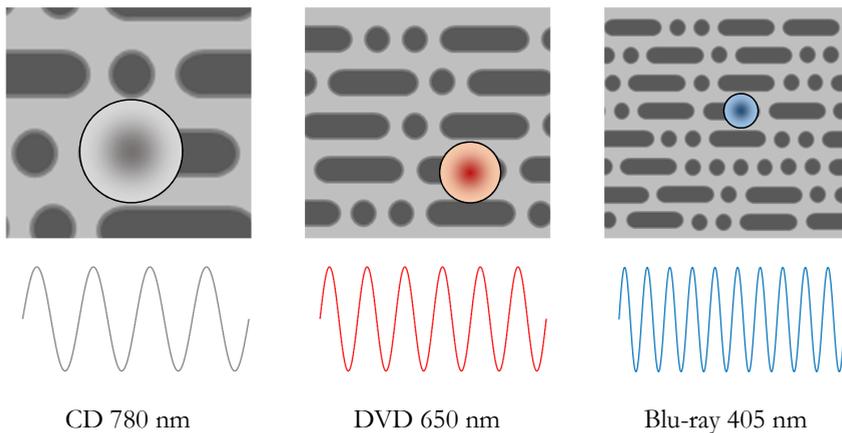


Figura VII.6: Láseres de semiconductor empleados en sistemas de almacenamiento de información en disco. Cuanto menor es la longitud de onda del láser, más pequeñas pueden ser las marcas en el disco, y por tanto, mayor la densidad de información.

La evolución de la tecnología de semiconductores explica a su vez la evolución de la capacidad de almacenamiento de los sistemas ópticos como los CD, DVD o Blu-ray. Todos ellos son discos del mismo tamaño (12 centímetros de diámetro) pero su capacidad es muy diferente: 700 MB el CD, 4.7 GB el DVD y

25 GB el Blu-ray⁸. En todos ellos, la información digital está codificada en forma de pequeñas cavidades en la superficie del disco (figura VII.6). El láser se emplea para leer⁹ la información detectando si en la superficie hay o no una cavidad a medida que el disco gira. El tamaño de la cavidad (que está relacionado con cuánta información cabe en el disco) está limitado por la longitud de onda del láser empleado para su lectura. Por ejemplo, un láser de color azul (con longitud de onda de 405 nm) podrá leer cavidades con tamaño menor que un láser rojo (con longitud de onda de 700 nm). Por tanto, el aumento de la capacidad de almacenamiento ha estado directamente ligado a la existencia de fuentes de luz láser de longitud de onda cada vez menor, lo que se ha conseguido gracias a la investigación en nuevos materiales semiconductores.

3 FOTODETECTORES

Los **fotodetectores** son dispositivos basados en el fenómeno de la **fotoconductividad**, que consiste en el *aumento de la conductividad (y en consecuencia de la corriente) por iluminación con luz de longitud de onda adecuada*.

Ello es posible gracias a la **generación** de portadores, fenómeno que ya estudiamos en el capítulo III. En este caso, un fotón impacta con un electrón en la banda de valencia, que absorbe la energía del fotón, y si esta energía es mayor que el valor del gap del semiconductor¹⁰, el electrón podrá alcanzar la banda de conducción, formándose un nuevo par electrón-hueco, aumentando la concentración de portadores y en consecuencia la conductividad del material.

La energía del fotón debe cumplir, por tanto:

$$\varepsilon_{ph} = h \frac{c}{\lambda} \geq \varepsilon_{gap} \quad (\text{VII.2})$$

Los fotodetectores habitualmente son **fotodiodos**, diodos especialmente diseñados para funcionar en polarización inversa, de modo que al recibir luz con la longitud de onda adecuada producen una corriente eléctrica gracias al principio de la fotoconductividad. En un diodo normal la corriente en polarización

⁸Existen variaciones en estas capacidades dependiendo de la densidad de información y del uso de las dos caras o una. MB y GB hacen referencia a *megabyte* y *gigabyte*, siendo el *byte* una unidad de información digital de 8 bits. Más adelante estudiaremos con detalle qué es un bit y su consideración física en los circuitos.

⁹En las unidades de lectura-escritura también puede emplearse para grabar información en discos grabables.

¹⁰Este es un factor importante: no toda la radiación luminosa es capaz de generar pares electrón-hueco.

inversa es prácticamente despreciable, pues se debe al arrastre de portadores minoritarios. Sin embargo, en un fotodiodo la iluminación provoca que se generen numerosos pares electrón-hueco, que son arrastrados por el campo en inversa. Al aumentar de manera significativa el número de portadores lo hace también la conductividad y por tanto la corriente.

$$I = \sigma EA = (\sigma_0 + \Delta\sigma)EA = \sigma_0EA + \Delta\sigma EA = I_0 + \Delta I \quad (\text{VII.3})$$

En este caso ΔI es el aumento de corriente producido por la fotoconductividad. La cantidad de corriente producida (que depende del número de portadores generados) estará influenciada por la intensidad de la radiación luminosa recibida pero también por la calidad del material semiconductor con el que esté construido el fotodiodo. Si la calidad es baja y existe un gran número de trampas y defectos, estos actúan como centros de recombinación que reducen la cantidad de portadores creados, y por tanto la corriente.

Los fotodiodos se fabrican de manera especial para aprovechar al máximo el fenómeno de la fotoconductividad (figura VII.7). La zona de la unión (donde se van a crear los portadores) debe tener por tanto un tamaño relativamente grande. Asimismo, el encapsulado del diodo debe tener una abertura (protegida por una lámina de material transparente, en ocasiones con forma de lente) para que el semiconductor pueda recibir la luz. Suelen ser dispositivos especialmente sensibles a la luz infrarroja o visible.

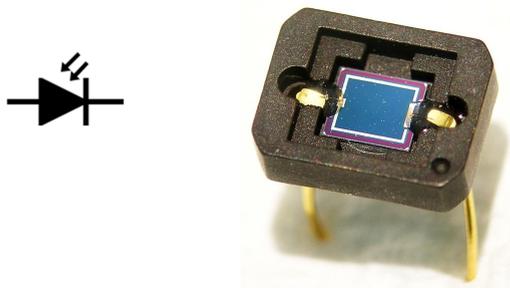


Figura VII.7: Símbolo de un fotodiodo (izquierda) e imagen de un fotodiodo real (<https://commons.wikimedia.org/wiki/File:Photodiode-closeup.jpg>), donde se puede observar la amplia abertura para recibir la luz.

Un **fototransistor** (figura VII.8) es un tipo especial de transistor bipolar en el que la corriente de base es generada gracias a la iluminación de la unión colector-

base. Esta corriente de base es amplificada por la ganancia del transistor, obteniéndose una corriente de colector apreciable. Gracias a ello, los fototransistores poseen una sensibilidad mucho mayor que la de los fotodiodos.

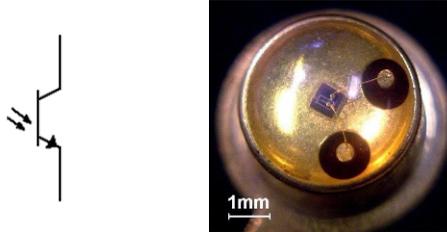


Figura VII.8: Símbolo de un fototransistor (izquierda) e imagen (<https://commons.wikimedia.org/wiki/File:Fototransb.jpg>) donde se puede apreciar el detalle de los cables que unen los terminales externos con el semiconductor.

Una **fotorresistencia** o **LDR** (de las siglas en inglés *Light-Dependent Resistor*, o resistencia dependiente de la luz) es una resistencia cuyo valor cambia dependiendo de la iluminación que reciba (figura VII.9).



Figura VII.9: Símbolo de una LDR (izquierda) y fotografía de una fotorresistencia de sulfuro de cadmio (CdS).

En ocasiones las fotorresistencias también se denominan **fotocélulas** o **células fotoeléctricas**. La variación de resistencia al iluminarse es debida en exclusiva al fenómeno de la fotoconductividad. Al ser dispositivos pasivos, debe realizarse un pequeño circuito para detectar eléctricamente los cambios de luz. En el caso de las fotorresistencias se pueden construir circuitos sensores a muy bajo coste utilizando una LDR como una de las resistencias de un divisor de voltaje¹¹.

¹¹Un divisor de voltaje son dos resistencias puestas en serie, en la que el valor del voltaje en el punto medio depende de los valores de las resistencias y la diferencia de potencial global entre ambas.

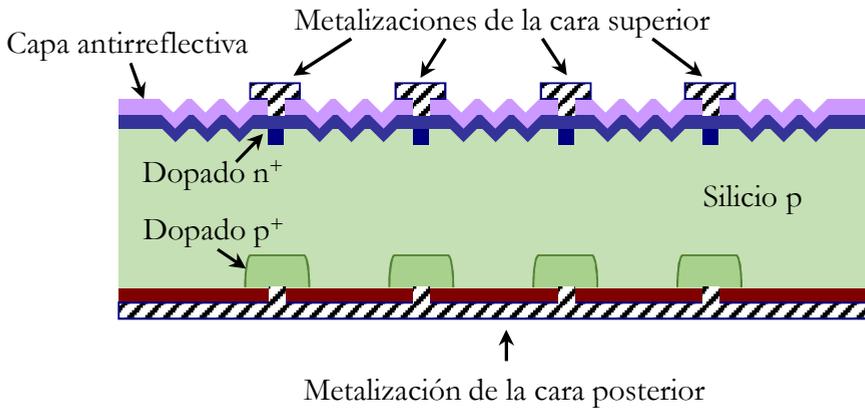


Figura VII.10: Esquema lateral de una célula solar.

4 CÉLULAS SOLARES

La producción de energía eléctrica de forma respetuosa para el medio ambiente es uno de los grandes retos de nuestra sociedad. En este contexto, los dispositivos optoelectrónicos son de gran importancia. En concreto las **células solares** cada vez tienen mayor peso en la producción de electricidad. Conceptualmente una célula solar puede entenderse como un fotodiodo de gran tamaño. En definitiva, se basan también en la estructura de diodo y en el fenómeno de la fotoconductividad, por lo que se diseñan para operar en polarización inversa (lo que permite extraer los pares de electrones y huecos creados por la luz, generando una corriente eléctrica). A la energía eléctrica producida por los paneles solares la denominamos **energía fotovoltaica**.

Una característica importante de las células solares (figura VII.10) es que su superficie se diseña específicamente para aprovechar al máximo la radiación solar. Por una parte, la superficie suele estar texturizada (es decir, no es completamente plana) para reaprovechar la luz reflejada. Además, se cubre con una capa de material antirreflexión para intentar captar la máxima radiación luminosa posible. También debe tenerse en cuenta que los contactos en la cara orientada al sol deben ser lo suficientemente delgados como para dejar pasar el máximo de radiación luminosa, pero suficientemente grandes como para recolectar los portadores generados antes de que se pierdan por recombinación.

La mayor parte de las células solares comerciales están fabricadas con silicio. A pesar de que este material, como ya hemos comentado, no es el más adecuado

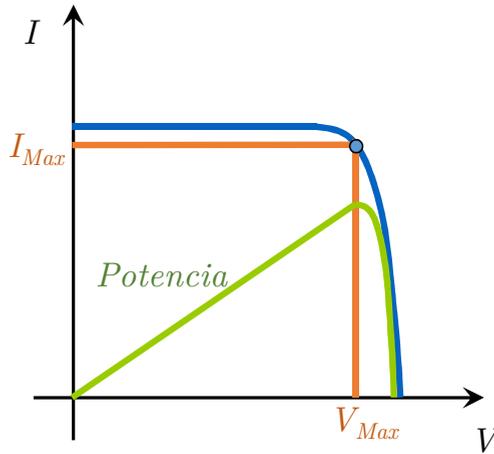


Figura VII.11: Característica $I - V$ de una célula solar y elección del punto de máxima potencia.

para aplicaciones optoelectrónicas debido a su gap indirecto, que dificulta los procesos de generación de portadores (y por tanto reduce su eficiencia), su bajo coste comparado con otros semiconductores y la facilidad para obtener células de gran tamaño (el tamaño estándar de una célula solar es de 15×15 cm) hacen que la relación eficiencia-coste sea la más favorable. Además, con frecuencia no se emplea silicio monocristalino, sino multicristalino¹² por su menor coste y su relativamente buen rendimiento.

Dado que las células solares se emplean para producir energía eléctrica, se elige de manera cuidadosa su punto de operación buscando la potencia máxima. Para ello se busca en la característica $I - V$ de la célula la combinación de ambas magnitudes que proporcione un producto IV (potencia) máximo (figura VII.11).

Las células solares producen energía eléctrica en continua (DC). La red industrial trabaja con corriente alterna (AC), por lo que la energía producida por la célula solar debe ser convertida de DC a AC antes de ser inyectada a la red. Se trata del proceso contrario a la rectificación: el circuito que lo implementa se denomina **inversor**.

¹²Recordemos que en este tipo de sólidos la orientación cristalina no es única, sino que existen en el material regiones de gran tamaño con diferentes orientaciones.

5 OPTOACOPADORES

Un **optoacoplador** es un *dispositivo que permite acoplar dos partes de un circuito manteniéndolas aisladas eléctricamente*. Para ello hace uso tanto del fenómeno de la electroluminiscencia como de la fotoconductividad.

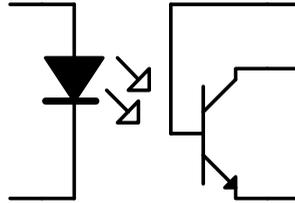


Figura VII.12: Símbolo de un optoacoplador.

El principio de funcionamiento de un optoacoplador es sencillo. Imaginemos que queremos enviar la información de una señal eléctrica desde una parte de un sistema que trabaja con voltajes y corrientes reducidos hacia otra parte que trabaja con voltajes y corrientes elevadas. Esto ocurre con frecuencia por ejemplo en los robots industriales. La parte de control del robot (el microcontrolador o el microprocesador que dé órdenes al brazo) trabajará con voltajes y corrientes no muy elevados (voltajes de como mucho una docena de voltios y corrientes siempre inferiores a un amperio). El robot, en cambio, funcionará con un motor que requerirá voltajes y corrientes elevados (decenas o incluso cientos de voltios, y corrientes de hasta decenas de amperios) para poder mover el peso del propio brazo robótico y la carga que porte.

Si conectásemos directamente ambas partes del circuito, muy probablemente dañáramos la parte de control. Por ello, debe existir un aislamiento eléctrico entre la zona de potencia elevada y la zona de potencia reducida. Con este fin se emplea el optoacoplador. Imaginemos que la información de control generada por el microprocesador está contenida en una señal eléctrica que a lo largo del tiempo toma un valor bajo (por ejemplo cero voltios) o un valor alto (por ejemplo 5 voltios), en un circuito con una configuración similar al de la figura VII.13. Aplicaremos esta señal a la entrada del optoacoplador, que consiste en un LED.

Cuando el LED de la parte inicial del optoacoplador reciba una señal en bajo, no se encenderá. En consecuencia, la base del fototransistor que conforma la salida del optoacoplador no recibirá iluminación, y por tanto no existirá corriente en la salida del optoacoplador.

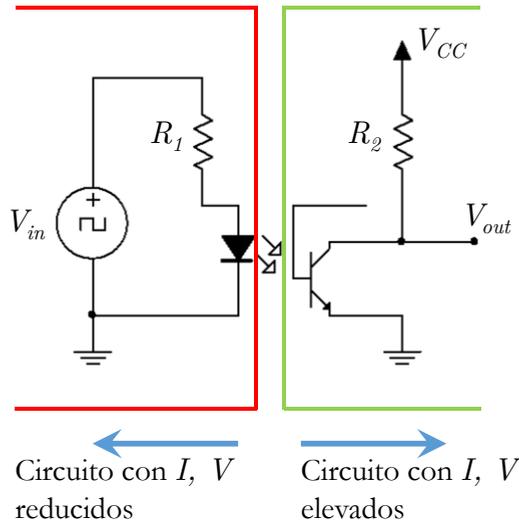


Figura VII.13: Ejemplo de funcionamiento de un optoacoplador.

Sin embargo, cuando el LED reciba una señal en alto, se iluminará. Dicha luz llegará a la base del fototransistor y producirá una corriente. De este modo, la señal alto/bajo de la entrada se podrá traducir en un paso/no paso de corriente en la salida, situación que podrá ser procesada para reconstruir la señal de partida y actuar en consecuencia.

La clave de todo el proceso radica en que en el interior del optoacoplador la información se transmite de manera óptica, y no eléctrica. Usualmente el optoacoplador tiene un encapsulado cerrado, sin partes expuestas, por lo que todo el proceso tiene lugar de forma aislada del exterior. También puede dejarse abierta la parte entre el LED y el fototransistor (conformando un **optointerruptor**), lo que puede ser útil en ciertas aplicaciones (por ejemplo en el conteo de vueltas, haciendo que una parte móvil del objeto que gira pase por la zona expuesta, de forma que se pueda detectar el cambio de señal eléctrica debida a la interrupción del paso de luz por el objeto).

Los optoacopladores se emplean con mucha frecuencia en robótica, control de motores, mecanismos, etc.

VIII.

AMPLIFICACIÓN

Tanto en el ámbito de la informática como en las telecomunicaciones la transmisión y procesamiento de información juega un papel fundamental. Esta información va a estar codificada en forma de señales eléctricas, cuyo valor irá cambiando en el tiempo. En el caso de las **señales digitales**, la señal eléctrica (típicamente un voltaje) variará entre dos posibles valores¹. Sin embargo, en las **señales analógicas** la información está contenida en el valor concreto de voltaje (o corriente), que puede variar de forma continua en un intervalo dado.

1 CONCEPTO DE AMPLIFICACIÓN

Especialmente en el caso de las señales analógicas, en ocasiones la señal puede ser débil (por ejemplo, una comunicación lejana recibida mediante una antena) y presentar dificultades para poder extraer la información contenida en ella. La **amplificación** de señales consiste en *tomar una cierta señal eléctrica variable en el tiempo y obtener una señal proporcional a ésta, pero con una amplitud de variación mayor*, que pueda ser tratada o procesada con mayor facilidad.

Un ejemplo claro de la necesidad de amplificar señales es la toma de datos de sensores: por su propia naturaleza (los sensores deben perturbar lo menos posible el entorno y la magnitud que se desea medir) la señal proporcionada suele ser muy débil y no puede ser procesada o digitalizada directamente: generalmente es necesario amplificarla para poder hacer uso de la información proporcionada por el sensor.

En este capítulo vamos a tratar de manera resumida algunos aspectos relacionados con la amplificación de señales analógicas. Si bien nuestro interés principal

¹En realidad, variará entre dos rangos de valores, uno de voltaje *alto* y otro de voltaje *bajo*, como veremos en el capítulo IX.



Figura VIII.1: Amplificación de una señal variable en el tiempo.

es el tratamiento de señales digitales, es interesante comprender los fundamentos básicos de la amplificación empleando transistores, así como conocer algunos circuitos de gran utilidad en campos como las telecomunicaciones o la adquisición de señales. Sin la amplificación sería extraordinariamente complicado capturar y mostrar información: dicho de otra manera, sería muy difícil obtener los datos con los que trabajan los ordenadores.

2 UN EJEMPLO: EL AMPLIFICADOR DE VOLTAJE EN FUENTE COMÚN

Consideremos señales variables en el tiempo, que variarán respecto a un cierto valor de continua. En nuestro caso, trabajaremos con magnitudes correspondientes a los voltajes y corrientes característicos de un transistor, por ejemplo un MOSFET. Así, en un instante dado, para un transistor MOSFET trabajando en torno a un punto de operación estacionario sabemos que su voltaje drenador-fuente, su corriente y su voltaje de puerta serán:

$$V_{ds}(t) = V_{DS} + v_{ds}(t) \quad (\text{VIII.1})$$

$$I_d(t) = I_D + i_d(t) \quad (\text{VIII.2})$$

$$V_{gs}(t) = V_{GS} + v_{gs}(t) \quad (\text{VIII.3})$$

Donde como ya vimos en el capítulo VI, V_{DS} , I_D y V_{GS} son los valores de continua del punto de operación estacionario (que asumimos constantes en el tiempo), y $v_{ds}(t)$, $i_d(t)$ y $v_{gs}(t)$ son las magnitudes en alterna, la parte variante de las señales con el tiempo.

Consideremos un circuito sencillo análogo al de la figura VI.10, que nos sirvió de ejemplo para el cálculo del punto de operación estacionario en un MOSFET. A este circuito vamos a incorporarle una señal oscilante en la puerta, $v_{gs}(t)$, que será la señal que amplificaremos, mientras que la señal amplificada será $v_{ds}(t)$: esta señal, sumada a V_{DS} , será la que obtengamos en el drenador del transistor en el circuito de la figura, es decir, en el drenador tendremos $V_{ds}(t)$ conforme a las expresiones anteriormente citadas, mientras que en la puerta del transistor tendremos $V_{gs}(t)$. El circuito con estas condiciones se muestra en la figura VIII.2.

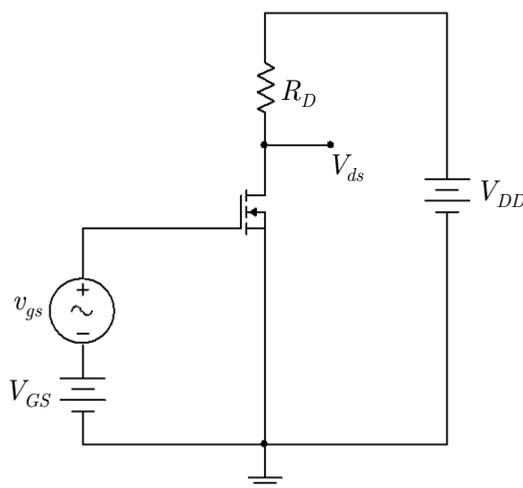


Figura VIII.2: Circuito amplificador de voltaje en fuente común.

Un determinado voltaje en la puerta del transistor debe producir, necesariamente, un cierto valor de voltaje en el drenador para una red de polarización dada (los valores concretos que estemos considerando para las fuentes de alimentación y la resistencia de drenador). Vamos a considerar como la entrada del circuito V_i el valor de voltaje en la puerta, y como la salida V_o el valor de voltaje en el drenador.

Si representamos el valor de la salida V_o en función de la entrada V_i obtendremos una curva como la mostrada en la figura VIII.3; una explicación más detallada de cómo se obtiene esta curva se presenta en el capítulo X.

A un voltaje de entrada de valor V_{GS} le corresponderá un voltaje de salida de valor V_{DS} , como se indica en la figura. Dado que trabajamos con una señal de entrada variable en el tiempo (recordemos que la parte variable y que queremos amplificar es $v_{gs}(t)$, que se superpone a V_{GS}), nuestra señal de entrada V_i variará

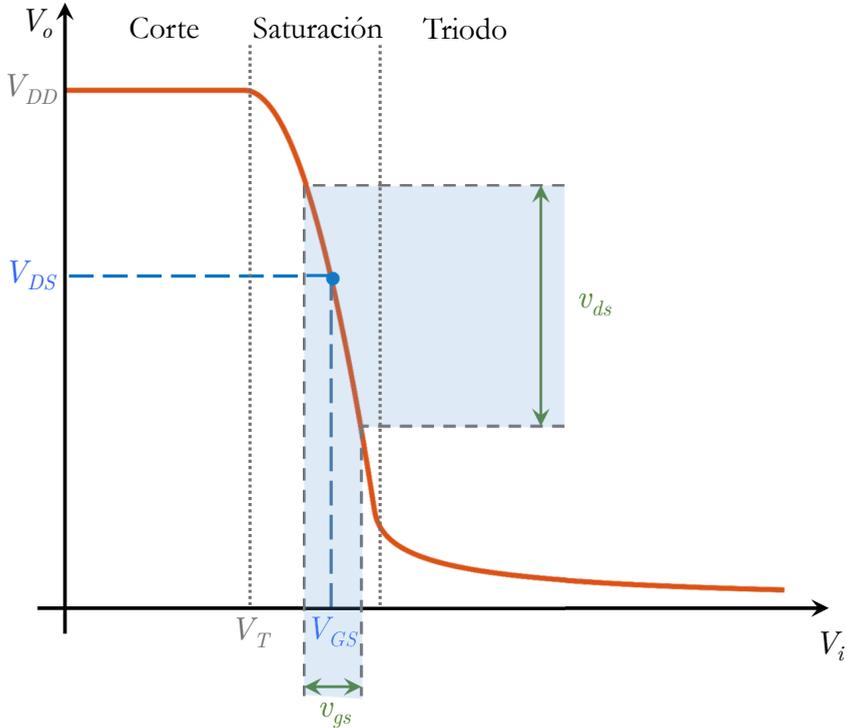


Figura VIII.3: Amplificación en un circuito: la señal de salida (v_{ds}) responde a la señal de entrada (v_{gs}) siguiendo su forma, pero con mayor amplitud.

una cierta cantidad, pequeña, en torno a V_{GS} . A ese rango de voltajes en el que varía la entrada le corresponderá, conforme a la curva, un rango de valores de voltaje de salida. Las zonas sombreadas en la figura indican los rangos de variación de las señales. Como vemos, una pequeña variación de voltaje en la entrada (v_{gs}) produce un amplio rango de variación del voltaje en la salida (v_{ds}), siempre en torno a los valores de polarización estacionarios V_{GS} y V_{DS} . Dentro de estos rangos, en cada instante de tiempo (asumiendo que la variación de la señal no fuese demasiado rápida) se podría obtener el valor concreto de voltaje de salida para un voltaje de entrada dado sin más que localizar en la curva el valor de V_o que corresponde a cada V_i .

El resultado cuando introducimos una señal variable en la puerta es por tanto una señal variable en la salida que reproducirá la forma de la señal de entrada, pero

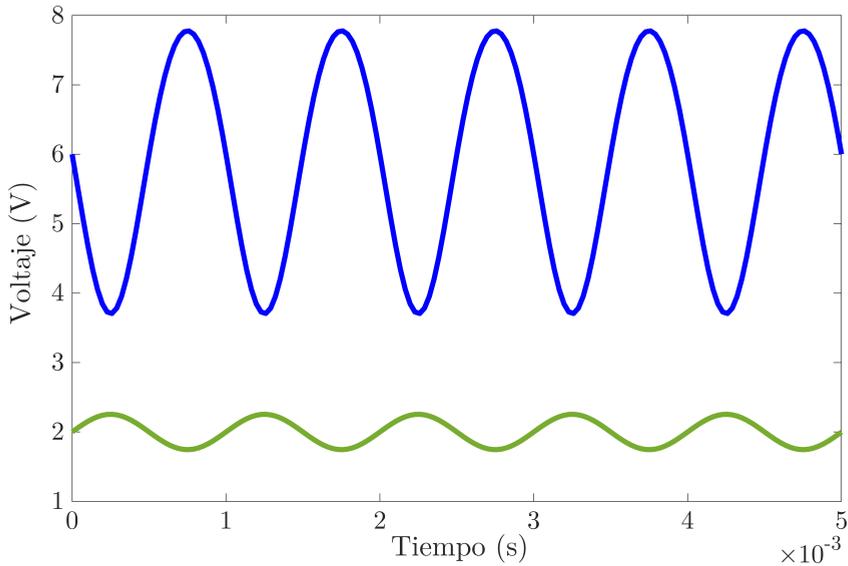


Figura VIII.4: Amplificación de una señal senoidal que oscila en torno a $V_{GS} = 2$ V, con una frecuencia de 1 KHz y amplitud de 0.2 V (curva inferior), obteniéndose en la salida una señal oscilante en torno a $V_{DS} = 6$ V. Nótese que la señal de salida está invertida respecto a la entrada.

con mayor amplitud: se produce en consecuencia una amplificación². En la figura VIII.4 se muestra la amplificación obtenida con este circuito.

Si la zona de la curva que corresponde a las partes sombreadas en la figura VIII.3 fuera completamente recta, la señal de salida reproduciría fielmente la entrada (con la salvedad de la inversión de la señal), con la única diferencia de una mayor amplitud de oscilación. Pero si la pendiente en la zona de saturación no es completamente recta (como es el caso en el ejemplo mostrado), existirá una cierta **distorsión** de la señal amplificada, es decir, no será perfectamente fiel a la señal de partida. En la figura VIII.4 esto se observa en el hecho de que la señal amplificada no oscila de forma simétrica en torno a $V_{DS} = 6$ V, como debería corresponder a una señal senoidal perfecta. Conseguir una distorsión mínima en los circuitos amplificadores es especialmente importante para realizar bien el pro-

²En el circuito de la figura, la señal variable en la salida estaría invertida respecto a la entrada, es decir, el valor máximo de v_{gs} produciría el valor mínimo de v_{ds} , y viceversa. Por ello decimos que el circuito es amplificador e inversor. Corregir este cambio de signo en la salida es sencillo haciendo pasar la señal de nuevo por otro circuito inversor.

ceso. Por ello, además de que la respuesta del transistor sea lo más lineal posible en el régimen de saturación, es fundamental que la señal a amplificar (v_{gs} en este caso) sea de pequeña señal, es decir, con una amplitud reducida. Si la amplitud de la señal de entrada fuera aún más grande, podría llevar al transistor a las fronteras con las regiones de corte y triodo, donde la respuesta es fuertemente no lineal y la distorsión sería aún más grande.

En un circuito amplificador un parámetro esencial es la **ganancia**, que se define como la relación entre la amplitud de la señal de salida y la señal de entrada. En el ejemplo que nos ocupa podemos calcular la ganancia en voltaje (G_V) a partir del circuito equivalente de pequeña señal del transistor MOSFET, que ya presentamos en el capítulo VI, figura VI.12.

Para ello vamos a suponer que la frecuencia de oscilación de la señal de entrada no es excesivamente elevada, de modo que podemos prescindir de las capacidades del circuito equivalente del MOSFET. En ese caso, reemplazando el MOSFET por su circuito equivalente de pequeña señal y baja frecuencia tendremos el circuito de pequeña señal mostrado en la figura VIII.5.

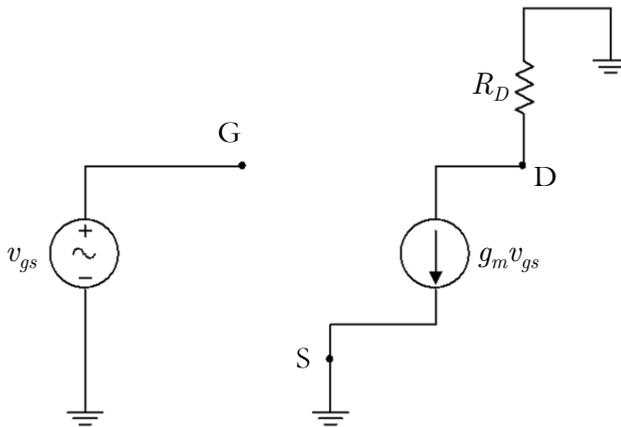


Figura VIII.5: Circuito de pequeña señal del amplificador de voltaje en fuente común.

En el circuito de pequeña señal las fuentes de voltaje DC (V_{DD} y V_{GS} en la figura VIII.2) deben considerarse como cero (tierra) en alterna (el valor en esos puntos del circuito está fijado por las fuentes y no oscila), dado que sólo tratamos con la parte variable de las señales.

Así, de este modo, podremos calcular el voltaje en el drenador (D), que a

partir de la ley de Ohm en la resistencia R_D será:

$$0 - v_d = g_m v_{gs} R_D \quad (\text{VIII.4})$$

Dado que la fuente (S) está conectada a tierra, tendremos que $v_{ds} = v_d$, y por tanto:

$$v_{ds} = -g_m v_{gs} R_D \quad (\text{VIII.5})$$

Y la ganancia en voltaje (G_V) será en consecuencia:

$$G_V = \frac{v_{ds}}{v_{gs}} = -g_m R_D \quad (\text{VIII.6})$$

La ganancia en voltaje es una magnitud adimensional, es decir, no tiene unidades³, y como vemos, depende de factores asociados al transistor (g_m) y a la red de polarización (R_D).

3 EL AMPLIFICADOR OPERACIONAL

Existen numerosísimas configuraciones de circuitos amplificadores basadas tanto en MOSFET como en BJT, que pueden emplear uno o varios transistores, y cuyo estudio detallado queda fuera de los objetivos de este libro. Sin embargo, por su importancia vamos a comentar con más detalle un circuito amplificador de especial interés, el **amplificador operacional**.

Un amplificador operacional es un circuito con entrada diferencial (en realidad dos entradas, una *no inversora* y otra *inversora*, cuya diferencia es la cantidad a amplificar) y una única salida. Tiene una característica particular, y es que su ganancia, G_V , es muy elevada (idealmente infinita). El símbolo del amplificador operacional, junto con sus entradas y salida, se muestra en la figura VIII.6.

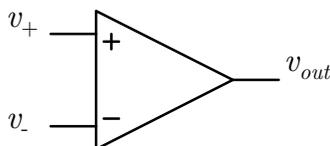


Figura VIII.6: Símbolo del amplificador operacional.

³Recordemos que la transconductancia se miden en siemens, que es el precisamente el inverso del ohmio.

La salida v_{out} del operacional será:

$$v_{out} = G_V(v_+ - v_-) \quad (\text{VIII.7})$$

El amplificador operacional idealmente tiene una impedancia de entrada infinita (lo que se traduce en que desde sus entradas no entra corriente hacia el amplificador) e impedancia de salida nula (es decir, se comporta como una fuente de voltaje ideal proporcionando una salida constante v_{out}). En la realidad su impedancia de entrada es muy elevada (de decenas de $M\Omega$) y la de salida es muy baja (de apenas unas decenas de ohmios). Asimismo, la ganancia real no es infinita, sino que tiene un valor muy grande (del orden de 10^5).

Para funcionar debemos proporcionar al amplificador operacional una alimentación adecuada. En uno de sus modelos más habituales, el $\mu A741$, deben proporcionarse dos voltajes en continua, V_{CC+} (de por ejemplo +10 V) y V_{CC-} (de por ejemplo -10 V). Estos voltajes marcan los límites de amplificación, dado que no puede obtenerse en la salida v_{out} ninguna señal superior o inferior a estos valores.

Este hecho nos permite disponer de una primera aplicación directa del amplificador operacional, y es emplearlo como circuito **comparador**. Si conectamos en la entrada inversora v_- la señal de referencia y en la no inversora v_+ la señal a comparar, en la salida v_{out} tendremos un valor igual a V_{CC+} cuando $v_+ > v_-$, mientras que tendremos un valor igual a V_{CC-} cuando $v_+ < v_-$, como se muestra en la figura VIII.7.

La mayor utilidad del amplificador operacional, sin embargo, se obtiene cuando empleamos un **lazo de realimentación**, es decir, conectamos la salida con alguna de las entradas, de modo que se establece un modo de funcionamiento en el cual podremos realizar múltiples aplicaciones. La realimentación tenderá a estabilizar el funcionamiento del amplificador operacional, haciendo que la salida no se vaya a los valores de saturación.

En el caso de disponer de un amplificador operacional con características ideales, si se encuentra realimentado podremos suponer, por una parte, que la condición de ganancia infinita implica que no debe existir diferencia de voltaje en sus entradas, o lo que es lo mismo, que al realimentarlo se fija la condición $v_+ = v_-$. Por otra parte, si la impedancia de entrada es infinita, no entrará corriente por ninguna de las entradas del amplificador operacional.

Con estas dos condiciones pueden realizarse numerosos circuitos muy interesantes. Un ejemplo es el circuito **amplificador no inversor** mostrado en la figura VIII.8.

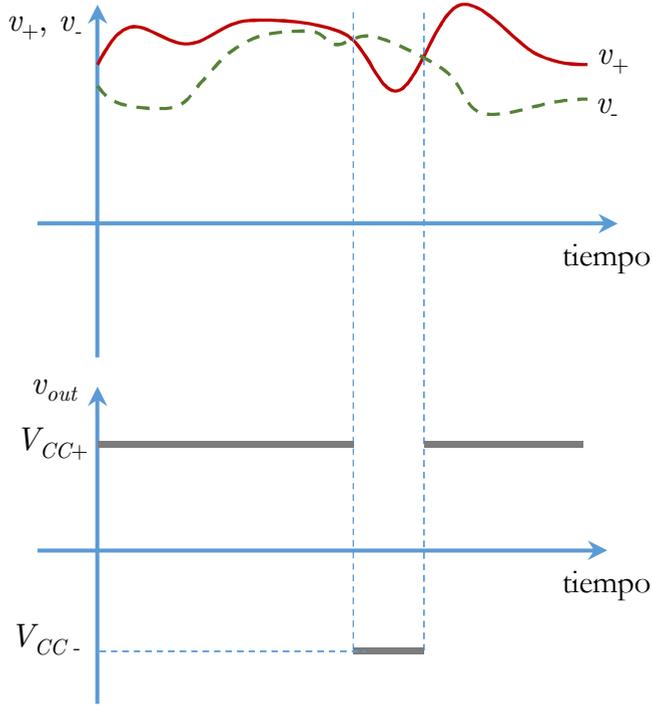


Figura VIII.7: El amplificador operacional puede ser usado como comparador sin más que conectar las dos señales a comparar en sus entradas inversora y no inversora. La salida es una señal cuadrada con valor V_{CC+} o V_{CC-} , dependiendo del resultado de la comparación.

En este circuito debemos tener en cuenta en primer lugar que la condición de ganancia infinita implica que $v_+ = v_-$, por lo que en la entrada inversora del operacional (que está conectada al punto medio entre las dos resistencias) tendremos el voltaje de entrada v_i . Por otra parte, en dicha entrada inversora no puede entrar corriente, por lo que la corriente que circula por R_1 y R_2 deberá ser necesariamente la misma, es decir, es como si se tratase de dos resistencias en serie, como vemos en la figura VIII.9.

Si aplicamos la ley de Ohm en las resistencias R_1 y R_2 , tendremos que:

$$I = \frac{v_i - 0}{R_1} = \frac{v_o - v_i}{R_2} \quad (\text{VIII.8})$$

De donde despejando se obtiene:

$$v_o = v_i \left(1 + \frac{R_1}{R_2} \right) \quad (\text{VIII.9})$$

La ganancia en voltaje de este circuito será por lo tanto:

$$G_V = \frac{v_o}{v_i} = \left(1 + \frac{R_1}{R_2} \right) \quad (\text{VIII.10})$$

Es decir, en este caso la ganancia obtenida no depende de parámetros propios de un transistor, sino que depende exclusivamente de los valores de las resistencias R_1 y R_2 , con lo que basta seleccionar adecuadamente su valor para obtener la ganancia deseada. Obviamente existirán unos límites respecto a los posibles valores del voltaje de salida v_o , que serán los voltajes de alimentación V_{CC+} y V_{CC-} .

Con el amplificador operacional pueden construirse circuitos que no solamente realicen funciones de amplificación con baja distorsión, sino también operaciones matemáticas, como sumar señales, restarlas, integrarlas, derivarlas, calcular exponenciales y logaritmos, etc. Es probablemente uno de los circuitos electrónicos más versátiles que existen.

Particularmente interesante es el *amplificador de instrumentación* (figura VIII.10). Se trata de un circuito muy empleado en la toma de datos de sensores, en los

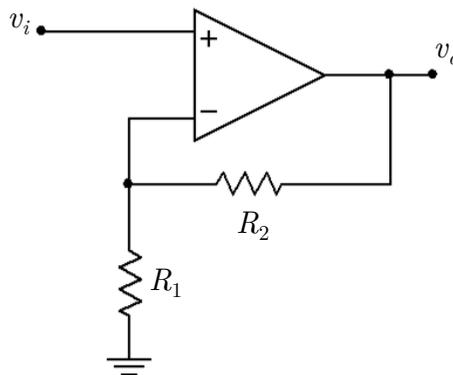


Figura VIII.8: Circuito amplificador no inversor, con entrada v_i y salida v_o .

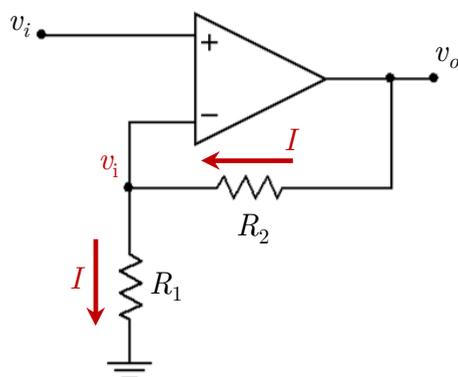


Figura VIII.9: La corriente en este circuito circula por R_1 y R_2 , dado que no pueden fluir corrientes hacia las entradas del amplificador operacional.

que es primordial amplificar la señal obtenida, por ser inicialmente muy débil. El amplificador de instrumentación amplifica la diferencia entre sus dos entradas con una elevada precisión y estabilidad, proporcionando una ganancia igual a:

$$G_V = \frac{v_{out}}{v_2 - v_1} = 1 + \frac{2R}{R_g} \quad (\text{VIII.11})$$

Como resistencia R_g suele tomarse una resistencia variable, de forma que así podemos controlar la ganancia del amplificador. El amplificador de instrumentación emplea tres amplificadores operacionales. Los dos situados a la izquierda constituyen la etapa de entrada. Al ir cada una de las entradas conectada directamente a un amplificador operacional, el circuito posee una elevada impedancia de entrada (es decir, que se opone en gran medida a que entre corriente hacia el mismo, y gracias a ello afecta poco a las señales que se quiere amplificar), mientras que el tercer amplificador operacional y sus resistencias asociadas constituyen la etapa diferencial. Esta configuración se caracteriza por lo que se conoce como un elevado grado de rechazo al modo común: amplifica la diferencia entre dos señales pero elimina las posibles componentes comunes que tengan estas entradas (por ejemplo las debidas a ruido e interferencias). Por ello permite amplificar con gran precisión señales débiles, y se emplea muy habitualmente en equipos médicos (como electrocardiogramas), sistemas de adquisición de datos, lectura de termopares, sensores, etc. Las señales así amplificadas se pueden procesar y

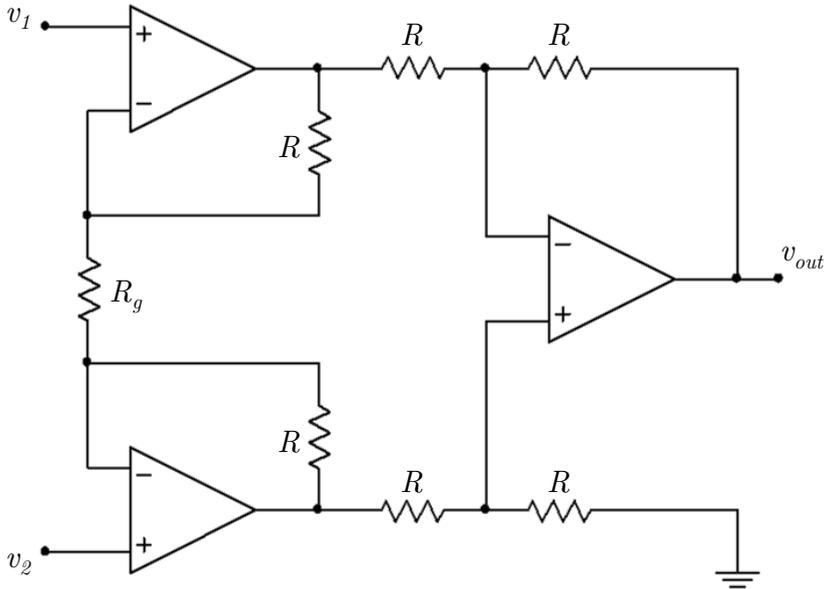


Figura VIII.10: Amplificador de instrumentación.

tratar adecuadamente.

Pero quizá la aplicación más relevante del amplificador operacional en el ámbito de la informática es la realización de **circuitos conversores de señal analógica a digital y la operación contraria, de digital a analógica**. Presentaremos este tipo de circuitos y los mostraremos con más detalle en el capítulo IX.

IX.

BASES DE
ELECTRÓNICA
DIGITAL

La práctica totalidad del procesamiento de información se realiza en la actualidad mediante circuitos basados en la **electrónica digital**. En el ámbito de las telecomunicaciones, la televisión digital terrestre (TDT) o las emisiones de radio digital han reemplazado de forma masiva a las emisiones analógicas, gracias a su mayor calidad y definición. La telefonía es también en la actualidad un ámbito eminentemente digital, con mejoras constantes y mayores capacidades respecto a las antiguas comunicaciones analógicas.

Los circuitos digitales son aquellos que funcionan con señales eléctricas a las que se le asignan únicamente dos valores, a diferencia de los analógicos, donde se trabaja con señales eléctricas que pueden tomar cualquier valor dentro de un amplio rango. Los microprocesadores, por su propia naturaleza, son circuitos digitales, y sólo por ello ya estaría justificado su estudio en el ámbito de las ingenierías informáticas; pero es que además, como se ha indicado, hoy en día los circuitos digitales son universalmente empleados en el ámbito de las telecomunicaciones por su mayor fiabilidad. En esencia, el mundo electrónico es en la actualidad un mundo digital.

1 ÁLGEBRA DE BOOLE

Antes de conocer cómo son las señales digitales en el mundo físico, es conveniente describir algunos conceptos esenciales respecto al **álgebra de Boole**¹. En el álgebra de Boole, las operaciones se llevan a cabo con variables **binarias**,

¹Por el matemático inglés George Boole (1815-1864). También podremos referirnos a ella como lógica booleana o lógica binaria. No realizaremos una definición matemática formal completa del álgebra de Boole, dado que no es el objetivo de este libro. Nos centraremos en los aspectos básicos prácticos necesarios para conocer la implementación de puertas lógicas y el diseño lógico.

es decir, que pueden tomar dos posibles valores, que llamaremos 0 y 1 (también podemos asignar los valores *falso* y *verdadero*; lo importante es que únicamente hay dos posibles valores).

Las operaciones básicas que podemos realizar con esas variables binarias son:

- **AND** (producto lógico): la representaremos por el símbolo \cdot
- **OR** (suma lógica): la representaremos por el símbolo $+$

Ambas operaciones son conmutativas, y se puede aplicar la propiedad distributiva para ambas², es decir, si tenemos tres variables lógicas a , b y c , se cumple que:

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c) \quad (\text{IX.1})$$

$$a + (b \cdot c) = (a + b) \cdot (a + c) \quad (\text{IX.2})$$

Existe además un elemento neutro para cada una de las operaciones, 1 para la operación AND y 0 para la operación OR, de modo que se cumple que:

$$a \cdot 1 = a \quad (\text{IX.3})$$

$$a + 0 = a \quad (\text{IX.4})$$

Para cada variable a existe su complementada o negada \bar{a} , que se obtiene mediante la operación **NOT**, de modo que se cumple que:

$$a \cdot \bar{a} = 0 \quad (\text{IX.5})$$

$$a + \bar{a} = 1 \quad (\text{IX.6})$$

Los resultados que se obtienen de las operaciones AND, OR y NOT se muestran en la figura IX.1 conforme a los posibles valores binarios de las variables implicadas.

²Nótese que para las operaciones AND y OR hemos empleado los símbolos \cdot y $+$, análogos a los del producto y suma aritméticos; sin embargo, no se trata de las mismas operaciones, dado que mientras que aritméticamente sólo aplicamos la propiedad distributiva del producto respecto a la suma, en lógica booleana se puede aplicar para ambas operaciones.

AND		
<i>a</i>	<i>b</i>	<i>a · b</i>
0	0	0
0	1	0
1	0	0
1	1	1

OR		
<i>a</i>	<i>b</i>	<i>a + b</i>
0	0	0
0	1	1
1	0	1
1	1	1

NOT	
<i>a</i>	\bar{a}
0	1
1	0

Figura IX.1: Operaciones AND, OR y NOT.

Una herramienta útil en lógica binaria son las **leyes de Morgan**³, que en este contexto se expresan como:

$$\overline{a + b} = \bar{a} \cdot \bar{b} \tag{IX.7}$$

$$\overline{a \cdot b} = \bar{a} + \bar{b} \tag{IX.8}$$

Existen varios principios y teoremas, como el principio de dualidad, el teorema de Shannon, etc. que son útiles en un estudio detallado del álgebra de Boole y que no abordaremos aquí, pues no es nuestro objetivo desarrollar con detalle el estudio matemático de esta estructura algebraica, sino conocer los aspectos básicos imprescindibles para entender los circuitos digitales fundamentales. Con esta premisa, sí es importante definir el concepto de **función lógica**: se trata de una *expresión que representa la combinación de un conjunto finito de símbolos, que denotarán constantes o variables, unidos por las operaciones AND, OR y NOT.*

Un ejemplo de una posible función lógica de tres variables podría ser:

$$f(a, b, c) = (\overline{a \cdot b}) + c$$

O por ejemplo:

$$f(a, b, c) = a + c \cdot (b + \bar{a})$$

En las funciones lógicas pueden aparecer **términos producto**, es decir, un conjunto de variables (o sus complementadas) unidas por la operación AND, o **términos suma**, un conjunto de variables (o sus complementadas) unidas por la operación OR.

³Por Augustus de Morgan (1806-1871), matemático inglés.

En las funciones lógicas debe tenerse en cuenta de la prioridad de las operaciones las marcan primero los paréntesis, y en caso de no existir, se considera que existe prioridad del producto AND respecto a la suma OR.

La **tabla de verdad** de una función lógica es una tabla en la que se muestran todas las posibles combinaciones de las variables lógicas y el resultado de la función. Por ejemplo, en la figura IX.1 se mostraban las tablas de verdad de las funciones lógicas $a \cdot b$ (operación AND), $a + b$ (operación OR) y \bar{a} (operación NOT).

NAND			NOR		
a	b	$\overline{a \cdot b}$	a	b	$\overline{a + b}$
0	0	1	0	0	1
0	1	1	0	1	0
1	0	1	1	0	0
1	1	0	1	1	0

XOR			XNOR		
a	b	$a \oplus b$	a	b	$\overline{a \oplus b}$
0	0	0	0	0	1
0	1	1	0	1	0
1	0	1	1	0	0
1	1	0	1	1	1

Figura IX.2: Operaciones NAND, NOR, XOR y XNOR.

Además de las operaciones básicas (AND, OR y NOT) existen otras operaciones derivadas de interés en electrónica digital, como son las siguientes:

- **NAND**: es la operación AND complementada, es decir, $\overline{a \cdot b}$
- **NOR**: es la operación OR complementada, es decir, $\overline{a + b}$
- **XOR**: operación OR exclusivo, dada por $a \oplus b = a \cdot \bar{b} + \bar{a} \cdot b$
- **XNOR**: operación NOR exclusivo, dada por $\overline{a \oplus b} = a \cdot b + \bar{a} \cdot \bar{b}$

Los circuitos que implementarán estas funciones reciben el nombre de **puertas lógicas**. Combinando puertas lógicas puede realizarse cualquier función lógica, aunque también puede realizarse un diseño directo para implementar la función, como veremos en el capítulo XI.

2 REPRESENTACIÓN FÍSICA DE SEÑALES DIGITALES

Las variables binarias pueden tomar valor 0 o 1. Estos son, sin embargo, conceptos matemáticos. Para que un circuito digital pueda trabajar, debemos traducir esas abstracciones matemáticas a magnitudes físicas, en nuestro caso, eléctricas. Usualmente lo que se emplea son rangos de valores de voltaje (denominados **niveles lógicos**) a los que asignaremos 0 o 1 según corresponda.

Así, tendremos un nivel lógico correspondiente a voltajes bajos y otro nivel correspondiente a voltajes altos. Si asignamos el 1 al nivel alto y el 0 al nivel bajo hablaremos de **lógica definida positiva** (véase la figura IX.3), mientras que si lo hacemos al revés, es decir, al nivel de voltaje alto le asignamos el 0 y al bajo el 1, hablaremos de **lógica definida negativa**. Usualmente se trabaja en la primera situación, es decir, con lógica definida positiva. Entre las dos regiones o rangos de voltaje existe una zona en la que la señal no estaría definida (no sería ni 0 ni 1), por lo cual no debe trabajarse en ella.

Los circuitos digitales tendrán por tanto una serie de entradas, que podrán valer 0 o 1 (es decir, que a ellas llegará un voltaje bajo o un voltaje alto, respectivamente), de modo que en función de esas entradas en la salida se proporcionará un 0 o un 1 (de nuevo, un voltaje bajo o un voltaje alto) dependiendo de la función lógica u operación que realice el circuito.

El hecho de trabajar con niveles lógicos (rangos de voltaje) y no con valores exactos de magnitudes eléctricas (como ocurre en las señales analógicas) presenta una serie de ventajas muy interesantes que hacen que los circuitos digitales sean, en la actualidad, la forma preferida de tratar y procesar información en la mayoría de aplicaciones, incluidas las telecomunicaciones. La principal de todas ellas es la relativa inmunidad al fenómeno del **ruido electrónico**. El ruido electrónico es una *perturbación involuntaria de la señal eléctrica* que puede originar un cambio no deseado en el circuito. El ruido constituye un problema que es universal a todos los circuitos, tanto analógicos como digitales. La perturbación de la señal puede tener diversos orígenes: en primer lugar, el simple hecho de encontrarnos a una temperatura por encima del cero absoluto ya implica necesariamente una cierta fluctuación de las señales eléctricas, al sufrir los electrones una agitación térmica

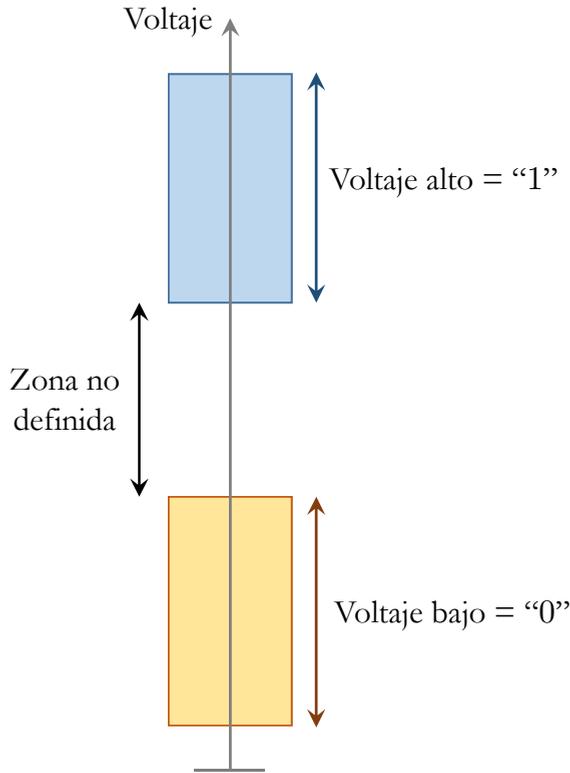


Figura IX.3: Niveles lógicos en lógica definida positiva. Los valores concretos de los rangos de voltaje dependerán del tipo de implementación física del circuito.

aleatoria. Por otra parte, el propio movimiento de los electrones en el interior del semiconductor, al estar influido por colisiones con átomos de la red cristalina o impurezas, incorpora también factores de aleatoriedad, y por tanto de ruido. Los defectos o trampas en los materiales también son una fuente de ruido. Además, también pueden existir perturbaciones externas, como campos magnéticos, que alteren las señales. En definitiva, es un fenómeno que, si bien en algunos casos se puede paliar (como las perturbaciones por campos externos, que se pueden reducir empleando cables con una configuración adecuada), es consustancial al transporte de carga eléctrica, y por tanto siempre va a estar presente en mayor o menor medida.

Las señales digitales, al codificar la información en niveles de voltaje y no en valores exactos, son menos sensibles al ruido que las señales analógicas, como estudiaremos cuando veamos las características de las familias lógicas en el capítulo

XI.

3 CONVERSIÓN DE SEÑALES ANALÓGICAS Y DIGITALES

La percepción de los sentidos del ser humano es, en esencia, analógica, como lo son las magnitudes medibles en la naturaleza (temperatura, presión, etc.). Cuando tomamos los datos de un sensor, inicialmente este proporciona la información en forma de una señal eléctrica analógica: la información está contenida en el valor exacto de la magnitud que proporciona el sensor (en la mayoría de las ocasiones, un voltaje).

La propia naturaleza de los sensores, sin embargo, hace que esas señales eléctricas analógicas sean débiles, ya que un sensor debe perturbar lo menos posible el entorno que mide. Por ello es necesario, como ya sabemos, amplificar la señal para poder transmitirla y procesarla. En este proceso pueden producirse alteraciones no deseadas de la señal: además de la altamente probable distorsión que aparecerá en el proceso de amplificación, el fenómeno del ruido electrónico puede afectar notablemente a las sucesivas etapas de tratamiento de la señal, haciendo que el último término se degrade la información o aparezcan interferencias.

Por ello cada vez con mayor frecuencia se recurre al tratamiento digital de las señales. Digitalizando adecuadamente la señal, se puede procesar y transmitir con mayores garantías de respeto a la integridad de la información, mejorando notablemente la calidad de las señales finales que llegan al receptor. Una vez recibida, con frecuencia la señal tendrá que ser procesada de nuevo para volver a convertirla en una señal analógica, que el ser humano pueda interpretar con sus sentidos. Un buen ejemplo es la emisión y recepción de radio digital, cuyo resumen se muestra en la figura IX.4.

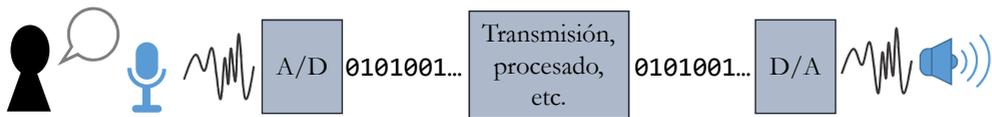


Figura IX.4: Ejemplo de procesamiento digital de la información.

Una señal audible se capta con un micrófono, produciendo una señal analógica. Esta se convierte a digital en el convertor A/D, para después transmitir o procesar la secuencia de ceros y unos. En la parte de recepción, el convertor de digital a analógico (D/A) reconstruye la señal, que se pasa a un altavoz para su escucha.

El proceso de digitalización debe realizarse con cuidado para garantizar que se obtiene una señal digital con la calidad suficiente para contener toda la información relevante que contenía la señal analógica. La elemento básico de información digital es el **bit**, acrónimo del inglés *Binary digiT*, o dígito binario. Empleando solamente un bit para representar la información, únicamente tendríamos dos posibles valores, 0 o 1. Sin embargo, podemos combinar varios bits para disponer de más posibilidades. Por ejemplo, empleando 2 bits tendremos cuatro posibles combinaciones de valores: 00, 01, 10 y 11.

Si empleásemos 3 bits, tendríamos hasta 8 combinaciones: 000, 001, 010, 011, 100, 101, 110 y 111. En general, si empleamos n bits las posibles combinaciones serán 2^n . Cuantos más bits consideremos, mayor será la resolución que conseguiremos al digitalizar la señal.

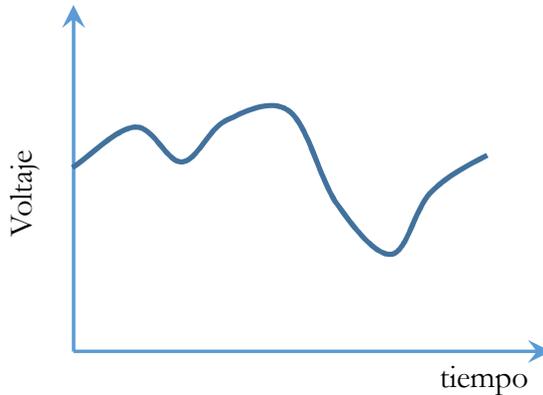


Figura IX.5: Señal analógica variable en el tiempo.

Consideremos una señal analógica como la mostrada en la figura IX.5. Como vemos, se trata de un valor de voltaje que varía gradualmente en el tiempo, pudiendo tomar cualquier valor dentro de un cierto rango.

El primer paso para la digitalización de la señal es decidir el número de bits que emplearemos. Supongamos inicialmente que empleamos 3 bits. Ello quiere decir que podremos distinguir hasta 8 posibles valores, dado que con 3 bits existen 8 posibles combinaciones de dígitos binarios. Consecuentemente, dividiremos el rango de voltaje en 8 zonas equiespaciadas, y le asignaremos a cada una un valor, en orden creciente, de números binarios, desde el 000 hasta el 111. El segundo paso es determinar la rapidez de muestreo, es decir, cada cuánto tiempo tomaremos los valores analógicos de la señal de partida y los convertiremos a

dígitos binarios. Supongamos que tomamos datos cada 2 segundos, de modo que tenemos una situación como la mostrada en la figura IX.6. En cada instante de tiempo en el que realicemos el muestreo, asignaremos como valor digital de la señal la combinación de bits correspondiente al valor instantáneo de la señal en ese momento (marcado con una cruz en la figura).

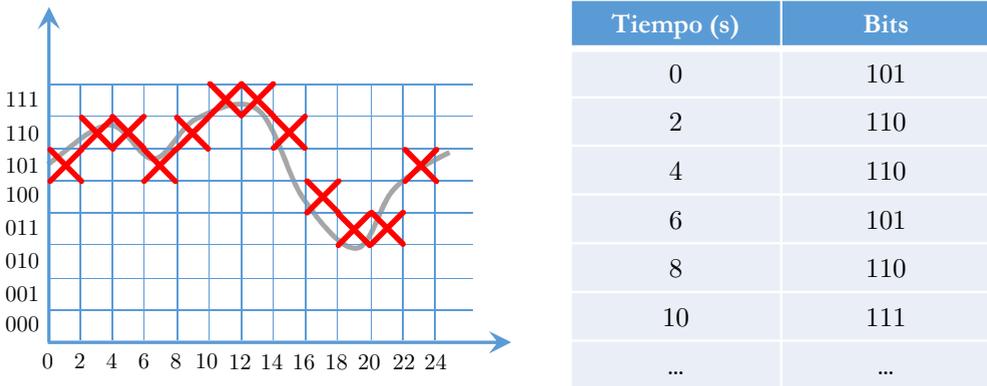


Figura IX.6: Digitalización empleando 3 bits y muestreando cada 2 segundos.

De este modo hemos convertido la información de la señal analógica en una serie de dígitos binarios para cada instante de muestreo. Obviamente, al tomar valores cada cierto tiempo y al haber discretizado los posibles valores de voltaje de la señal (recordemos que en este caso hemos dividido el rango en 8 regiones, las posibles combinaciones correspondientes a 3 bits) perderemos parte de la información que originalmente contenía la señal, pero a cambio, será mucho más sencillo conservar la integridad de los valores binarios obtenidos cuando transmitamos o procesemos la información.

Una vez que disponemos de los valores de la señal digitalizada, sería posible intentar reconstruir la señal original a partir de ellos. En el caso que nos ocupa, con 3 bits y el muestreo cada dos segundos, obtendríamos una señal como la mostrada en la figura IX.7. Como podemos observar, en cada franja temporal hemos considerado el valor medio de voltaje en el rango correspondiente.

Esta señal no reproduce con demasiada fidelidad la señal original. Ello es debido, por una parte, a que el número de bits que hemos empleado es pequeño (sólo 3) y por otra parte, a que la frecuencia de muestreo ha sido demasiado lenta. Si empleásemos 4 bits, podríamos discretizar el rango de voltaje en $2^4 = 16$ posibles valores, con lo que mejoramos la precisión; si además duplicamos la

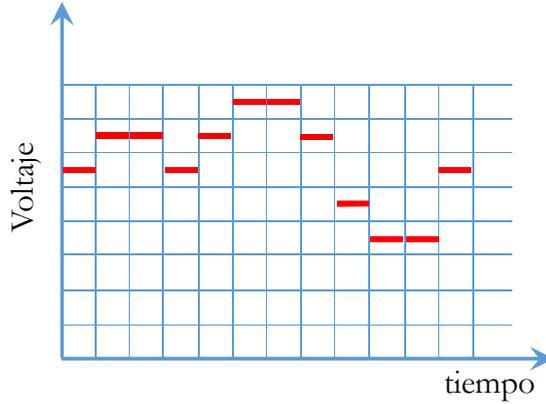
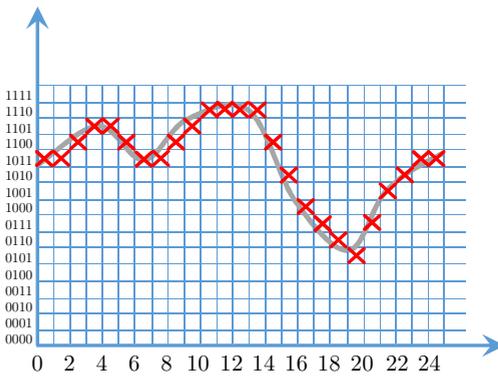


Figura IX.7: Señal reconstruida, con digitalización de 3 bits y muestreo cada 2 segundos.

velocidad de muestreo, es decir, tomamos datos cada segundo, obtendremos un resultado como el que se muestra en la figura IX.8.



Tiempo (s)	Bits
0	1011
1	1011
2	1100
3	1101
4	1101
5	1100
...	...

Figura IX.8: Digitalización empleando 4 bits y muestreando cada segundo.

El número de muestras lógicamente se ha duplicado (tendremos que manejar un volumen de información mucho mayor) y además, cada unidad de información tendrá un bit más que en el caso anterior, es decir, 4 en vez de 3. Ello supone que los circuitos digitales que procesen esta información deberán ser capaces de tratar con la mayor velocidad a la que se generan los datos, por una parte, y por otra parte aumentará su complejidad al tratar con un número mayor de bits. A cambio,

hemos conseguido que la señal reconstruida sea ahora más fiel a la original, como se observa en la figura IX.9.

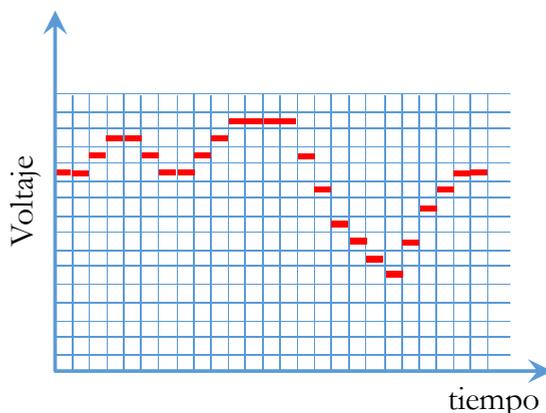


Figura IX.9: Señal reconstruida, con digitalización de 4 bits y muestreo cada segundo.

Obviamente, aumentando el número de bits y la frecuencia de muestreo conseguiríamos que la señal reconstruida fuera cada vez más fidedigna a la original, pero se incrementaría notablemente el coste y los requerimientos del sistema digital para poder tratar un mayor volumen de información en menor tiempo. Debe existir por lo tanto un compromiso entre emplear valores adecuados para garantizar una calidad adecuada de la señal digitalizada y las limitaciones en cuanto a la cantidad de datos a transmitir o procesar por unidad de tiempo, es decir, el **ancho de banda**⁴.

Como hemos visto en este ejemplo, el paso de señal analógica a digital consiste esencialmente en traducir a bits un valor concreto de voltaje, tomado en un cierto instante de tiempo. Los circuitos que realizan la conversión de señales analógicas a digitales reciben el nombre de **convertidores analógico/digital** o simplemente **ADC**. La operación contraria, es decir, convertir dígitos en valores concretos de voltaje (como hacíamos por ejemplo al reconstruir la señal) la realizan los **convertidores digital/analógico**, o **DAC**.

⁴Como ejemplo de parámetros empleados en digitalización de señales analógicas, en el caso de los discos compactos de audio (CD), que son un medio de almacenamiento digital de sonido, el estándar es considerar 16 bits y una frecuencia de muestreo de 44.1 KHz, es decir, tomar 44100 muestras por segundo.

La conversión de digital a analógico se basa en los postulados de la representación binaria de cantidades. Imaginemos que tenemos un número binario compuesto por tres bits (por ejemplo 101, o 011). De forma genérica lo expresaremos como $a_2a_1a_0$, donde los diferentes a_i pueden valer 0 o 1. Si queremos calcular qué número decimal representa el número binario compuesto por n bits debemos realizar la operación:

$$N = \sum_{i=0}^{n-1} a_i 2^i \quad (\text{IX.9})$$

Por ejemplo, el número binario de 3 bits 110 corresponde al decimal $1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 6$. El bit que está situado más a la izquierda será por tanto el **bit más significativo**, es decir, el que mayor peso tendrá en la suma (2^2 en este caso), mientras que el bit situado más a la derecha será el **menos significativo**, el que menos peso tenga (2^0).

En los circuitos conversores DAC las entradas del circuito serán los bits del número que se quiere convertir. Estos bits corresponderán físicamente a un valor de voltaje bajo (cerca o igual a 0 V) o a un valor de voltaje alto (cerca o igual a un valor de referencia V_R). Los circuitos DAC realizan una operación en la que la contribución de cada bit al voltaje final de salida estará ponderada por su posición, dando más peso al bit más significativo y menos al menos significativo.

Este tipo de operaciones son sencillas de realizar empleando amplificadores operacionales. En la figura IX.10 mostramos dos ejemplos de DAC, el conversor de resistencias ponderadas y el conversor en escalera R - $2R$. En ambos casos el voltaje obtenido en la salida, V_{out} , será proporcional al valor decimal del número binario que dispongamos en las entradas.

Vamos a tomar como ejemplo el conversor de resistencias ponderadas. En primer lugar, al encontrarse el amplificador operacional realimentado, la condición de ganancia infinita implica que $v_+ = v_-$, por lo que al estar v_+ conectado a tierra, en v_- tendremos un valor fijo de cero voltios, lo que se conoce como **tierra virtual**. Esta situación la indicamos en la figura IX.11 como una tierra añadida, que sin ser estrictamente una tierra, corresponde al voltaje nulo. Por otra parte, en las entradas los bits corresponderán a valores físicos de voltaje, por ejemplo cero voltios para representar un 0 y un cierto valor de voltaje, V_R , para representar un 1. Dado que cada a_i toma valor 0 o 1, podemos expresar el valor de voltaje en cada entrada como $a_i V_R$.

En consecuencia, la corriente debida a cada una de las entradas viene dada,

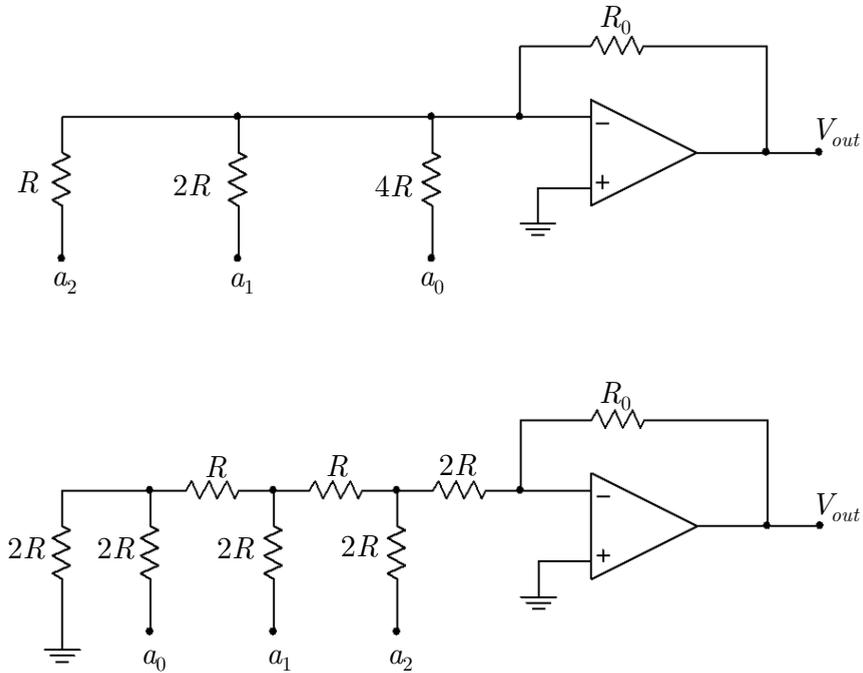


Figura IX.10: Circuito conversor DAC de resistencias ponderadas (arriba) y conversor DAC en escalera R - $2R$ (abajo).

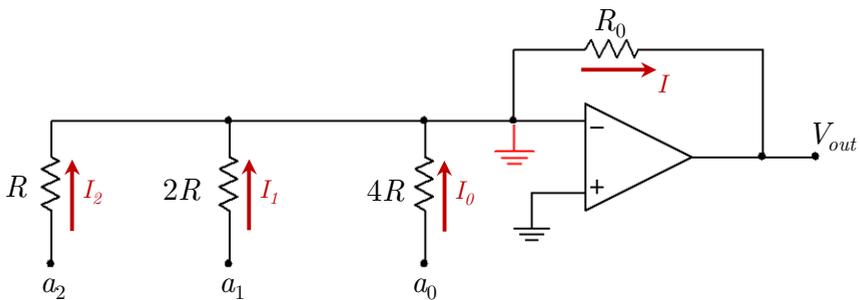


Figura IX.11: Circuito conversor DAC de resistencias ponderadas, con indicación de la tierra virtual y las corrientes consideradas.

conforme a la ley de Ohm, por:

$$\begin{aligned} I_0 &= \frac{a_0 V_R - 0}{4R} = \frac{a_0 V_R}{4R} = \frac{a_0 V_R}{2^2 R} \\ I_1 &= \frac{a_1 V_R - 0}{2R} = \frac{a_1 V_R}{2R} = \frac{a_1 V_R}{2^1 R} \\ I_2 &= \frac{a_2 V_R - 0}{R} = \frac{a_2 V_R}{R} = \frac{a_2 V_R}{2^0 R} \end{aligned} \quad (\text{IX.10})$$

De manera genérica, por tanto, podemos expresar la corriente debida a una entrada i como:

$$I_i = \frac{a_i V_R}{2^{2-i} R} = \frac{a_i 2^i V_R}{2^2 R} = \frac{a_i 2^i V_R}{4R} \quad (\text{IX.11})$$

Atendiendo a los sentidos de las corrientes que hemos dibujado en la figura IX.11, debe cumplirse que:

$$I_0 + I_1 + I_2 = \sum_{i=0}^2 I_i = I = \frac{0 - V_{out}}{R_0} \quad (\text{IX.12})$$

Y por tanto tendremos que el voltaje de salida, V_{out} , será:

$$V_{out} = -R_0 \sum_{i=0}^2 I_i = -R_0 \sum_{i=0}^2 \frac{a_i 2^i V_R}{4R} = -\frac{R_0 V_R}{4R} \sum_{i=0}^2 a_i 2^i \quad (\text{IX.13})$$

Como vemos, la parte del sumatorio corresponde al valor decimal del número binario con el que estemos tratando, teniendo en cuenta que hemos considerado 3 bits en este circuito. Es definitiva, como ya se había indicado, el voltaje de salida será proporcional al número que queremos convertir.

La operación de conversión de analógico a digital es en cambio más compleja, y para realizarla debemos ayudarnos de un conversor de digital a analógico. La idea básica es disponer de un sistema de control que actúe sobre las entradas de un conversor DAC (por ejemplo, puede comenzar poniendo todas las entradas digitales a 0). La salida del conversor DAC se lleva a una de las entradas de un comparador (circuito que, como ya vimos en el capítulo VIII puede realizarse de manera sencilla con un amplificador operacional), mientras que en la otra entrada colocamos el valor de voltaje analógico que queremos digitalizar. Si el resultado de la comparación es que el voltaje analógico es mayor que el valor de

voltaje proporcionado por el DAC, el sistema de control aumenta en una unidad el número digital de entrada del DAC, hasta que el resultado de la comparación cambie, momento en el que se tomará como salida digital el número binario que indiquen las entradas del DAC. El esquema de un conversor ADC se muestra en la figura IX.12.

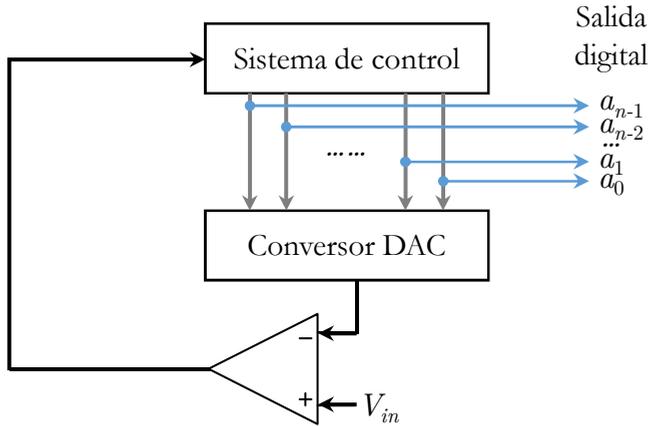


Figura IX.12: Esquema de un circuito conversor ADC.

La conversión de analógico a digital va a ser más lenta que la de digital a analógico, ya que dependerá del tiempo que emplee el sistema de control en encontrar el número binario que se adecúe al valor de la entrada analógica V_{in} .

X.

CONMUTACIÓN

Como ya hemos visto en el capítulo IX, el tratamiento digital de la información se realiza a partir de señales binarias, usualmente valores de voltaje, que toman dos posibles valores: en el caso más habitual, los 0 están representados por un rango de voltajes bajos y los 1 por un rango de voltajes altos. Para realizar este tipo de circuitos, los dispositivos electrónicos con los que se fabricarán los circuitos digitales trabajarán en dos posibles estados, uno en el que no permitirán el paso de corriente (**estado off**) y otro en el que lo harán de la mejor forma posible (**estado on**). La **conmutación** es precisamente *el modo de operación de un dispositivo electrónico entre dos estados*, uno que corresponde al estado de corte (*off*) y otro a conducción con una resistencia muy baja, idealmente nula (*on*).

El tránsito entre los dos estados es particularmente importante en la operación de los circuitos digitales, dado que por ejemplo influye notablemente en la rapidez de operación y el consumo de potencia de los microprocesadores.

1 CONMUTACIÓN DE DIODOS

El dispositivo semiconductor más sencillo, el diodo, posee dos estados a los que les podemos asignar de manera clara la etiqueta *on* y *off*. En la zona subumbral (voltaje aplicado inferior al umbral del diodo, V_γ), el diodo prácticamente no conduce (existe una pequeña corriente, la corriente inversa de saturación, que es despreciable en la práctica). Para un voltaje aplicado superior al umbral, en cambio, el diodo conduce muy bien, como una resistencia de valor muy bajo.

Los diodos por tanto pueden ser utilizados para fabricar circuitos digitales basándonos en esta posibilidad de conducción/no conducción, y de hecho algunos de los primeros circuitos digitales realizados con semiconductores se implementaron empleando este tipo de dispositivos de dos terminales.

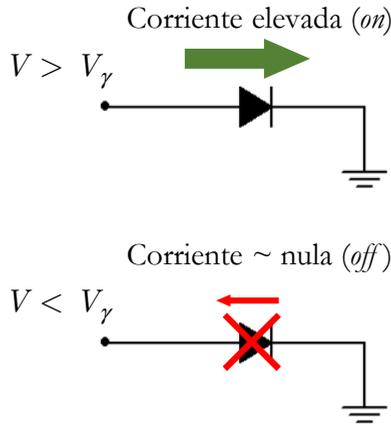


Figura X.1: Estados *on* y *off* de un diodo.

Sin embargo, los diodos presentan una serie de problemas para su uso en circuitos digitales. En primer lugar, la ausencia de un terminal de control específico que sea el que regule el paso de un estado a otro, pues la conmutación en el diodo se produce por el cambio en el valor del voltaje entre sus terminales de conducción. En segundo lugar, si bien es posible realizar circuitos digitales con diodos, ello implica necesariamente el paso de corriente en el estado *on* (y por tanto, un consumo de potencia). Finalmente, un problema que resulta crítico: al ser el diodo un dispositivo pasivo (es decir, no es capaz de amplificar señales) existirá una pérdida del nivel de la señal que se irá acumulando a medida que se atraviesen diferentes circuitos. En consecuencia, no resulta práctico emplear los diodos para realizar circuitos digitales complejos y que además sean eficientes energéticamente, por lo que su uso hoy en día está descartado para este fin.

2 CONMUTACIÓN DE TRANSISTORES

En la actualidad los circuitos digitales se fabrican de forma prácticamente exclusiva empleando MOSFETs, por lo que vamos a centrarnos principalmente en este tipo de transistores (en la parte final de esta sección examinaremos brevemente el comportamiento de los transistores bipolares en conmutación). En el caso del transistor MOSFET, los estados *on* y *off* los relacionaremos con las regiones de **triado** y **corte**, respectivamente. El estado de triado se caracteriza por el hecho de que la respuesta del transistor es de tipo óhmico, es decir, se

comporta de manera similar a una resistencia, dado que corresponde a la parte lineal inicial de las características de salida, como ya estudiamos en el capítulo VI. En cambio, en el estado de corte el voltaje de puerta no alcanzaba el umbral, V_T , y no se daban las condiciones para que existiese corriente en el dispositivo¹.

En definitiva, es la puerta del transistor (el voltaje aplicado en ella) la que actúa como terminal de control, llevando al transistor desde un régimen en el que el dispositivo no conduce a otro en el que conduce razonablemente bien, y viceversa. El paso de una situación a otra debe hacerse necesariamente recorriendo la recta de carga, que como ya sabemos corresponde al conjunto de todos los posibles puntos de operación de un transistor una vez colocado en un circuito. En ese camino, deberemos atravesar necesariamente la región de saturación, que era de interés para amplificación de señales, pero no aplicaciones digitales. Por ello, es conveniente que la transición entre triodo y corte (o viceversa) sea lo más rápida posible.

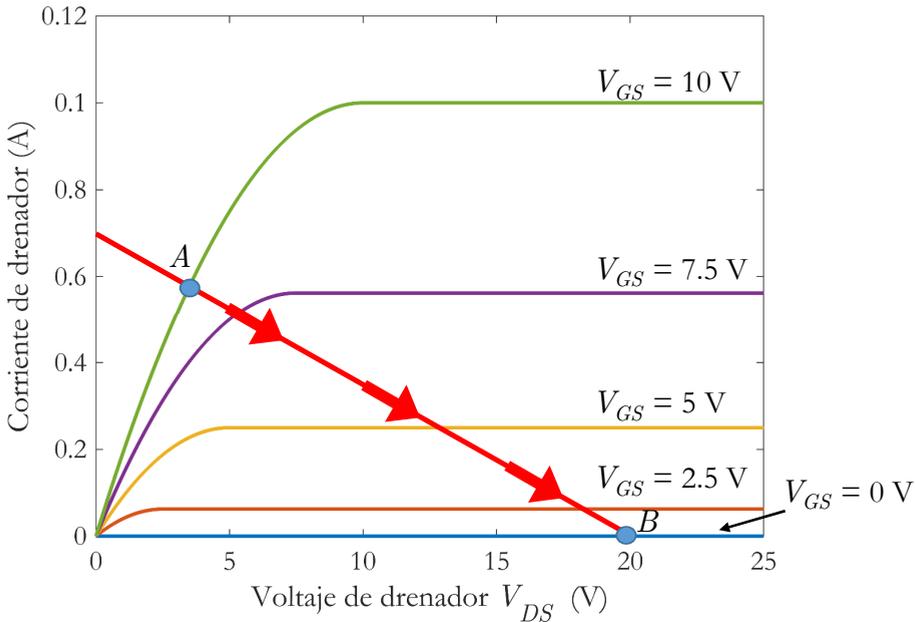


Figura X.2: Proceso de conmutación en un transistor MOSFET (paso de triodo, punto A, a corte, punto B) recorriendo la recta de carga.

¹Recordemos que en ese régimen de funcionamiento no se ha formado la capa de inversión, por lo que no existe un canal conductor entre drenador y fuente.

Consideremos que inicialmente tenemos un valor de V_{GS} que implica un punto de operación estacionario en triodo, como en el punto A de la figura X.2, que corresponde a un voltaje de puerta de 10 V. En este punto de operación, como ya sabemos, existe una capa de inversión en el transistor que ocupa todo el canal, de modo que disponemos de un canal conductor que se comporta de forma parecida a una resistencia.

Supongamos que, en un instante dado, el voltaje de puerta conmuta a un valor de 0 V, por debajo del voltaje umbral del transistor. A este nuevo valor del voltaje de puerta debe corresponder un punto de operación estacionario en el extremo de la recta de carga de corriente nula (punto B).

Sin embargo, el transistor necesitará un cierto tiempo (tiempo de paso al estado *off*, o t_{off}) para poder acomodarse a esta nueva situación. Deberá desalojar a los portadores que forman la capa de inversión, conforme a lo correspondiente a una situación de corte (*off*), en la que el transistor no conducirá la corriente eléctrica, y se comportará de forma parecida a una situación de circuito abierto. Durante este tiempo, el transistor atravesará la región de saturación de las curvas. Resulta conveniente, por tanto, que el tiempo empleado en alcanzar el nuevo punto de operación estacionario sea lo menor posible. Ello dependerá de diversos factores, como el tipo de semiconductor, la longitud del canal, la concentración inicial de la capa de inversión en triodo, etc.

El proceso contrario, es decir, el paso de corte a triodo, lógicamente tampoco será instantáneo. Partiremos de una situación donde no existen portadores minoritarios en la zona adyacente al óxido de puerta (corte) a una situación en la que debe existir una capa de inversión a lo largo de todo el canal (triodo). El tiempo necesario para formar esa capa de inversión y pasar al estado *on* lo denotaremos por t_{on} .

Las situaciones transitorias, de paso de *on* o *off* y viceversa, serán de gran importancia cuando determinemos la velocidad máxima a la que podrán operar los circuitos digitales. El hecho de que estos tiempos no sean nulos implica que existen límites a la **frecuencia de reloj**² con la que operarán. Por tanto, la física de los transistores (el material semiconductor, la geometría del transistor, etc.) es

²Los circuitos digitales complejos, como los microprocesadores, funcionan siguiendo un ritmo pautado por una señal cuadrada, llamada señal de reloj, que actúa como un metrónomo, marcando la cadencia de las operaciones a realizar. Los ordenadores actuales trabajan con frecuencias de reloj en el rango de los gigahercios (GHz), es decir, les corresponden periodos del orden de nanosegundos (ns). En consecuencia, los tiempos de respuesta de los transistores que conformen ese circuito deben ser inferiores para garantizar un correcto funcionamiento.

un factor determinante a la hora de establecer la rapidez de un circuito digital.

Más allá de las implicaciones de la duración de los transitorios entre ambos regímenes, en aplicaciones digitales lo esencial es el comportamiento del transistor en corte y triodo. En el primer caso esperamos que idealmente el transistor impida el paso de corriente; en el segundo, que facilite al máximo posible la circulación de corriente. Podemos por tanto pensar en un modelo simplificado del MOSFET en conmutación. En el estado de corte lo trataremos como un circuito abierto, y en el estado de triodo lo reemplazaremos por una resistencia (la resistencia equivalente en el régimen lineal, que dependerá de la pendiente de la curva de salida en ese régimen óhmico). Así, podemos establecer un **modelo en conmutación** del MOSFET, que mostramos en la figura X.3.

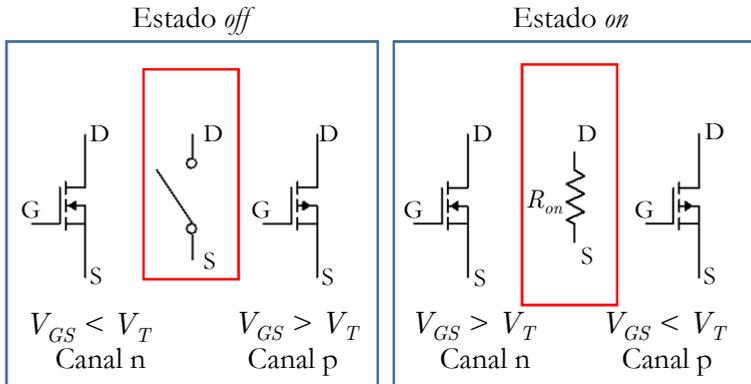


Figura X.3: Modelo en conmutación del transistor MOSFET.

Para emplear el modelo simplemente reemplazaremos el transistor entre drenador y fuente por un circuito abierto o por una resistencia R_{on} , según sea el caso (*off/on*). El valor de R_{on} puede calcularse a partir de la expresión de la corriente de drenador del transistor MOSFET en triodo, que recordemos viene dada por:

$$I_D = K(2(V_{GS} - V_T)V_{DS} - V_{DS}^2) \tag{X.1}$$

En la zona fuertemente lineal (es decir, para valores de V_{DS} pequeños), el término V_{DS}^2 es despreciable frente al resto, por lo que la corriente I_D se puede aproximar por:

$$I_D \approx 2K(V_{GS} - V_T)V_{DS} \tag{X.2}$$

La resistencia R_{on} se podrá determinar a partir de la ley de Ohm, ya que en el régimen lineal debe cumplirse que $V_{DS} = I_D R_{on}$, y por tanto:

$$R_{on} = \frac{1}{2K(V_{GS} - V_T)} \quad (\text{X.3})$$

Como puede comprobarse, se trata de una resistencia controlada por voltaje. Cuanto mayor sea el voltaje de puerta V_{GS} , menor será el valor de R_{on} , por lo que nos interesará trabajar con un V_{GS} lo más elevado posible en el contexto de alimentación del circuito en el que nos encontremos³, con el objetivo de que el comportamiento del transistor se acerque lo más posible al ideal (que sería una situación de resistencia equivalente nula, es decir, que se comportase entre fuente y drenador como un cable conductor perfecto). No debemos olvidar, sin embargo, que un valor elevado de V_{GS} implica una mayor concentración de la capa de inversión, y en consecuencia, mayor tiempo para crearla y desalojarla, es decir, valores mayores de t_{on} y t_{off} .

Desde el punto de vista aplicado, deben tenerse en cuenta una serie de factores. Cuando se trabaja en la región de corte, la corriente debería ser nula, pero en un MOSFET real no va a ser exactamente cero, sino que tendrá un valor por lo general extremadamente pequeño y despreciable en la práctica. La potencia consumida en ese estado por tanto será prácticamente nula. También debe tenerse en cuenta que no se debe superar un voltaje de ruptura máximo en V_{DS} ⁴.

En la región triodo, el voltaje V_{GS} debe ser lo mayor posible, pero obviamente sin superar el valor máximo especificado por el fabricante (lo que podría suponer la ruptura del óxido). En este estado, si existe una corriente I_D a través del transistor el consumo de potencia vendría dado por:

$$P_{MOSFET(on)} = V_{DS} I_D = R_{on} I_D^2 \quad (\text{X.4})$$

El trabajo con un transistor real supone que deberán tenerse en cuenta las especificaciones del dispositivo concreto empleado, indicadas por el fabricante en la hoja de especificaciones o *datasheet*. En dicho documento se indican tanto las características geométricas del transistor como los valores máximos admitidos de voltaje, potencia, etc. Respecto a la potencia consumida, es importante respetar siempre las condiciones que garantizan el trabajo en el área de operación segura o

³Generalmente, se tratará del voltaje de alimentación en continua que denotaremos por V_{DD} .

⁴Si se aplica un valor excesivamente elevado de V_{DS} podría ocurrir un fenómeno de conducción directa entre fuente y drenador que no estuviera controlado por la puerta, lo que no es deseable.

safe operating area (SOA). Dicha área corresponde a los valores de voltaje y corriente soportados por el transistor sin exceder la potencia máxima. Esta área es mayor cuanto menor sea el tiempo que se somete al transistor a dichas condiciones: es decir, un transistor puede soportar unas condiciones más agresivas si tienen lugar durante un intervalo de tiempo pequeño.

De manera completamente análoga al transistor MOSFET, el transistor bipolar o BJT también puede emplearse para construir circuitos digitales. En este caso nos interesará trabajar en las regiones de corte y saturación (recordemos que por nomenclatura la región de saturación en un BJT es la equivalente a la región triodo del MOSFET).

Para definir el **modelo en conmutación del BJT** tendremos que considerar que las uniones base-emisor y base-colector son uniones de tipo diodo, por lo que su equivalencia vendrá determinada por el estado de polarización (directa o inversa) en el que se encuentren en cada caso. En el estado de corte en un BJT, tanto la unión base-emisor como la unión base-colector están polarizadas en inversa, por lo que puede considerarse que se comportan como circuitos abiertos, como se muestra en la figura X.4.

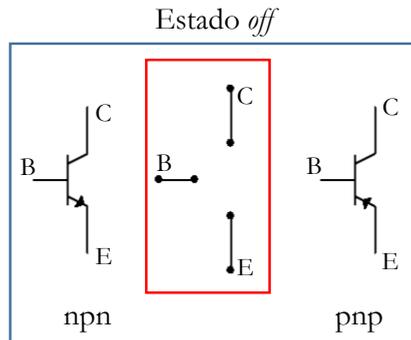


Figura X.4: Modelo del BJT en corte.

En cambio, en la región de saturación del BJT el comportamiento puede representarse mediante dos fuentes de voltaje, una fuente $V_{BE(on)}$ (cuyo valor en transistores reales ronda los 0.7 V) y otra fuente $V_{CE(sat)}$ (que en un transistor real es de aproximadamente 0.2 V), con la orientación que se indica en la figura X.5 para cada tipo de transistor⁵.

⁵Para disponer de un modelo más preciso, se puede incluir una resistencia en serie en el colector.

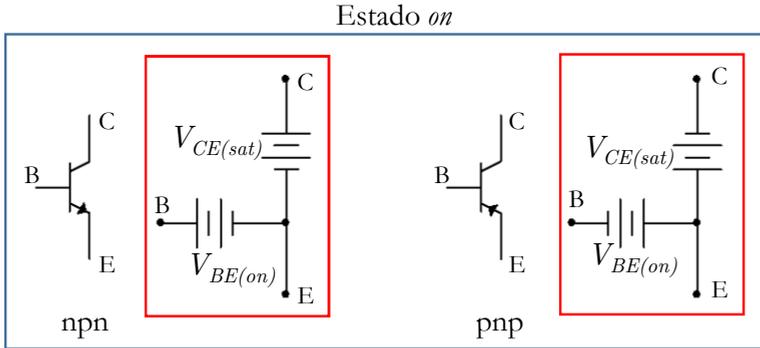


Figura X.5: Modelo del BJT en saturación.

Haciendo uso de estos modelos en conmutación es muy sencillo realizar el análisis de circuitos digitales.

3 UN EJEMPLO: CIRCUITO INVERSOR EN FUENTE COMÚN

Con el fin de analizar desde un punto de vista práctico el proceso de conmutación, vamos a considerar en primer lugar un circuito digital sencillo, un **inversor**, que realiza la siguiente función: cuando en su entrada tenemos un voltaje elevado, la salida tomará un valor bajo. Y viceversa, cuando en la entrada hay un valor de voltaje bajo, en la salida tendremos un valor alto.

Una posible implementación de un circuito inversor es el MOSFET en fuente común, que ya estudiamos en el caso de amplificación en el capítulo VIII. En la figura X.6 mostramos el circuito y el valor de la salida del mismo (V_o) en función de su entrada (V_i).

Si el voltaje de entrada V_i (que coincide con V_{GS} al encontrarse la fuente conectada a tierra) es pequeño e inferior al voltaje umbral (V_T) del transistor, éste se encontrará necesariamente en la región de corte. En ese caso, el transistor se comportará como un circuito abierto. Al no existir corriente que circule por R_D , la diferencia de potencial entre sus extremos será, a partir de la ley de Ohm, igual a 0 V, o lo que es lo mismo, el voltaje en los extremos de R_D deberá ser el mismo. Dado que en un extremo tenemos el voltaje fijado por la fuente de alimentación V_{DD} , en el otro extremo de la resistencia R_D , que es precisamente la salida V_o , deberemos tener también V_{DD} . Es lo que ocurre en el tramo inicial en el que la salida V_o toma un valor *alto*, precisamente cuando la entrada toma un valor *bajo*.

Al aumentar el voltaje de entrada, una vez superado el voltaje umbral el tran-

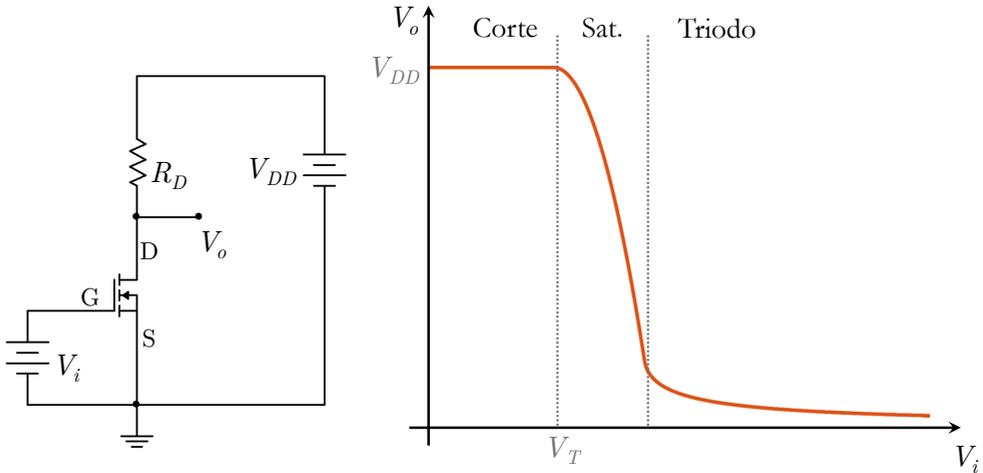


Figura X.6: Circuito inversor en fuente común (izquierda) y salida del circuito en función de su entrada.

El transistor entra en la región de saturación y empezamos a recorrer la recta de carga. A partir de ese momento el voltaje de salida V_o va decreciendo. Cuando alcanzamos la región de triodo, el transistor se comportará como una resistencia de valor R_{on} , que será más pequeña cuanto mayor sea V_{GS} (V_i en este caso). De manera simplificada, en la rama en la que están el drenador y fuente del transistor tendríamos dos resistencias, de valor R_D y R_{on} , con un voltaje V_{DD} en un extremo y tierra en el otro. En el punto medio el voltaje V_o será:

$$V_o = \frac{R_{on}}{R_{on} + R_D} V_{DD} \tag{X.5}$$

Dado que en las condiciones de trabajo habituales R_{on} va a tener un valor mucho más pequeño que R_D , el valor de V_o será bastante pequeño, aunque no nulo. En definitiva, un voltaje *alto* en la puerta del transistor produce en la salida un voltaje *bajo*. Se trata en consecuencia de un circuito inversor, que produce en la salida lo contrario de lo que se tenga en la entrada.⁶

El tiempo de respuesta del transistor (tiempos t_{on} y t_{off}) ante cada cambio de las condiciones de entrada va a influir en el comportamiento del circuito. En la figura X.7 mostramos cualitativamente una representación del valor del voltaje

⁶En el contexto de las señales digitales, donde un voltaje alto representa un 1 y un voltaje bajo un 0, el circuito realiza la operación NOT, que trataremos con más detalle en el capítulo XI.

de salida V_o ante una señal cuadrada V_i en la entrada⁷.

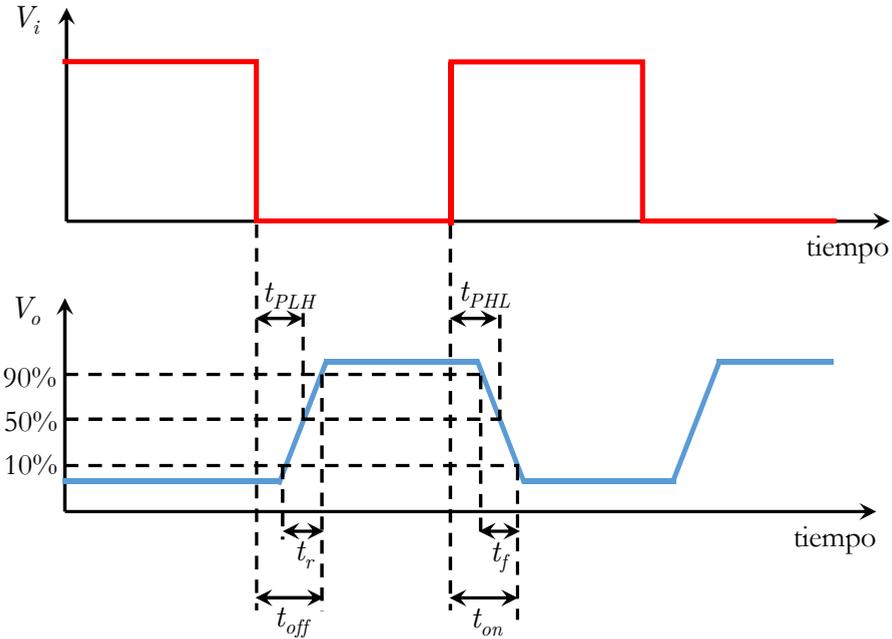


Figura X.7: Tiempos característicos en la conmutación.

Si bien lo mostrado en la figura es una representación aproximada, nos sirve para definir varios tiempos característicos en este circuito inversor, así como para describir con mayor precisión los tiempos t_{on} y t_{off} :

- t_{on} : tiempo que tarda el voltaje de salida en pasar del valor máximo al correspondiente al 10% de la diferencia entre el máximo y el mínimo en salida, desde el momento en el que conmuta la entrada.
- t_{off} : tiempo que tarda el voltaje de salida en pasar del valor mínimo al correspondiente al 90% de la diferencia entre el máximo y el mínimo en salida, desde el momento en el que conmuta la entrada.

⁷La salida mostrada no corresponde a una simulación o medida real de un transistor, sino que simplemente es una representación aproximada de lo que ocurre cuando los tiempos de conmutación de la señal de entrada son muy rápidos. En condiciones normales de operación, la señal de salida sería prácticamente cuadrada. En una situación real este tipo de señal aparecería a frecuencias elevadas, y no sería simétrico el comportamiento en subida y en bajada de la señal de salida.

- t_{PHL} : tiempo de propagación de alto a bajo (*high-to-low*). Es el tiempo en pasar del valor máximo del voltaje de salida al correspondiente al 50% de la diferencia entre el máximo y el mínimo en salida, desde el momento en el que conmuta la entrada.
- t_{PLH} : tiempo de propagación de bajo a alto (*low-to-high*). Es el tiempo en pasar del valor mínimo del voltaje de salida al correspondiente al 50% de la diferencia entre el máximo y el mínimo en salida, desde el momento en que conmuta la entrada.
- t_f : tiempo de caída (*fall*). Es el tiempo que tarda la salida en pasar del valor correspondiente al 90% de la diferencia entre el máximo y el mínimo en salida al 10%.
- t_r : tiempo de subida (*rise*). Es el tiempo que tarda la salida en pasar del valor correspondiente al 10% de la diferencia entre el máximo y el mínimo en salida al 90%.

Teniendo en cuenta la forma de la salida del circuito, es interesante analizar el consumo de potencia en el transistor a lo largo del tiempo. Por una parte, cuando la entrada está en un valor bajo, el transistor está en corte, de modo que idealmente la corriente será nula, y en consecuencia, no existirá consumo de potencia.

Sin embargo, cuando la entrada está en alto, el transistor conduce corriente, y además la diferencia de potencial V_{DS} no es nula debido a que R_{on} tampoco lo es. Además, en los tiempos transitorios de paso de corte a triodo y de triodo a corte, el transistor atraviesa la región de saturación, en la que también circulará corriente y V_{DS} no será nula, por lo que existirá un consumo de potencia. En la figura X.8 se muestra la forma simplificada del voltaje de entrada, el de salida, la corriente en el circuito y la potencia disipada en el transistor.

Como hemos señalado anteriormente, únicamente existe consumo de potencia cuando el transistor proporciona una salida en bajo (cuando está en triodo) y durante el tiempo transitorio de cambio de estado, es decir, mientras se completa el proceso de conmutación. A la potencia disipada durante la operación en triodo la denominaremos **potencia estática**, dado que nos encontraríamos en un punto de operación estacionario estable, mientras que la potencia consumida en los transitorios recibe el nombre de **potencia dinámica**.

Determinar la potencia estática es sencillo, pues al estar el transistor en triodo basta con considerar que según el modelo del MOSFET en conmutación

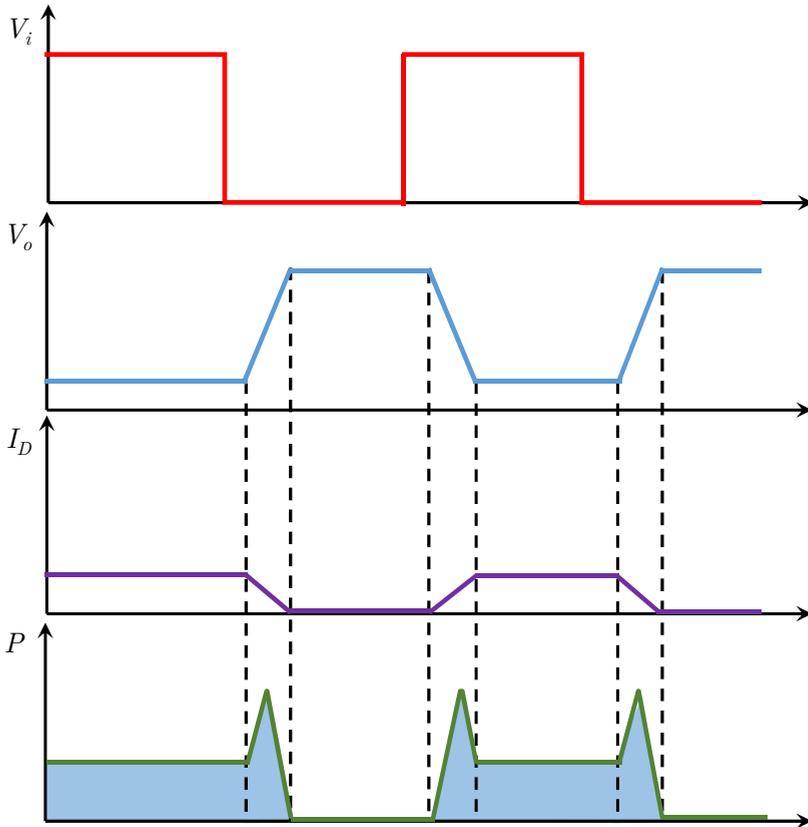


Figura X.8: Representación del consumo de potencia en un circuito inversor básico en fuente común (zonas sombreadas en la gráfica inferior). Las pérdidas dinámicas tienen lugar durante la conmutación, mientras que las pérdidas estáticas ocurren en la situación estacionaria en triodo.

podemos reemplazarlo por una resistencia entre fuente y drenador de valor R_{on} . Tendremos entonces que en el transistor la potencia disipada (potencia estática, P_S en este caso) será:

$$P_S = V_{DS}I_D = I_D^2 R_{on} = \left(\frac{V_{DD}}{R_D + R_{on}} \right)^2 R_{on} \quad (\text{X.6})$$

El consumo dinámico de potencia, por su parte, está relacionado con el proceso de creación y desalojo de la capa de inversión en el transistor, según pasemos de corte a triodo o de triodo a corte, respectivamente. Cuando está formada la

capa de inversión, ésta corresponde a una carga eléctrica situada bajo el óxido, carga que tiene su reflejo en otra igual y de signo contrario en el metal de la puerta. Es decir, la estructura MOS (metal-óxido-semiconductor) se va a comportar de forma similar a un condensador cargado (dos cargas opuestas e iguales separadas por un aislante). Por tanto, también puede interpretarse la conmutación como el proceso de carga (paso a *on*) y descarga (paso a *off*) de un condensador.

La capacidad del condensador asociado a la puerta del MOSFET la vamos a denotar por C_G , que se puede dividir a su vez en las capacidades puerta-fuente (C_{GS}) y puerta-drenador (C_{GD}) que aparecen en el modelo de pequeña señal del MOSFET, siendo $C_G = C_{GS} + C_{GD}$.

Así, podemos considerar que el trabajo que es necesario realizar para crear la capa de inversión (es decir, *cargar* el condensador de la estructura MOS) será:

$$U = C_G V_{GS}^2 \tag{X.7}$$

Existirá por tanto un consumo de potencia asociado, la potencia dinámica P_D , que vendrá dada por:

$$P_D = C_G V_{GS}^2 \frac{1}{T} = C_G V_{GS}^2 f \tag{X.8}$$

donde f es la frecuencia de la señal de entrada (asumiendo una señal cuadrada que conmuta con un periodo T).

Es decir, existe un consumo de potencia asociado al proceso de conmutar la señal, y que será mayor cuanto mayor sea la frecuencia de los cambios, es decir, cuanto más elevada sea la frecuencia de operación del circuito (en un microprocesador, por ejemplo, su señal de reloj).

Pero una frecuencia de operación elevada no presenta únicamente el problema de un alto consumo de potencia dinámica. El hecho de que los tiempos de conmutación no sean nulos implica que existe un límite respecto a la frecuencia máxima a la que puede trabajar un circuito digital. Si ordenamos al transistor que conmute de bajo a alto (o viceversa) y volvemos a requerir un cambio antes de que se haya completado el primer proceso, la señal de salida puede sufrir distorsiones que hagan que el funcionamiento del circuito sea inadecuado, como podemos observar en la figura X.9.

Puede obtenerse una estimación aproximada de la frecuencia máxima de operación a partir de los tiempos t_{on} y t_{off} , de la forma:

$$f_{max} < \frac{1}{t_{on} + t_{off}} \tag{X.9}$$

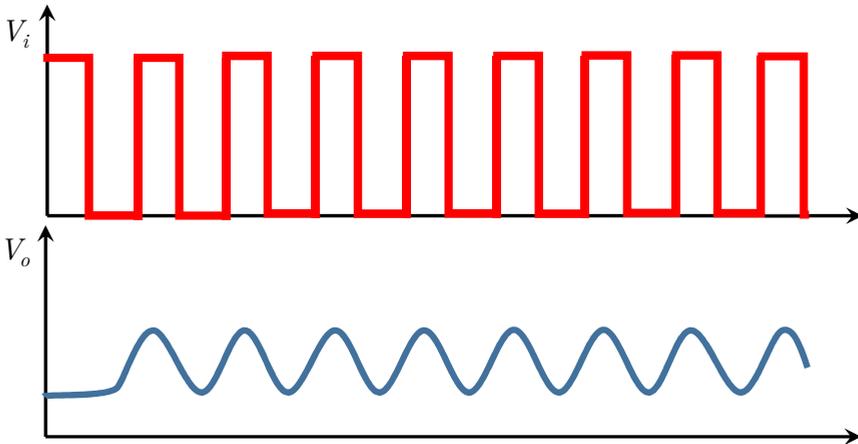


Figura X.9: Ejemplo de mal funcionamiento del inversor al aplicar una señal de entrada de frecuencia excesivamente elevada: en la salida ya no aparece una señal cuadrada invertida.

El circuito que hemos analizado en esta sección se puede realizar de manera sencilla empleando componentes discretos. Sin embargo, en un circuito integrado podría realizarse de forma más compacta reemplazando adecuadamente la resistencia R_D . Cuando se implementa un determinado circuito en la forma de circuito integrado⁸, el espacio que ocupa el circuito en la oblea⁹ de semiconductor es un parámetro fundamental, dado que siempre debe procurarse que sea lo menor posible (cuanto menos ocupe un circuito dado más cabrán en la oblea y el coste económico por unidad será menor). Integrar una resistencia requiere de una superficie grande, ya que suele realizarse con una región no dopada de semiconductor que tenga una longitud relativamente larga. Para minimizar la ocupación en superficie se suelen realizar con forma de serpentin plano, pero aún así el espacio ocupado es apreciable. Con el fin de evitar este problema, una posible solución es reemplazar la resistencia por un transistor MOSFET trabajando en la región de saturación (figura X.10), en la que es posible la conducción de corriente pero comportándose como una resistencia relativamente elevada (a diferencia de triodo, donde la resistencia es baja). Un transistor MOSFET integrado ocupa mucho menos espacio que una resistencia.

⁸En el capítulo XII discutiremos con más detalle las tecnologías de fabricación de circuitos integrados.

⁹Una oblea de semiconductor es un disco delgado de semiconductor sólido cristalino en el

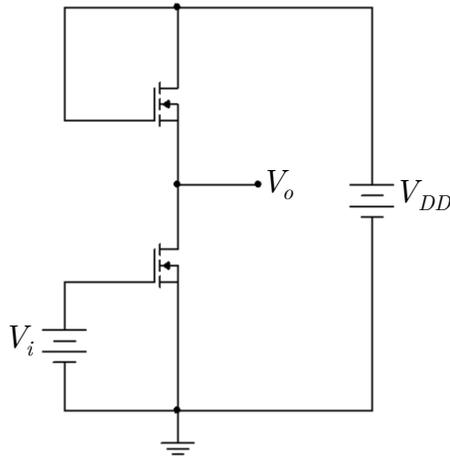


Figura X.10: Circuito inversor con carga integrada.

El MOSFET que reemplaza a la resistencia tiene la fuente y el drenador cortocircuitados (por tanto, $V_{GS} = V_{DS}$). Al ser canal n y de realce ello implica que necesariamente $V_{DS} > V_{GS} - V_T$, cumpliéndose la condición de saturación.

Esta solución no resuelve el principal problema del inversor estudiado con anterioridad, que es el elevado consumo de potencia cuando el transistor principal está en el régimen de triodo. Al comportarse el transistor de carga como una resistencia sigue existiendo una corriente no nula en esas condiciones, lo que implica un consumo estático de potencia.

4 EL INVERSOR CMOS

Para solucionar el problema del consumo de potencia cuando la salida toma un valor bajo (es decir, vale 0 en el contexto de la electrónica digital), se emplea una configuración inversora conocida como **CMOS** (*Complementary MOS* o MOS complementario), que se basa en el uso de dos transistores MOSFET de realce, uno canal n y otro canal p adecuadamente acoplados, como se muestra en el circuito de la figura X.11 (en el circuito la fuente de voltaje V_{DD} se muestra de forma simplificada, y la entrada se indica sin dibujar explícitamente la fuente correspondiente; por simplicidad a partir de ahora seguiremos este criterio).

Dado que el transistor MOSFET canal n trabaja con corrientes que entran

que se fabricarán varios microchips, como se explicará en el capítulo XII.

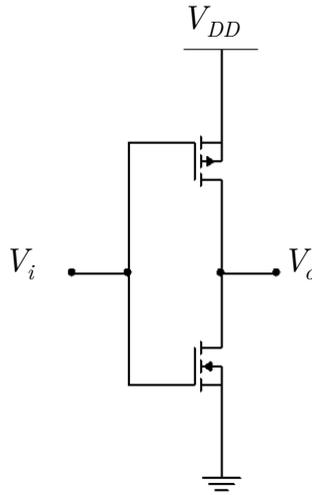


Figura X.11: Circuito inversor CMOS.

hacia el drenador, mientras que el MOSFET canal p trabaja con corrientes que salen desde el drenador, la conexión se realiza poniendo en común el drenador de ambos transistores (que constituirá la salida del circuito), de modo que la corriente que salga del MOSFET canal p entre al MOSFET canal n (la corriente I_D , de existir, debe ser la misma en los dos transistores). Además, como el transistor canal p trabaja con voltajes negativos su fuente se conecta a la alimentación positiva (el valor más grande de potencial que habrá en el circuito) para que su V_{DS} sea negativo, mientras que la fuente del transistor canal n se conecta a tierra, lo que garantiza que su voltaje V_{DS} será positivo. La puerta de los dos transistores se conecta en común, constituyendo la entrada del circuito.

Para el MOSFET canal n se cumplirá que su voltaje de puerta V_{GSn} será:

$$V_{GSn} = V_{Gn} - V_{Sn} = V_i - 0 = V_i \quad (\text{X.10})$$

Es decir, que coincide con el voltaje de entrada, por lo que el MOSFET canal n conducirá si $V_i > V_{Tn}$, siendo V_{Tn} el voltaje umbral de dicho MOSFET.

Para el MOSFET canal p, en cambio, tendremos que su voltaje de puerta V_{GSp} será:

$$V_{GSp} = V_{Gp} - V_{Sp} = V_i - V_{DD} \quad (\text{X.11})$$

y por tanto el transistor canal p conducirá si se cumple que $V_i - V_{DD} <$

$-|V_{Tp}|$, de donde, despejando, la condición sobre la entrada es que $V_i < V_{DD} - |V_{Tp}|$, siendo V_{Tp} el voltaje umbral del transistor canal p¹⁰.

Si en este circuito representamos el voltaje de salida V_o en función del de entrada V_i obtendremos una gráfica como la de la figura X.12, en la que se ha considerado una alimentación V_{DD} de 5 V, con una entrada V_i que varía entre 0 y 5 V.

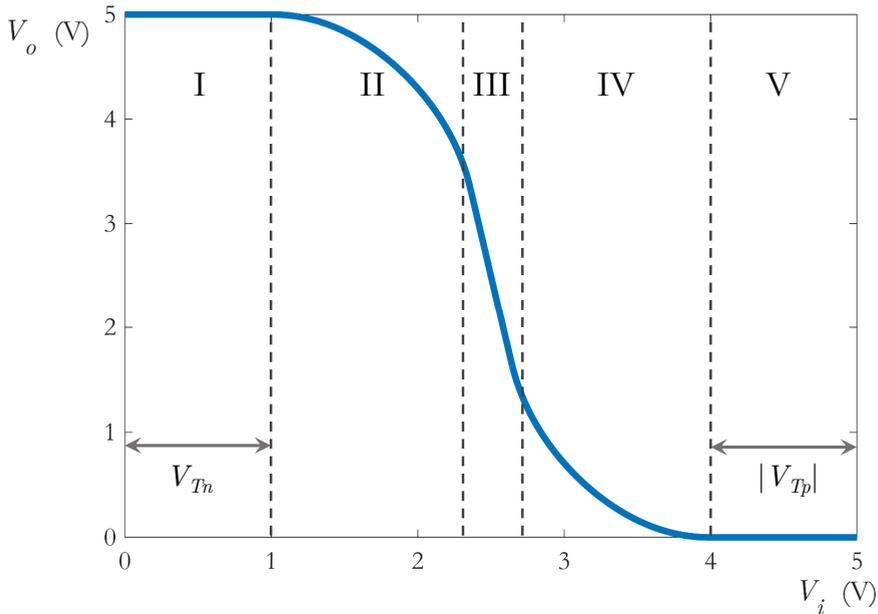


Figura X.12: Salida V_o del inversor CMOS en función de la entrada V_i .

Como vemos, existen cinco regiones diferentes. En la región I, que corresponde a un voltaje de entrada inferior a V_{Tn} (es decir, un voltaje bajo), el MOSFET canal n está en estado *off* (corte) y el MOSFET canal p en estado *on* (triodo)¹¹. Por tanto, el MOSFET canal n impedirá que circule corriente, y al ser un

¹⁰Para evitar posibles confusiones, hemos tenido en cuenta que como V_{Tp} es negativo, se cumple que $V_{Tp} = -|V_{Tp}|$.

¹¹En realidad, al estar el transistor canal n en corte no existiría corriente, por lo que no podríamos decir en propiedad que el transistor p está operando en triodo: en realidad lo que ocurre es que existe una capa de inversión formada y homogénea en todo el canal (como ocurriría en el caso de triodo) pero no hay una diferencia de potencial entre drenador y fuente que haga circular corriente (al ser mucho más resistivo el transistor canal n todo el voltaje cae entre sus extremos). Por ello, ejerciendo un cierto abuso del lenguaje diremos que el transistor p está en triodo, aunque no circule corriente por él.

elemento muy resistivo (idealmente un circuito abierto) absorberá toda la caída de voltaje entre la fuente de alimentación V_{DD} y la referencia (tierra) de la parte inferior. En consecuencia, en la salida del circuito V_o tendremos un valor igual a V_{DD} , es decir, una salida en alto.

En cambio, en la región V (entrada en alto, con un valor cercano a V_{DD}), la situación es la opuesta. El transistor canal n está trabajando con un voltaje de entrada muy elevado, lo que lo sitúa en una situación de fuerte inversión (estado *on*, en triodo si hubiera conducción). Sin embargo, el transistor canal p no alcanza el voltaje mínimo necesario para conducir y se encuentra en estado *off*, lo que bloquea el paso de corriente. Al ser este último muy resistivo (idealmente se comportaría como un interruptor abierto) toda la caída de potencial tiene lugar en él, de modo que en la salida V_o tendremos prácticamente 0 voltios, es decir, un voltaje en bajo.

En definitiva, en las regiones I y V (es decir, para entradas en bajo y alto, respectivamente), el circuito se comporta como un inversor, produciendo una salida opuesta (salida en alto y bajo para cada caso), con la característica principal de que no existe corriente I_D , y por tanto, el circuito no consumirá potencia, a diferencia de los ejemplos anteriores en los que en uno de los estados (salida en bajo) circulaba corriente y existía un consumo. Ello hace que los circuitos CMOS sean especialmente interesantes, al confluir en ellos dos características vitales en circuitos integrados: por una parte, espacio reducido, al no requerir de resistencias, y por otra, consumo de potencia prácticamente nulo en situación estacionaria (cuando la salida se mantiene estable en alto o en bajo).

Sin embargo, en el proceso de conmutación (el paso de la región I a la V o viceversa) la situación es diferente. En la región II el MOSFET canal n está en saturación y el canal p se mantiene en triodo, por lo que los dos pueden conducir corriente, y al existir un voltaje de alimentación V_{DD} y dos elementos conductores (los dos transistores) existirá corriente I_D . En la región III ambos transistores están en saturación¹², y en la región IV el canal n está en triodo y el canal p en saturación, por lo que de nuevo, para ambas regiones, existirá una corriente I_D . Es decir, a lo largo de todo el proceso de conmutación (en el paso de un transistor en corte a que esté en corte el otro) existirá corriente y en consecuencia un consumo de potencia. En consecuencia el circuito CMOS consume energía casi exclusivamente en el proceso de cambio de estado.

¹²Esta circunstancia sólo tendrá lugar si los transistores son no ideales (por ejemplo, con canales muy cortos) y la corriente en saturación no es constante sino que presenta un ligero incremento con V_{DS} .

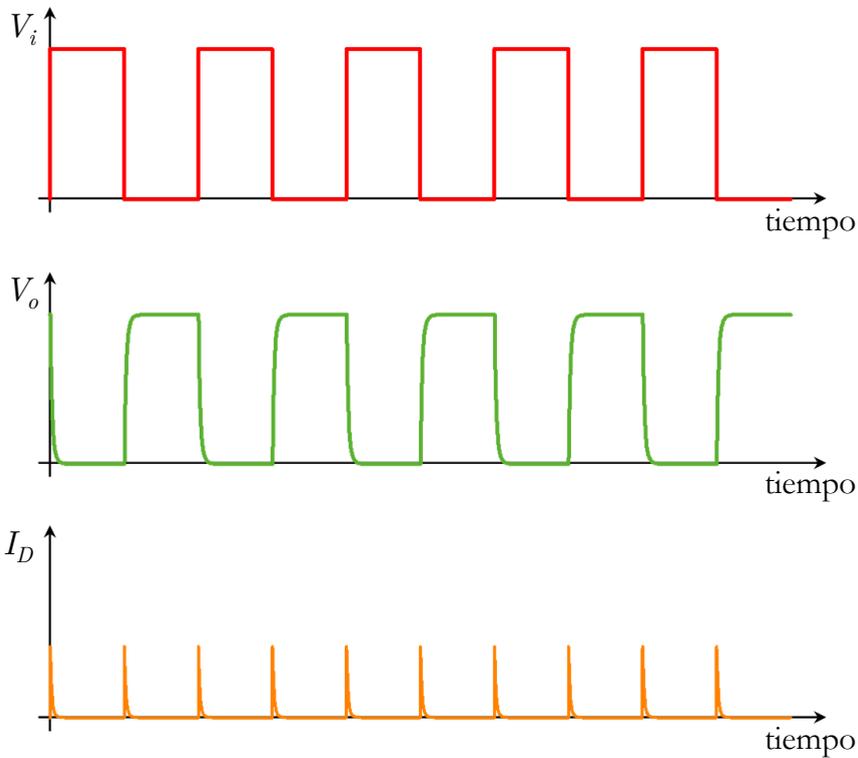


Figura X.13: Entrada, salida y corriente en un inversor CMOS en función del tiempo.

En la figura X.13 se muestra un voltaje de entrada V_i en función del tiempo. Se trata de una señal cuadrada que oscila entre dos estados, alto y bajo. La salida del circuito V_o tiene la forma opuesta, como corresponde a un circuito inversor. Además, se muestra la corriente I_D . Como hemos mencionado con anterioridad, sólo existen picos de corriente (y por tanto consumo de potencia) en los cambios de estado, es decir, únicamente se da un consumo dinámico, siendo el estático prácticamente nulo. Por tanto, en este tipo de circuitos el consumo de potencia eléctrica irá ligado a la frecuencia de operación, ya que a mayor frecuencia más cambios de estado se producirán en un determinado periodo de tiempo¹³.

¹³Una práctica habitual, aunque no recomendable, para mejorar el rendimiento de los ordenadores es el llamado *overclocking*, que consiste en aumentar la frecuencia de reloj por encima de la

En la actualidad la inmensa mayoría de circuitos digitales se realizan empleando esta técnica que combina los dos tipos de transistores, como veremos en el capítulo XI, donde estudiaremos con detalle la implementación de funciones lógicas. Para facilitar la lectura de los diagramas de los circuitos digitales, con frecuencia se utilizan símbolos simplificados para los MOSFET canal n y canal p, como se muestra en la figura X.14 (téngase en cuenta que el MOSFET canal p, en los circuitos CMOS, se coloca con la fuente en la parte superior, para ser conectada a la alimentación V_{DD}).

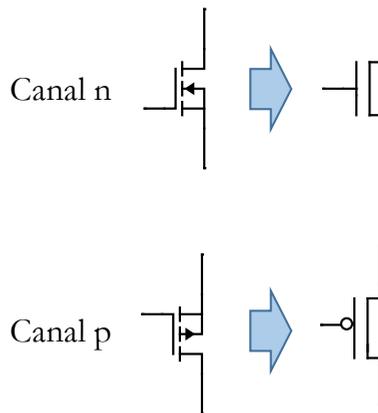


Figura X.14: Símbolos simplificados del transistor MOSFET canal n y canal p para circuitos digitales.

5 APLICACIONES DEL MOSFET EN CONMUTACIÓN

Además de la construcción de puertas y funciones lógicas, que detallaremos en el siguiente capítulo, en el ámbito de la informática existen numerosas aplicaciones en las que se trabaja con un transistor en conmutación.

Las memorias **DRAM** (*Dynamic Random Access Memory*, o memoria dinámica de acceso aleatorio) son un tipo de memorias en el que cada celda (que corresponde a un bit) consiste en un condensador que almacena o no carga, según guarde un 1 o un 0. En la figura X.15 se muestra una representación de una memoria DRAM de 16 bits (organizados en una matriz de celdas con cuatro filas y cuatro

recomendada. Ello supone una mejora de la velocidad del sistema, pero implica un mayor consumo de potencia y mayor calor generado en el funcionamiento del ordenador, lo que podría acabar dañando al microprocesador si se realiza de manera prolongada.

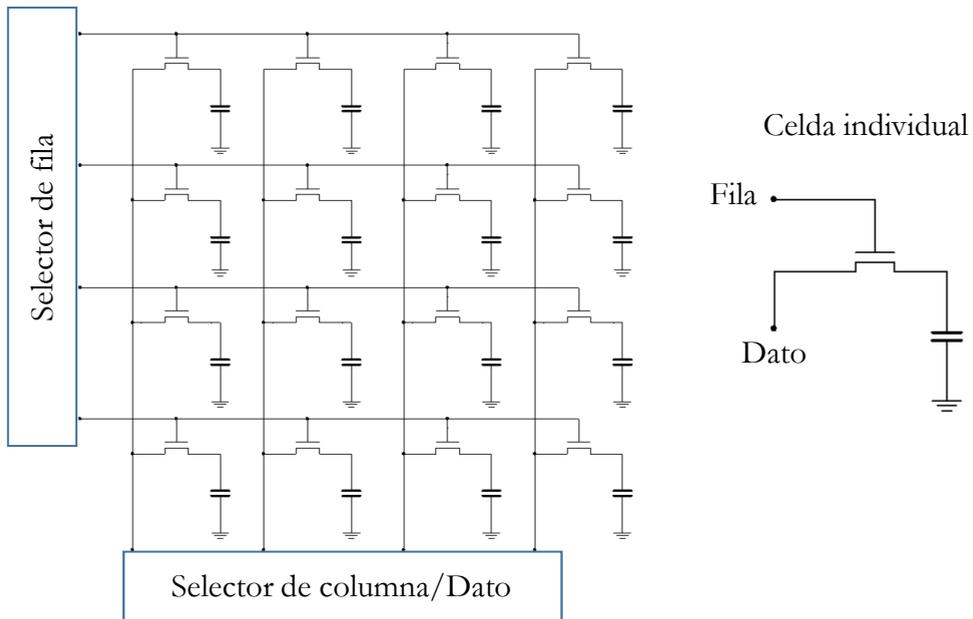


Figura X.15: Estructura de una memoria dinámica de acceso aleatorio (DRAM) con una matriz de 4x4 bits, y detalle de una de las celdas de almacenamiento de un bit.

columnas). Para acceder a la carga almacenada, se emplea un transistor MOSFET funcionando en conmutación, de forma que normalmente estará en estado *off*, y cuando se quiera efectuar la lectura de la celda, se activará la puerta (el transistor de acceso pasará a estado *on*) mediante una señal en alto aplicada en la fila correspondiente, y se aplicará simultáneamente un voltaje en la columna de la celda seleccionada para extraer la carga del condensador¹⁴.

Este tipo de memorias tiene la gran ventaja de que únicamente son necesarios un transistor y un condensador MOS para almacenar un bit (el condensador para almacenar la carga y el transistor para acceder a ella), y por ello suelen constituir la memoria principal de los ordenadores. Sin embargo, la carga almacenada en el condensador se pierde progresivamente (no permanece almacenada de manera indefinida) y además el proceso de lectura es destructivo (se pierde la carga), por lo que es necesario refrescar la memoria de forma periódica para no perder los datos almacenados.

¹⁴El condensador en una memoria integrada puede realizarse de manera sencilla con una estructura MOS.

Las memorias **SRAM** (*Static Random Access Memory*, o memoria estática de acceso aleatorio), en cambio, no necesitan ser refrescadas. Poseen una mayor rapidez de lectura que las DRAM, dado que no se basan en la descarga de un condensador, sino en la lectura de un estado almacenado (un voltaje).

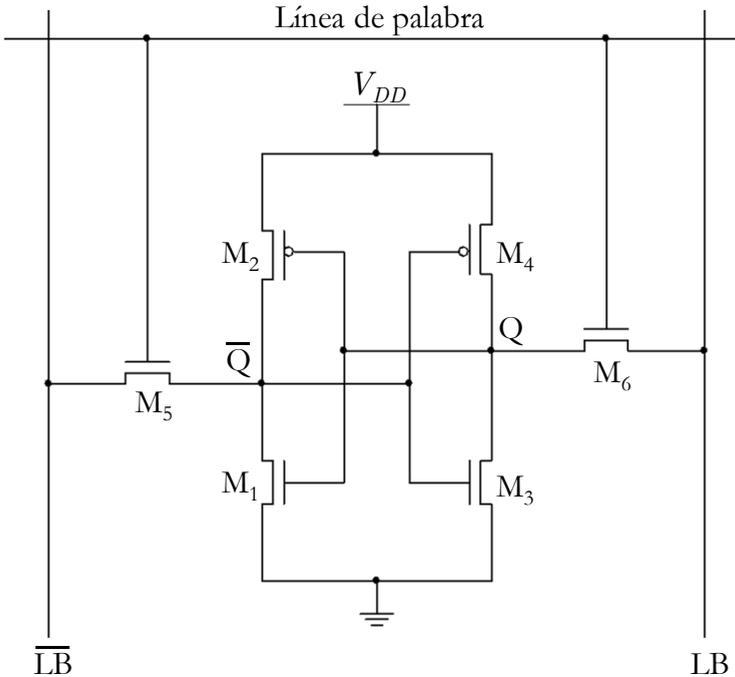


Figura X.16: Estructura de una celda SRAM.

Cada celda de almacenamiento de un bit (véase la figura X.16) se compone de seis transistores, cuatro que conforman un circuito biestable (que tiene dos posibles estados estables), que en la imagen corresponden a dos inversores CMOS acoplados formados por los transistores M₁/M₂ y M₃/M₄, y dos transistores para acceder a la lectura/escritura del bit, M₅ y M₆. La celda almacena tanto el valor del bit, Q, como su complementario, \bar{Q} . De nuevo, estos transistores funcionan en modo de conmutación. El acceso a la lectura o escritura se realiza a través de la línea de palabra, activando los transistores M₅ y M₆, mientras que la lectura o escritura del bit y su valor complementario se realiza mediante las líneas de bit LB. La densidad de estas memorias es mucho menor que la de las DRAM, y se emplean por ejemplo en memorias caché, almacenamiento de registros, etc. Es decir, aplicaciones de acceso rápido y que no requieren relativamente de una gran

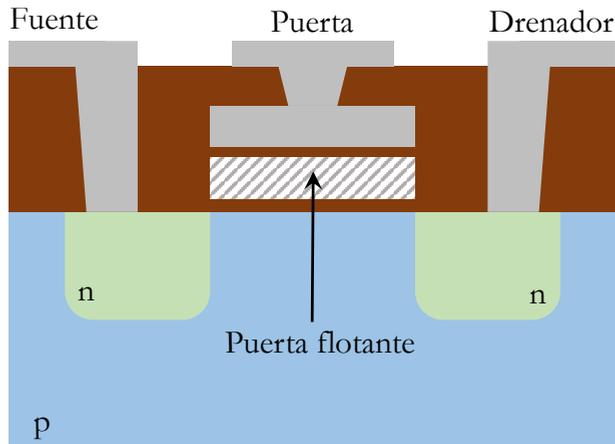


Figura X.17: Estructura de un transistor de puerta flotante, empleados en las memorias de tipo flash.

cantidad de memoria.

Las memorias **flash** también hacen uso de la conmutación de un transistor para funcionar. En este tipo de memorias (presentes en las conocidas como memorias USB, en las memorias SD, discos de estado sólido SSD, etc.) cada celda está constituida por un tipo de transistor especial con dos puertas (figura X.17), una puerta de control y una puerta flotante (un trozo de metal rodeado de óxido, sin conectar eléctricamente con ninguna otra parte del circuito). La puerta flotante actúa de la siguiente manera: si está llena de carga (electrones), apantalla a la puerta de control, de modo que si se da la orden de pasar al estado *on* no se observan cambios. Sin embargo, si está vacía, la puerta de control puede actuar sobre el canal conductor y realizar la conmutación. De este modo podemos asignar un 0 o un 1 a cada situación. El proceso de carga o descarga de la puerta flotante se realiza aplicando un voltaje más elevado a la puerta de control, de modo que por el denominado **efecto túnel**¹⁵ pueden acceder los electrones a la puerta flotante.

Pero los usos de transistores en conmutación se extienden más allá de las memorias. En las **pantallas planas TFT-LCD** (*Thin Film Transistor Liquid Cristal Display*, o pantalla de cristal líquido con transistor de capa delgada, concepto que incluye a los monitores y televisores LED convencionales) cada pixel está controlado por un transistor funcionando en conmutación: según el voltaje aplicado

¹⁵El efecto túnel es un efecto de naturaleza cuántica por el cual un electrón es capaz de atravesar la barrera de energía que conforma una capa delgada de óxido.

a su puerta, actuará sobre un cristal líquido que obstruirá o permitirá el paso de luz en cada caso.

Las **cámaras fotográficas digitales** también hacen uso de transistores en conmutación. Disponen de sensores de imagen en cada pixel, que pueden ser de dos tipos, CMOS o CCD¹⁶¹⁷. Por cada pixel existe una parte fotosensible (figura X.18) en la que se genera carga eléctrica en función de la luz que llegue durante el tiempo de exposición de la fotografía. La lectura de la carga de cada pixel puede realizarse activando un transistor de control funcionando en conmutación (conceptualmente de forma similar a como se efectuaba la lectura de una celda de memoria DRAM).

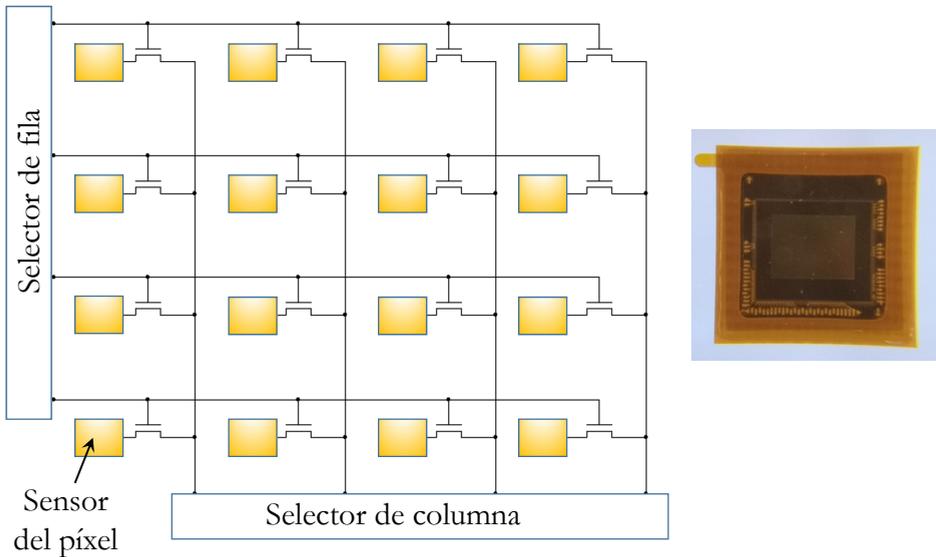


Figura X.18: Estructura simplificada de píxeles del sensor de una cámara digital (izquierda) e imagen de un sensor de cámara CMOS (derecha; imagen cortesía de Tomás González Sánchez).

En definitiva, los usos y aplicaciones de los transistores MOSFET en conmutación van mucho más allá de la fabricación de circuitos digitales, siendo la base de numerosas tecnologías de almacenamiento, visualización o captura de imágenes.

¹⁶CCD son las siglas de *Charge-Coupled Device*, o dispositivo de carga acoplada.

¹⁷Las diferencias entre los sensores CCD y CMOS se encuentran principalmente en el modo de procesar la información colectada por las celdas sensoras de cada pixel.

XI.

FAMILIAS LÓGICAS Y
DISEÑO CMOS

La electrónica es en gran parte una disciplina modular: a partir de una serie de circuitos que realizan una determinada función, es posible construir circuitos más complejos, que a su vez pueden ser bloques funcionales para otros circuitos de complejidad aún mayor. Esta característica es aún más evidente en el caso de los circuitos digitales que constituyen los microprocesadores. Si examinásemos con detalle el esquema de un microprocesador, comprobaríamos que en su nivel más elemental existen numerosos transistores que conforman **puertas lógicas**, es decir, circuitos digitales sencillos que implementan una determinada operación lógica; dichas puertas se combinan adecuadamente para realizar funciones más complejas, y sucesivamente se va creciendo en escala de complejidad hasta completar la arquitectura del microprocesador.

En último término, cualquier operación que realiza un ordenador o un teléfono inteligente, por ejemplo, puede ser reducida a operaciones sobre bits. En definitiva, son circuitos digitales que se basan en la implementación de diferentes operaciones mediante lógica binaria, utilizando, preferentemente, transistores.

1 PRINCIPALES FAMILIAS LÓGICAS

Una **familia lógica** es un *conjunto de circuitos digitales simples que implementan las operaciones lógicas básicas (AND, OR, NOT) y derivadas (NAND, NOR, XOR, etc.), empleando una determinada tecnología de fabricación y tipo de dispositivo*. Ello nos permite sintetizar cualquier función lógica, en primer lugar, y de forma general, cualquier circuito digital. Cada familia lógica corresponde a una forma específica de diseñar los operadores lógicos básicos, caracterizándose por un conjunto de parámetros estáticos y dinámicos.

Las familias lógicas se pueden clasificar en dos grandes grupos:

- Familias lógicas pasivas: Las que emplean como elementos constituyentes resistencias y diodos.
- Familias lógicas activas: Son aquellas que emplean como constituyentes dispositivos electrónicos activos, en concreto transistores, que al proporcionar ganancia ofrecen un rendimiento mucho mejor.

Las familias lógicas pasivas hoy día están en desuso por sus pobres prestaciones, aunque técnicamente es posible realizar funciones lógicas haciendo uso de la capacidad de los diodos de actuar de manera similar a un interruptor. En la figura XI.1 mostramos una puerta lógica AND de dos entradas implementada con diodos. Cuando cualquiera de las dos entradas A y B vale 0 (un valor de voltaje bajo) el diodo correspondiente estará polarizado en directa, haciendo que en la salida S se obtenga un voltaje bajo (en realidad, el voltaje umbral del diodo, que será aproximadamente 0.6 V, más el valor de entrada *bajo*); sin embargo, cuando las dos entradas están en alto ambos diodos estarán polarizados en inversa, no podrá haber circulación de corriente y en la salida se obtendrá un valor de voltaje alto (idealmente, V_{CC}).

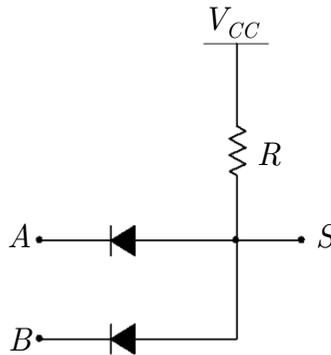


Figura XI.1: Puerta lógica AND de dos entradas realizada con diodos y una resistencia.

Esta implementación tiene, sin embargo, una serie de problemas. En primer lugar, el valor de salida en bajo (el 0) corresponde a la entrada en bajo más el voltaje umbral del diodo. Si encadenásemos varias puertas, se incrementaría sucesivamente el voltaje *bajo* y rápidamente dejaría de serlo. Además, cualquier circuito que acoplásemos en la salida tendría repercusión de vuelta en la entrada, por lo que en realidad no sería posible de manera realista acoplar varias puertas

para realizar funciones más complejas. Por último, sería un circuito demasiado lento para ser útil. En definitiva, todos estos problemas hacen que en la actualidad la consideración de familias lógicas realizadas con elementos pasivos esté completamente descartada.

Las familias lógicas activas, que emplean transistores, presentan prestaciones muy superiores. Históricamente algunas de las primeras fueron las familias RTL (siglas de *Resistor-Transistor Logic*, o lógica resistencia-transistor) y las familias DTL (siglas de *Diode-Transistor Logic*, o lógica diodo-transistor). A modo de curiosidad, el sistema de guiado del *Apollo XI*, la primera misión tripulada que alcanzó la superficie de la Luna en julio de 1969, realizaba los cálculos pertinentes empleando funciones lógicas NOR de tres entradas implementadas con resistencias y transistores bipolares (lógica RTL): en la figura XI.2 se muestra un esquema de estas puertas. Si cualquiera de las entradas toma un valor alto, activa al transistor bipolar (lo lleva a su región de saturación, la región lineal en el BJT) y fuerza a la salida a tomar un valor bajo, cercano a cero. En cambio, si todas las entradas toman valor bajo todos los transistores estarían cortados, no existiría corriente y la salida sería aproximadamente V_{CC} , es decir, un valor alto. En definitiva, efectúa la operación NOR.

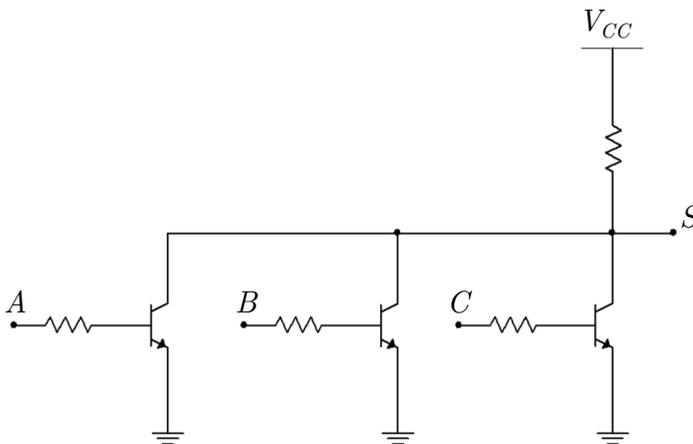


Figura XI.2: Puerta lógica NOR de tres entradas con tecnología RTL, similar a las empleadas en el sistema de guiado del *Apollo XI*.

El principal problema de las familias RTL era su elevado consumo energético, debido a la potencia disipada en las resistencias que se emplean, por lo que pronto fueron reemplazadas por otras familias lógicas más eficientes. Entre ellas,

podemos señalar las fabricadas con transistores bipolares, en concreto la familia **TTL** (siglas de *Transistor-Transistor Logic*, o lógica transistor-transistor), que fue muy popular en los años 70. En la figura XI.3 se muestra el esquema de una puerta NAND de dos entradas realizada con esta tecnología.

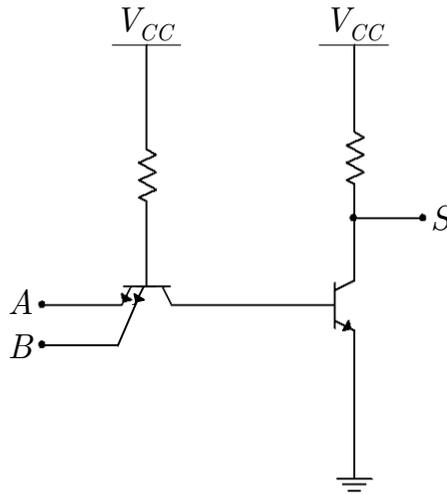


Figura XI.3: Puerta lógica NAND de dos entradas con tecnología TTL.

Como podemos ver en la figura, el aspecto más llamativo es el uso de transistores **multiemisor**, es decir, con tantos contactos de emisor como entradas vaya a tener la puerta lógica. Cuando en alguna de estas entradas se conecta un valor de voltaje bajo, ello lleva a la unión emisor-base del transistor multiemisor a estar polarizada en directa, de modo que la corriente fluye desde la base hacia el emisor que tenga el 0, y no existirá flujo de corriente hacia la base del segundo transistor, por lo que éste estará en corte. La salida en ese caso será aproximadamente igual a V_{CC} , es decir, corresponderá a un valor alto.

En cambio, si todas la entradas toman valor alto, la unión emisor-base del transistor multiemisor estará polarizada en inversa, la unión base-colector estará en directa (trabaja en la región activa inversa) y fluirá corriente hacia la base del segundo transistor, llevándolo a saturación y activándolo, y haciendo que la salida tome un valor bajo. El circuito realiza por tanto la función NAND (la salida únicamente valdrá 0 cuando todas las entradas valgan 1).

Las familias TTL se caracterizan por una gran velocidad de respuesta ante cambios en las entradas, así como por la posibilidad de alimentar con su salida

a otros circuitos, al proporcionar una corriente. Sin embargo, su consumo de potencia no es óptimo, pues al trabajar con corrientes no nulas en ciertos estados estacionarios (cuando no cambian las entradas) es inevitable el gasto de energía en estas situaciones.

En la actualidad las familias lógicas más empleadas son las **CMOS**, que se basan en transistores MOSFET, y que constituyen la inmensa mayoría de circuitos digitales que se fabrican. Hoy en día en algunas aplicaciones se siguen empleando circuitos TTL, pero con una salvedad. Ya no se fabrican con transistores bipolares, sino con transistores MOSFET. Es decir, en realidad son circuitos **compatibles TTL**, que trabajan con los niveles de voltaje y corriente de los circuitos TTL originales, pero utilizando transistores MOSFET. Las familias CMOS presentan una gran ventaja frente a las TTL, consistente en un reducido consumo de potencia (sólo tiene lugar durante el proceso de conmutación, y no en estado estacionario). Además, dado que a lo largo de varias décadas la tecnología de fabricación de circuitos integrados con transistores MOSFET ha evolucionado de forma espectacular, es posible realizar circuitos extraordinariamente complejos, con un número muy elevado de transistores, empleando esta tecnología CMOS. Son en la actualidad las familias dominantes, con gran diferencia, por lo

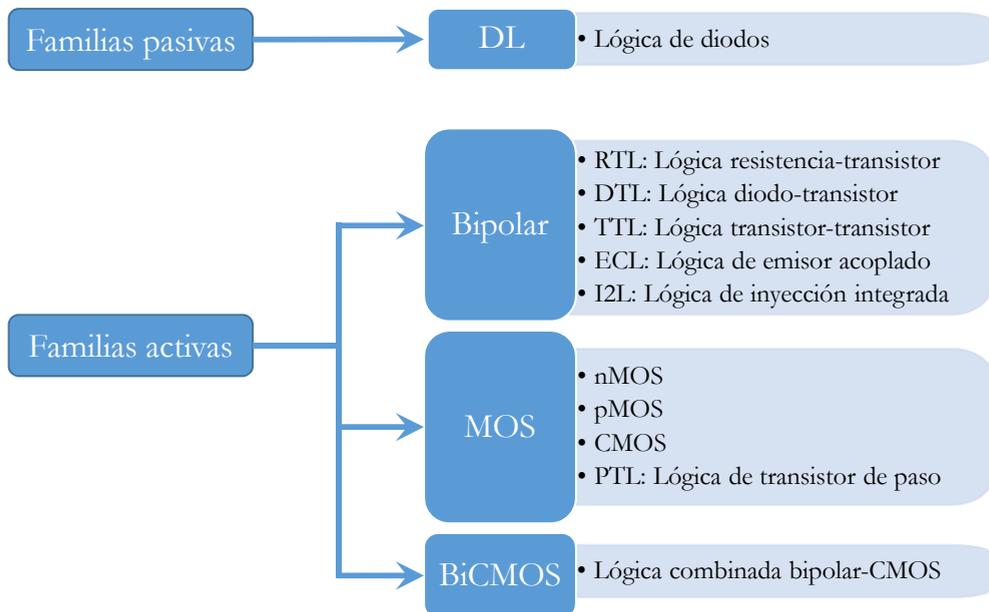


Figura XI.4: Algunas de las principales familias lógicas.

que estudiaremos con mayor detalle las puertas lógicas básicas construidas con transistores MOSFET complementarios.

En la figura XI.4 se muestra una clasificación de las principales familias lógicas, donde se debe tener en cuenta, como acabamos de señalar, la enorme preponderancia actual de las familias CMOS.

2 FUNCIONAMIENTO DE PUERTAS LÓGICAS CMOS BÁSICAS

A continuación vamos a ver cómo se pueden implementar las funciones lógicas básicas (NOT, AND, OR, NAND, NOR) empleando la tecnología CMOS. Presentaremos los esquemas de los circuitos y explicaremos su funcionamiento en función de las posibles entradas. Por simplicidad, consideraremos en general dos entradas en las puertas AND, OR, NAND y NOR.

Puerta NOT

El circuito inversor CMOS ya ha sido estudiado con detalle en el capítulo X, siendo en esencia la base de la función NOT (es decir, complementa -invierte- una variable). La clave es el efecto que tiene un mismo voltaje en la entrada sobre las puertas de los dos transistores canal n y canal p, y la conexión de la fuente del transistor canal n a tierra y la del canal p a la fuente de alimentación V_{DD} .

Cuando la entrada está en alto (voltaje elevado, por ejemplo V_{DD} , correspondiente a un 1), el transistor MOSFET canal n pasa a estar en estado *on*, mientras que el transistor canal p está en estado *off*. En el modelo más sencillo posible para un MOSFET en conmutación, éste se comporta como un interruptor, abierto o cerrado según esté *on* u *off* respectivamente¹. Por tanto, en este caso en la salida tendremos aproximadamente 0 V (figura XI.5), el valor de referencia, y en consecuencia el estado de salida será bajo (un 0 lógico).

En cambio, si la entrada está en bajo (aplicamos por ejemplo un voltaje de 0 V), el MOSFET canal n está en estado *off* (interruptor abierto) y el MOSFET canal p en estado *on* (interruptor cerrado). El camino desde tierra a la salida estará bloqueado, pero existirá un camino desde la fuente V_{DD} , que será el valor que tendremos en la salida (alto, un 1 lógico). Por tanto, como ya hemos visto, el circuito efectúa la función NOT, es decir, cambia el valor del bit a la entrada (de 0 a 1 y de 1 a 0).

¹Sabemos ya que, en el modelo en conmutación, en el estado *on* el transistor podría sustituirse por una resistencia equivalente R_{on} de valor pequeño, pero por simplicidad supondremos que se comporta de forma ideal, con resistencia nula, en dicho estado.

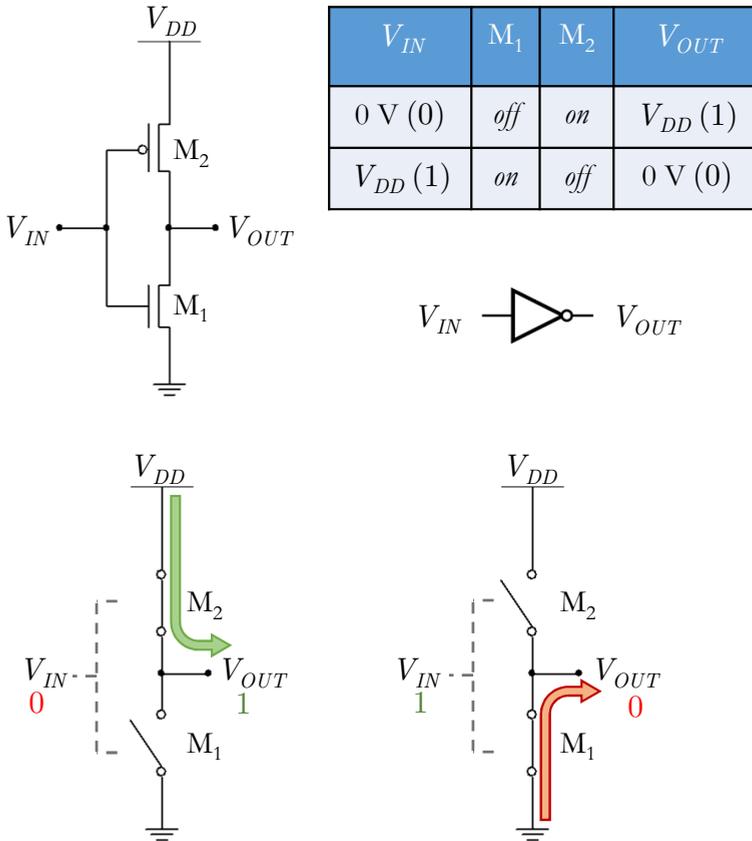


Figura XI.5: Implementación de la función NOT en tecnología CMOS. Se incluye el símbolo de la puerta lógica, los estados de los transistores (*on/off*) para cada valor de la entrada, así como la situación para los dos estados del circuito.

Puerta NAND

El circuito correspondiente a una puerta NAND realizada con tecnología CMOS se muestra en la figura XI.6. Por simplicidadm a partir de ahora vamos a referirnos a los voltajes bajos como L (*low*) y a los altos como H (*high*), asumiendo que trabajaremos con lógica definida positiva (el estado bajo corresponde al 0 y el alto al 1).

Como vemos, la salida del circuito está en bajo únicamente cuando las dos entradas están en alto. En todos los demás casos la salida toma un valor alto (1).

Si las dos entradas están en bajo, los dos transistores canal n, que están en

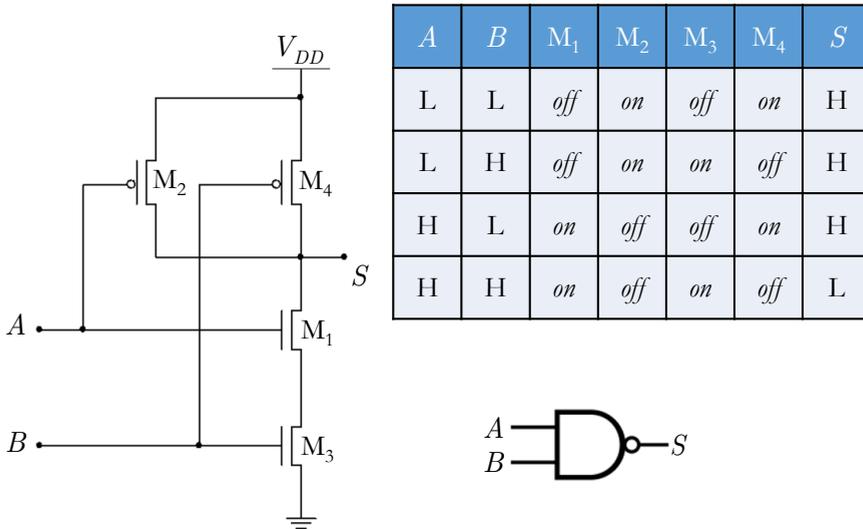


Figura XI.6: Esquema de una puerta NAND de dos entradas en tecnología CMOS, junto con la tabla de valores y estados de los transistores, y su símbolo.

serie, se encuentran en estado *off*, es decir, se comportan como un interruptor abierto (figura XI.7). Los dos transistores canal p, en cambio, están en estado *on*, y de este modo llevan el voltaje V_{DD} hasta la salida, que estará por tanto en alto.

Si la entrada A está en alto y B en bajo, en la parte de los MOSFET canal n el camino hasta tierra estará bloqueado por el transistor que tiene entrada B. Por la parte superior, el transistor canal p correspondiente permitirá que a la salida lleguen V_{DD} voltios, por lo que de nuevo la salida estará en alto. Algo análogo ocurrirá si la entrada A está en bajo y B en alto.

Finalmente, si las dos entradas están en alto, llevan a los transistores canal n al estado *on* simultáneamente, por lo que hay un camino eléctrico desde tierra hasta la salida, que estará en bajo. Los transistores canal p estarán en estado *off* y bloquearán el camino hasta la fuente V_{DD} , para evitar cortocircuitos entre la alimentación y tierra. En definitiva, el circuito realizará la operación NAND.

En caso de que queramos diseñar una puerta NAND con más entradas el circuito es fácilmente escalable, simplemente es necesario añadir un transistor canal n en serie en la parte inferior y un transistor canal p en paralelo en la superior, ambos con sus puertas conectadas a la nueva entrada. Por tanto, bastan un par de transistores por cada entrada adicional del circuito (figura XI.8).

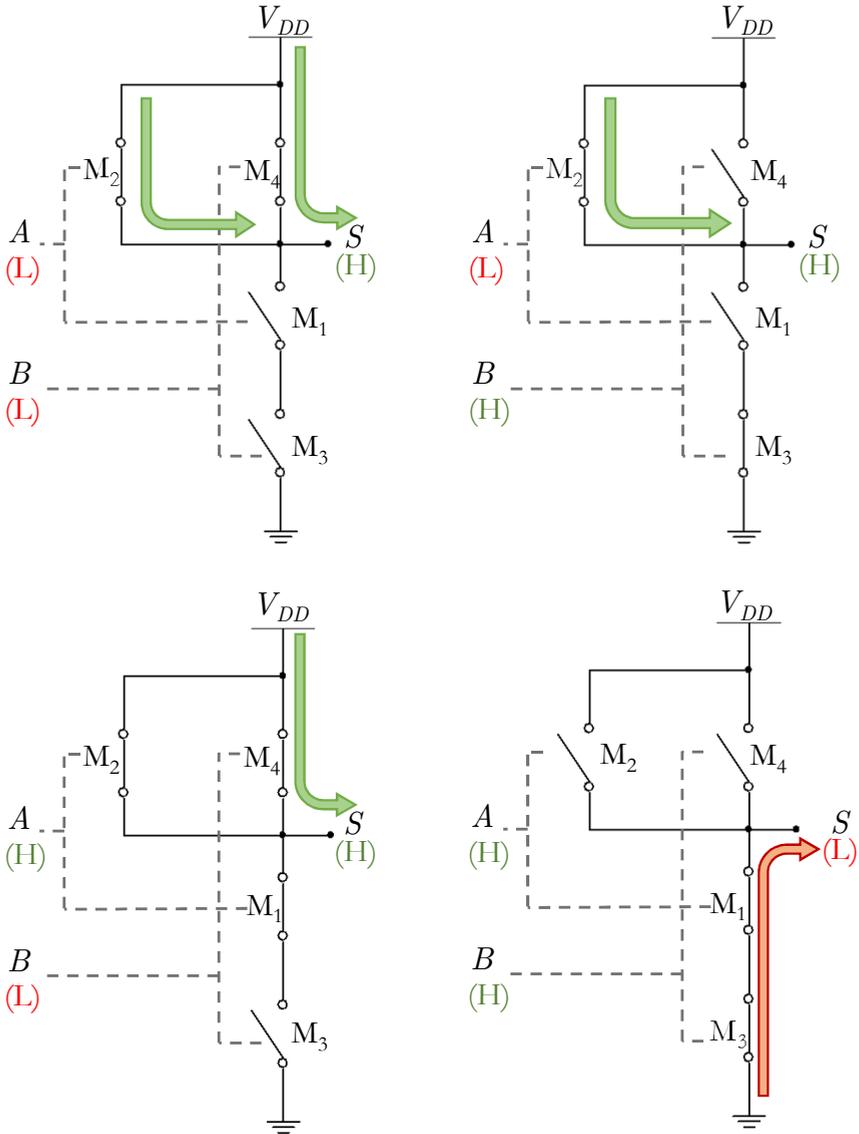


Figura XI.7: Funcionamiento de una puerta NAND de dos entradas para las cuatro posibles combinaciones.

Puerta AND

A partir de los dos circuitos anteriores, es sencillo realizar el diseño de una puerta AND. Basta con acoplar la salida de una puerta NAND a un inversor, y

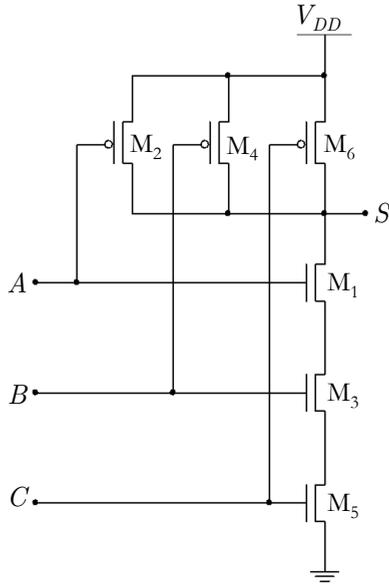


Figura XI.8: Puerta NAND de tres entradas en tecnología CMOS.

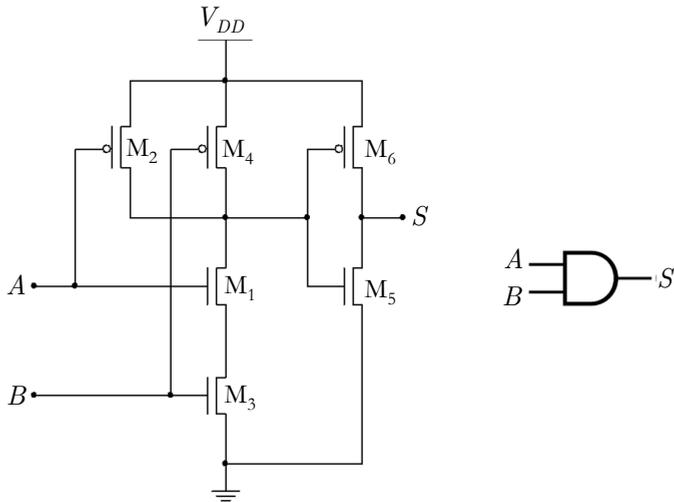


Figura XI.9: Puerta AND de dos entradas en tecnología CMOS, y símbolo circuital (derecha). Los transistores M_1 , M_2 , M_3 y M_4 corresponden a una puerta NAND, cuya salida se acopla a la entrada del inversor formado por M_5 y M_6 .

obtendremos la función buscada, que proporcionará una salida en alto (1) únicamente cuando todas las entradas están en alto (figura XI.9).

Puerta NOR

Una puerta NOR proporciona una salida en alto únicamente cuando todas las entradas están en bajo. Es posible realizar una puerta NOR de dos entradas en tecnología CMOS empleando cuatro transistores (dos canal n y dos canal p), conectados como se muestra en la figura XI.10.

Cuando las dos entradas están en bajo, L (es decir, 0), los dos MOSFET canal n están en estado *off*, mientras que los dos canal p están en estado *on*, por lo que existe un camino eléctrico entre la salida y la alimentación V_{DD} , proporcionando un valor alto H (1).

Sin embargo, si alguna de las entradas (o las dos) toman un valor alto, se forma un camino eléctrico entre la salida y tierra a través del transistor (o transistores) canal n activo, mientras que el camino hasta la alimentación V_{DD} está bloqueado por el transistor (o transistores) canal p con su entrada en alto. La salida por tanto

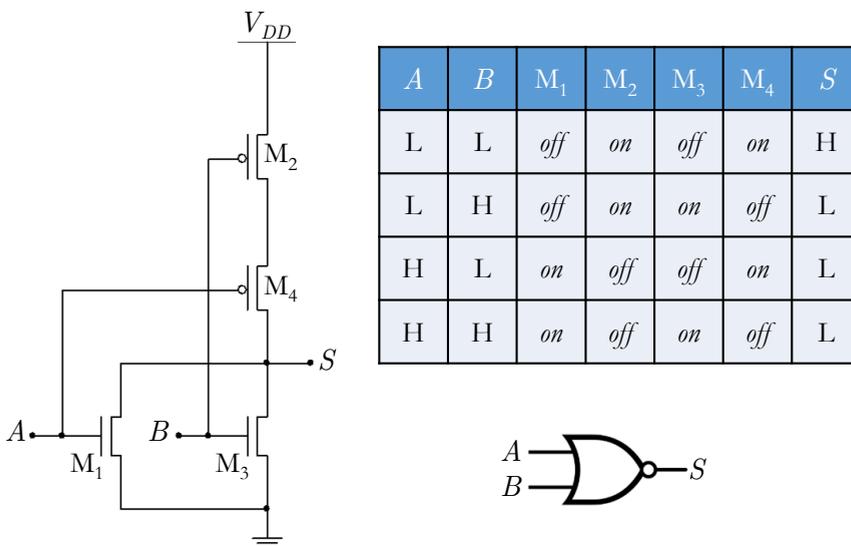


Figura XI.10: Puerta NOR de dos entradas en tecnología CMOS, junto con los estados de los transistores para cada una de las combinaciones de las entradas, y símbolo circuital.

estará en bajo, L (0). Queda el lector emplazado a realizar un análisis detallado de cada situación como el presentado para la puerta NAND en la figura XI.7.

Al igual que en el caso de la puerta NAND, el diseño es fácilmente escalable a más entradas. En este caso, si queremos añadir una entrada adicional necesitamos dos transistores, uno canal n conectado en paralelo a sus homólogos y otro canal p conectado en serie en la parte superior del circuito.

Puerta OR

La realización de una puerta OR es sencilla tomando una puerta NOR y acoplando su salida a un inversor, como se indica en la figura XI.11. De este modo se realiza la función deseada, es decir, un circuito que da salida en alto (1) cuando alguna o varias de sus entradas está en alto, o si se prefiere, que únicamente proporciona una salida en bajo (0) cuando todas las entradas toman valor bajo.

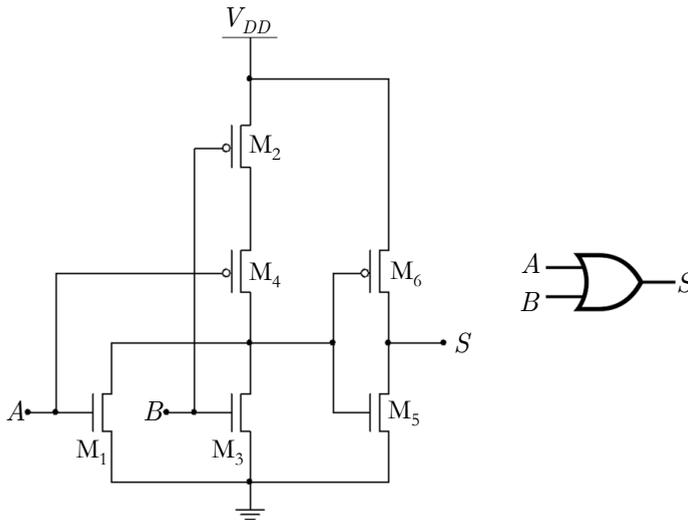


Figura XI.11: Puerta OR de dos entradas en tecnología CMOS, y símbolo circuital (derecha). Los transistores M_1 , M_2 , M_3 y M_4 corresponden a una puerta NOR, cuya salida se acopla a la entrada del inversor formado por M_5 y M_6 .

Otras funciones lógicas, como XOR o XNOR, pueden construirse también combinando las puertas anteriores. Circuitos de mayor complejidad, como biestable, multiplexores, codificadores, etc. pueden también realizarse empleando

puertas lógicas sencillas, sin más que combinándolas adecuadamente. En particular, es interesante saber que puede realizarse cualquier función lógica empleando únicamente puertas NAND (o puertas NOR), conforme a la propiedad denominada **completitud funcional**.

3 NIVELES LÓGICOS Y MÁRGENES DE RUIDO

Como hemos indicado con anterioridad, los valores matemáticos 0 y 1 corresponden físicamente a rangos de voltaje², que denominamos **niveles lógicos**. Así, definiremos un nivel bajo o *low* (L) y un nivel alto o *high* (H), conforme a la notación empleada cuando hemos descrito las puertas lógicas CMOS (figura XI.12). Entre ambos niveles debe existir una zona de separación suficientemente amplia, en la cual la entrada y la salida no tendrán valores definidos. Serán zonas prohibidas, donde no debe darse la operación de los circuitos.

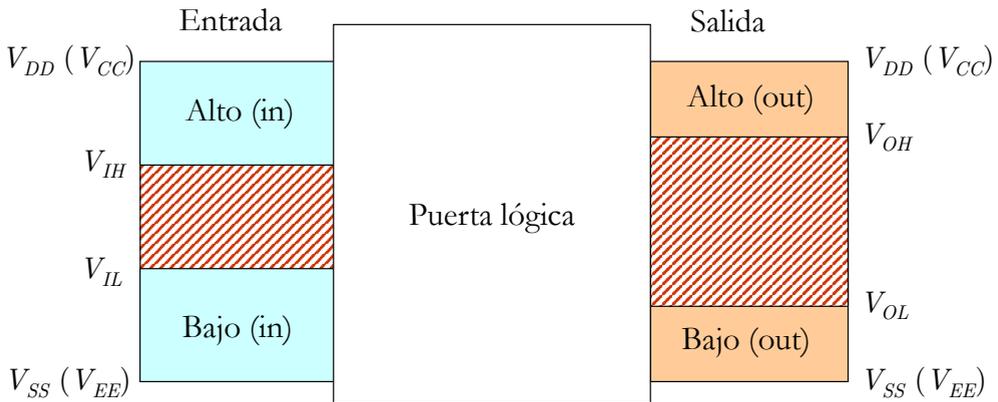


Figura XI.12: Niveles lógicos.

Como vemos, los niveles se definen por separado, siendo más amplios en la entrada de la puerta lógica que en su salida. Como ya hemos señalado, usualmente se emplea la lógica definida positiva, en la que se asigna el 1 al rango alto y el 0 al

²A la vista de los ejemplos anteriores, cabría preguntarse por qué deben considerarse rangos de voltaje y no estrictamente los valores de V_{DD} y tierra. Ello se debe a que en el análisis que hemos efectuado con anterioridad consideramos una imagen completamente ideal de los transistores, comportándose como interruptores perfectos, que en la realidad no se da. Además, el fenómeno del ruido electrónico, que estudiaremos en este apartado, puede afectar a los valores de voltaje. Por ello son necesarios los rangos de voltaje para definir los valores lógicos y trabajar de manera adecuada con los circuitos digitales.

rango bajo. Así, cualquier valor de voltaje en la entrada que esté entre V_{DD} (V_{CC} en familias TTL) y V_{IH} valdrá 1, y cualquier valor entre V_{IL} y V_{SS} (o V_{EE} en familias TTL; en cualquier caso normalmente estará conectado a tierra) valdrá 0. De manera análoga, a la salida del circuito un valor de voltaje entre V_{DD} y V_{OH} representará un 1 y un valor entre V_{SS} y V_{OL} representará un 0.

Que los niveles de entrada y salida sean distintos es importante para minimizar el problema del ruido electrónico, que ya se discutió en el capítulo IX. La relativa **inmunidad al ruido** es una de las características que hacen que en la actualidad los circuitos digitales tengan mayores ventajas que los analógicos. Ello es debido precisamente a la diferencia entre los niveles lógicos de entrada y de salida. Dichas diferencias reciben el nombre de **márgenes de ruido**: de este modo podremos definir un margen de ruido en alto (NMH, de *noise margin high*) y un margen de ruido en bajo (NML, de *noise margin low*).

$$NMH = V_{OH} - V_{IH} \quad (\text{XI.1})$$

$$NML = V_{IL} - V_{OL} \quad (\text{XI.2})$$

La figura XI.13 nos muestra la importancia de los márgenes de ruido. Supongamos que tenemos una puerta lógica que en su salida proporciona un 1 en la parte inferior del rango alto, es decir, un valor de V_{OH} . En la línea de transmisión hasta la siguiente puerta, este voltaje puede sufrir una fluctuación que haga que se reciba en la siguiente puerta con un valor ligeramente inferior a V_{OH} . Si el mínimo del rango de entrada, V_{IH} , fuera igual a V_{OH} , se recibiría un valor en el rango prohibido. Sin embargo, al ser inferior y existir un cierto margen de ruido, siempre que la perturbación no sea excesivamente fuerte se recibirá correctamente como un 1.

Cada familia lógica se caracteriza por trabajar con unos niveles y márgenes de ruido diferentes. En la figura XI.14 se muestran los niveles para varias familias lógicas, la TTL, la CMOS con alimentación de 5 V y la CMOS con alimentación de 3.3 V. Como puede comprobarse, existen diferencias apreciables entre ellas.

Estas diferencias tienen implicaciones importantes desde el punto de vista práctico. Por ejemplo, si pretendemos conectar una puerta que funciona con niveles TTL con una puerta que funciona con niveles CMOS, podemos encontrarlos con problemas, como ilustra la figura XI.15.

Si una puerta TTL produce una salida en alto dentro de su nivel, puede ocurrir que el valor de voltaje sea de hasta 2.7 V. Si ese voltaje se acopla directamente a la

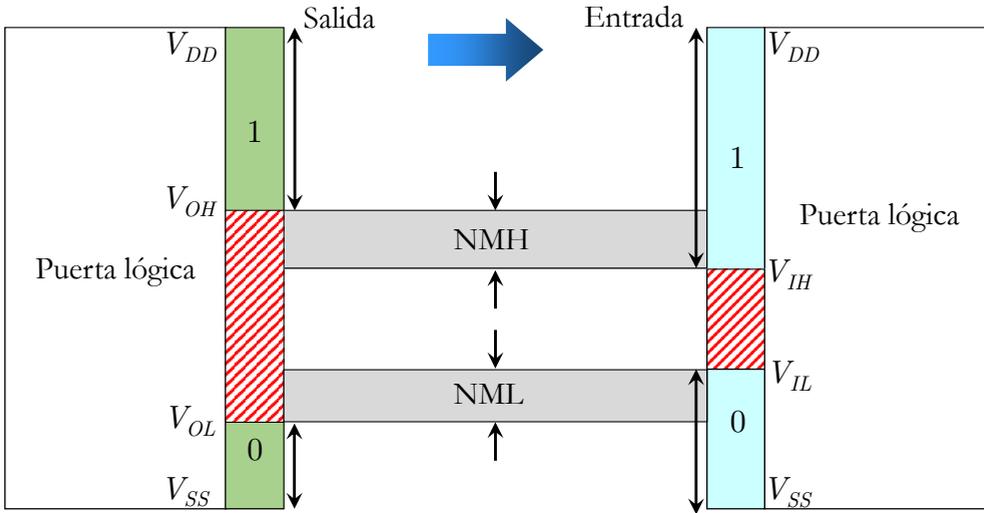


Figura XI.13: Márgenes de ruido.

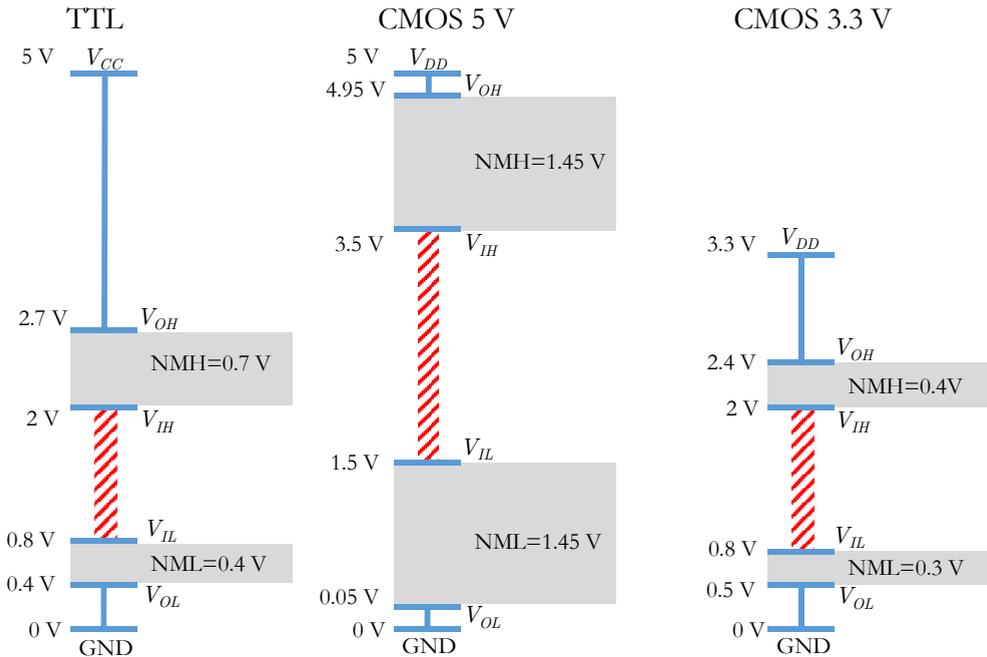


Figura XI.14: Niveles de ruido de diferentes familias lógicas.

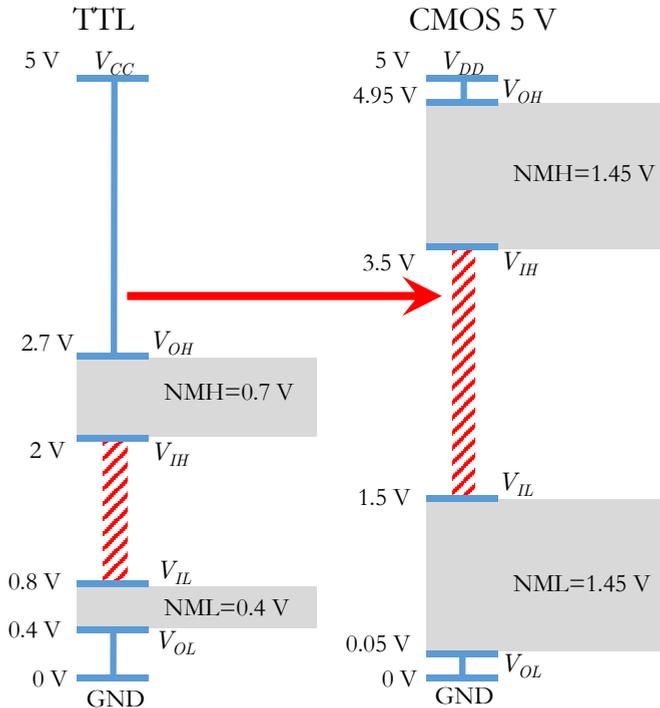


Figura XI.15: Problemas de compatibilidad entre familias lógicas: un valor 1 admisible para la salida de una puerta TTL puede no serlo para la entrada de una CMOS.

entrada de una puerta CMOS, estaría por debajo del mínimo de aceptación de esa familia (3.5 V), por lo que el segundo circuito no sabría interpretar si ha recibido un 1 o un 0, y muy probablemente se produciría un error.

Para evitar estos problemas es necesario emplear circuitos adaptadores de nivel, que garanticen que la transmisión de datos entre puertas de diferentes familias se produce de forma correcta.

4 CARACTERÍSTICAS GENERALES Y CONSUMO DE POTENCIA

Cuando tratamos con puertas lógicas, existen una serie de parámetros relevantes a considerar. El primero es el número de entradas de que disponen. Se denomina **abanico de entrada** o *fan-in* al *número máximo de entradas de una puerta lógica*. Debe tenerse en cuenta que en caso de no utilizar alguna entrada, ésta debe conectarse al voltaje correspondiente al elemento neutro de la operación que

realice la puerta. Por ejemplo, si es una puerta OR o NOR, deberá conectarse a tierra (0), y si es una operación AND o NAND, a la alimentación V_{DD} (1).

El **abanico de salida** o *fan-out* es el *número máximo de entradas de circuitos que se pueden conectar a la salida de una puerta* (siempre dentro de la misma familia).

Los factores que limitan el *fan-out* son diferentes dependiendo de cada familia lógica, su principio de funcionamiento y los transistores que se empleen en su fabricación. Por ejemplo, en el caso de las familias TTL, una característica importante es que pueden entregar corriente a las puertas conectadas a su salida (las puertas de carga). Cuantas más puertas se conecten, más corriente entregará la puerta inicial, y ello producirá una caída de potencial interna mayor, hasta llegar a un punto en el que no puedan mantenerse los niveles lógicos. Por tanto, la corriente es el mayor factor limitante del abanico de salida en un circuito TTL.

En cambio, en las familias CMOS, cuando una puerta alimenta a otras y cambia el valor de la salida (la entrada de las siguientes puertas), físicamente tiene lugar un proceso de creación o eliminación de la capa de inversión de los transistores que constituyen las puertas de carga, que como ya sabemos, podemos asimilar a la carga y descarga de un condensador. Por tanto, existe un proceso dinámico asociado, que requiere de un cierto tiempo para realizarse, asociado a los retardos y los tiempos de respuesta de los transistores. En ese caso, el *fan-out* estará limitado principalmente por la frecuencia de operación (a mayor frecuencia menor abanico de salida).

Otro aspecto especialmente importante a tener en cuenta en relación a las familias lógicas es el consumo de potencia. En las familias TTL existe una corriente, que denominaremos I_{CCH} para el estado alto e I_{CCL} para el estado en bajo. Con el fin de estimar el consumo de potencia, asumiendo que estamos la mitad de tiempo en un estado y la otra mitad en el otro, la corriente promedio será:

$$I_{CC} = \frac{I_{CCH} + I_{CCL}}{2} \quad (\text{XI.3})$$

Y en consecuencia el consumo de potencia estático vendrá dado por:

$$P_S = V_{CC} I_{CC} \quad (\text{XI.4})$$

siendo V_{CC} el voltaje de alimentación. Se trata por tanto de un consumo de potencia que no va a variar sustancialmente con la frecuencia de operación, ya que existe una corriente relevante en condiciones estacionarias, que es la principal responsable del consumo.

En cambio, en la familia CMOS la situación es diferente. Existirá por una parte un consumo de corriente estático, como en el caso anterior, pero con la diferencia de que en este caso la corriente estática es muy pequeña³, por lo que la potencia asociada (P_S) es prácticamente despreciable.

Sin embargo, existe un importante consumo de potencia dinámica, asociado al cambio de estado en los transistores. Dicho consumo dependerá de las capacidades internas de la puerta (C_P) y de la carga conectada (C_L), siendo por tanto:

$$P_D = f(C_P + C_L)V_{DD}^2 \quad (\text{XI.5})$$

siendo f la frecuencia de operación. Así, la potencia total será la suma de las partes estática y dinámica, es decir:

$$P_T = P_S + P_D \quad (\text{XI.6})$$

La dependencia cualitativa del consumo de potencia con la frecuencia para una familia TTL y para una familia CMOS se muestra en la figura XI.16.

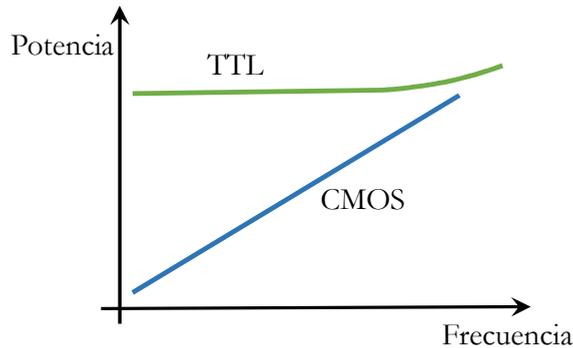


Figura XI.16: Dependencia del consumo de potencia con la frecuencia de operación.

Otro aspecto a tener en cuenta son los **retardos de propagación**. Cuando una señal se propaga a través de un circuito digital, siempre experimenta un retardo temporal. Es decir, un cambio en las condiciones de entrada no se traduce en un cambio inmediato en la salida, sino que debe transcurrir un cierto tiempo. Dicho tiempo está directamente relacionado con el tiempo necesario para realizar

³ Idealmente debería ser nula, pero en los transistores MOSFET en estado *off* siempre existe una corriente de valor prácticamente despreciable, que lleva aparejada un pequeño consumo de potencia.

la conmutación de los transistores que constituyen la puerta lógica, como ya hemos estudiado en el capítulo X. Considerando el tiempo propagación t_{PHL} , que es el tiempo que tarda una puerta en cambiar su estado de salida de alto a bajo, y el tiempo t_{PLH} , que es el tiempo que tarda una puerta en cambiar el estado de salida de bajo a alto, el retardo de propagación t_P será el promedio de ambos:

$$t_P = \frac{t_{PHL} + t_{PLH}}{2} \quad (\text{XI.7})$$

Cuando se trabaja con varias puertas, los retardos de propagación (que, no olvidemos, son debidos al tiempo necesario para producir la conmutación de los transistores) pueden afectar a la coherencia temporal de los resultados de una operación lógica, produciendo los denominados **riesgos lógicos**.

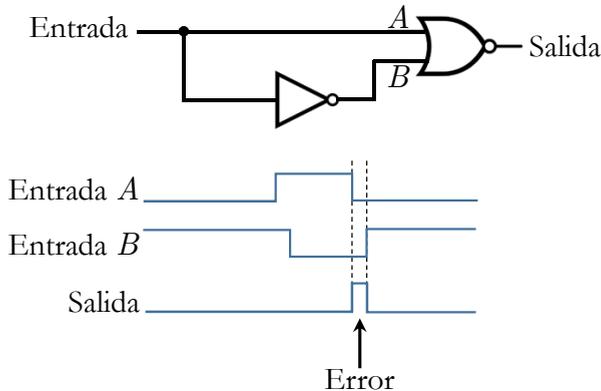


Figura XI.17: Aparición de riesgos lógicos debido a los retardos de propagación.

A modo de ejemplo, en la figura XI.17 se muestra un circuito digital sencillo, consistente en una puerta NOR en cuyas entradas tendremos un bit y su complementario (este último obtenido tras pasar la entrada por una puerta NOT inversora), es decir, o bien 0 y 1 o bien 1 y 0. En todo caso, ello debería producir una salida final igual a 0, conforme a la operación NOR. Ahora bien, la entrada de la puerta NOR que pasa por el inversor sufre un retardo respecto a la entrada sin invertir, por lo que si se producen cambios en ésta, tardarán un tiempo en reflejarse en la señal invertida. Así, puede ocurrir que en un instante dado, las dos entradas estén a 0, lo que daría durante un breve periodo de tiempo un valor 1 en la salida, que es erróneo.

A la hora de trabajar con un circuito digital, por tanto, debe tenerse en cuenta este problema, y trabajar a frecuencias de operación adecuadas, tomando los

valores de los resultados transcurrido un tiempo suficiente para que la señal sea estable. Para conseguir tiempos de propagación cortos, además de factores tecnológicos⁴ pueden emplearse condiciones que favorezcan la conmutación de manera más rápida, lo que implica un mayor consumo de potencia. Es decir, existe un compromiso entre rapidez y consumo energético.

Se suele definir así el llamado **producto velocidad-potencia**, que es una figura de mérito adecuada para poder comparar el rendimiento de diferentes familias lógicas, y viene dado por el *producto del tiempo de propagación promedio y la potencia media disipada*:

$$PVP = t_P P_T \quad (\text{XI.8})$$

Se mide habitualmente en picoJulios, y cuanto menor sea, mejor será el rendimiento.

5 DISEÑO LÓGICO CMOS

Cualquier función lógica puede realizarse combinando adecuadamente algunas puertas básicas. Sin embargo, es posible diseñar circuitos digitales específicos realizando su **diseño lógico**. El diseño de funciones sencillas mediante tecnología CMOS se basa en la consideración de circuitos que utilicen únicamente transistores MOSFET canal y canal p. Los circuitos así diseñados van a tener dos partes, que denominamos **árbol de pull up** y **árbol de pull down** (figura XI.18).

El árbol de *pull up* está constituido por transistores MOSFET canal p, y en condiciones normales se encuentra situado entre la salida y el voltaje de alimentación V_{DD} . Es el responsable de llevar los 1 a la salida, por lo que cuando se dé la combinación de entradas adecuada deberá abrir un camino de conducción entre el voltaje de alimentación y la salida.

El árbol de *pull down* se compone de transistores MOSFET canal n, y en condiciones normales está situado entre la salida y la referencia de tierra. Es, por tanto, la parte responsable de poner la salida a 0 cuando corresponda, según los valores de las entradas: en ese caso deberá abrirse un camino de conducción entre la salida y la tierra.

Un aspecto importante es que ambas secciones del circuito (árboles de *pull up* y *pull down*) deben ser duales, es decir, para una cierta combinación de entradas

⁴En el caso de transistores MOSFET, por ejemplo, la longitud de puerta es un factor clave para determinar el tiempo de propagación: cuanto más pequeño es el transistor, menor será.

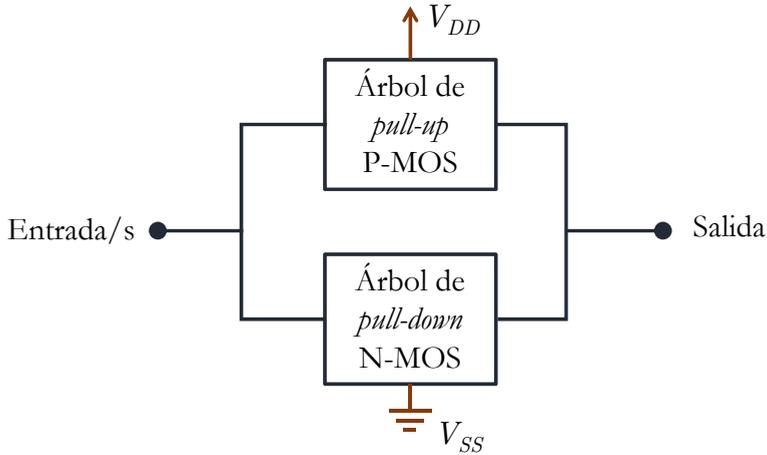


Figura XI.18: Esquema general de un circuito CMOS.

si uno está activo (abriendo un camino entre su voltaje de referencia y la salida) el otro no debe estarlo (los caminos deben estar bloqueados). Ello garantiza que no haya ambigüedades ni cortocircuitos en la salida.

Para implementar una determinada función lógica, optaremos por realizar primero una de las dos secciones del circuito (árbol de *pull down* o árbol de *pull up*) y, una vez diseñada, la complementaria la obtendremos a partir de ella.

Implementación por árbol de *pull down*

En caso de que queramos realizar la implementación de una función lógica por árbol de *pull down*, realizaremos los siguientes pasos:

- Se complementa la función.
- En caso necesario, se simplifica para que no haya operaciones que estén complementadas (es decir, que no haya por ejemplo operaciones del tipo $\overline{A \cdot B}$; en ese caso se reduciría a $\overline{A} + \overline{B}$ conforme a las leyes de Morgan vistas en el capítulo IX).
- Una vez simplificada la función, se procede a la implementación del árbol de *pull down* de la siguiente forma: **los términos producto (función AND, \cdot) se realizan como transistores canal n en serie, y los términos suma (función OR, $+$) como transistores canal n en paralelo.**

- En caso de que al simplificar la función alguna variable haya quedado complementada, ello significa que debe incluirse un inversor antes de conectarla al circuito.
- Una vez implementado el árbol de *pull down*, se procede a implementar el árbol de *pull up*. Para ello se realiza el circuito dual: **en vez de transistores canal n, se emplean transistores canal p, y los elementos que estén conectados en serie en el *pull down* se conectan en paralelo en el *pull up*, y viceversa.** La entrada de cada transistor en el árbol de *pull up* será la misma que la del transistor correspondiente en el árbol de *pull down*.

Una vez realizados estos pasos, ya tendremos la implementación del circuito CMOS.

Implementación por árbol de *pull up*

Si la implementación del circuito la queremos hacer comenzando por el árbol de *pull up*, los pasos son los siguientes:

- En caso necesario, se simplifica la función para que no haya operaciones que estén complementadas (es decir, que no haya por ejemplo operaciones del tipo $\overline{A \cdot B}$; en ese caso se reduciría a $\overline{A} + \overline{B}$ de acuerdo con las leyes de Morgan).
- Una vez simplificada la función, se procede a implementar el árbol de *pull up* de la siguiente forma: **los términos producto (función AND, \cdot) se realizan como transistores canal p en serie, y los términos suma (función OR, $+$) como transistores canal p en paralelo.**
- En caso de que al simplificar la función alguna variable haya quedado sin complementar, ello significa que debe incluirse un inversor. En cambio, si la variable está complementada no es necesario hacer nada (se pone simplemente como entrada sin complementar).
- Una vez implementado el árbol de *pull up*, se procede a implementar el árbol de *pull down*. Para ello, al igual que en el caso anterior, se realiza el circuito dual: **en vez de transistores canal p, se emplean transistores canal n, y los elementos que estén conectados en serie en el *pull up* se conectan en paralelo en el *pull down*, y viceversa.** La entrada de cada

transistor en el árbol de *pull down* será la misma que la del transistor correspondiente en el árbol de *pull up*. El resultado final debería ser el mismo que por el procedimiento anterior.

Para comprender mejor el procedimiento, vamos a realizar el diseño de una función lógica con cuatro variables, A , B , C y D . Sea la función Z :

$$Z = \bar{A} \cdot B + \bar{C} \cdot \bar{D} \quad (\text{XI.9})$$

Vamos a realizar la función por árbol de *pull down*⁵. En primer lugar, debemos complementar la función Z y simplificarla hasta que no haya operaciones complementadas, aplicando en caso necesario las leyes de Morgan:

$$\bar{Z} = \overline{\bar{A} \cdot B + \bar{C} \cdot \bar{D}} = \overline{(\bar{A} \cdot B)} \cdot \overline{(\bar{C} \cdot \bar{D})} = (A + \bar{B}) \cdot (C + D) \quad (\text{XI.10})$$

Una vez simplificada la función, comenzaremos la implementación propiamente dicha. En primer lugar, implementamos las operaciones de mayor prioridad⁶, en nuestro caso, la operación $(A + \bar{B})$, por un lado, y $(C + D)$ por el otro. Al tratarse de operaciones OR, deben colocarse los transistores (canal n, por ser el árbol de *pull down*) en paralelo, tal y como se indica en la figura XI.19.

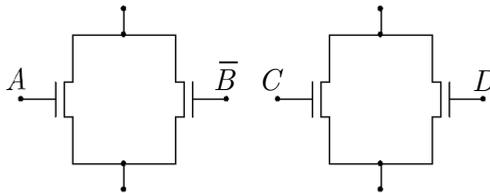


Figura XI.19: Implementación en árbol de *pull down* de las operaciones $(A + \bar{B})$ (izquierda) y $(C + D)$ (derecha).

A continuación, debemos colocar estas dos agrupaciones en serie, de modo que realicen la operación AND entre sí, tal y como indicaba la función complementada. El resultado se muestra en la figura XI.20.

⁵La realización por árbol de *pull up* debería dar el mismo resultado final.

⁶Para ello seguiremos los mismos criterios que con las operaciones aritméticas, es decir, primero se realizan las operaciones entre paréntesis, y en ausencia de ellos, primero las multiplicaciones -operación AND en este caso- y en segundo lugar las sumas -operación OR en las funciones lógicas-

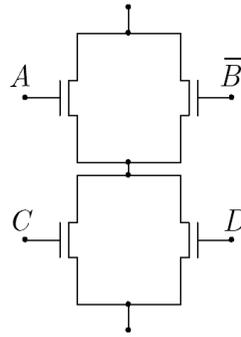


Figura XI.20: Realización de la operación $(A + \overline{B}) \cdot (C + D)$ para completar el árbol de *pull down*.

Una vez completado el árbol de *pull down*, debemos proceder a realizar el árbol de *pull up*, teniendo en cuenta que este último se obtiene como el dual del primero. Es decir, las agrupaciones de transistores que estaban en paralelo pasarán a estar en serie, y viceversa. Además, los transistores deben ser de tipo canal p. El resultado para el ejemplo considerado se muestra en la figura XI.21.

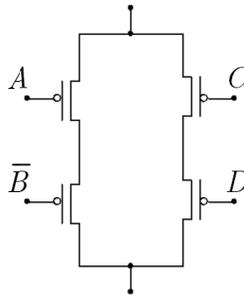


Figura XI.21: Realización del árbol de *pull up*.

El último paso consiste en unir los dos árboles, colocando el árbol de *pull up* en la parte superior y el de *pull down* en la parte inferior. En el punto de conexión entre ambos tendremos la función buscada, Z . No debemos olvidar la fuente de alimentación V_{DD} en la parte superior (que proporcionará los 1 cuando se active el árbol de *pull up*) y la tierra en la parte inferior (que proporcionará los 0 cuando se active el árbol de *pull down*). Si el diseño está bien realizado, no se puede producir la activación simultánea de los dos árboles. El resultado final se muestra en la figura XI.22. Téngase en cuenta que para evitar confusiones hemos indicado

explícitamente las entradas de todos los transistores, pero que en una implementación real existiría una única entrada para cada variable, que se lleva hasta las puertas de los transistores que corresponda empleando los cables o conexiones que sean necesarios. En el caso de la variable que aparece complementada, \bar{B} , debe incluirse previamente un inversor CMOS, que no se muestra en el diagrama de la figura XI.22.

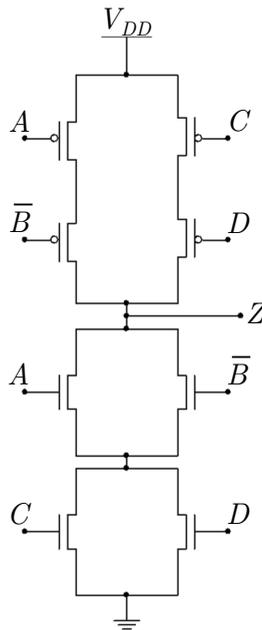


Figura XI.22: Circuito CMOS que implementa la función Z buscada.

En definitiva, empleando tecnología CMOS es posible realizar funciones lógicas que funcionen de manera altamente eficiente: combinando las diferentes operaciones se pueden realizar circuitos más complejos, que acaben conformando la unidad aritmético lógica (ALU) de una unidad central de proceso (CPU).

A modo de ejemplo de la realización de una función aritmética, en la figura XI.23 mostramos un circuito, realizado en tecnología CMOS, que implementa un semisumador de dos bits (realiza la suma de dos dígitos de un bit, proporcionando un acarreo en caso necesario). Al ser las entradas 0 o 1, el resultado de la suma podrá ser 0, 1 o 2 en decimal. Por ello, necesitamos representar la salida con dos dígitos binarios, que nos den como resultado 00 (0 en decimal), 01 (1

en decimal) o 10 (2 en decimal) según el caso. De este modo, tendremos dos salidas, una denominada *acarreo*, que denotaremos con la letra C , y otra *suma*, que denotaremos con S . La primera se puede implementar con una simple función AND, mientras que la segunda se puede implementar con la operación XOR, operaciones que ya estudiamos en el capítulo IX. En la figura XI.23 mostramos el circuito implementado con tecnología CMOS, así como el esquema empleando puertas lógicas, la tabla de posibles valores del circuito y las funciones lógicas implementadas. Son necesarios 18 transistores (9 canal n y 9 canal p) para poder realizar esta operación relativamente sencilla. Ello nos indica que para hacer las diferentes operaciones que realiza un ordenador, un teléfono inteligente, una videoconsola o cualquier otro aparato electrónico complejo, será necesario un circuito con una cantidad elevadísima de transistores.

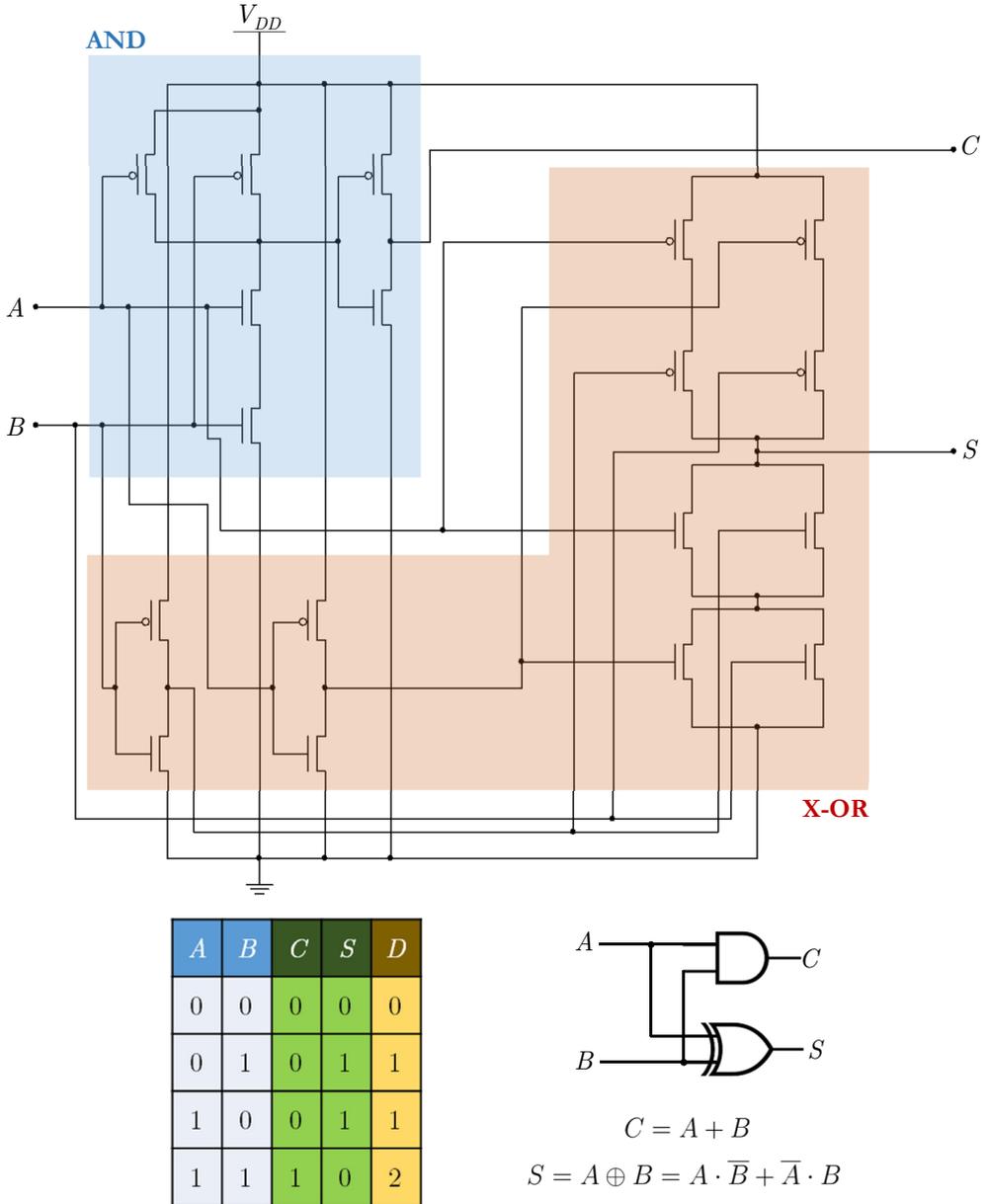


Figura XI.23: Esquema circuital de un semisumador de dos bits, junto con su representación funcional con puertas lógicas, funciones lógicas implementadas (acarreo, C , y suma, S) y tabla de valores, donde se incluye el valor decimal (D) del resultado.

XII.

TECNOLOGÍA DE
FABRICACIÓN DE
CIRCUITOS
INTEGRADOS

En los primeros años tras la invención del transistor en 1947, los circuitos se fabricaban empleando componentes discretos. Rápidamente se observó que esta forma de realizar los circuitos suponía graves problemas para el desarrollo de la tecnología de semiconductores. Una de las grandes ventajas de los transistores respecto a los tubos de vacío era su tamaño, mucho más reducido. Sin embargo, ello suponía también un inconveniente: cuando se quisieran realizar circuitos que contuviesen cientos o miles de transistores, la interconexión de todos los componentes iba a resultar muy compleja de realizar para ser efectuada por un operario, debido al elevado número de cables y al tamaño tan reducido de los transistores. Las posibilidades del transistor como reemplazo a los tubos de vacío se veían severamente afectadas por este problema, que fue conocido en su época como la **tiranía de los números**: la realización de circuitos complejos estaba limitada por el número de conexiones necesarias y el pequeño tamaño de los transistores.

La solución al problema llegó a finales de la década de los años 50 del siglo XX, con la invención del **circuito integrado**, es decir, un circuito en el que *todos sus componentes (transistores, resistencias, condensadores, diodos, etc.) y sus interconexiones están fabricados en un mismo trozo de semiconductor*. De manera independiente, Jack Kilby¹ y Robert Noyce², entre septiembre de 1958 y marzo de 1959, desarrollaron los principales métodos necesarios para fabricar circuitos en una única pieza de semiconductor.

La invención del circuito integrado hizo posible el aprovechamiento de las

¹Jack S. Kilby (1923-2005), ingeniero electrónico estadounidense. Recibió el premio Nobel de física en 2000 por la invención del circuito integrado.

²Robert N. Noyce (1927-1990), físico estadounidense. Cofundador de Intel y protagonista fundamental en el desarrollo de Silicon Valley, no pudo compartir con Kilby el premio Nobel de física debido a su prematuro fallecimiento a los 62 años.

enormes posibilidades de los transistores y el excepcional desarrollo de la era digital que llega hasta nuestros días, basándose en el **microchip** o simplemente **chip**³. A medida que se perfeccionaban los métodos de fabricación, iba aumentando el número de transistores que podían ser integrados en un chip. Existen diferentes **escalas de integración** dependiendo de la complejidad del chip:

- **SSI:** *Small Scale of Integration*. Son chips que contienen menos de 12 puertas lógicas en su interior, por ejemplo los circuitos que implementan puertas lógicas básicas y biestables (*flip-flops*).
- **MSI:** *Medium Scale of Integration*. Chips que contienen entre 12 y 100 puertas lógicas. Por ejemplo, codificadores, registros, contadores, multiplexores, etc.
- **LSI:** *Large Scale of Integration*. Chips que contienen entre 100 y 10000 puertas lógicas. Microprocesadores de 8 y 16 bits, por ejemplo.
- **VLSI:** *Very Large Scale of Integration*. Chips con entre 10000 y 100000 puertas lógicas. Microprocesadores de 32 bits, microcontroladores, sistemas de adquisición de datos, etc.
- **ULSI:** *Ultra Large Scale of Integration*. Chips con más de 100000 puertas lógicas, como son los microprocesadores actuales.

La complejidad de los circuitos digitales integrados ha crecido de manera exponencial. Ello ha sido posible gracias al desarrollo sin precedentes de las tecnologías de fabricación, que presentaremos brevemente en este capítulo.

1 OBLEAS DE SEMICONDUCTOR

El primer paso en la fabricación de un circuito integrado es la obtención del material de partida, el semiconductor. En la actualidad la práctica totalidad de circuitos digitales se realiza utilizando silicio, el semiconductor de referencia, por lo que nos centraremos principalmente en este caso.

En primer lugar, es necesario obtener silicio puro, con un nivel de pureza desconocido en otros ámbitos tecnológicos. En concreto, el denominado **silicio de grado electrónico** es silicio con una pureza de al menos el 99,9999999%. Ello

³Nombre que recibe un circuito integrado debidamente encapsulado para poder ser utilizado en un aparato electrónico.



Figura XII.1: Un trozo de silicio puro.

implica que sólo se permite un átomo de un elemento no deseado entre mil millones de átomos de silicio: un nivel de pureza absolutamente extraordinario. Esto se consigue en diferentes etapas y mediante procesos químicos⁴, cuyo resultado final es silicio altamente puro pero sin una estructura cristalina uniforme.

El segundo requisito es que el material esté cristalizado y tenga una orientación uniforme en todo su interior. A partir del silicio purificado, se sigue un proceso de cristalización denominado **método de Czochralski**⁵ (figura XII.2). El procedimiento consiste en disponer el silicio puro en forma líquida (fundido) en un crisol. Seguidamente, se toma un pequeño trozo de silicio sólido que esté adecuadamente cristalizado, es decir, cuyos átomos constituyentes se encuentren formando una red regular y periódica a lo largo de todo el material. Este trozo de semiconductor recibe el nombre de **semilla**. La semilla se pone en contacto con la superficie líquida y se comienza a tirar de ella, rotando de forma opuesta la semilla y el crisol. El resultado es que al tirar de la semilla se forma progresivamente una barra cilíndrica, denominada **lingote** (figura XII.3), de silicio puro y monocristalino, conforme a la orientación cristalina de la semilla. Los lingotes de silicio pueden alcanzar tamaños de casi dos metros de largo y pesos de cientos de kilos, con diámetros cercanos al medio metro.

Estos lingotes se cortan en **obleas** (discos extraordinariamente delgados, de espesores inferiores a 1 mm), a partir de las cuales se fabrican microchips y me-

⁴En primer lugar, para extraer el silicio de los óxidos en que se encuentran en la naturaleza, como el cuarzo. En segundo lugar, para obtener silicio lo más puro posible.

⁵Jan Czochralski (1885-1953), químico polaco descubridor de este método de cristalización.

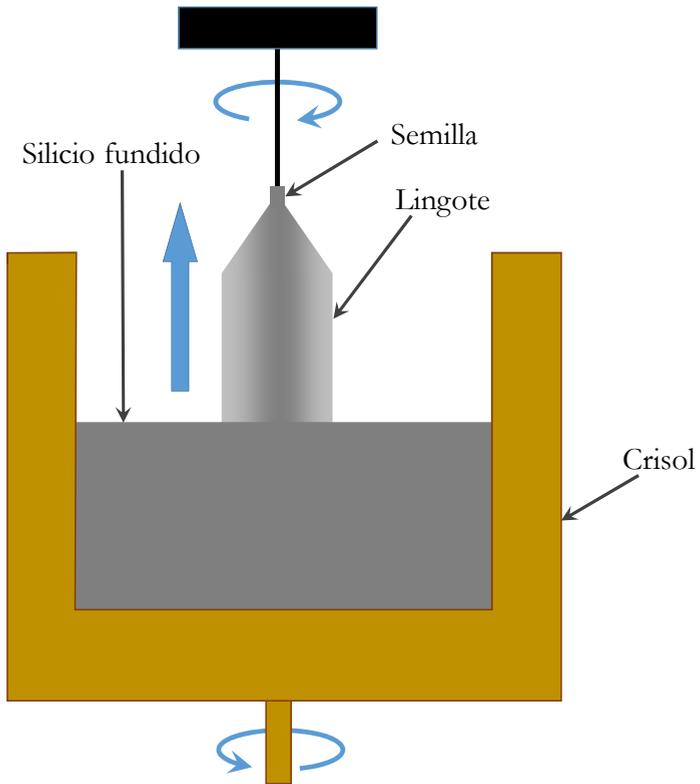


Figura XII.2: Método de Czochralski para la obtención de lingotes monocristalinos.



Figura XII.3: Un lingote de silicio (imagen cortesía de Tomás González Sánchez).

morias RAM, por ejemplo (figura XII.4). El grosor del lingote es un aspecto fundamental, pues determina el diámetro de la oblea. Cuanto mayor sea el diámetro de la oblea, mayor será el número de microchips que podremos fabricar con ella.

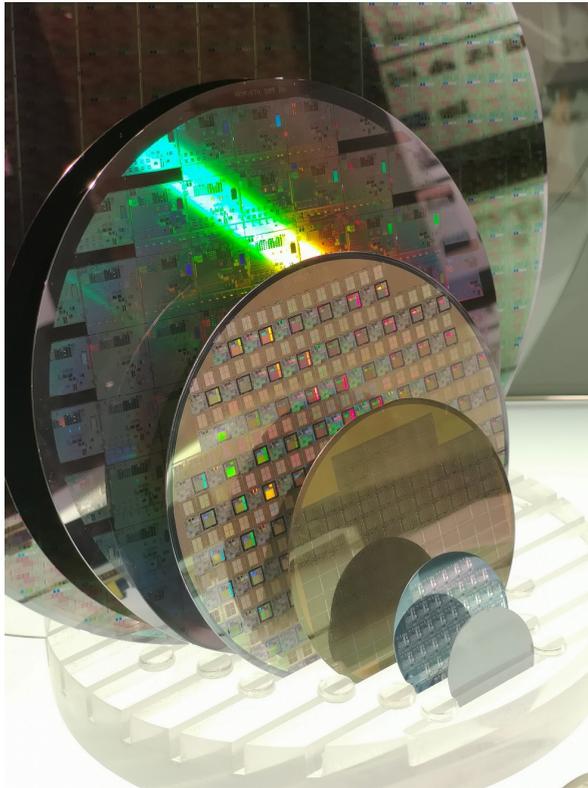


Figura XII.4: Obleas procesadas de diferentes tamaños. Pueden distinguirse en ellas los circuitos integrados correspondientes a los futuros chips (imagen cortesía de Tomás González Sánchez).

2 EL ENTORNO: LA SALA BLANCA

Una vez que disponemos de obleas de material semiconductor puro y monocristalino, estamos en condiciones de proceder a la fabricación de los microchips. Todo este proceso debe realizarse en un entorno de extrema limpieza, en las denominadas **salas limpias o salas blancas**, entornos de trabajo donde las

condiciones ambientales, tanto de temperatura, presión o presencia de partículas contaminantes en el ambiente, están controladas de forma exhaustiva.

Para ello, las salas blancas deben disponer de un control riguroso de los contaminantes en el ambiente, mediante sistemas de filtrado y purificación del aire. En el interior de la sala suele existir una sobrepresión, para el aire salga hacia el exterior y no se introduzcan partículas no deseadas. Los flujos de aire en el interior de la sala también están controlados. Los operarios deben emplear trajes especiales y mascarillas para evitar la contaminación de las obleas. El nivel de limpieza ambiental, en general, es muy superior al de cualquier otro entorno de trabajo. Las salas blancas de las grandes empresas fabricantes de circuitos integrados tienen superficies de miles de m^2 , y requieren fuertes inversiones económicas para su puesta en marcha y funcionamiento.

Clase		Número máximo de partículas/ m^3					
ISO	FS	>0.1 μm	>0.2 μm	>0.3 μm	>0.5 μm	>1 μm	>5 μm
ISO 3	1	1000	237	102	35	8	
ISO 4	10	10000	2370	1020	352	83	
ISO 5	100	100000	23700	10200	3520	832	29
ISO 6	1000	1000000	237000	102000	35200	8320	293
ISO 7	10000				352000	83200	2930

Figura XII.5: Categorías de salas blancas conforme a las normas ISO (14644-1) y FS (209E).

Las salas blancas se pueden clasificar en función de la cantidad de partículas de un cierto tamaño que pueden estar presentes en el ambiente de trabajo. En la figura XII.5 se muestra la clasificación de los tipos de salas blancas conforme a las normas ISO 14644-1 y FS 209E, indicándose el número máximo de partículas permitidas, según tamaños, por metro cúbico. En una sala blanca ISO 3 (o de clase 1 según la norma FS) no pueden existir en el aire más de 1000 partículas de tamaño superior a 0.1 micras por metro cúbico. Lógicamente, a menor cantidad de partículas permitidas mayores son los costes y las restricciones.

3 LITOGRAFÍA

Una vez que se dispone del material de partida (la oblea de semiconductor) y del entorno e instalaciones con los equipos adecuados (la sala blanca) nos encontramos en condiciones de poder fabricar un circuito integrado.

El proceso comienza con el diseño del circuito asistido por ordenador (CAD, *Computer Aided Design*). Inicialmente se realiza el esquema circuital (empleando los símbolos de los componentes de manera similar a como hemos descrito los circuitos en este texto), para posteriormente realizar el **layout** o diseño de cómo se implementará físicamente ese circuito. Este es un aspecto fundamental, ya que en la realización física del circuito debe tenerse en cuenta que el proceso de fabricación se realiza mediante **tecnología planar**, es decir, que todos los procesos de fabricación tienen lugar en la superficie de la oblea. El *layout* (figura XII.6) debe traducir el esquema del circuito al mundo físico, señalando las partes en las que se deben fabricar los transistores, resistencias, diodos, interconexiones, y llegando al detalle de indicar los tamaños de las regiones dopadas, anchura de las conexiones metálicas, etc. En general, toda la fase de diseño del chip es crítica para asegurar tanto las especificaciones del circuito como la viabilidad de su fabricación.

En el proceso de diseño físico se consideran también los diferentes procesos tecnológicos (dopado, oxidación, metalización, etc.) que, de manera sucesiva, se realizarán para conseguir el producto final. Para cada una de las etapas relevantes del proceso de fabricación se procede al diseño de una **máscara de litografía** (figura XII.7). La palabra **litografía** procede del griego *lithos* (piedra) y *graphe* (dibujo), y hace referencia en este caso al proceso de transferir patrones a la superficie del semiconductor. En una etapa dada, por ejemplo, deberá realizarse el dopado de las regiones de semiconductor correspondientes a las fuentes y drenadores de los transistores que se estén fabricando. La máscara litográfica contendrá el dibujo de esas regiones, siendo por ejemplo transparente en las zonas que correspondan a las partes del chip que serán dopadas y opaca en las regiones que no hay que dopar.

El proceso de transferencia del patrón de la máscara a la superficie del semiconductor se realiza de la siguiente manera. En primer lugar, se oxida uniformemente la superficie de la oblea y, seguidamente, se cubre con una sustancia denominada **resina**. La resina es un polímero fotosensible, cuyas propiedades cambian cuando es iluminada con luz de una determinada longitud de onda. A continuación se proyecta la imagen de la máscara sobre la superficie del semicon-

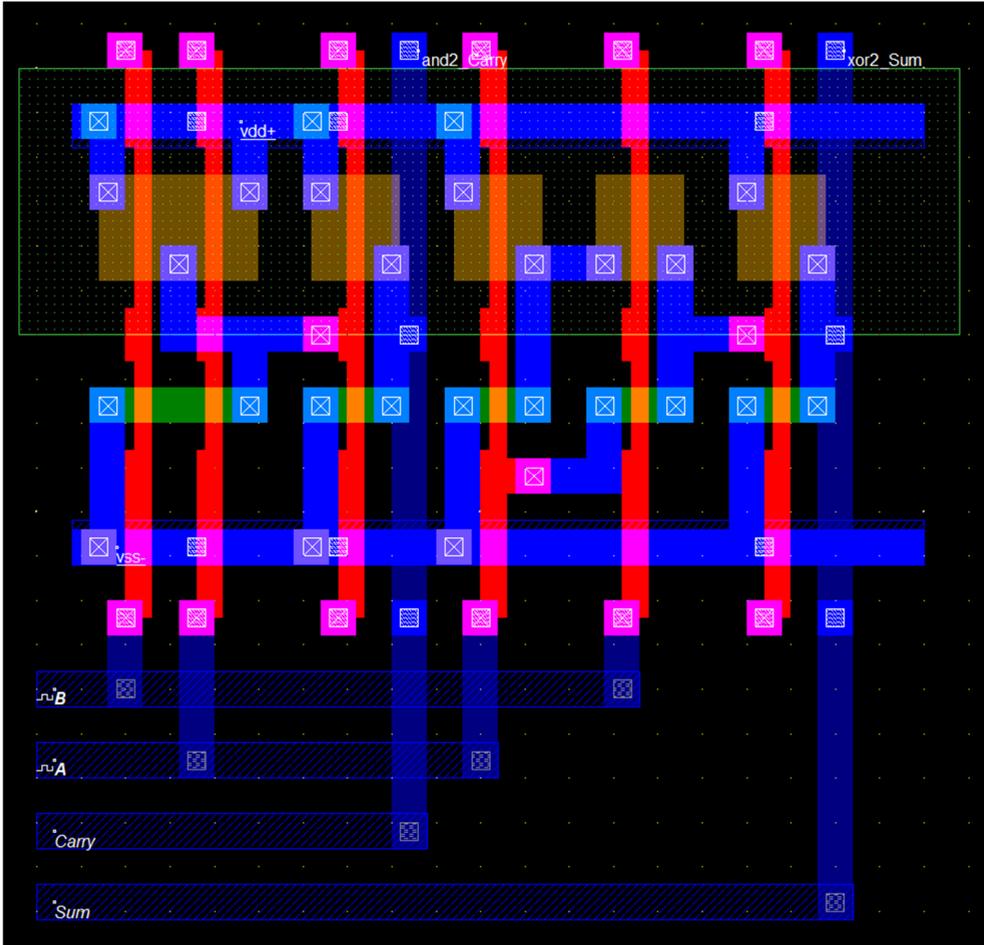


Figura XII.6: Ejemplo de *layout* de un circuito semisumador en tecnología CMOS.

ductor, mediante un sistema óptico adecuado⁶ y empleando luz de longitud de onda que active la resina. De este modo, en la superficie de la oblea la resina quedará sensibilizada en las zonas correspondientes a las partes transparentes (véase la figura XII.8).

La resina de estas zonas podrá ser eliminada mediante un ataque químico, para posteriormente eliminar el óxido que estaba debajo. Así, el semiconductor de

⁶La máscara suele ser 4 o 5 veces más grande que el patrón que se quiere dibujar, por lo que debe proyectarse con un sistema de lentes para que en la superficie de la oblea la imagen proyectada tenga el tamaño adecuado.



Figura XII.7: Máscara de litografía (izquierda) y detalle de la misma (derecha), donde se aprecian los patrones que se van a transferir al semiconductor. Como puede observarse, en la máscara existen patrones correspondientes a múltiples futuros chips (imagen cortesía de Tomás González Sánchez).

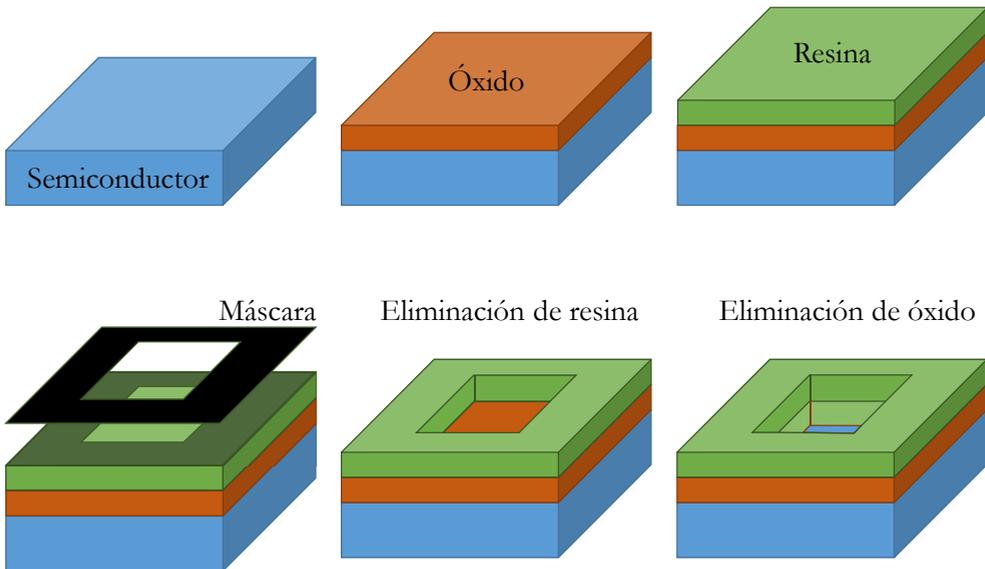


Figura XII.8: Proceso litográfico. Inicialmente se crece una capa de óxido y se deposita una resina fotosensible. Tras iluminar con luz de longitud de onda adecuada a través de una máscara, es posible eliminar las partes de resina y óxido iluminadas, dejando así abiertas zonas específicas en la superficie del semiconductor sobre las que realizar el proceso tecnológico que sea pertinente.

estas zonas queda al descubierto, mientras que el correspondiente a las partes no iluminadas queda tapado por el óxido. De esta manera puede realizarse el proceso tecnológico que corresponda (dopado, grabado, etc.) únicamente en las zonas seleccionadas de la superficie de la oblea. Realizado el proceso, se elimina el material protector sobrante⁷ y se puede comenzar una nueva etapa de litografía, para abordar la siguiente etapa de fabricación. El proceso se repite tantas veces como sea necesario hasta completar la fabricación del chip. Es importante destacar que la máscara se proyecta sucesivamente a lo largo y ancho de toda la superficie de la oblea, de modo que con una misma oblea fabricaremos de manera simultánea numerosos chips⁸.

4 PROCESOS TECNOLÓGICOS

El proceso litográfico permite transferir los patrones de las máscaras en cada etapa del proceso de fabricación. En cada una de dichas etapas se realizará un proceso tecnológico diferente. En esta sección resumimos los principales, junto con las técnicas empleadas para realizarlos.

Dopado

El dopado del semiconductor es un proceso fundamental en la fabricación de un circuito integrado, dado que es imprescindible efectuarlo para realizar los dispositivos semiconductores, como transistores o diodos, que estarán presentes en el circuito.

En el caso del silicio, las impurezas dopantes más comunes son boro, para el dopado tipo p, y fósforo, arsénico o antimonio para el dopado tipo n. Existen dos técnicas principales para dopar un semiconductor.

En el **dopado por difusión** se coloca la oblea (en la que previamente se ha realizado un proceso litográfico, por lo que únicamente estarán expuestas las áreas que serán dopadas) en una atmósfera controlada, que contenga átomos del elemento dopante que queremos incorporar. Aplicando temperaturas elevadas, los átomos de impurezas se difundirán hacia el interior del semiconductor en las

⁷Cada proceso de eliminación de un tipo de material se realiza mediante el ataque químico con una sustancia reactiva a ese material en concreto: la química juega un papel fundamental en todo este proceso. La química del silicio y sus materiales auxiliares en la fabricación de microchips se encuentra extraordinariamente desarrollada.

⁸Por ello el diámetro de la oblea es un factor determinante en el proceso, ya que cuanto más grande sea mayor será el número de chips que se obtendrán de ella.

zonas expuestas y se activarán, es decir, reemplazarán a átomos de silicio en la red cristalina. Regulando el tiempo, la temperatura y la cantidad de átomos presentes en la atmósfera controlada de los equipos en los que se realiza la difusión es posible obtener la concentración de impurezas deseada.

En la **implantación iónica**, mediante la aplicación de campos magnéticos y eléctricos, se disparan iones de impurezas hacia la superficie de la oblea. Estos iones impactarán en la superficie y en las zonas expuestas recorrerán un cierto camino hacia el interior del semiconductor. Controlando la energía de implantación (la energía con que se aceleran los iones) y el número de iones se puede controlar la profundidad y la concentración de impurezas. Tras la implantación se realiza un tratamiento de la oblea a alta temperatura durante un breve periodo de tiempo para reparar los daños en la red cristalina producidos por el bombardeo de impurezas. Este tratamiento permite también que las impurezas se activen.

Oxidación

Una de las ventajas del silicio frente a otros semiconductores es la existencia de un óxido propio, el dióxido de silicio (SiO_2), de gran calidad. La oxidación para obtener capas de SiO_2 es uno de los procesos fundamentales en la fabricación de un chip, y se puede llevar a cabo principalmente de dos formas.

En la **oxidación seca** se coloca la oblea en una atmósfera con oxígeno puro (O_2) a temperatura elevada. El oxígeno reacciona con el silicio formando el óxido, que tendrá una gran calidad, aunque el proceso es lento. Este proceso se puede emplear, por ejemplo, para obtener el óxido de puerta de los transistores MOSFET, que debe poseer la mayor calidad posible.

En la **oxidación húmeda** se coloca la oblea en una atmósfera con vapor de agua (H_2O) a temperatura alta, produciéndose de nuevo una reacción química que acabará oxidando el silicio. Los óxidos así formados son de peor calidad que con la oxidación seca, pero el proceso es más rápido.

Metalización

Otro proceso primordial en la fabricación de un circuito integrado es la deposición de metal para formar las interconexiones en los circuitos. Una técnica empleada en numerosas ocasiones es el **pulverizado** o *sputtering*. Consiste esencialmente en bombardear un trozo de metal con átomos de un gas noble para arrancar átomos del metal, que saltarán disparados del mismo, proyectándolos hacia la superficie de la oblea.

Existen otras técnicas más avanzadas, como la **deposición de capas atómicas** o **ALD** (*Atomic Layer Deposition*) que permiten controlar con mucha mayor precisión la metalización, y en particular, su grosor y uniformidad.

Los metales habitualmente empleados son tungsteno, cobre, plata, etc.

Otros procesos

La fabricación de un circuito integrado involucra también otros procesos además de los mencionados hasta ahora. En general, suele comenzarse con la **epitaxia**, que consiste en el *crecimiento de una capa adicional de semiconductor*. Se realiza mediante técnicas como la deposición química de vapor (reacciones químicas en la superficie del semiconductor) o la epitaxia de haces moleculares⁹ (consistente en disparar hacia la superficie de la oblea átomos del material que se quiere crecer). En esas capas recién creadas de semiconductor es en las que se procede a la fabricación de los dispositivos.

Otro proceso clave es el **grabado**, que consiste en la *eliminación de materiales*. El grabado se realiza en general mediante reacciones químicas, en las que se utilizan compuestos que atacan al material que se quiere eliminar, no reaccionando con los materiales que no se quieren grabar, aunque también puede realizarse por métodos físicos (conceptualmente parecidos al pulverizado).

La **deposición** de óxidos y nitruros también es un proceso habitual a lo largo del ciclo de fabricación del circuito integrado.

Al final de las etapas de fabricación del circuito integrado el resultado es un trozo de material semiconductor, en el que se han fabricado todos los componentes y sus interconexiones, que de cara al exterior presenta únicamente algunas conexiones metálicas, estando el resto de la superficie protegido. El paso final es cortar en la oblea los trozos de semiconductor que corresponderán a cada chip y colocarlos en un encapsulado adecuado para su utilización. En dicho encapsulado será necesario realizar las conexiones entre las metalizaciones finales en la superficie del circuito integrado y los terminales metálicos del encapsulado. Este proceso se realiza de forma automática con máquinas de gran precisión, dada la delgadez de los hilos de conexión y el pequeño tamaño de los circuitos¹⁰.

Finalmente, se completa el encapsulado (usualmente plástico o cerámico), con lo que el chip ya está listo para ser utilizado, generalmente soldándolo en una

⁹La epitaxia de haces moleculares es más costosa y lenta, pero permite controlar con enorme precisión el crecimiento de nuevas capas de átomos sobre la oblea.

¹⁰Por ejemplo, en un microprocesador el tamaño del semiconductor correspondiente al circuito integrado es de poco más de 2 cm².

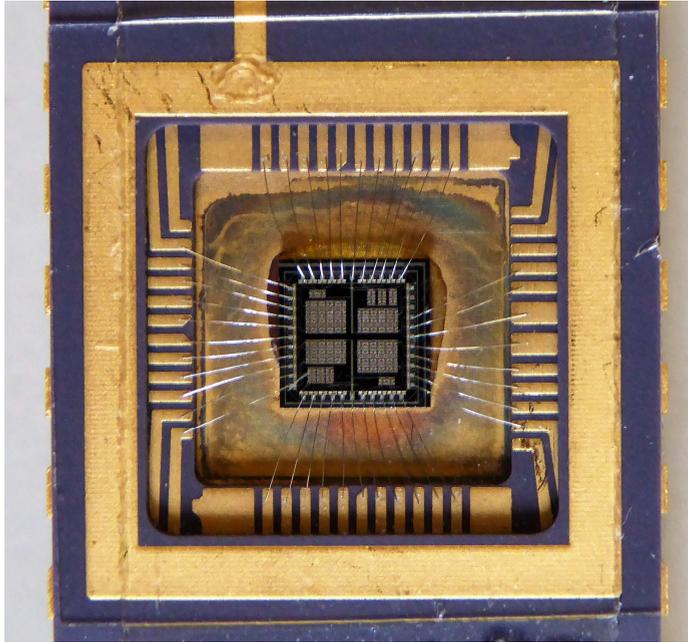


Figura XII.9: Detalle de un circuito integrado: pueden apreciarse los delgados hilos que conforman las interconexiones con el encapsulado de soporte. El último paso que resta es completar el encapsulado para proteger el semiconductor (imagen cortesía de Tomás González Sánchez).

placa de circuito impreso o PCB (*Printed Circuit Board*).

5 EL ESCALADO DEL TRANSISTOR

En 1965, apenas unos años tras la invención del circuito integrado, Gordon Moore¹¹ analizó las tendencias tecnológicas en la fabricación de circuitos integrados hasta ese momento. Moore observó que la mejora en los procesos llevaba a un incremento del número de componentes que se podían integrar en un circuito, y que se podían optimizar los costes de fabricación implementando el número adecuado de componentes en el circuito integrado. El incremento del número de componentes llevaría a mayores funcionalidades de los circuitos integrados,

¹¹Gordon E. Moore (n. 1929), químico estadounidense. Fue junto con Robert Noyce uno de los principales responsables del desarrollo de la tecnología de circuitos integrados y de los microprocesadores.

que acarrearían beneficios económicos a los fabricantes. Invirtiendo en mejoras de los procesos de fabricación, se podría seguir incrementando el número de componentes en un circuito, alcanzándose así un círculo virtuoso de mejora del rendimiento de los dispositivos electrónicos y de las ganancias económicas. Moore dio forma a esta idea en lo que hoy se conoce como la **ley de Moore**, que en su formulación tradicional dice que el número de transistores integrados en un chip se puede duplicar cada dos años aproximadamente.

La ley de Moore no es una ley física, sino una predicción a partir de una observación empírica sobre datos de circuitos fabricados, que la industria semiconductor siguió a rajatabla durante décadas. La principal herramienta para conseguir su cumplimiento fue la reducción progresiva del tamaño de los transistores, en un proceso que se conoce como **escalado**. El escalado implica no sólo la reducción de las dimensiones (largo y ancho) del transistor, sino también la variación de otras magnitudes: incremento de los valores de dopaje, reducción del espesor del óxido de puerta, reducción de los voltajes aplicados, etc.

Para lograr este objetivo, se introdujo el concepto de **nodo tecnológico**, *una forma simplificada de denominar a las diferentes generaciones de circuitos integrados* que permitieron el cumplimiento de la ley de Moore. En las primeras décadas del proceso de escalado, existía una cierta correspondencia entre el tamaño de la puerta (determinado por su longitud, L_G) y la mitad de la separación promedio entre metalizaciones de los transistores en el circuito integrado (llamada *half-pitch*). La longitud de puerta es crítica para determinar la velocidad de los transistores; el *half-pitch*, por su parte, determina en gran medida la densidad de empaquetamiento de los transistores. En aquella época, esa longitud daba nombre al nodo: así, existía el nodo de $5\ \mu\text{m}$, de $2\ \mu\text{m}$, de $1\ \mu\text{m}$, etc. Dicha denominación se ha mantenido hasta nuestros días, llegándose a nodos inferiores a la decena de nanómetro, a pesar de que **en la actualidad este número no corresponde a ninguna dimensión física del transistor**: es simplemente una forma de referirse a las sucesivas generaciones de dispositivos.

La ley de Moore ha permitido un crecimiento exponencial del número de transistores en los microprocesadores (figura XII.10), llevando a estos a funcionalidades no soñadas apenas unas décadas atrás. El primer microprocesador, el 4004 de Intel, contenía aproximadamente 2300 transistores. Los microprocesadores actuales contienen miles de millones, o incluso decenas de miles de millones de transistores en su interior.

El camino para llegar hasta estas elevadas densidades de empaquetamiento de transistores ha sido muy complicado. A medida que se reducen las dimensiones de

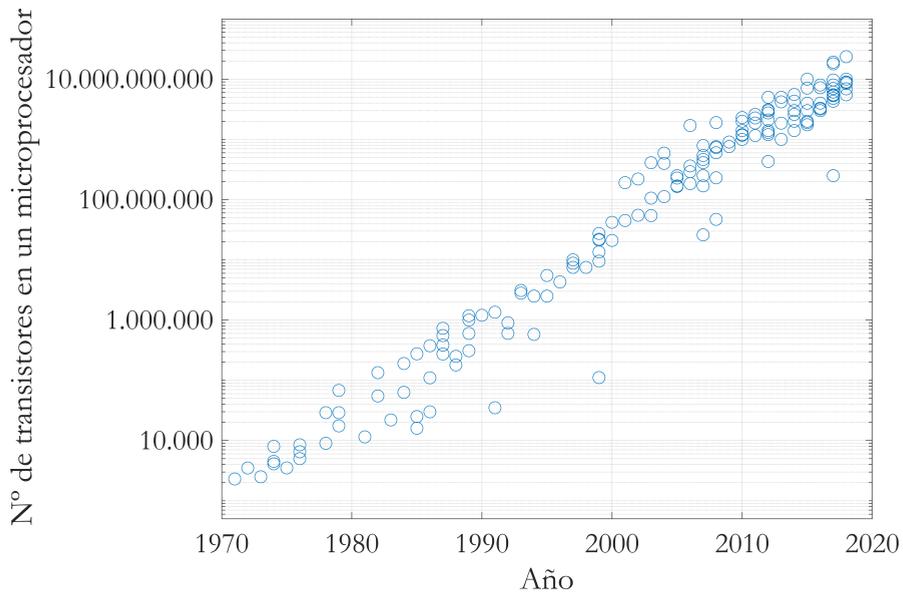


Figura XII.10: Número de transistores en microprocesadores comerciales

los dispositivos aparecen múltiples problemas físicos y retos tecnológicos, entre los que podemos destacar los siguientes:

- **Retos tecnológicos:**

- Disponer de herramientas de fabricación de dimensiones nanométricas, en particular fuentes de luz adecuadas para el proceso litográfico de impresión de circuitos. Cuanto menor sea el tamaño de los transistores, menor debe ser la longitud de onda de la luz empleada en la litografía.
- Problemas de variabilidad: a la hora de fabricar un transistor existe siempre un cierto margen de variación entre la dimensión planteada y la real obtenida (es imposible transferir de forma totalmente fidedigna los patrones de diseño de las máscaras). A medida que se reducen las dimensiones, este problema es más grave y hace que transistores sobre el papel idénticos puedan presentar comportamientos significativamente distintos.

- **Problemas físicos:**

- Efecto túnel: al reducir el grosor del óxido de puerta (requisito necesario en el proceso de escalado) se puede producir una corriente de fugas no deseada a través del mismo.
- Calentamiento: al reducirse el tamaño e incrementar la velocidad de operación, aumenta la disipación de potencia y los dispositivos se calientan en exceso.
- Efectos extremos de canal corto: tamaños de puerta muy pequeños implican una cercanía excesiva entre fuente y drenador, por lo que el control de la corriente por parte del terminal de puerta se complica. Puede existir un salto directo de portadores entre fuente y drenador en condiciones en las que no debería existir corriente.

Respecto al problema relacionado con la litografía, se han empleado sucesivamente fuentes de luz de menor longitud de onda, hasta alcanzar en la actualidad el rango del ultravioleta extremo, es decir, valores de λ ligeramente superiores a los 10 nm. Debemos señalar que el problema no está limitado a encontrar fuentes adecuadas de longitud de onda, sino que también es necesario disponer de resinas sensibles a esas nuevas longitudes de onda, los productos químicos necesarios para disolverlas, etc. La variabilidad, sin embargo, continua siendo uno de los principales problemas de la escala nanométrica, y un factor decisivo a la hora de establecer los límites del escalado.

Para paliar el efecto túnel a través del óxido, se ha reemplazado el óxido de silicio por otros materiales dieléctricos de permitividad más elevada, de modo que aplicando los mismos voltajes en la puerta se pueden emplear óxidos más gruesos, lo que reduce el problema.

La solución al problema del calentamiento al aumentar la velocidad de operación pasó, desde mediados de los años 2000, por no incrementar la frecuencia de reloj de los microprocesadores. Recordemos que el consumo de energía en la tecnología CMOS tiene lugar fundamentalmente en el cambio de estado de las salidas de las puertas, y que este consumo de potencia dinámica es proporcional a la frecuencia de operación. A partir de ese momento, para mejorar el rendimiento de las nuevas generaciones se optó por modificaciones a nivel de arquitectura funcional de la unidad central de proceso, introduciéndose los procesadores multinúcleo.

Los problemas asociados a la pérdida de control de la puerta son, por su parte, la razón de que la arquitectura tradicional del MOSFET, que vimos por ejemplo en la figura VI.3, haya sufrido una serie de modificaciones a lo largo

del tiempo, con especial intensidad a partir de comienzos del siglo XXI (véase la figura XII.11). En primer lugar, se planteó el uso de sustratos de silicio sobre aislante (*Silicon on Insulator* o **SOI**), de modo que el espesor del canal conductor está limitado, con el fin de paliar las fugas directas por el sustrato y mejorar el control de la puerta. Sin embargo, esta solución no fue definitiva: el MOSFET sufrió su modificación más radical en la segunda década del siglo, con la aparición del denominado **FinFET**¹². El canal conductor se encuentra ahora rodeado en tres de sus lados por la puerta, mejorando notablemente el control.

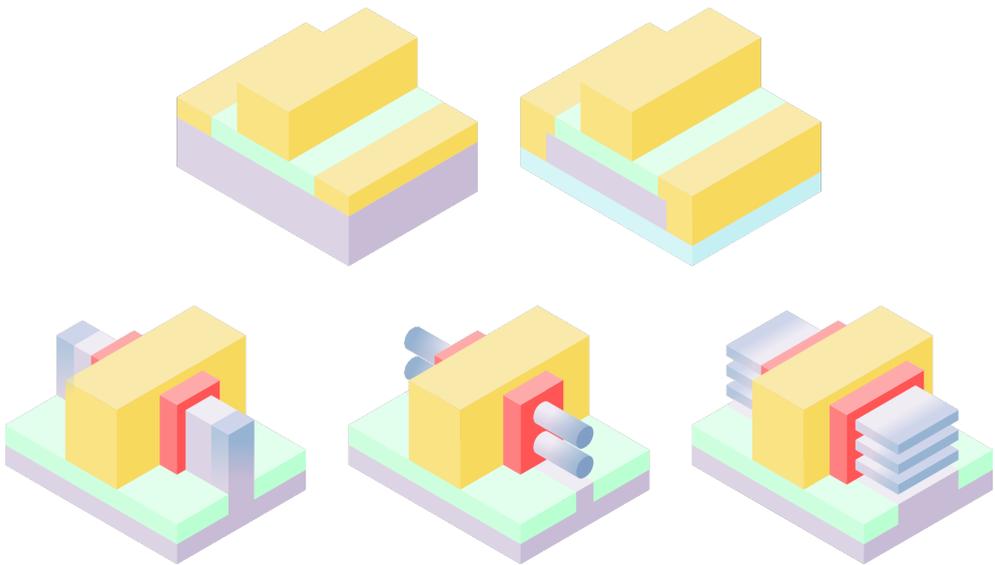


Figura XII.11: Evolución de arquitectura del transistor MOSFET. De izquierda a derecha y de arriba a abajo: MOSFET tradicional, SOI MOSFET, FinFET, MOSFET de nanohilos, MOSFET de nanoláminas (imagen cortesía de José Manuel Iglesias Pérez).

La introducción del FinFET supuso, por una parte, el final del escalado de la longitud de puerta de los transistores (que quedó en torno a 10 nm) para mejorar la densidad jugando con otros factores, como la separación entre aletas o su anchura, permitiendo la extensión de la ley de Moore en la década de 2010. Inevitablemente el FinFET no fue la solución definitiva, y por ello se siguen explorando continuamente nuevas soluciones, como el uso de canales con forma de

¹²La palabra *fin*, aleta en inglés, hace referencia a la forma del canal conductor.

nanoláminas, que permiten un control mayor en los cuatro lados del canal conductor, o de nanohilos, donde la puerta también rodea a todo el semiconductor. En ambos casos se obtiene una mejora notable del control de la puerta.

El crecimiento exponencial del número de transistores indicado por la ley de Moore es insostenible de manera indefinida, por lo que la fabricación de las nuevas generaciones de microprocesadores no puede estar guiada exclusivamente por este patrón. En las próximas décadas la mejora de las prestaciones previsiblemente estará orientada hacia el desarrollo de nuevas arquitecturas funcionales de los microprocesadores, combinaciones de subcircuitos integrados en tres dimensiones, y, en general, nuevas y mejores formas de procesar la información en los circuitos integrados. En ese escenario, el transistor MOSFET evolucionado, con ya más de medio siglo a sus espaldas de desarrollo y perfeccionamiento en la fabricación de circuitos integrados a escala industrial, seguirá jugando un papel esencial en la electrónica de consumo en las décadas venideras, en las que el silicio probablemente continuará siendo el material de referencia.

BIBLIOGRAFÍA

Existen numerosas referencias bibliográficas que pueden permitir ampliar conocimientos sobre los diferentes temas que se han tratado en este libro, así como información complementaria en multitud de páginas web. A continuación se indican algunas referencias seleccionadas, junto con la mención de los capítulos del libro para los que son relevantes. Varios de estos libros han sido consultados en la elaboración del presente texto.

Albella, J. M., Martínez-Duart, J. M., y Agulló-Rueda, *Fundamentos de microelectrónica, nanoelectrónica y fotónica*, editorial Prentice Hall, 2005. Un libro centrado principalmente en la física de los semiconductores y dispositivos, incluye también contenidos interesantes sobre amplificación e información sobre tecnologías de fabricación. Capítulos III, IV, V, VI, VII y XII.

Alexander C. K., y Sadiku, M. N. O., *Fundamentos de circuitos eléctricos*, 6ª edición, editorial McGraw-Hill Interamericana, 2018. La primera parte del libro es interesante para profundizar en los contenidos de los capítulos dedicados a la teoría de circuitos (II) y amplificación (VIII).

Bachiller Mayoral, M., *Fundamentos físicos de la informática*, editorial UNED, 2015. El libro puede ser un buen complemento para los capítulos I y II, así como los dedicados a dispositivos (IV, V, VI y VII) y familias lógicas (XI).

Boylestad, R. L., *Introducción al análisis de circuitos*, 13ª edición, editorial Pearson, 2017. Un texto muy completo sobre teoría de circuitos eléctricos para conocer con mayor detalle los contenidos del capítulo II.

Floyd, T. L., *Fundamentos de sistemas digitales*, 11ª edición, editorial Pearson, 2016. Un texto esencial para el aprendizaje de la electrónica digital, en nuestro contexto puede ser una buena referencia en relación a los capítulos IX y XI.

Gómez Vilda, P., Nieto Lluís, V., Álvarez Marquina, A., y Martínez Olalla, R., *Fundamentos físicos y tecnológicos de la informática*, editorial Pearson, 2007. Libro que trata gran parte de los contenidos aquí estudiados, pero con un enfoque didáctico distinto, lo que puede ser de utilidad para complementar los conocimientos adquiridos.

Jewett J. W., y Serway, R. A., *Física: Electricidad y Magnetismo*, 10ª edición, editorial Cengage, 2019 (los contenidos de este libro también pueden encontrarse como parte del segundo volumen de *Física para ciencias e ingeniería* de los mismos autores). De interés para el capítulo I.

Malvino, A., y Bates, D. J., *Principios de electrónica*, 7ª edición, editorial McGraw-Hill Interamericana, 2007. Un libro con una visión muy práctica y aplicada de los circuitos y sus componentes, especialmente indicado para ampliar conocimientos de los capítulos II, IV, V, VI y VIII.

Pardo Collantes, D., y Bailón Vega, L. A., *Fundamentos de electrónica digital*, ediciones Universidad de Salamanca, 2006. Este libro puede ser un buen apoyo para los contenidos estudiados. Especialmente interesante en relación a los capítulos dedicados a la electrónica digital, es decir, IX, X y XI, y a la extensión hacia el estudio de circuitos combinacionales y secuenciales.

Prat Viñas, Ll. (ed.), *Circuitos y dispositivos electrónicos. Fundamentos de electrónica*, ediciones UPC, 1999. Un texto muy didáctico y de utilidad para complementar los conocimientos relativos a los capítulos desde el I al VII, junto con algunos de los procesos tecnológicos estudiados en el capítulo XII.

Sedra A. S., y Smith K. C., *Circuitos microelectrónicos*, 5ª edición, editorial McGraw-Hill Interamericana, 2006. Un libro que trata con gran detalle el estudio de circuitos electrónicos. En el contexto de lo estudiado aquí puede ser útil para profundizar especialmente en el comportamiento del transistor en conmutación, las etapas CMOS y las familias lógicas de los capítulos X y XI.

Tipler, P. A., y Mosca, G., *Física para la ciencia y la tecnología volumen 2A: Electricidad y magnetismo*, 6ª edición, editorial Reverté, 2010 (los contenidos de este libro también pueden encontrarse como parte del segundo volumen de *Física para la ciencia y la tecnología* de los mismos autores). De interés para el capítulo I.

ÍNDICE ALFABÉTICO

- Abanico de entrada, 254
- Abanico de salida, 255
- AC, *véase* Corriente alterna
- Aceptadora, 92
- Agitación térmica, 84
- Aislante, 81
- Álgebra de Boole, 195
- ALU, 263
- Amperio, 25
- Amplificación, 181
- Amplificador de instrumentación, 190
- Amplificador operacional, 187
- Ancho de banda, 205
- AND, 196
- Ánodo, 108
- Análisis de mallas, 57
- Aproximación semiclásica, 96
- Árbol de pull down, 258
- Árbol de pull up, 258
- Área de operación segura, 218
- Arrastre, 97
- Arseniuro de galio, 104
- Átomo, 73

- Banda de conducción, 80
- Banda de energía, 80

- Banda de valencia, 80
- Bit, 202
- BJT, 127
 - Características de salida, 133
 - Circuito equivalente de pequeña señal, 138
 - Ganancia en corriente, 134
 - Modelo en conmutación, 219
 - Región activa directa, 131, 133
 - Región de corte, 131, 133
 - Región de saturación, 130, 134
 - Regiones de funcionamiento, 133
- Bobina, 40
- Boltzmann, constante de, 84
- Byte, 172

- CAD, 275
- Campo electromagnético, 42
- Campo eléctrico, 20
- Campo magnético, 35
- Canal conductor, 145
- Capa de inversión, 145
- Capacidad, 31
- Carga eléctrica, 16
- Cátodo, 108
- Célula fotoeléctrica, *véase*

- Fotorresistencia
- Célula solar, 175
- Chip, 270
- Circuito abierto, 54
- Circuito compatible TTL, 243
- Circuito eléctrico, 47
- Circuito integrado, 125, 269
- CMOS, 227
- Código de colores, 49
- Coefficiente de difusión, 99
- Coefficiente de transconductancia, 154
- Comparador, 188
- Compleitud funcional, 251
- Componente discreto, 48
- Concentración intrínseca, 88
- Condensador, 30
- Condensador plano-paralelo, 30
- Conductividad, 29
- Conductor, 81
- Conmutación, 213
- Contacto óhmico, 146
- Convertor ADC, 205
- Convertor DAC, 205
- Corriente alterna, 39
- Corriente continua, 39
- Corriente eléctrica, 24
- Corriente inversa de saturación, 112
- Cortocircuito, 53
- Coulomb, ley de, 18
- CPU, 263
- Cuantización de la carga, 17
- Culombio, 18
- Czochralski, método de, 271
- DC, *véase* Corriente continua
- Deposición, 280
- Deposición de capas atómicas, 280
- Diferencia de potencial, 24
- Difusión, 99
- Diodo, 107
- Aproximaciones, 114
- Avalancha, 120
- Barrera de energía, 110
- Característica I-V, 113
- Efecto Túnel, 120
- Voltaje de ruptura, 119
- Voltaje umbral, 112
- Diodo Schottky, 121
- Diodo Zener, 120
- Dipolo eléctrico, 20
- Diseño lógico CMOS, 258
- Distorsión, 185
- Donadora, 91
- Dopado (material), *véase* Extrínseco (semiconductor)
- Dopado (proceso)
- Difusión, 278
- Implantación iónica, 279
- DRAM, 232
- Dualidad de la carga, 17
- Efecto de campo, 144
- Efecto Joule, 29
- Efecto túnel, 235
- Einstein, relación de, 100
- Electroluminiscencia, 165
- Electrónica de estado sólido, 73
- Electrónica de vacío, 73
- Electrónica digital, 195
- Emisión estimulada, 169
- Emisor común, 135
- Energía fotovoltaica, 175
- Energía térmica, 84

- Epitaxia, 280
Escala de integración, 270
Escalado, 282
Extrínseco (semiconductor), 90

Familia lógica, 239
Familia lógica CMOS, 243
Fan-in, *véase* Abanico de entrada
Fan-out, *véase* Abanico de salida
Faraday, ley de, 37
Faradio, 32
Fick, ley de, 99
Filtrado, 118
FinFET, 285
Flujo magnético, 36
Fotocélula, *véase* Fotorresistencia
Fotoconductividad, 172
Fotodetector, 172
Fotodiodo, 172
Fotón, 102
Fotorresistencia, 174
Fototransistor, 173
Frecuencia angular, 39
Frecuencia lineal, 39
Fuente de corriente, 54
Fuente de voltaje, 54
Fuerza de Lorentz, 35
Fuerza electromotriz, 36
Función lógica, 197

GaAs, *véase* Arseniuro de galio
Ganancia, 47, 186
Gap, 80
Gap directo, 167
Gap indirecto, 166
Generación, 101
Generación de energía eléctrica, 38
Germanio, 103

GND *véase* Tierra 116
Grabado, 280
Grado electrónico, 270

Heteroestructuras semiconductoras,
170
Hoja de especificaciones, 168
Hueco, 87

Impedancia, 67
Impureza, 90
Inducción electromagnética, 36
Ingeniería del gap, 167
Inmunidad al ruido, 252
Intrínseco (semiconductor), 86
Inversor, 176, 220
Isótopo, 74

Kennelly, Teorema de, 50
Kirchhoff, leyes de, 55
LCK, 55
LVK, 55

Láser, 169
Layout, 275
LCD, 235
LDR, *véase* Fotorresistencia
LED, 163
LED RGB, 168
Ley de acción de masas, 93
Líneas de campo, 20
Lingote, 271
Litografía, 275
Lógica definida negativa, 199
Lógica definida positiva, 199, 251
Longitud de onda, 42

Malla, 53
Márgenes de ruido, 252

- Masa efectiva, 96
- Máscara, 275
- Maxwell, ecuaciones de, 41
- Mayoritarios, 93
- Memoria flash, 235
- Minoritarios, 93
- Moore, ley de, 282
- Morgan, leyes de, 197
- MOSFET, 143
 - Acumulación, 145
 - Característica de salida, 152
 - Característica de transferencia, 152
 - Circuito equivalente de pequeña señal, 158
 - Drenador, 146
 - Ecuaciones, 154
 - Estructura MOS, 143
 - Fuente, 146
 - Inversión, 145
 - Modelo en conmutación, 217
 - Puerta, 144
 - Realce o normally off, 149
 - Región de corte, 150
 - Región de saturación, 152
 - Región triodo, 151
 - Sustrato, 144
 - Vaciamiento, 145
 - Vaciamiento o normally on, 149
 - Voltaje umbral, 149
- Movilidad, 96
- NAND, 198
- Neutralidad de carga, 93
- Niveles lógicos, 199, 251
- Nodo, 53
- Nodo tecnológico, 282
- NOR, 198
- Norton, Teorema de, 61
- NOT, 196
- Notación fasorial, 68
- Núcleo atómico, 73
- Número atómico, 74
- Oblea, 271
- Offset, 39
- Ohm, ley de, 28
- Ohmio, 28
- Onda electromagnética, 42
- Optoacoplador, 177
- Optoelectrónica, 163
- Optointerruptor, 178
- OR, 196
- Orbital, 74
- Overclocking, 231
- Oxidación húmeda, 279
- Oxidación seca, 279
- P.O.E., *véase* Punto de Operación estacionario
- Par electrón-hueco, 101
- Paralelo, conexión en, 48
- Pauli, principio de exclusión de, 79
- PCB, 281
- Periodo, 39
- Pinzamiento, 152
- Planck, constante de, 102
- Polaridad, 108
- Polarización directa, 111
- Polarización inversa, 112
- Portador, 88
- Potencia, 29
- Potencia dinámica, 223
- Potencia disipada, 29
- Potencia estática, 223

- Potencial eléctrico, 23
 Principio de conservación de carga,
 18
 Principio de exclusión, *véase* Pauli
 Producto velocidad-potencia, 258
 Puerta lógica, 199, 239
 Puertas lógicas CMOS
 AND, 247
 NAND, 245
 NOR, 249
 NOT, 244
 OR, 250
 Pulverizado, 279
 Punto de operación estacionario,
 135

 Rama, 53
 Realimentación, 188
 Recombinación, 101, 165
 Recta de carga, 135
 Rectificación, 116
 Rectificador de media onda, 116
 Rectificador de onda completa, 118
 Red de polarización, 135
 Redes de Bravais, 78
 Resina, 275
 Resistencia, 28, 47
 Resistividad, 28
 Retardo de propagación, 256
 Riesgo lógico, 257
 Ruido electrónico, 199, 252

 Sala blanca, 273
 Semiconductor, 82
 Semilla de semiconductor, 271
 Señal analógica, 181
 Señal digital, 181
 Serie, conexión en, 48

 Shockley, ecuación de, 113
 Silicio, 103
 SOI, 285
 Sólido amorfo, 76
 Sólido monocristalino, 77
 Sólido policristalino, 77
 SRAM, 234

 Tabla periódica, 76
 Tecnología planar, 275
 Término producto, 197
 Término suma, 197
 TFI, 235
 Thévenin, Teorema de, 60
 Tiempo de vida media, 102
 Tiempo de vuelo libre, 96
 Tierra, 116
 Tiranía de los números, 269
 Transformación triángulo-estrella,
 50
 Transistor, 125
 Transistor Bipolar de Unión, *véase*
 BJT
 Triboelectricidad, 16
 TTL, 242
 Tubo de vacío, 125

 Unión p-n, 107

 Valencia atómica, 76
 Valor eficaz, 40
 Varactor, 122
 Vatio, 29
 Velocidad de deriva, 25, 96
 Voltio, 24

 Watt *véase* Vatio 29
 Wifi, 42

XNOR, 198
XOR, 198

Zona de vaciamiento, 110

Semblanza



RAÚL RENGEL ESTÉVEZ.

Licenciado y Doctor en Física (especialidad Electrónica) por la Universidad de Salamanca, en la que ejerce como profesor titular del área de Electrónica. En los últimos veinte años ha impartido docencia de asignaturas relacionadas con los fundamentos físicos de la informática,

el comportamiento electrónico de los materiales, la instrumentación electrónica, la fabricación de circuitos integrados, la electrónica digital y analógica, etc. en las titulaciones de ingeniería informática, ingeniería de materiales, ingeniería mecánica o física, entre otras. Su labor investigadora se ha centrado en el estudio del transporte electrónico y del ruido a alta frecuencia en materiales y dispositivos semiconductores, con particular interés en transistores MOSFET de silicio, arquitecturas MOSFET avanzadas, y más recientemente, grafeno y otros materiales bidimensionales.

Documentos Didácticos, 167

En este libro se abordan los fundamentos que permiten comprender la física subyacente a la operación de los circuitos digitales que constituyen la base de los aparatos electrónicos que utilizamos a diario. Ordenadores personales, teléfonos inteligentes, cámaras digitales, videoconsolas... La nuestra es una *sociedad de la información*, en la que cada día empleamos multitud de dispositivos. Todo ello ha sido posible gracias a la increíble evolución de la electrónica, y en particular, a los circuitos electrónicos de estado sólido. La informática tal y como la conocemos hoy en día es posible gracias a la física de semiconductores, a los transistores de silicio y a los *microchips*, temas que aquí se tratan, junto con un repaso a los fundamentos de la teoría de circuitos y las bases del electromagnetismo. Aunque el libro está orientado a estudiantes de primer curso de grados universitarios en informática, se ha intentado que los conceptos explicados puedan ser comprendidos, sin excesiva dificultad, por cualquier persona interesada con formación a nivel de secundaria en física y matemáticas.



ISBN: 978-84-1311-386-9



9 788413 113869