

NATURALEZA INTRINSECA DEL
TRANSISTOR DE EFECTO DE CAMPO

BIBLIOTECA
DE LA
UNIVERSIDAD DEL VALLE DE GUATEMALA

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades

BIBLIOTECA
DE LA
UNIVERSIDAD DEL VALLE DE GUATEMALA

NATURALEZA INTRINSECA DEL
TRANSISTOR DE EFECTO DE CAMPO

JOSE LUIS COFIÑO SAMAYOA

Trabajo de investigación monográfica
presentada para optar al grado
académico de Licenciado en Física

Guatemala

1980

TRABAJO QUE DEDICO:

A mi padre
DOCTOR ERNESTO COFIÑO UBICO


A la memoria de mi madre
CLEMENCIA SAMAYOA DE COFIÑO

A la memoria de mi abuela
SOFIA RUBIO DE SAMAYOA

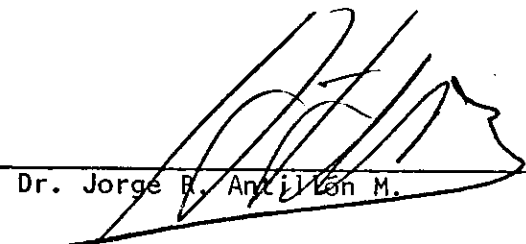
A mi madrina
MARIA RASKIN DE PINOL


A mis profesores, en especial a
DOCTOR JORGE R. ANTILLON MATTA

Vo. Bo. :

(f) 
Ing. Carlos Enrique Zaparolli P.
Asesor

Tribunal:

(f) 
Dr. Jorge R. Antillon M.

(f) 
Ing. Carlos Enrique Zaparolli P.

(f) 
Lic. Marco Eugenio Cabrera V.

Fecha de aprobación: marzo 25 de 1981

CONTENIDO		Página
	SIMBOLOGIA	ix
	INTRODUCCION	1
1.	PRINCIPIOS FISICOS DE LOS MATERIALES SEMICONDUCTORES.	3
	Densidades de portadores.	12
	Generación y recombinación de portadores.	18
	Difusión.	22
	Altos campos eléctricos en semiconductores.	26
2.	TEORIA DE JUNTAS.	30
	Potencial de contacto interno de equilibrio.	33
	Junta abrupta.	33
	Teoría del rectificador de unión P-N.	37
	Relación corriente-voltaje en una junta abrupta.	40
	Transistores de unión.	44
	Ganancia de corriente.	49
	Ganancia de corriente en función de la frecuencia.	52
	Circuito de emisor a tierra.	54
	Circuito de base abierta.	54
3.	EL TRANSISTOR DE EFECTO DE CAMPO.	57
	Corriente en el FET de unión: discusión cuantitativa.	60

	Página
Características estáticas y parámetros dinámicos del FET de unión.	65
Factor de amplificación.	67
Transconductancia.	67
Resistencia de drenador.	69
Efectos de alta frecuencia en el FET de unión.	70
El modelo lineal de alta frecuencia del FET.	70
Amplificador FET de fuente común.	72
Amplificador FET de drenador común.	74
4. DISCUSION FINAL.	77
5. BIBLIOGRAFIA.	80
APENDICES.	82
A. Desarrollo de las ecuaciones 1-10 y 1-11.	83
B. Efecto Hall.	87
C. Ecuación de continuidad. Leyes de Fick. Relaciones de Einstein.	90

SIGNIFICADO DE ALGUNOS SIMBOLOS
QUE SE UTILIZARAN EN ESTE TRABAJO

\hbar = constante de Planck dividida 2π .

$\Psi(x, y, z)$ = la función de onda.

U = energía potencial.

$|\Psi|^2$ = densidad de probabilidad.

f = función de distribución de Fermi-Dirac

k = constante de Boltzmann.

T = temperatura.

E_F = nivel de energía de Fermi.

$N(E)$ = densidad de niveles de energía por rango de energía.

E_c = energía en el borde inferior de la banda de conducción.

E_v = energía en el borde superior de la banda de valencia.

N = densidad de electrones.

P = densidad de agujeros.

N_c = densidad efectiva de los estados en la banda de conducción.

N_v = densidad efectiva de los estados en la banda de valencia.

E_g = anchura de la banda prohibida.

N_i = densidad de portadores en el caso intrínseco en que
 $N = P = N_i$. Es constante para cada semiconductor.

D = densidad de donores o donadores = N_d .

A = densidad de aceptores o receptores = N_a .

σ = conductividad eléctrica.

J = densidad de corriente eléctrica.

μ_n = movilidad en una muestra tipo N.

- μ_p = movilidad en una muestra tipo P.
 \vec{v} = velocidad de arrastre de los portadores.
 D = constante de difusión.
 \vec{E} = campo eléctrico.
 ρ = densidad de carga real.
 τ = tiempo de vida medio.
 G_p, G_n = razón de generación de portadores.
 \vec{j} = densidad de flujo de partículas.
 e = carga del electrón.
 m = masa del electrón.
 ϕ = potencial ($e\phi$ = energía potencial).
 U = potencial electrostático.
 ρ = densidad de carga.
 L_p = longitud de difusión de agujeros.
 L_n = longitud de difusión de electrones.
 J_{nr} = flujo de recombinación en lado N (en el lado P, J_{er}).
 J_{ng} = flujo de generación en el lado N (en el lado P, J_{eg}).
 b = razón de las movilidades.
 I, V = notación de fasor.
 v_{os} = voltaje instantáneo de salida.
 V_{osa} = voltaje de salida en punto estático de operación.
 V_{oo} = voltaje de salida de fuente directamente polarizada.
 v_{oa} = voltaje de salida instantáneo de la componente alterna.
 V_{oas} = voltaje de salida rms de la componente alterna.
 i_o = corriente de salida instantánea.

I_{o0} = corriente de salida en punto estático de operación.

i_o = corriente de salida instantánea de la componente alterna.

I_o = corriente de salida rms de la componente alterna.

v_{es} = voltaje instantáneo de entrada.

V_{es0} = voltaje de entrada en punto estático de operación.

V_{es} = voltaje de entrada de fuente directamente polarizada.

v_{es} = voltaje de entrada instantáneo de la componente alterna.

V_{es} = voltaje de entrada rms de la componente alterna.

I. INTRODUCCION

Durante los últimos años, los elementos semiconductores han avanzado hasta el grado de sustituir a los tubos de vacío, no tan sólo reemplazándolos en gran número de aplicaciones sino además adquiriendo múltiples usos antes no imaginados.

Este cambio gradual de componentes de vacío a componentes de estado sólido, ha ocurrido debido a que éstos últimos son más ventajosos por su tamaño reducido, poco consumo de potencia, larga vida, incrementos de resistencia mecánica, bajo costo y en muchos casos mejores características intrínsecas.

Recientemente, gracias al acelerado avance que ha tenido el conocimiento de los fenómenos intrínsecos en sólidos, la tecnología asociada con la fabricación de semiconductores de uso específico y aprovechando la idea expresada por W. Shockley en 1952, salió de los laboratorios de Estado Sólido un transistor denominado "Transistor de efecto de campo". Este elemento tiene las características propias de un amplificador de voltaje, presenta una impedancia de entrada sumamente grande, tiene características de transferencia similares a las del tubo de vacío y ofrece ventajas adicionales.

Al Físico siempre le ha interesado la medición de parámetros, preocupándose por medir en la forma más precisa y simple posible. En particular, a la detección de voltajes y corrientes se le ha dado mucha importancia, sobre todo cuando éstos son muy pequeños.

La detección de pequeños voltajes y corrientes suele llevarse a cabo por medio de microvoltímetros, los cuales emplean tubos de vacío, debido a la necesidad implícita de tener una gran impedancia de entrada para no consumir potencia del circuito en el que se está llevando a cabo la detección.

La construcción de dichos equipos de detección, empleando tan sólo elementos de estado sólido es de gran valor para el Físico, debido a las ventajas intrínsecas que presentan los elementos de estado sólido sobre aquellos elementos de vacío.

La idea del transistor de efecto de campo surgió antes que la del transistor de unión, pero debido a sofisticada tecnología involucrada en su fabricación, su desarrollo se postergó varios años, utilizándose el transistor de unión. El funcionamiento del transistor de efecto de campo es mucho más simple que el de el transistor de unión, como veremos posteriormente.

El transistor de efecto de campo no sólo es importante por la potencialidad de sus aplicaciones, sino también porque su advenimiento y lo simple de su mecanismo, nos hace abrigar esperanzas de que en el futuro, profundizando nuestros conocimientos del comportamiento microscópico de

los elementos de estado sólido, podremos hallar nuevos dispositivos que beneficien a la ciencia y a la Humanidad.

La electrónica debe considerarse parte integral de la Física, especialmente después del advenimiento de los nuevos elementos semiconductores, que nos permiten visualizar la correlación existente entre los diferentes campos científicos y la electrónica.

Deseo hacer notar, que en el desarrollo de este trabajo, se incluyen algunos términos en inglés, que no sólo, no tienen una traducción exacta, sino que al tratar de traducirlos, perderían la funcionalidad que poseen por ser parte de un lenguaje técnico.

1. PRINCIPIOS FISICOS DE
LOS MATERIALES SEMICONDUCTORES

Los electrones, quienes forman una parte esencial de los átomos, se encuentran en un continuo y rápido movimiento alrededor del núcleo. Es un hecho bien conocido de la física atómica moderna, que los valores de la energía (cinética más potencial) que un electrón puede tener en su movimiento orbital, no forman un rango continuo, sino que se encuentran restringidos a un conjunto perfectamente definido de niveles o estados posibles de energía. Transiciones entre dichos niveles, ocurrirán a "brincos", es decir, una partícula no puede radiar o absorber energía gradualmente y por tanto ir a un estado energético inferior o superior, sino más bien, la partícula deberá de absorber o de emitir un "cuanto" de energía igual a la diferencia entre los estados final e inicial en que se encuentra, debido a la transición.

Este comportamiento, no pudo ser explicado por las leyes de la física clásica. Lo que se necesitó fue una ecuación de movimiento que en las dimensiones atómicas, solamente tuviese soluciones para un conjunto discreto de valores para la energía de la partícula. Schrödinger en 1926, formuló una ecuación de movimiento de las partículas atómicas, donde la energía total de la partícula (E) es el eigen-valor o valor característico. Bajo ciertas condiciones de frontera, dicha ecuación suele tener solución para tan sólo un conjunto discreto de energías, las cuales se identifican con las energías permitidas para las partículas en movimiento.

La ecuación de Schrodinger en su forma más simple, para el caso independiente del tiempo es:

$$-\frac{\hbar^2}{2m} \nabla^2 \psi + U\psi = E\psi \quad 1-1$$

donde \hbar es la constante de Planck dividido 2π ; M = masa de la partícula (el electrón); U = energía potencial de la partícula como función de las coordenadas x, y, z ; E = energía total de la partícula; $\psi(x, y, z)$ = la función de onda.

La interpretación de la función de onda ψ es que la cantidad $|\psi|^2$ nos da la densidad de probabilidad de que la partícula se encuentre en cierto elemento diferencial de volumen alrededor de (x, y, z) . Esto es todo lo que se puede determinar en la mecánica cuántica, en contraste con la posición exacta de la partícula como función del tiempo que se obtiene de las ecuaciones clásicas de movimiento. La ecuación de Schrödinger nos da la información acerca del movimiento de la partícula, dándonos la probabilidad de su posición y dándonos el conjunto completo de estados energéticos posibles, que la partícula puede adquirir. Nuestro interés se centra en este último punto. Algunos de estos niveles energéticos posibles, se encuentran ocupados, es decir, existirán electrones que posean dicha energía particular. Habrá sin embargo, niveles energéticos que se encuentren vacíos, ya que no habrá electrones que posean esa energía.

Ahora, de acuerdo con el principio de exclusión de Pauli, sólo será posible que dos electrones de un sistema, ocupen un mismo nivel (tomando en cuenta el espín del electrón).

Los diferentes elementos de la tabla periódica, tendrán diferentes configuraciones para los niveles de energía de sus electrones y por lo tanto, la cantidad de energía absorbida o emitida por los electrones de cada elemento en una transición, varía no sólo de nivel a nivel sino de elemento a elemento y a ello se deben los diferentes espectros característicos emitidos por los elementos de la tabla. Ver figura F1-1.

Al crecer un cristal, los átomos se acomodan en un arreglo periódico, con espaciamentos reducidos, causando así interacción entre los electrones de los diferentes átomos. Por lo tanto, los niveles energéticos discretos de los electrones en un átomo aislado, se dispersan sobre rangos discretos, dentro de los cuales los niveles se encuentran sumamente juntos y fuera de los cuales no existan niveles permitidos. A estos rangos de niveles sumamente juntos se les llama "bandas de energía" y al espacio o región que se encuentra entre ellos y que no contiene niveles permitidos, se les llama "bandas prohibidas". Ver figura F1-2.

El comportamiento eléctrico de un sólido, se encuentra determinado en gran parte, por la manera en que los electrones se encuentran ocupando los niveles permitidos, es decir, por la distribución de energía de los electrones en el sólido. Los electrones tienen una tendencia a caer en los niveles energéticos disponibles más bajos, de modo que la probabilidad de encontrar ocupado un nivel de baja energía, es mucho mayor que la probabilidad de encontrar ocupado un nivel de alta energía. Sin tomar en cuenta la energía térmica de la red cristalina, que se encuentra siempre presente a temperaturas superiores al cero absoluto, podemos esperar que debido a esa tendencia de los electrones a establecerse en estados de baja energía, todos los niveles debajo de un cierto nivel dado estarán ocupados, y todos los niveles por encima de dicho nivel dado, estarán desocupados. Sin embargo, debido a la agitación térmica a temperaturas finitas, algunos de los electrones tendrán energías superiores que las esperadas y un cierto porcentaje de dichos niveles de baja energía, estarán vacíos. Dicho porcentaje aumenta con un incremento de la temperatura. La descripción matemática cuantitativa de la ocupación de estados, es un problema de mecánica estadística, la cual nos proporciona funciones de distribución. Estas funciones nos dan la probabilidad de que un cierto nivel se encuentre ocupado. Ya que de la mecánica cuántica conocemos la densidad de estados energéticos disponibles, es posible calcular el número promedio de electrones en un cierto rango de energía, es decir, en toda una banda o en parte de ella, como por ejemplo en su borde superior o en el inferior. La función de distribución que se aplica a los electrones en los sólidos, es la llamada "función de distribución de Fermi-Dirac" y está dada por:

$$f = \frac{1}{1 + e^{(E - E_f)/kT}} \quad 1-2$$

donde k = constante de Boltzmann; T = temperatura absoluta; E_f = nivel de Fermi, que es el único parámetro de esta función de distribución.

La función f nos da la probabilidad de que un nivel de energía E

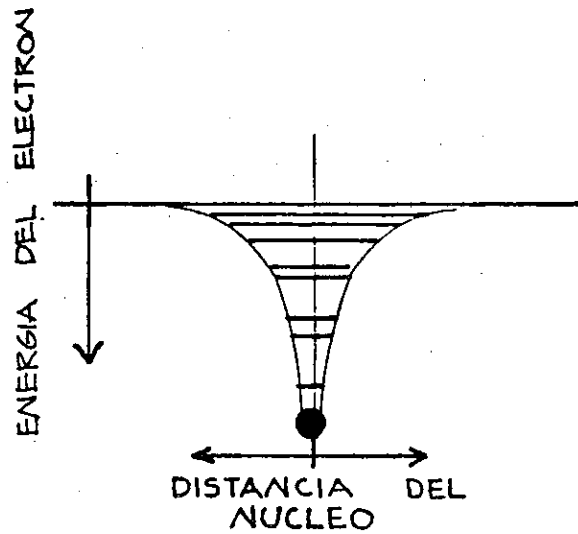


FIGURA F1-1: algunos de los niveles permitidos de energía, para un átomo aislado.

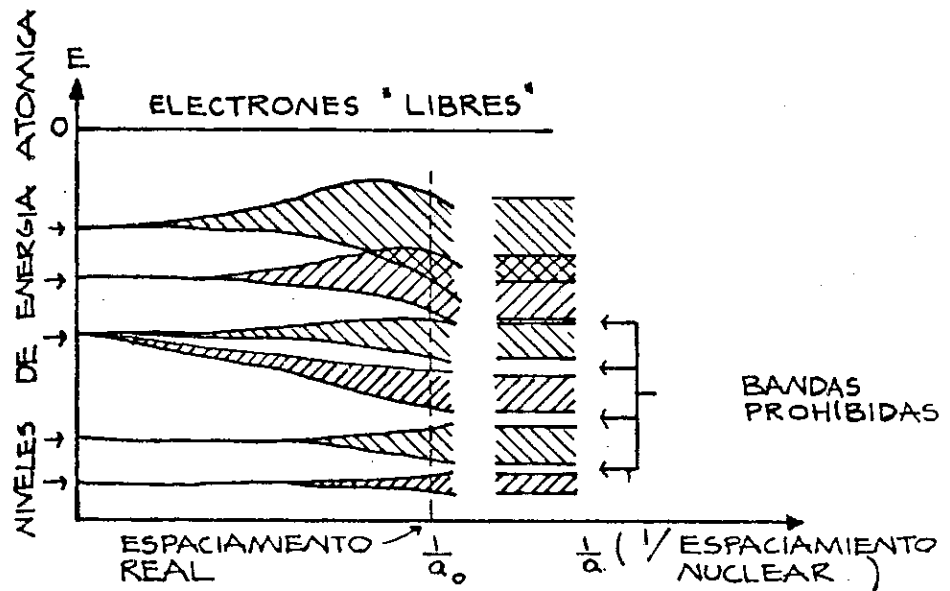


FIGURA F1-2: los niveles de energía de los átomos libres (flechas) se dispersan cuando se juntan los átomos.

se encuentre ocupado por un electrón a la temperatura T . Así pues, el nivel de Fermi es el nivel cuya energía nos da la probabilidad de ocupación de $1/2$. Si la energía es mayor que E_f , entonces la probabilidad de ocupación es menor que $1/2$. Por otro lado, si la energía es menor que E_f , entonces la probabilidad es mayor que $1/2$. Cuando $(E - E_f)$ tiende a infinito, es decir, cuando el nivel considerado se encuentra muy arriba del nivel de Fermi, la probabilidad de ocupación se aproxima a cero o sea que el nivel se encuentra casi siempre vacío. Si por otro lado, el nivel considerado está muy por debajo del nivel de Fermi, es decir, si $(E - E_f)/kT \ll -1$, entonces la probabilidad se aproxima a 1, o sea que el nivel está casi siempre ocupado. Resumiendo tendremos: los niveles de energía inferiores al nivel de Fermi, están predominantemente ocupados, mientras que los niveles superiores al nivel de Fermi, son predominantemente vacíos. La función de distribución f , en cualquier sólido, es independiente de la distribución de los niveles permitidos, y únicamente nos da la probabilidad de que un nivel permitido se encuentre ocupado. La función de distribución f , solamente se aplica en equilibrio térmico, para cuyo caso fue definido el nivel de Fermi y es entonces constante a través del sólido.

La ocupación en equilibrio, depende primordialmente de: la temperatura absoluta y la posición del nivel de Fermi. La dependencia explícita de la función f con respecto a la temperatura, es muy sencilla: en el cero absoluto de temperatura, todos los niveles debajo de E_f están ocupados y todos los niveles arriba de E_f están vacíos, es decir, todos los electrones se encuentran en su mínimo estado de energía, compatible con el principio de exclusión de Pauli. Al aumentar la temperatura, se aumentará el porcentaje de niveles vacíos debajo de E_f y se aumentará el de niveles llenos arriba de E_f . Ver figura F1-3.

Es posible aumentar o disminuir el número de electrones libres en un cristal, ya sea variando la temperatura o introduciendo impurezas dentro de la red del cristal, tal y como veremos posteriormente con gran detalle. En este último caso, se introducen también nuevos niveles posibles de energía. Debido a dichos cambios en el número de electrones y de niveles disponibles de energía, la ocupación de niveles cambia y el nivel de Fermi se autoajusta (ya que es el único parámetro variable de la distribución), para describir correctamente la probabilidad de ocupación de todos los niveles de energía a la temperatura dada. Al variar la temperatura variará también el nivel de Fermi, pues la probabilidad de ocupación y de electrones libres también ha cambiado.

El nivel de Fermi puede encontrarse en medio de una banda de energía permitida. Cuando aplicamos un campo eléctrico, las energías ligeramente superiores que los electrones adquieren, caen en estados permitidos y disponibles, ya que debido a la posición del nivel de Fermi, existen niveles ligeramente superiores y que se encuentran desocupados, de manera que los electrones pueden subir de nivel. El incremento de energía electrónica, toma la forma de un aumento en la velocidad del electrón en la dirección opuesta al campo, resultándonos así un flujo de corriente. Este

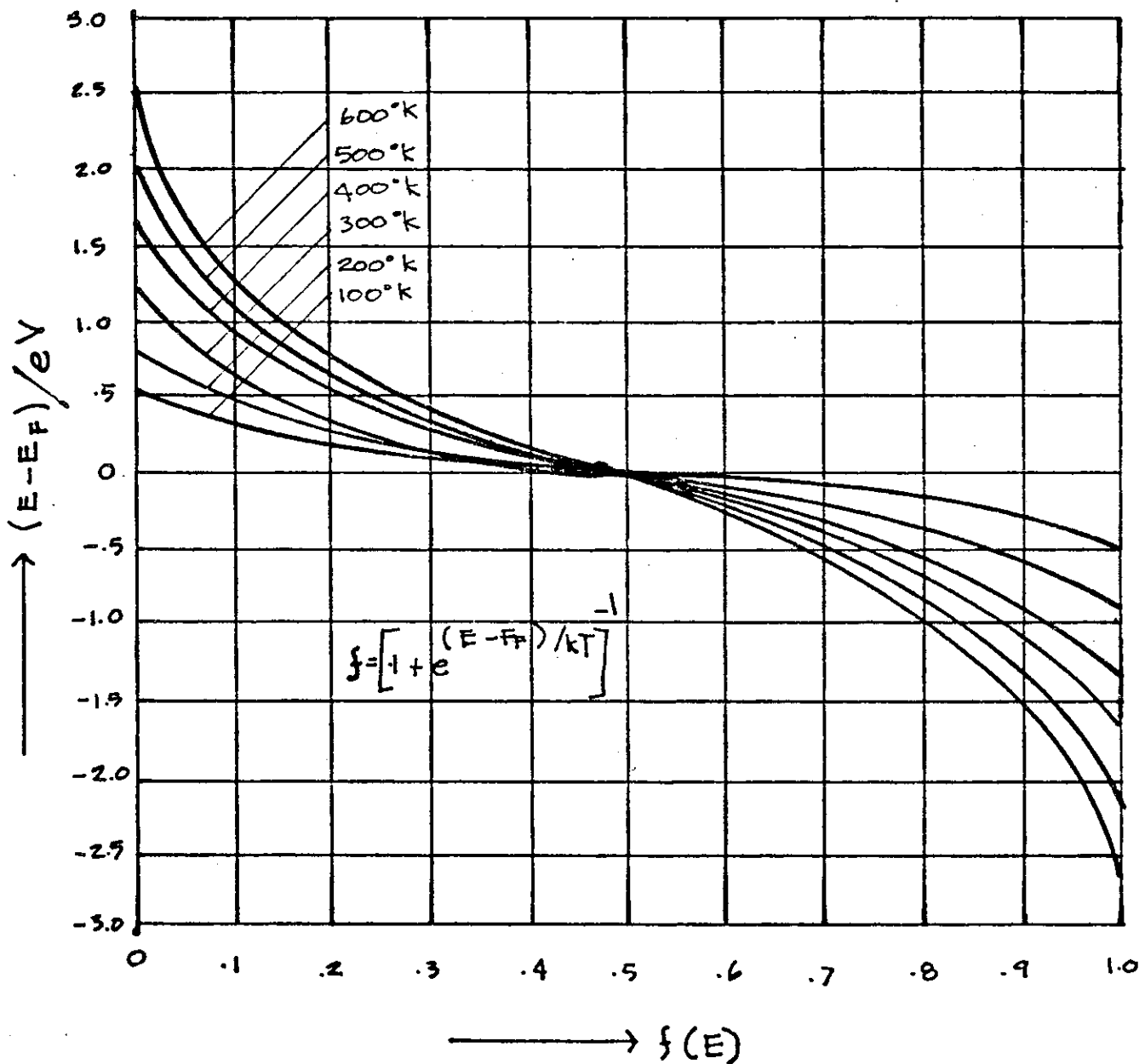


FIGURA F1-3: gráfica de la función de distribución de Fermi $f(E)$ vs. $(E - E_f)/kT$ (en eV) para diversas temperaturas.

es el caso de los metales y por tanto de la conducción metálica.

Existen casos en que el nivel de Fermi se encuentra muy por encima de una banda permitida que está completamente llena, es decir que todos sus niveles se encuentran ocupados, pero al mismo tiempo, la banda permitida inmediata superior, se encuentra suficientemente arriba del nivel de Fermi de modo tal que está completamente vacía, es decir no hay electrones que ocupen alguno de sus niveles. Obviamente la banda permitida vacía no puede contribuir a ninguna conducción. Lo mismo le sucede a la banda permitida completamente llena ya que debido al principio de exclusión de Pauli, no es posible que dos electrones se encuentren en el mismo nivel de energía y por tanto si un electrón toma energía del campo y sube de nivel (dentro de la misma banda permitida), el electrón que ocupaba dicho nivel deberá caer al estado inicial del primero, de modo que no se viole el principio de exclusión de Pauli, así pues, la ganancia neta de energía de todo el sistema es cero y por tanto no habrá flujo neto de corriente. Este caso es el que se conoce como aislador.

Entre los dos modelos anteriormente introducidos, el del metal y el del aislador, se encuentra el de los semiconductores. En este caso la banda prohibida entre las bandas permitidas no es tan ancha como en los aisladores y el nivel de Fermi se encuentra en algún sitio de la banda prohibida. Entonces, la banda permitida inferior está casi llena y la banda permitida superior casi vacía. Los relativamente pocos electrones de la banda permitida superior, adquirirán energía bajo el efecto de un campo eléctrico y la conducción será análoga a la de los metales, excepto que en este caso tenemos una densidad electrónica mucho menor. Esta banda permitida casi vacía de un semiconductor, se llama "banda de conducción". La banda casi llena que se encuentra debajo de la banda prohibida, que en el caso del germanio y el silicio se encuentra ocupada por los cuatro electrones de valencia, recibe el nombre de "banda de valencia".

En la banda de valencia casi llena también puede haber conducción, pero por un proceso diferente al de la conducción electrónica en la banda de conducción. El cristal como un todo es eléctricamente neutro. Entonces si removemos un electrón de la banda, exitándolo fuera de la banda de valencia, dejará en su lugar una carga positiva que llamaremos un "agujero". Ahora, otro electrón puede caer dentro del agujero y recombinarse con él o si no otro electrón de la banda de valencia puede ocupar el agujero sin necesidad de ser exitado fuera de la banda. Este proceso es entonces equivalente al transporte de carga positiva del sitio donde se encontraba el agujero al sitio donde se encuentra ahora.

Así pues, tenemos dos procesos distintos aunque simultáneos de conducción en las respectivas bandas. La banda de valencia se encuentra casi llena de electrones y la conducción en ella se lleva a cabo por el transporte de agujeros. La banda de conducción conduce por medio de electrones, de modo similar que la conducción metálica.

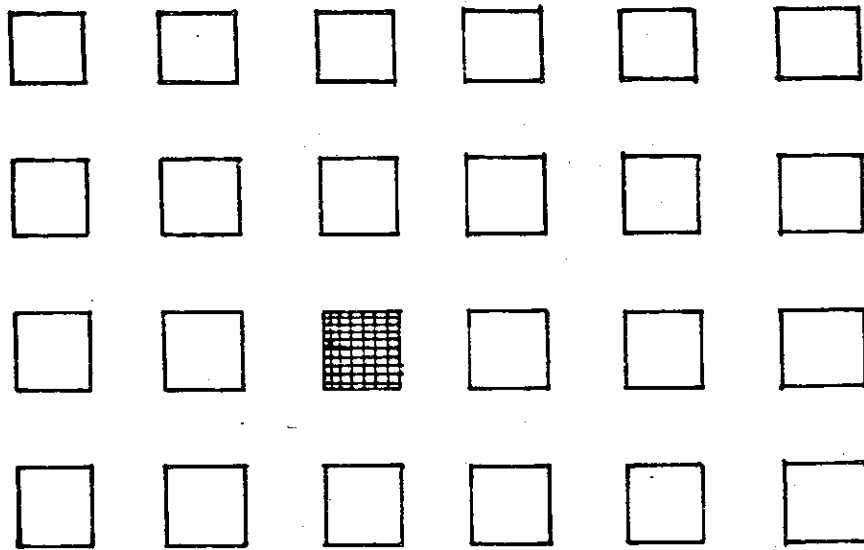
La representación de bandas descritas anteriormente, se aplica en

sentido estricto, únicamente a cristales ideales con dimensiones infinitas. Cualquier variación de la periodicidad absoluta del cristal, recibe el nombre de imperfección o defecto y tendremos que introducirá niveles de energía adicionales, los cuales pueden estar en una banda permitida o prohibida. En este último caso, tienen una pronunciada influencia en el comportamiento eléctrico del semiconductor, aunque se encuentren presentes en pequeñas cantidades. Si las imperfecciones son átomos extraños incorporados dentro del cristal, se les llama átomos de impureza, aunque se les suele introducir intencionalmente.

Nos interesarán particularmente aquellos niveles que se producen de átomos impureza que tienen un electrón de valencia más o menos que los átomos básicos del cristal. Si los átomos del cristal son de un elemento que tiene cuatro electrones de valencia, como el germanio y el silicio, que son elementos de la cuarta columna de la tabla periódica, entonces tendrán un electrón más que los elementos de la tercera columna (B, Al, Ga, In) y un electrón menos que aquellos elementos que se encuentran en la quinta columna (P, As, Sb). Los átomos de impureza pueden ser incorporados dentro del cristal ya sea en un sitio de la red ocupado normalmente por los átomos propios del cristal (substitucional) o ya sea en posiciones impropias de la red cristalina, quedando entre los átomos propios del cristal y deformándolo un poco en sus proximidades (intersticial). Ver figura F1-4.

Si introducimos un elemento de la quinta columna, substitucionalmente, dentro de un cristal de germanio o de silicio, tendremos que cuatro de los cinco electrones de la impureza formarán un enlace covalente con el germanio. Sin embargo el quinto electrón estará débilmente ligado al átomo de impureza y con una pequeña cantidad adicional de energía, será liberado y podrá moverse a través del cristal, contribuyendo así a la conducción. Los niveles extra que introducen las impurezas que tienen un electrón en exceso, se encuentran localizados en la banda prohibida, muy próximos a la banda de conducción. Para temperaturas bajas, estos niveles de impurezas se encuentran ocupados, pero a la temperatura ambiente, la energía térmica del cristal es suficiente para excitar el exceso de electrones hacia la banda de conducción, donde contribuirán al proceso de conducción electrónica antes descrito. Las impurezas que proporcionan e lectrones de conducción, reciben el nombre de impurezas tipo N o "donadores" y se dice que el cristal está "drogado" con impurezas de tipo N.

Un cristal que tiene impurezas de un elemento con un electrón de valencia menor que los átomos constituyentes de la red, en posición substitucional, formará solamente tres enlaces covalentes, sobrándole al átomo de la red, un electrón de valencia, entonces la impureza actuará como un sitio posible para otro electrón más. Es más sencillo tratar el problema del electrón de valencia faltante, si lo representamos por un agujero (cargado positivamente). Un electrón que tiene un electrón de valencia menos que los átomos del cristal, induce niveles adicionales de energía dentro de la banda prohibida, muy cerca de la banda de valencia. A bajas temperaturas tendremos que dichos niveles se encuentran vacíos, pero a tem



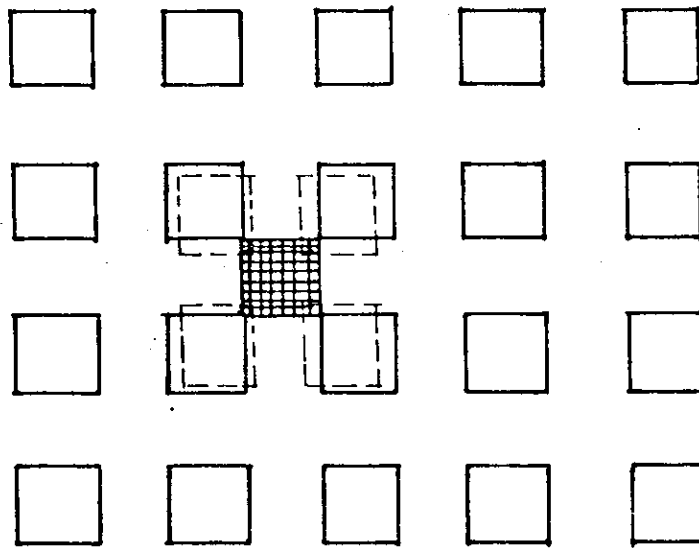
(a)



ATOMO NORMAL DEL CRISTAL



ATOMO IMPUREZA



(b)

FIGURA F1-4: a) Impurezas sustitucionales; b) Impurezas intersticiales.

peratura ambiente se excitarán los electrones de la banda de valencia hacia dichos niveles de impurezas, dejando así un agujero, el cual queda disponible a conducir. Este tipo de impurezas se llaman "impurezas del tipo P" o "aceptores".

La energía necesaria para elevar a un electrón de un nivel de donador hacia la banda de conducción o de la banda de valencia a un nivel de aceptor se llama energía de ionización.

Dependiendo de cual sea el tipo de impureza predominante, se dice que el cristal exhibe conducción tipo N o tipo P, es decir, conducción por electrones o por agujeros. Al tipo más abundante de portadores se le llama "portadores mayoritarios" y a los menos abundantes se les llama "portadores minoritarios". En un semiconductor tipo N, el nivel de Fermi se encuentra cerca de la banda de conducción y en un semiconductor tipo P, se encuentra cerca de la banda de valencia.

Si no tenemos impurezas presentes y por lo tanto los electrones y agujeros de conducción tienen las mismas densidades (cada electrón de conducción solamente puede provenir de la banda de valencia, por excitación a través de la banda prohibida, dejando exactamente un agujero), entonces el semiconductor se llama "intrínseco". Se emplea el mismo término, si el efecto de las impurezas añadidas es despreciable a la temperatura considerada, en comparación con la densidad de electrones excitados a través de la banda prohibida. En todos los demás casos se dice que el semiconductor es "extrínseco" y la densidad de alguno de los portadores será mayor que la del otro.

Densidades de portadores

Las densidades de portadores, es decir la densidad de electrones en la banda de conducción y de agujeros en la banda de valencia, pueden ser calculados de la siguiente manera. Ya que los niveles permitidos de energía en las bandas se encuentran muy poco espaciados, puede ser definida una densidad de niveles de energía por unidad de rango de energía, $N(E)$. El número de niveles entre dos valores de energía dados por E y $E+dE$ es: $N(E)dE$

El número promedio de niveles ocupados entre las energías E y $E+dE$ se calcula como el producto de la densidad de niveles disponibles, por la probabilidad de que los niveles se encuentren ocupados (dada por la distribución de Fermi $f(E)$):

$$N(E) \cdot f(E) dE \quad 1-3$$

El número de electrones (número de niveles ocupados) con una energía entre E_1 y E_2 , será entonces:

$$\int_{E_1}^{E_2} N(E) f(E) dE \quad 1-4$$

La densidad de estados $N(E)$, no se conoce exactamente en los casos prácticos pero se han hecho aproximaciones que nos permiten calcular el número de electrones, por la ecuación anterior. Otra simplificación que suele utilizarse es para cuando $E - E_f \gg kT$ y en tal caso la distribución de Fermi queda:

$$f \cong e^{-(E - E_f)/kT} \quad 1-5$$

que es idéntica con la distribución clásica de Maxwell-Boltzmann. Esta condición es válida para niveles en la banda de conducción, si el nivel de Fermi (ese suele ser el caso general) se encuentra varios kT por debajo de la banda.

Ahora nos es posible calcular la densidad de electrones en la banda de conducción:

$$N = e^{E_f/kT} \int_{E_c}^{E_2} N(E) e^{-E/kT} dE \quad 1-6$$

donde E_c es el borde inferior de la banda de conducción y E_2 se encuentra en alguna parte bastante alta de la región totalmente desocupada (se suele hacer que $E_2 = \infty$).

La probabilidad de que un nivel no se encuentre ocupado, es desde luego $(1-f)$, la cual es por tanto la función de distribución para los agujeros. El número de agujeros en el rango de energía de E_1 a E_2 , es entonces:

$$\int_{E_1}^{E_2} N(E)(1-f) dE \quad 1-7$$

para $E_f - E \gg kT$, la función de distribución de Fermi se reduce a:

$$(1-f) \cong e^{(E - E_f)/kT} \quad 1-8$$

que es nuevamente una distribución de Maxwell-Boltzmann. Esta condición es válida si el nivel de Fermi se encuentra varias veces kT sobre la banda de valencia (éste es el caso usual de interés). Así pues, la densidad de agujeros en la banda de valencia está dada por:

$$P = e^{-E_f/kT} \int_{E_1}^{E_v} N(E) e^{E/kT} dE \quad 1-9$$

E_v es el borde superior de la banda de valencia, y E_1 se encuentra en alguna parte baja de la región totalmente ocupada (se suele hacer $E_1 = 0$).

Bajo ciertas suposiciones se calculan las integrales de las ecuaciones de N y de P :

$$N = N_c e^{-(E_c - E_f) / kT} \quad 1-10$$

$$P = N_v e^{-(E_f - E_v) / kT} \quad 1-11$$

Ver apéndice A.

A N_c y N_v se les suele llamar las densidades efectivas de los estados en las bandas de conducción y de valencia respectivamente. Sus valores numéricos a temperatura ambiente son del orden de $10^{17} / \text{cm}^3$ y su variación con la temperatura es relativamente pequeña (aproximadamente proporcional a $T^{3/2}$).

Es interesante notar que el producto NP es independiente de la posición del nivel de Fermi, es decir:

$$NP = N_c N_v e^{-E_c / kT} = N_i^2 \quad 1-12$$

donde $E_c = E_c - E_v$ es la anchura de la banda prohibida; N_i es la densidad de portadores en el caso intrínseco en que las densidades de electrones y agujeros son exactamente iguales: $N = P = N_i$ (entonces el nivel de Fermi se encuentra cerca del centro de la banda prohibida). Así pues, la cantidad N_i^2 es obviamente una constante para cada semiconductor.

La utilidad de la ecuación anterior surge del hecho de que en la mayoría de las muestras a temperatura ambiente, prácticamente todos los donores o aceptores se encuentran ionizados y la densidad de electrones de conducción en una muestra tipo N, es prácticamente igual a la densidad de donores (D) y la densidad de agujeros en una muestra tipo P, es prácticamente igual a la densidad de aceptores (A). Las densidades de portadores minoritarios, puede calcularse directamente de esta ecuación:

$$P = N_i^2 / D \quad 1-13$$

$$N = N_i^2 / A \quad 1-14$$

Esto facilita grandemente el cálculo de la dependencia de las densidades de portadores con respecto a la temperatura, en el rango extrínseco, donde la temperatura es tan baja que muy pocos portadores serán excitados a través de la banda prohibida. Entonces, las densidades de portadores mayoritarios son prácticamente constantes con respecto a la temperatura y son iguales a las densidades respectivas de impurezas, mientras que las densidades de portadores minoritarios, de acuerdo a las ecuaciones 13 y 14 se incrementan con N_i^2 .

A temperaturas superiores ocurre una transición a comportamiento in-

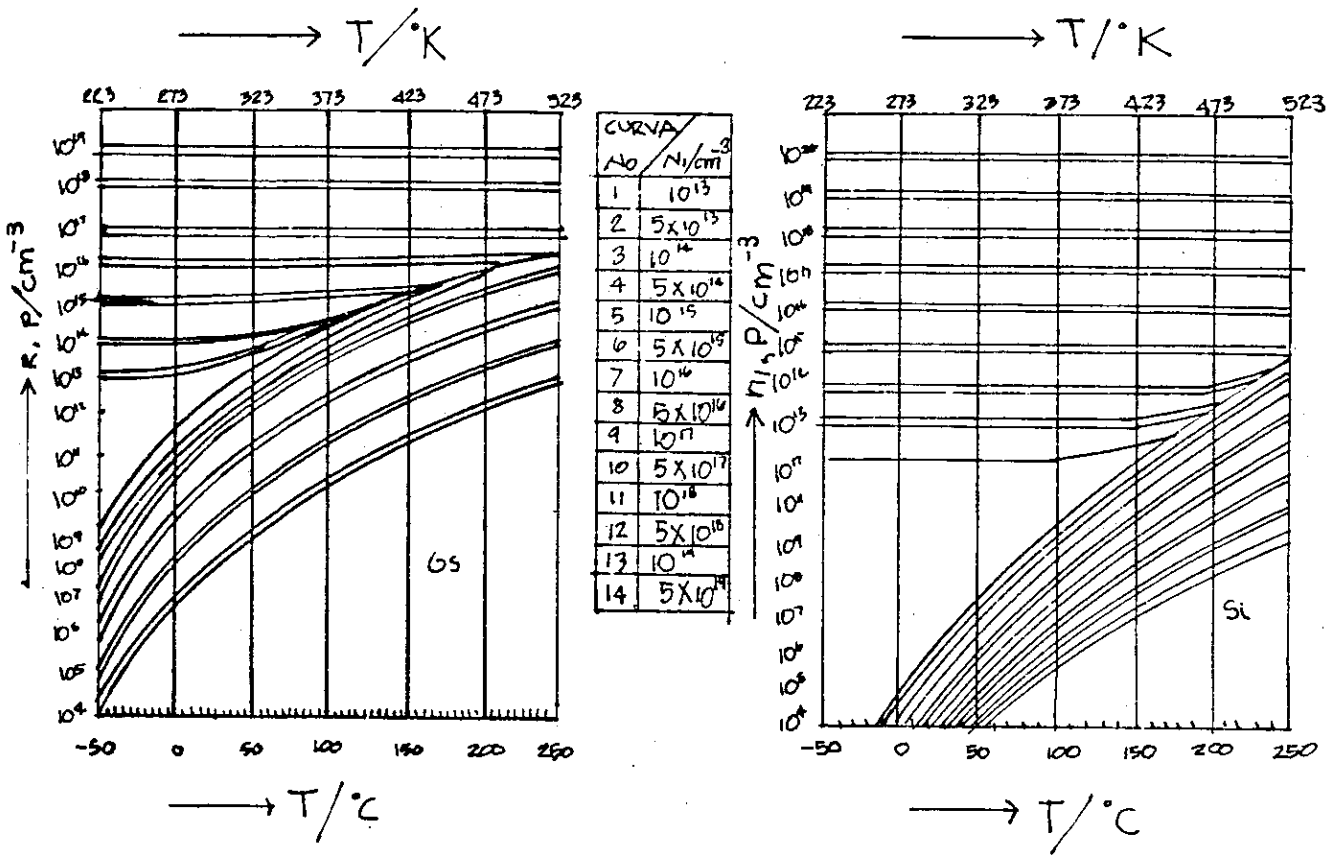


FIGURA F1-5: portadores mayoritarios y minoritarios N y P, en germanio y silicio con varias concentraciones N_i de impurezas, como función de la temperatura.

trínseco y una cantidad apreciable de electrones son excitados a través de la banda prohibida, siendo estos comparables o aún más, que la densidad de impurezas presentes. Ver figura F1-5. En tal caso se sigue el siguiente procedimiento para calcular las densidades de portadores como función de la temperatura: la ecuación 12 sigue siendo válida, pero ahora contiene dos incógnitas que son N y P . La segunda ecuación necesaria nos la da la condición de neutralidad eléctrica en el cristal homogéneo, es decir, la densidad de cargas positivas deberá ser igual a la densidad de cargas negativas,

$$N + A = P + D \quad 1-15$$

donde D y A han sido definidos como las densidades de donores y aceptores ionizados, es decir, simplemente cargados; para temperaturas superiores a la temperatura ambiente, ambas son constantes independientes de la temperatura. Las ecuaciones 12 y 15, permiten entonces calcular la dependencia de la densidad de portadores con la temperatura en el rango de 0°C hacia arriba, el cual es de nuestro interés, si elevamos la temperatura tan alto como para que los electrones térmicamente generados a través de la banda prohibida excedan por una buena cantidad a los portadores contribuidos por los átomos de impureza, entonces:

$$N = P = N_i \quad 1-16$$

y de acuerdo con la ecuación 12 es entonces posible determinar la banda prohibida de la pendiente de una gráfica semilogarítmica de la densidad de portadores contra el recíproco de la temperatura. Ver figura F1-6.

La movilidad de los portadores en un semiconductor puede determinarse experimentalmente midiendo, por efecto Hall (ver apéndice B), la densidad de portadores N y P . Y la conductividad eléctrica σ de la muestra, la cual se define por la ley de ohm como: $\vec{J} = \sigma \vec{E}$, donde \vec{J} es la densidad de corriente eléctrica y \vec{E} es la intensidad de campo eléctrico. En una muestra de tipo N donde la conducción ocurre únicamente por electrones, la movilidad μ_n es igual a:

$$\mu_n = \frac{\sigma_n}{qN} \quad 1-17$$

y en una muestra tipo P:

$$\mu_p = \frac{\sigma_p}{qP} \quad 1-18$$

donde q es el valor absoluto de la carga electrónica, N y P las densidades de electrones y agujeros y μ_n y μ_p denotan las movilidades de electrones y agujeros, definidas por:

$$\vec{u}_n = -\mu_n \vec{E} \quad 1-19$$

$$\vec{u}_p = -\mu_p \vec{E} \quad 1-20$$

donde \vec{u}_n y \vec{u}_p son las velocidades de arrastre de los portadores. Para calcular la densidad de portadores a partir de una medida de conductividad, debemos conocer las movilidades como función de la conductividad.

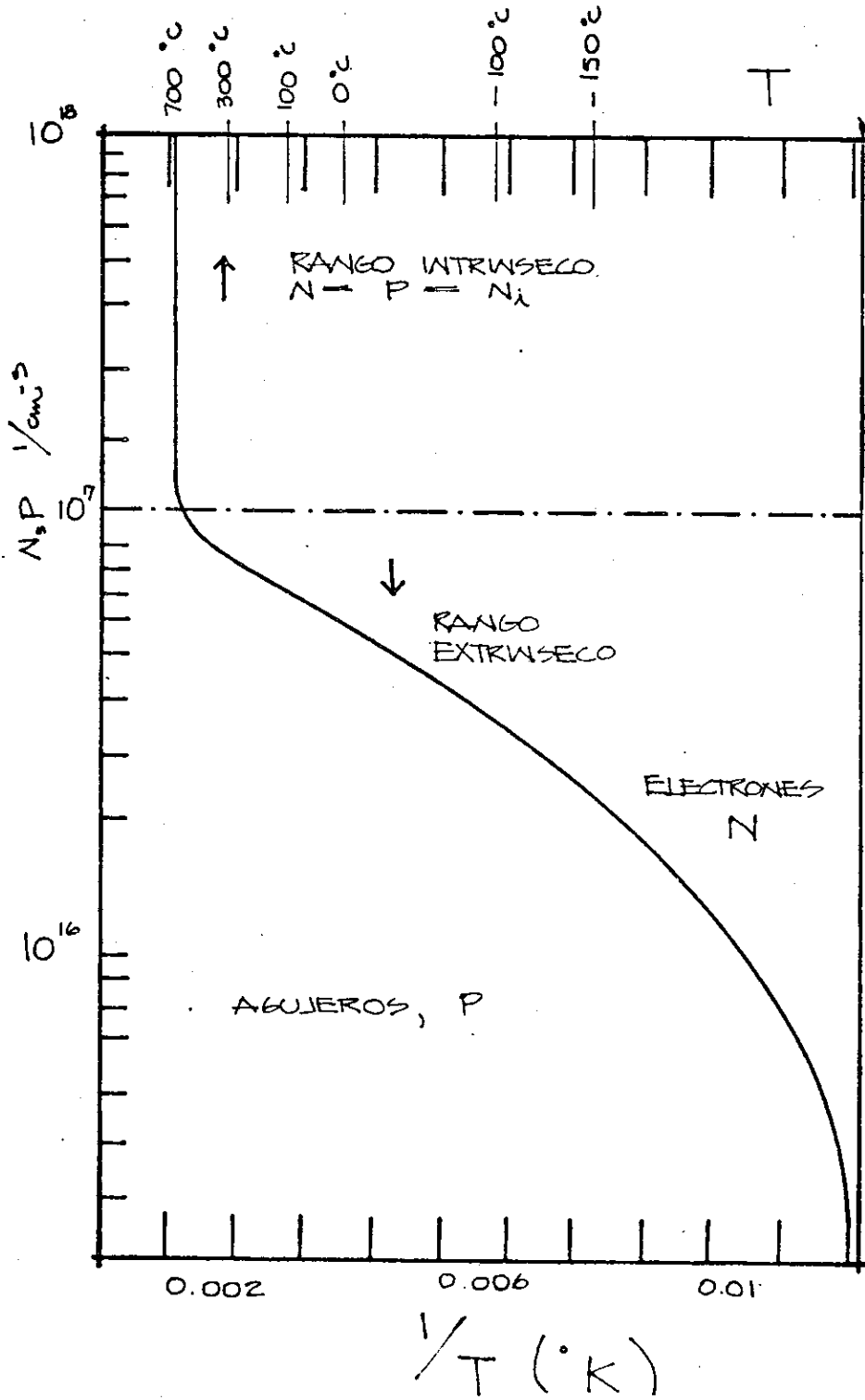


FIGURA F1-6: gráfica semilogarítmica de la densidad de portadores vs. el recíproco de la temperatura absoluta. Las curvas mostradas son para el silicio con impurezas de fósforo.

Cuando tenemos una conducción debida a electrones y agujeros, la conductividad será la suma de las dos contribuciones. Así pues, si asignamos una movilidad μ_n a todos los electrones y μ_p a todos los agujeros, y si tenemos N electrones y P agujeros por unidad de volumen, la conductividad es

$$\sigma = q (\mu_n N + \mu_p P) \quad 1-21$$

Esta relación se aplica a todos los semiconductores en general.

Generación y recombinación de portadores

Las concentraciones de electrones y agujeros se mantienen en equilibrio térmico, por los procesos opuestos de ionización térmica y de recombinación, entre bandas y estados de impurezas. Sin embargo, en la operación normal de los elementos semiconductores, las concentraciones de portadores se alteran de sus valores de equilibrio. Es muy importante el poder analizar ciertas desviaciones del equilibrio.

Cuando electrones se mueven a estados en la banda de conducción, o a estados fuera de la banda de valencia, se dice que los portadores han sido "generados". La mayor fuente de generación de portadores en muchos elementos, es la ionización de niveles o estados de impurezas por fonones térmicos. El proceso de generación de portadores por fonones térmicos, en un semiconductor, actúa principalmente para mantener la densidad de portadores mayoritarios. Sin embargo en muchos elementos, los portadores minoritarios son más importantes. Como veremos, éstos se introducen de varias formas, generalmente de regiones adyacentes de semiconductor de tipo opuesto. En muchos elementos es esencial que éstos portadores minoritarios puedan moverse a través del semiconductor sin una fuerte posibilidad de ser removidos de la banda en que se están moviendo. En el lenguaje de semiconductores, la "pérdida" de un portador, simplemente significa el retorno de un electrón de la banda de conducción a un estado localizado o a la banda de valencia. Aquí debemos distinguir entre "atrapamiento" de portadores y recombinación. El atrapamiento, es un removimiento temporal de un portador, de una banda a un estado localizado. Una trampa retornará eventualmente al portador a la banda; sin embargo, si el portador representa una señal de información, puede que regrese demasiado tarde para representar la señal deseada. A este respecto, podemos distinguir entre trampas "rápidas" y "lentas"; las primeras regresan al portador o la banda en un corto tiempo del orden de 10^{-7} seg., mientras que las segundas pueden retener a los portadores durante segundos, o inclusive horas. Estas trampas son de hecho, las que hacen posible la luminescencia de algunos materiales fosforescentes, tiempo después de que la excitación ha cesado.

Por otra parte, la recombinación implica la pérdida permanente del portador. Recombinación directa, significa que un electrón de la banda de conducción cae directamente a la banda de valencia. Generalmente la energía se libera en la forma de un fotón. Mientras que este proceso es

la mayor fuente de recombinación en algunos compuestos semiconductores (Ga As, In As), tenemos que la recombinación directa es muy poco probable en germanio y silicio, ya que la mayoría de los electrones y estados desocupados de la banda de valencia (agujeros), no pueden satisfacer la condición de conservación de energía-momentum. Entonces los procesos predominantes de recombinación, hacen uso intermedio de estados localizados en la banda prohibida. Un electrón de la banda de conducción cae primeramente a dicho estado intermedio, liberando energía en forma de uno o más fonones de baja energía. Después de cierto intervalo de tiempo, en el estado intermedio, el electrón cae a un estado desocupado de la banda de valencia, liberando nuevamente uno o más fonones. Este último paso puede describirse igualmente, como la caída de un agujero dentro del estado intermedio. Al estado localizado se le suele llamar "centro de recombinación". Ver figura F1-7.

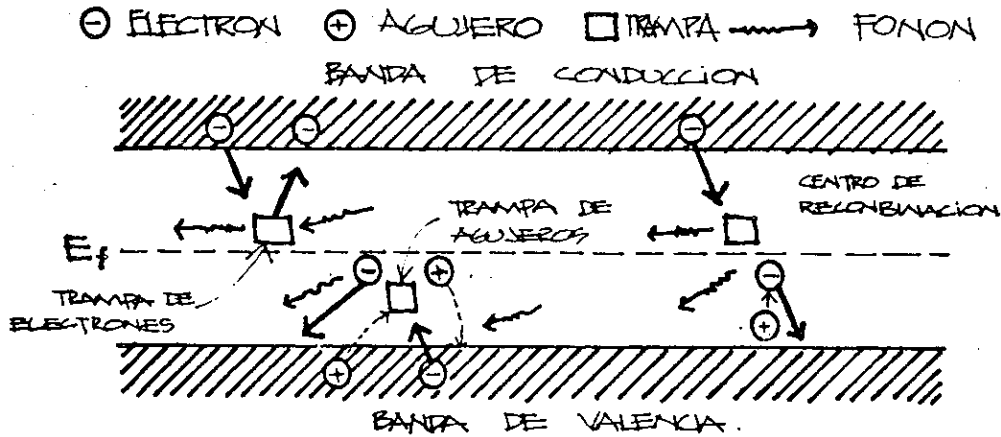
El grado de localización de un estado en la banda prohibida, depende de que tan distante se encuentre de los bordes de las bandas permitidas. Los estados más distantes, cerca del centro de la banda prohibida, tienen generalmente la función electrónica de onda más localizada, mientras que los estados poco distantes, como los estados de donador y aceptor, tienen funciones de onda que se extienden sobre varias constantes de la red. Esta localización controla la probabilidad de atrapamiento o de recombinación, ya que un electrón debe acercarse a la imperfección representada por un estado distante para interactuar con ella.

Estados de atrapamiento y de recombinación aparecen aún en semiconductores puros, como resultados de los defectos de la red. Monocristales altamente defectuosos o policristales semiconductores, suelen tener tantos de esos estados que hasta es difícil identificar su naturaleza semiconductor. Las dislocaciones son más efectivas que los defectos puntuales en lo que respecta a trampas. Las fronteras de grano y las superficies reales de los cristales constituyen regiones mucho mayores de atrapamiento. Sin embargo, la tecnología moderna del crecimiento de cristales, se ha desarrollado de modo que podemos obtener cristales de germanio, silicio, u otro semiconductor, de muy alta calidad, eliminando así el problema de las trampas y los centros de recombinación hasta el grado deseado. De hecho, en algunos elementos se introducen intencionalmente estados distantes, para remover portadores indeseables o para atrapar carga espacial dentro de una región determinada.

La influencia dinámica de las trampas y de los centros de recombinación se puede comprender a través de argumentos probabilísticos muy sencillos. Con fines explicativos, consideremos un semiconductor intrínseco que tiene un conjunto de trampas y centros de recombinación, localizados en la banda prohibida y en un nivel E_s de energía. Definiremos las diversas probabilidades de transición, por segundo, de la siguiente manera: ver figura F1-8

B_{cs} = probabilidad de que un electrón de la banda de conducción haga una transición a los estados localizados, por unidad de tiempo.

FIGURA F1-7: esquema de los procesos más importantes de recombinación y atrapamiento en semiconductores. A la izquierda, un electrón es atrapado, emitiendo un fonón. Es liberado por la absorción de otro fonón. En el centro, un agujero sufre un proceso similar (flechas punteadas); una descripción alternativa



es en términos de un electrón (flechas sólidas). A la derecha, un electrón cae sucesivamente de la banda de conducción a un estado localizado y a la banda de valencia, liberando fonones en ambos pasos. En el último paso aniquila un agujero.

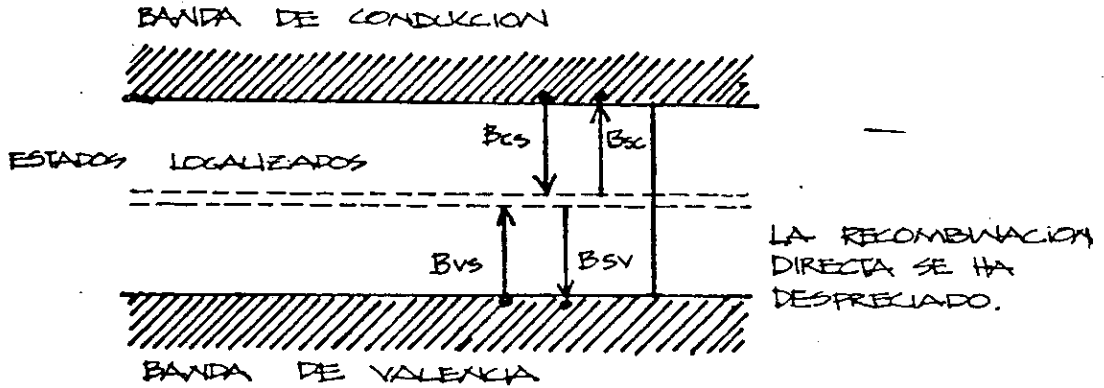


FIGURA F1-8: definición de las probabilidades de transición empleadas en el texto.

B_{vs} = probabilidad de que un electrón de la banda de valencia haga una transición a los estados localizados, por unidad de tiempo.

B_{sv} = probabilidad de que un electrón de un estado localizado dado, haga una transición a la banda de valencia, por unidad de tiempo.

Debemos notar que B_{cs} y B_{vs} tienen una interpretación un poco diferente de B_{sc} y B_{sv} , ya que hay infinidad de estados disponibles en la banda de conducción y muchos electrones en la banda de valencia aunque pocos estados disponibles localizados. Los parámetros B_{cs} y B_{vs} también implican electrones moviéndose en las proximidades de estados localizados, mientras que B_{sc} y B_{sv} , no.

Con estas definiciones y una densidad S de estados localizados (de la cual S_u están vacíos y S_f están llenos por electrones), la razón de tiempo de transición de la banda de conducción a estados localizados, es justamente $N B_{cs} S_u$. Habrá a su vez un flujo inverso de electrones a una razón $S_f B_{sc}$. Despreciando la recombinación directa, tal y como es más apropiado en los semiconductores elementales, la razón de cambio neta de la densidad N de electrones, es:

$$\frac{dN}{dt} = -N B_{cs} S_u + S_f B_{sc} \quad 1-22$$

Similarmente, hay una razón respecto al tiempo, del cambio de la densidad P de agujeros, dada por:

$$\frac{dP}{dt} = -P B_{sv} S_f + B_{vs} S_u \quad 1-23$$

La razón de cambio de la densidad de estados localizados llenos es justamente la diferencia de aquellos:

$$\frac{dS_f}{dt} = N B_{cs} S_u - B_{sc} S_f + B_{vs} S_u - P B_{sv} S_f \quad 1-24$$

En la condición de equilibrio térmico, la razón del cambio de la densidad de electrones y agujeros debe de ser cero y de las ecuaciones 22 y 23, obtenemos:

$$\frac{N B_{cs}}{B_{sc}} = \frac{S_f}{S_u} = \frac{B_{vs}}{P B_{sv}} \quad (\text{en equilibrio}) \quad 1-25$$

Así pues, la probabilidad relativa de transición entre pares de estados, están relacionados.

Como siguiente punto, habremos de ver las densidades no uniformes de portadores. En los metales, las densidades de electrones son tan altas que cualquier cambio sustancial en la densidad se encuentra prohibido por los grandes campos de carga espacial que les acompañan. En los semiconductores, la concentración de portadores (particularmente portadores minoritarios en materiales extrínsecos) puede ser tan pequeña, que campos eléctricos modestos permiten cambios en la densidad de portadores del orden de varios órdenes de magnitud. También son importantes las variaciones en la densidad de portadores, cuando la introducción de impurezas al mono

crystal semiconductor, varían de tipo N a tipo P, dentro de un pequeño elemento.

Difusión

La difusión es un mecanismo importante del flujo de portadores (minoritarios) en los semiconductores. La difusión ocurre siempre que exista una concentración no uniforme de partículas móviles de cualquier tipo en un ambiente térmico. El movimiento al azar lleva a las partículas a través del sistema.

La ecuación principal para la difusión, recibe el nombre de "primera ley de Fick" y está dada por:

$$\vec{J}_M = -e D \nabla M \quad 1-26$$

donde M es la densidad o concentración de partículas, D es la constante de difusión y \vec{J}_M es la densidad de corriente de las partículas consideradas. La constante D es dependiente de la temperatura y en el caso de difusión de electrones está dada por:

$$D_N = \frac{\mu_N k T}{e} \quad 1-27$$

Análogamente, para los agujeros:

$$D_P = \frac{\mu_P k T}{e} \quad 1-28$$

La segunda ley de Fick expresa simplemente el hecho de que la derivada parcial respecto al tiempo de la densidad de cualquier partícula indestructible, es igual a la divergencia de la corriente de la partícula, es decir:

$$\frac{\partial M}{\partial t} = -\nabla \cdot \vec{J}_M = e D \nabla^2 M \quad 1-29$$

Ver apéndice C

En un semiconductor que tiene una densidad no uniforme N de electrones y P de agujeros, tendremos que las corrientes de electrones y de agujeros estarán dadas por dos términos, el primero de los cuales resulta del campo eléctrico externo \vec{E} aplicado al semiconductor y el segundo término se debe a la difusión respectiva de electrones y agujeros. Así pues tendremos:

$$\vec{J}_N = e (\mu_N N \vec{E} + D_N \nabla N) \quad 1-30$$

$$\vec{J}_P = e (\mu_P P \vec{E} - D_P \nabla P) \quad 1-31$$

Aquella parte de la corriente debida al campo se llama "corriente de arrastre" y aquella otra debida al gradiente de concentración es la "corriente de difusión". El gradiente de concentración de electrones es positivo en la dirección \hat{x} [$(\partial N / \partial x) > 0$], la redistribución resultante a

través de difusión, acarrea a los electrones a la izquierda y la corriente será a la derecha (positiva). El mismo gradiente de concentración para agujeros, acarreará a los agujeros y a la corriente hacia la izquierda (negativa).

Si sacamos divergencia a las ecuaciones de corriente 1-30 y 1-31 obtendremos:

$$\nabla \cdot \vec{J}_n = [\mu_n \nabla \cdot (N \vec{E}) + D_n \nabla^2 N] e \quad 1-32$$

$$\nabla \cdot \vec{J}_p = [\mu_p \nabla \cdot (P \vec{E}) - D_p \nabla^2 P] e \quad 1-33$$

La ecuación de continuidad para la conservación de carga es:

$$\nabla \cdot \vec{J} = -\frac{\partial \rho}{\partial t} \quad 1-34$$

donde ρ es la densidad de carga real. En este sentido, sin embargo, los portadores de los semiconductores no siempre se conservan, sino que se pueden perder por recombinación.

En equilibrio térmico el único proceso de generación es el térmico. La generación a la que se generan los pares y a la que se recombinan debe ser idéntica, de otra manera, la concentración de electrones y agujeros crecería o declinaría como función del tiempo. La velocidad de generación térmica de equilibrio G_0 es el número de pares electrón-agujero generados por unidad de volumen en unidad de tiempo, mediante la ruptura térmica de enlaces covalentes. Esta cantidad es función de la temperatura y ciertos parámetros del cristal, pero es independiente de la concentración de electrones y agujeros. La velocidad de recombinación se relaciona con el tiempo medio que transcurre entre la generación de un portador y su recombinación subsecuente. Esta cantidad se conoce como tiempo de vida medio del portador. Cuando se tienen N portadores por unidad de volumen, la velocidad de recombinación (número de eventos de recombinación por unidad de volumen por unidad de tiempo) es N/τ . Las velocidades de generación deben ser iguales a las de recombinación y las de generación de agujeros iguales a las de generación de electrones (ca electrón que se genera, deja un agujero). Asimismo, la de recombinación de agujeros debe ser igual a la de recombinación de electrones.

Consideremos ahora una región del cristal cuyas dimensiones son dx , dy , dz . Supóngase que hay unas densidades de flujo de partículas \vec{J}_p y \vec{J}_n que fluyen dentro y fuera de esta región. La componente x de \vec{J}_p en $x + dx$ se puede expresar en función de \vec{J}_p en x , haciendo una expansión de Taylor de $J_{px}(x + dx)$ en el límite donde dx se hace pequeño, y tomando sólo los dos primeros términos,

$$J_{px}(x + dx) = J_{px}(x) + \frac{\partial J_{px}}{\partial x} dx$$

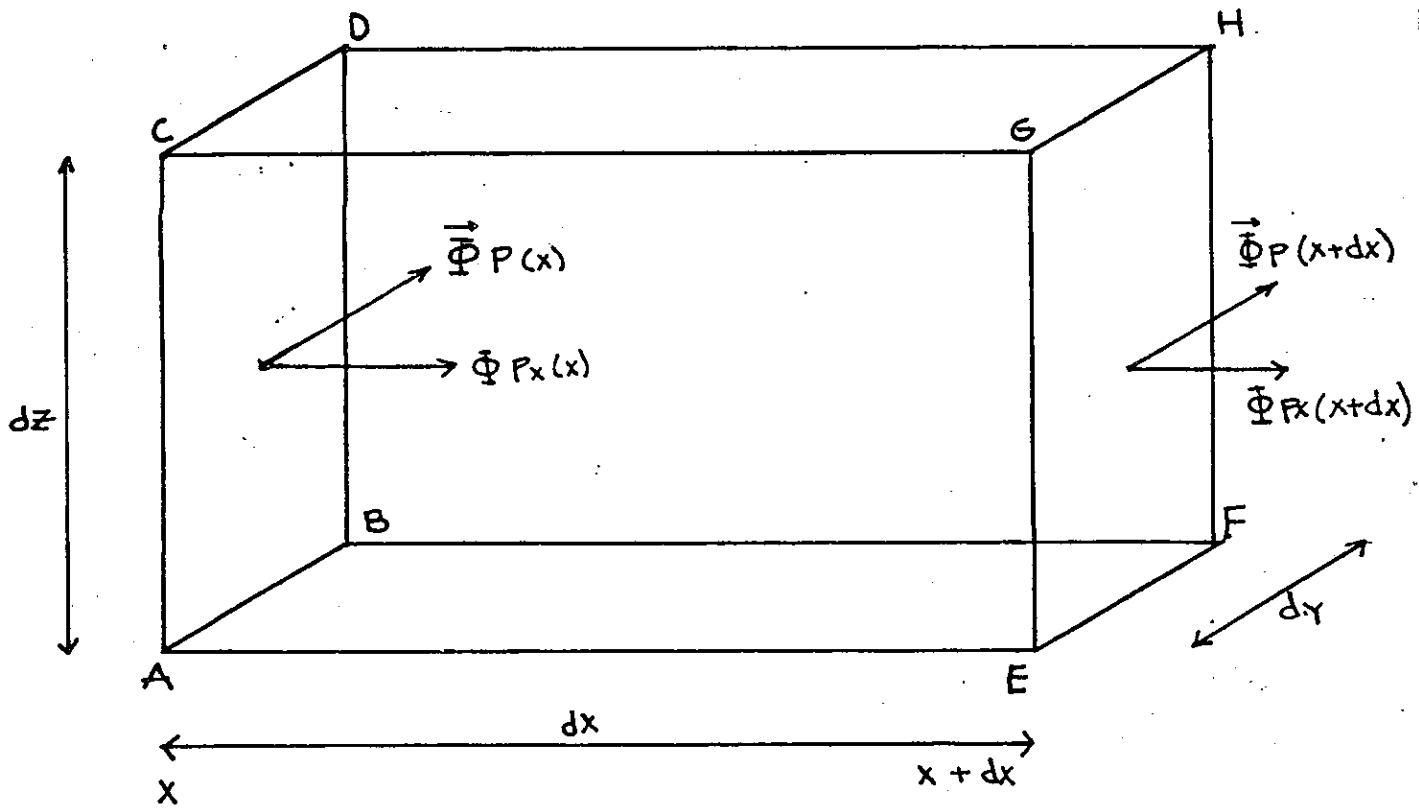


FIGURA F1-9

El incremento neto en el número de agujeros dentro de la región, por unidad de tiempo, originados en una diferencia en las componentes x de $\vec{\Phi}_p$ en las dos caras ABCD y EFGH, es entonces

$$[\Phi_{px}(x) - \Phi_{px}(x+dx)] dy dz = \frac{\partial \Phi_p}{\partial x} dx dy dz$$

Se tendrán términos similares para las componentes en "y" y en "z", con lo cual, el incremento total por unidad de tiempo en el número de agujeros dentro de la región, que se origina debido a estos términos es $-\nabla \cdot \vec{\Phi}_p dx dy dz$. Si los procesos de generación y recombinación son activos dentro del volumen, habrá $G_p dx dy dz$ agujeros generados y $(P/\tau_p) dx dy dz$ que se pierden por recombinación por unidad de tiempo dentro de la región de interés. La suma algebraica de todas estas cantidades representa el incremento neto total del número de agujeros dentro del volumen $dx dy dz$ por unidad de tiempo que es, evidentemente, $(\partial P/\partial t) dx dy dz$. Si se saca la suma algebraica de las diferentes contribuciones, se iguala a $(\partial P/\partial t) dx dy dz$ y se anula el elemento de volumen en ambos lados de la ecuación, se obtiene:

$$\frac{\partial P}{\partial t} = -\nabla \cdot \vec{\Phi}_p + G_p - \frac{P}{\tau_p}$$

mientras que con un cálculo similar para los electrones dentro del mismo elemento de volumen, se obtendrá:

$$\frac{\partial N}{\partial t} = -\nabla \cdot \vec{\Phi}_n + G_n - \frac{N}{\tau_n}$$

Ahora, la densidad de flujo ($\vec{\Phi}$) está relacionada a la densidad de corriente (\vec{J}) por: $\frac{1}{e} \vec{\Phi} = \vec{J}$, entonces:

$$\frac{\partial N}{\partial t} = -\frac{N}{\tau_n} + \frac{1}{e} \nabla \cdot \vec{J}_n + G_n \quad 1-35$$

$$\frac{\partial P}{\partial t} = -\frac{P}{\tau_p} - \frac{1}{e} \nabla \cdot \vec{J}_p + G_p \quad 1-36$$

donde G_n y G_p son (generalmente iguales), las razones de generación de portadores. Para incluir el fenómeno de atrapamiento, necesitamos añadir términos que expresen la pérdida de portadores en los diversos estados trampa, así como también términos adicionales que representen el retorno de portadores anteriormente atrapados. Entonces es necesario tratar con las densidades adicionales de portadores P_t y N_t que se encuentran atrapados.

Combinando las ecuaciones 1-32, 1-33 y 1-35, 1-36 y expandiendo los términos de las divergencias tenemos:

$$\frac{\partial N}{\partial t} = -\frac{N}{\tau_n} + \mu_n (N \nabla \cdot \vec{E} + \vec{E} \cdot \nabla N) + D_n \nabla^2 N + G_n \quad 1-37$$

$$\frac{\partial P}{\partial t} = -\frac{P}{\tau_p} - \mu_p (P \nabla \cdot \vec{E} + \vec{E} \cdot \nabla P) + D_p \nabla^2 P + G_p \quad 1-38$$

El campo eléctrico en el semiconductor \vec{E} , se determina por la densidad neta de carga espacial, a través de la ecuación de Poisson: $\nabla \cdot \vec{E} = \rho/\epsilon$. Si representamos las densidades de electrones y de agujeros atrapados por N_t y P_t respectivamente, tendremos:

$$\nabla \cdot \vec{E} = \frac{e}{\epsilon} [(P-N) + (P_t - N_t) + (D-A)] \quad 1-39$$

La combinación de las ecuaciones 37, 38 y 39 hace posible en principio, la solución de las configuraciones del campo de los elementos semiconductores, a partir de las soluciones conocidas de equilibrio. Entonces las ecuaciones 30 y 31 nos hacen encontrar la corriente de portadores que fluye.

Las relaciones se han vuelto no lineales por los términos producto que incluyen densidad de portadores y campo; para grandes variaciones, inclusive los tiempos de vida son dependientes de las concentraciones de portadores. Esto hace que las soluciones exactas sean tediosas y obtenibles únicamente por análisis numérico. Sin embargo, por lo general se reduce la complejidad introduciendo aproximaciones que simplifican y linealizan las ecuaciones.

Altos campos eléctricos en semiconductores

Como último punto, hablaremos de los fenómenos de altos campos en los semiconductores. En los metales la concentración de electrones es tan grande que durante la conducción, la velocidad de arrastre de los electrones es tan sólo una pequeña fracción de la velocidad térmica. Esto nos explica el alto grado de linealidad óhmica que exhiben los metales. Los semiconductores extrínsecos también pueden tener concentraciones de portadores bastante grandes, de modo que nuevamente se necesita tan sólo un pequeño campo para establecer densidades sustanciales de corriente óhmica. Al igual que en los metales, la iniciación de la conducción no lineal (en materiales homogéneos) ocurrirá para densidades de corriente y potencia tan altas que la disipación de potencia bajo operación continua, es destructiva. Sin embargo en los semiconductores con baja densidad de impurezas habrá tan pocos portadores que podremos aplicar muy altos campos sin riesgo alguno. De aquí que será posible observar conductividad no óhmica así como efectos debidos a la carga espacial de portadores de rompimiento eléctrico.

Primeramente consideremos el flujo de corriente en un aislador o en un semiconductor intrínseco, con portadores introducidos de material extrínseco adyacente o de metal. Por simplicidad supondremos que únicamente estará involucrado un tipo de portadores (electrones). La presencia de la densidad de portadores opuestos depende de los procesos de generación y atrapamiento y de la naturaleza de los contactos en los extremos del material intrínseco. La ecuación 30 que nos representa el transporte de corriente, se reduce a:

$$\vec{J} = \vec{J}_n = eN\mu_n \vec{E} \quad 1-40$$

El término de difusión se puede omitir si los gradientes de concentración son pequeños; en general aumentará el flujo de corriente hasta cierto grado. La ecuación de carga espacial es simplemente:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon} = \frac{eN}{\epsilon} \quad 1-41$$

Ya que se supuso que teníamos donores, aceptores, agujeros o portadores atrapados. Combinando ambas expresiones (en una dimensión) tenemos:

$$\epsilon \frac{dE}{dx} = \frac{J}{\mu_n \epsilon} \quad 1-42$$

En una solución unidimensional, de estado estacionario, la densidad de corriente debe ser independiente de x . Multiplicando por dx e integrando directamente la ecuación, tenemos:

$$E^2 - E_0^2 = \frac{2J}{\mu_n \epsilon} (x - x_0) \quad 1-43$$

Las condiciones de frontera más apropiadas son que en $x_0=0$, $E_0=0$. Esta es la condición correcta si los electrones entran en $x_0=0$ con velocidad cero. Entonces tendremos que el campo está dado por:

$$E = \sqrt{\frac{2Jx}{\mu_n \epsilon}} \quad 1-44$$

Si definimos el potencial en $x_0=0$, tendremos que el potencial en x se encuentra por integración mediante la ecuación $E = -dV/dx$:

$$V(x) = -\frac{2}{3} \sqrt{\frac{2J}{\mu_n \epsilon}} x^{3/2} \quad 1-45$$

Por lo tanto la densidad de corriente a través de un grueso x dado de un semiconductor intrínseco es proporcional al cuadrado del voltaje aplicado, es decir:

$$J = \frac{q}{8} \frac{\mu_n \epsilon}{x^3} V^2 \quad 1-46$$

En todo el análisis precedente hemos supuesto que la movilidad de los electrones es independiente del campo. Sin embargo, cuando los campos son del orden de varios miles de volts por centímetro, esto no seguirá siendo válido.

La movilidad se puede expresar como $\mu = e\bar{\tau}/m$ donde $\bar{\tau}$ es el tiempo medio libre entre colisiones. En los metales, la velocidad adquirida del campo eléctrico entre colisiones ($\vec{u} = \mu \vec{E} = e\bar{\tau} \vec{E}/m$) es pequeña en comparación con la velocidad térmica. En los bordes de las bandas de los semiconductores, las velocidades de los electrones y agu

jeros son cero. Ya que las distribuciones de velocidad son esencialmente distribuciones de Boltzmann, tendremos que la energía promedio de los electrones es $3kT/2$ y la velocidad térmica promedio está dada aproximadamente por la expresión $[m\langle v^2 \rangle / 2] = 3kT/2$. Al igual que en los metales, el tiempo medio libre de los electrones y sus movilidades a bajos campos, se determinan por colisiones de fonones. Se puede justificar un tiempo medio libre constante, basándonos en que si el electrón se está moviendo lentamente, un fonon coincidirá con él aproximadamente cada τ segundos. Para altos campos aplicados, la distancia que el portador viaja en este intervalo constante, se vuelve larga. Inclusive un portador que es conducido al reposo por una colisión previa, viaja una distancia proporcional a:

$$s = at^2/2 = (eE/m) \tau^2, \text{ en un tiempo } \tau. \text{ Para } \tau = 10^{-14} \text{ seg}$$

en un campo de 10^3 volts/cm, ésta distancia es del orden de $50,000 \text{ \AA}$. Con distancias de esta longitud, los portadores deberán chocar más frecuentemente con impurezas fijas o defectos que con fonones. Por lo tanto la descripción para altos campos deberá de ser en términos de una trayectoria libre media l_0 , en vez del tiempo libre medio. Entonces el tiempo libre medio se vuelve dependiente del campo, si hacemos

$$l_0 = \frac{eE}{2m} \tau^2 \quad \therefore \tau = \sqrt{\frac{2ml_0}{eE}} \quad \text{y como } \mu = v/E$$

donde $\mu = \sqrt{2as} = \sqrt{(2eEl_0)/m}$ entonces,

$$\mu = \sqrt{\frac{2eEl_0}{mE}} = \sqrt{\frac{2el_0}{mE}} \quad 1-47$$

Existe un rango de campos en el que la movilidad es aproximadamente proporcional al inverso de la raíz cuadrada del campo, y la velocidad de arrastre de los portadores, a la raíz cuadrada del campo. Tal y como de bemos esperar de la ecuación 47, la reducción en la movilidad es mayor con portadores de masas mayores.

Para campos eléctricos aún mayores, se observa una reducción posterior de la movilidad, culminando en lo que aparenta ser un valor límite de la velocidad de arrastre, cerca de la iniciación del proceso rompimiento. Esto se suele explicar como debido a la interacción de electrones con fonones del rango de frecuencia de la luz visible que a bajas energías de los electrones no juegan un papel importante. En los altos campos que estos procesos tienen lugar, las descripciones estadísticas usuales de los procesos de conducción se vuelven inexactas. Los fonones serán arrastrados por el flujo de electrones. La temperatura efectiva de los electrones se volverá mucho mayor que aquella que posee la red. Se dirá entonces que los portadores están "calientes". La estadística de colisión y recombinación también será bastante diferente de aquella que se utiliza para los portadores "fríos" usuales y ninguno de los análisis simples podrán aplicarse. Estas partículas de alta energía sufrirán colisiones que deflectarán muy ligeramente sus trayectorias; así pues una trayectoria libre media entre colisiones ya no será una medida significativa del avance de un portador.

Los semiconductores intrínsecos y los aisladores experimentan procesos de rompimiento para campos del orden de 10^5 hasta 10^7 volts por cm. Para estos campos los electrones y agujeros en los cristales casi perfectos pueden adquirir, entre colisión y colisión con fonones y defectos energías mayores que E_g , es decir, que la anchura de la banda prohibida. Le es permitido a dicho portador, ionizar directamente a un electrón de la banda de valencia a la banda de conducción, creando así un par electrón-agujero.

2. TEORIA DE JUNTAS

Centraremos ahora la atención en el estudio de sustancias dentro de las cuales la densidad de impureza puede variar de un punto a otro y, en particular, en muestras que contienen una región extrínseca tipo N y una región extrínseca tipo P separadas por una zona de transición relativamente angosta, denominada unión P-N.

La transición entre las regiones tipo P y tipo N puede ser abrupta, en cuyo caso se encuentra una región, que contiene una región neta más o menos constante de impurezas donadoras junto a otra que posee una concentración neta más o menos constante de impurezasceptoras. La unión puede ser también graduada en donde las concentraciones de impurezas son función de la distancia a la unión. Ver figura F2-1.

Daremos en este estudio más énfasis a la unión abrupta puesto que ilustra la mayoría de las propiedades importantes relacionadas con el comportamiento de las uniones P-N en general.

Supóngase que formamos una unión P-N juntando una muestra uniforme tipo P con otra tipo N para formar un sólo cristal. En el instante de formación existe una concentración uniforme N_{no} de electrones libres móviles y P_{no} de agujeros libres móviles en el lado N, extendiéndose hasta la unión, y en el lado P, una concentración uniforme P_{po} de agujeros móviles y N_{po} de electrones móviles que se extiende también hasta la unión. En ambos lados, las densidades de electrones y agujeros satisfacen:

$$N_{no} P_{no} = P_{po} N_{po} = N_i^2 \quad 2-1$$

Puesto que la concentración N_{no} de electrones del lado N es mucho mayor que la concentración de electrones N_{po} del lado P, en el instante de formación existe un gradiente enorme en la concentración de electrones en la unión entre ambas regiones. La misma situación existe con respecto a la concentración de agujeros en la misma unión. Los grandes gradientes de concentración iniciales establecen corrientes de difusión que hacen que los electrones de la región N y los agujeros de la región P fluyan descendiendo por los gradientes de concentración respectivos hasta la región de conductividad de tipo opuesto y dejan a la región cercana a la unión, vacía de portadores mayoritarios. No obstante este flujo de difusión inicial no puede continuar indefinidamente debido a que en las regiones cercanas a la unión hay deficiencia de portadores mayoritarios, las cargas de los iones fijos donadores y aceptores cercanos a la unión ya no están balanceadas por las cargas de los portadores libres móviles que estaban allí inicialmente, de modo que se establece un campo eléctrico. La dirección de este campo eléctrico es tal que se opone al flujo de los electrones que salen de la región N y al flujo de los agujeros que salen de la región P, y la magnitud de campo se desarrolla hasta el punto en que su efecto contrarresta exactamente la tendencia de los portadores mayoritarios a difundirse, descendiendo por la "pendiente" de concentración hacia la región de conductividad opuesta.

Entonces se establece una condición de equilibrio dinámico en la que la región cercana a la unión queda vacía de portadores mayoritarios y

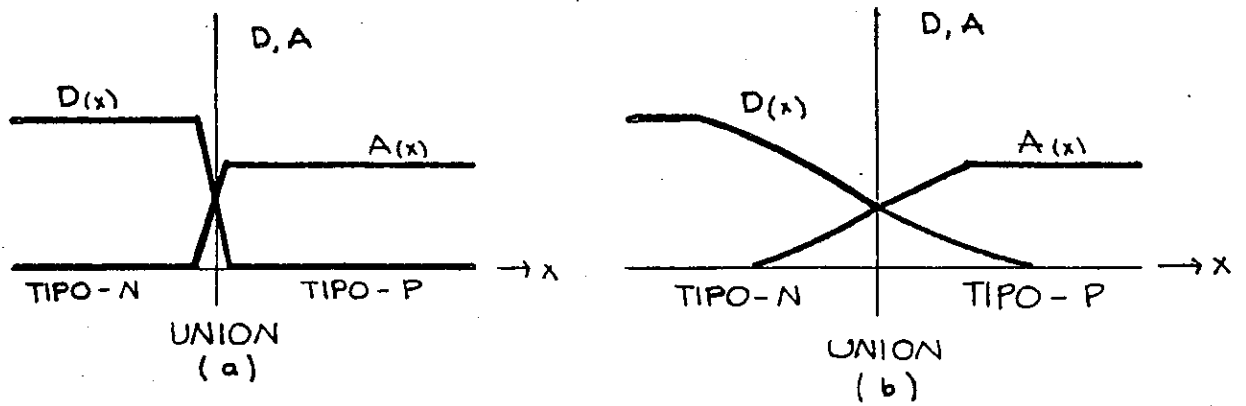


FIGURA F2-1: concentraciones de átomos de impureza en la cercanía de (a) una unión P-N abrupta y (b) una unión P-N gradual.

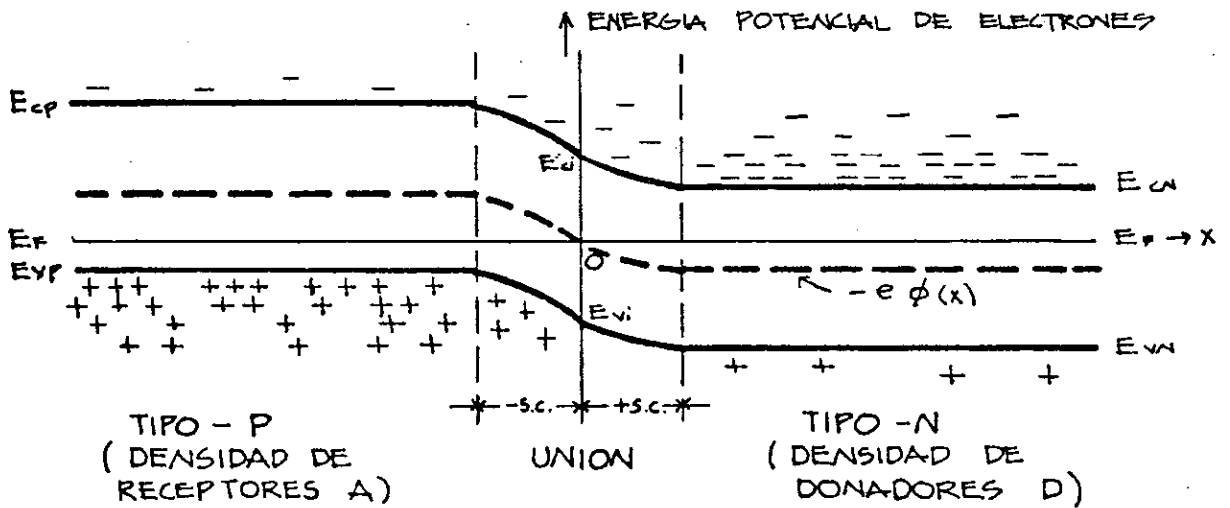


FIGURA F2-2: Diagrama de la energía potencial en el que se muestra la configuración de bandas de energía cerca de la unión P-N en ausencia de un voltaje aplicado.

en la que se forman fuertes capas de carga espacial que contienen campos eléctricos altos cerca de la unión. La configuración de carga espacial es una capa dipolar eléctrica con iones donadores no compensados en el lado N y iones aceptores no compensados en el lado P, proporcionando los componentes de carga positiva y negativa respectivamente.

Potencial de contacto interno de equilibrio.

De la fórmula 1-10 podemos escribir:

$$N_{n0} = N_c e^{-(E_{cn} - E_f)/kT} \quad 2-2$$

para la densidad de electrones en el lado N y, para la densidad de electrones en el lado P,

$$N_{p0} = N_c e^{-(E_{cp} - E_f)/kT} \quad 2-3$$

De estas dos ecuaciones, se puede despejar E_{cn} y E_{cp} , cuya diferencia es la energía potencial interna,

$$e\phi_0 = E_{cp} - E_{cn} = kT \ln \frac{N_{n0}}{N_{p0}} \quad 2-4$$

Haciendo uso de 2-1 podemos expresar el potencial de contacto interno de 2-4 como,

$$\phi_0 = \frac{kT}{e} \ln \frac{N_{n0} P_{p0}}{N_i^2} \quad 2-5$$

A continuación se examinará lo que sucede cuando se aplica un voltaje externo a la muestra que contiene una unión P-N. Es obvio, que puesto que las regiones de carga espacial a ambos lados de la unión tienen una deficiencia de portadores, estas regiones poseen una resistividad mucho mayor que cualquier otra parte del cristal. Esto significa que cuando se aplica una fuente de voltaje externo a la muestra en que se tiene una región P-N, casi toda la caída de voltaje se producirá en estas regiones.

Si la región P-N se usa como rectificador, se obtiene una condición de baja impedancia cuando la región N se conecta a la terminal negativa y la región P a la positiva de la fuente externa de voltaje. Esta polaridad se conoce como estado de polarización directa. Se acostumbra considerar que el signo del potencial externamente aplicado en este caso es positivo. Cuando se aplica un voltaje externo en esta forma, el efecto es reducir la altura de la barrera de potencial. Si se invierte la polaridad de la fuente de potencial externo (en cuyo caso su signo es considerado negativo), la altura de la barrera de potencial interno se incrementará (ver figura F2-3). En este caso, se dice que el voltaje aplicado es una polarización inversa.

Junta abrupta

Aunque una junta abrupta es realmente una mera idealización matemática, es sin embargo una aproximación razonable a la más abrupta de las juntas físicas.

La densidad de carga sin considerar atrapamiento (ver ecuación 1-39),

viene dada por, $\rho(x) = e(P - N + D - A)$ 2-6
 donde P y N son concentraciones de agujeros y electrones y D y A son las concentraciones de los iones de impureza. En la ecuación de Poisson,

$$\frac{d^2 V_i}{dx^2} = -\frac{e}{\epsilon} (P - N + D - A) \quad 2-7$$

Las densidades D y A se conocen como función de la posición, mientras que P y N dependen del potencial interno V_i y deben expresarse en función de ese potencial, antes de resolver 2-7. De la ecuación 1-10, tenemos,

$$N = N_c e^{-(E_c - E_f)/kT} \quad 2-8$$

Tomando en cuenta que $E_c = E_v + E_g$ y multiplicando por $e^{E_g/2kT}$,

$$N = N_c e^{-E_g/2kT} e^{(E_f - E_v - E_g/2)/kT} \quad 2-9$$

La cantidad $N_c e^{-E_g/2kT}$ se puede conocer como la concentración intrínseca de electrones, N_i , ya que cuando $E_f = E_v + E_g/2$, es decir, cuando el nivel de Fermi está en el centro de la banda prohibida, el segundo exponencial de la ecuación 2-9 se convierte en la unidad y las concentraciones de electrones y agujeros son iguales entre sí e iguales a N_i . Además, el cero de potencial interno se puede seleccionar de tal forma que,

$$eV_i = E_f - E_v - E_g/2 \quad 2-10$$

En otras palabras, se hace que el potencial interno sea cero, en donde el nivel de Fermi queda en la mitad de la banda prohibida o en donde el material es intrínseco. La ecuación 2-9 se transforma, ahora, en

$$N = N_i e^{eV_i/kT} \quad 2-11$$

que es la relación deseada entre N y V_i .

La forma más sencilla de expresar P en función de V_i , es usar el producto NP en la forma

$$NP = N_i^2 \quad 2-12$$

Al sustituir N de ésta ecuación en 2-11 y despejar P, se tiene que

$$P = N_i e^{-eV_i/kT} \quad 2-13$$

Sustituyendo 2-11 y 2-13 en la ecuación de Poisson 2-7 se ve que

$$\frac{d^2 V_i}{dx^2} = \frac{e}{\epsilon} N_i \left(e^{eV_i/kT} - e^{-eV_i/kT} + \frac{A-D}{N_i} \right) \quad 2-14$$

Podemos reacomodar esta ecuación, definiendo una nueva variable dependiente adimensional

$$y \equiv \frac{eV_i}{kT} \quad 2-15$$

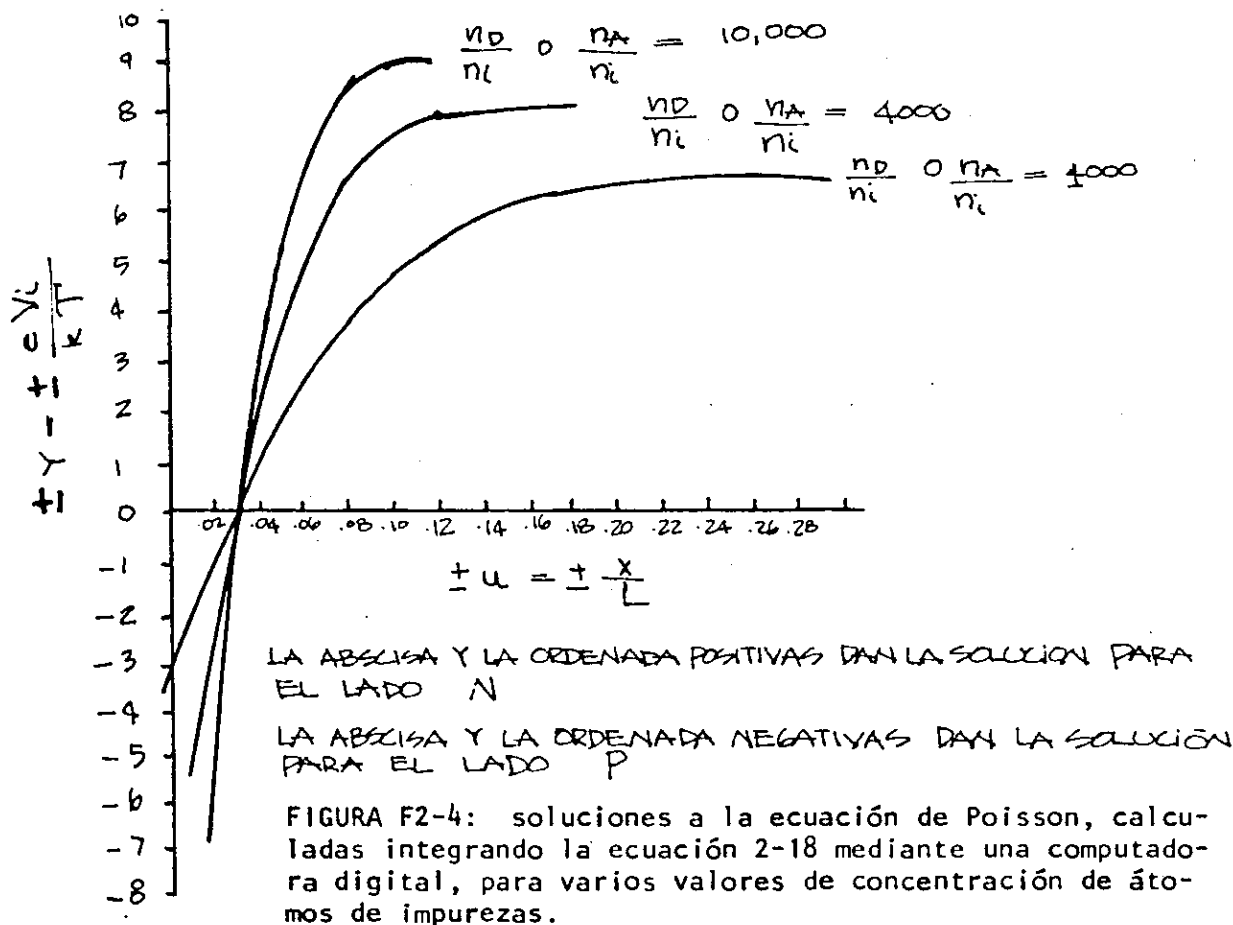
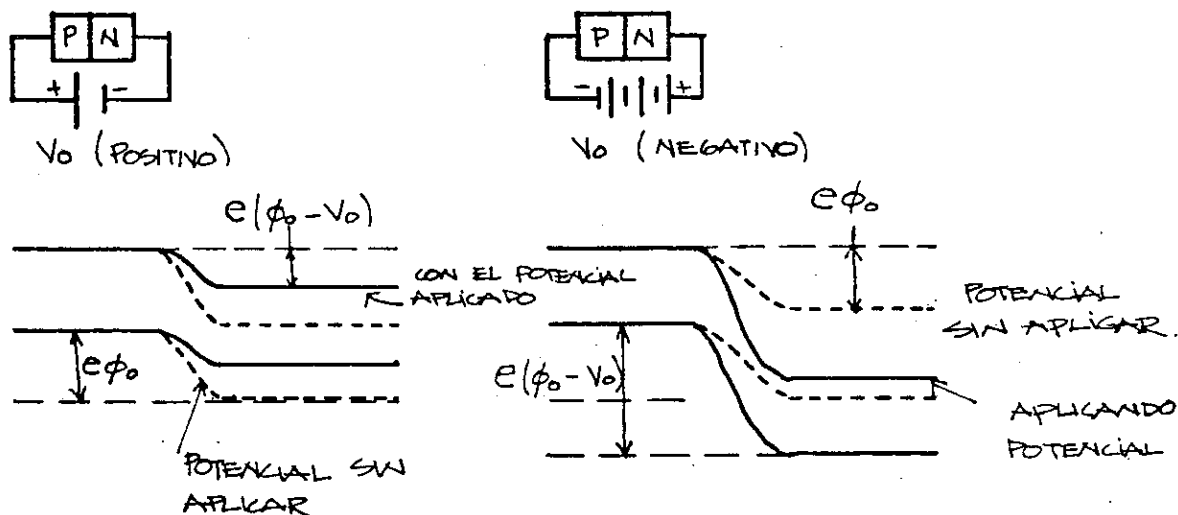
y una nueva variable independiente adimensional,

$$u = \frac{x}{L} \quad 2-16$$

en donde,

$$L = \sqrt{\epsilon kT / e^2 N_i} \quad 2-17$$

FIGURA F2-3: diagrama de la energía potencial que presenta el efecto de (a) un voltaje de polarización directa y (b) de polarización inversa en la configuración de bandas de energía en una unión P-N.



La cantidad L se conoce como la longitud intrínseca de Debye. Sustituyendo 2-15, 2-16 y 2-17 en 2-14,

$$\frac{dy}{du} = \frac{\partial y}{\partial V_i} \frac{\partial V_i}{\partial x} \frac{\partial x}{\partial u} \quad \text{pero} \quad \frac{\partial y}{\partial V_i} = \frac{e}{kT} \quad \text{y} \quad \frac{\partial x}{\partial u} = L$$

entonces
$$\frac{\partial y}{\partial u} = \frac{eL}{kT} \frac{\partial V_i}{\partial x} \quad \text{sacando la segunda derivada,}$$

$$\frac{\partial^2 y}{\partial u^2} = \frac{eL}{kT} \frac{\partial}{\partial u} \left(\frac{\partial V_i}{\partial x} \right) = \frac{eL}{kT} \frac{\partial}{\partial x} \left(\frac{\partial V_i}{\partial x} \right) \frac{\partial x}{\partial u}$$

$$= \frac{eL}{kT} \frac{\partial^2 V_i}{\partial x^2} \frac{\partial x}{\partial u} = \frac{eL^2}{kT} \frac{\partial^2 V_i}{\partial x^2}$$

$$= \frac{eL^2}{kT} \frac{eN_i}{\epsilon} \left(e^{eV_i/kT} - e^{-eV_i/kT} + \frac{A-D}{N_i} \right)$$

$$= \frac{eL^2}{kT} \frac{eN_i}{\epsilon} \left(e^y - e^{-y} + \frac{A-D}{N_i} \right) = \frac{e}{kT} \frac{\epsilon kT}{e^2 N_i} \frac{eN_i}{\epsilon} \left(e^y - e^{-y} + \frac{A-D}{N_i} \right)$$

por lo tanto,

$$\frac{d^2 y}{du^2} = e^y - e^{-y} + \frac{A-D}{N_i} \quad 2-18$$

Esta es una ecuación diferencial no lineal y no se puede resolver en forma cerrada.

Las condiciones de frontera, en esta ecuación, son que el campo eléctrico es cero lejos de la unión, ya que si la barra es lo suficientemente larga, la unión no puede tener ningún efecto en sus extremos. En el extremo del lado N de la unión, el término e^y , que es proporcional a la concentración de electrones, dominará al término e^{-y} , que es proporcional a la concentración de agujeros. Además, en el lado N, $A=0$ y, puesto que el campo eléctrico interno es cero lejos de la unión, $d^2 y/du^2$, que es proporcional a la razón de cambio del campo eléctrico interno con la distancia, debe ser cero. Por lo tanto, cuando u es grande y positiva,

$$e^y = \frac{D}{N_i} \quad \text{ó} \quad y = \ln \frac{D}{N_i} \quad 2-19$$

Es posible seguir un razonamiento semejante y demostrar que, lejos de la unión, en el lado P, es decir, para una u grande negativa,

$$e^{-y} = \frac{A}{N_i} \quad \text{ó} \quad y = -\ln \frac{A}{N_i} \quad 2-20$$

Usando la condición de frontera 2-19 la ecuación diferencial 2-18 se puede resolver por métodos numéricos para la región N a diferentes valores de D/N_i . Las soluciones de 2-18 para el lado P de la unión tienen la misma forma que las del lado N excepto para un cambio de signo de y y de U . Ver figura F2-4.

Teoría del rectificador de unión P-N

Ahora trataremos el flujo de corriente asociado con la unión en condiciones de desequilibrio y la característica de corriente-voltaje de un simple rectificador de unión P-N.

El modelo de unión que se adoptará se ilustra en estado de equilibrio en la figura F2-5. Se supone que las regiones tipo P y N son homogéneas y se extienden al infinito en las direcciones $+x$ y $-x$. En el estado de equilibrio habrá cierto número de electrones presentes como portadores mayoritarios en el lado N, que tienen suficiente energía para saltar la barrera de potencial y difundirse en el lado P de la unión. En la región P, estos electrones constituyen portadores minoritarios y pueden, después de cierto tiempo, desaparecer mediante su recombinación con agujeros. Por tanto, en equilibrio se tiene una corriente de electrones en la región N, pasando sobre la unión, hasta la región P, en donde se pierden por recombinación. Este flujo electrónico de equilibrio se citará como flujo de recombinación J_{nr} . Sin embargo, es obvio que en la condición de equilibrio no puede haber un flujo neto de corriente; es más, de acuerdo con el principio de balance detallado, en estado de equilibrio, cualquier proceso microscópico de transporte y su inversa deben proceder a la misma velocidad. La inversa, en este proceso en particular es la generación de pares electrón-agujero en el lado P de la unión, y la difusión subsecuente de electrones a través de la unión hacia el lado N, en donde se convierten en portadores mayoritarios. Este proceso inverso conduce a una corriente electrónica que fluye de la región P a la región N y que se denominará corriente de generación. En estado de equilibrio, el flujo de partículas resultante J_{ng} debe ser exactamente igual a $-J_{nr}$. Estas mismas observaciones se aplican a los agujeros por igual; existe un flujo de recombinación $-J_{pr}$ que se origina cuando los agujeros de la región P se desplazan sobre la barrera hasta la región N, en donde se recombinan con electrones. Esto va acompañado con un flujo de generación J_{pg} que se crea mediante la generación térmica de pares en la región N, seguida de la difusión de agujeros térmicamente generados a través de la unión. En estado de equilibrio se debe tener $J_{pg} = -J_{pr}$. Por tanto, en estado de equilibrio los flujos totales de agujeros y electrones a través de la unión son igual a cero.

La base para la acción de una unión P-N como rectificador es como sigue. En condiciones de polarización inversa, la altura de la barrera

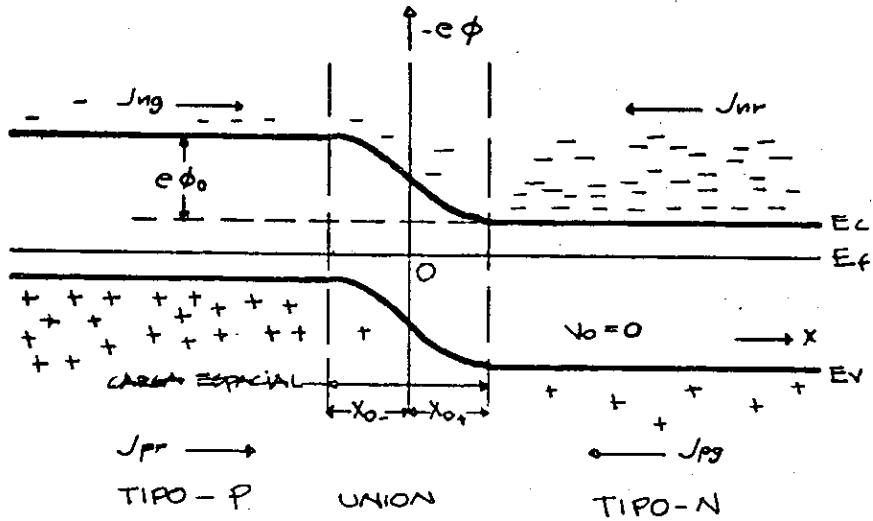


FIGURA F2-5: diagrama de potencial de una unión P-N en condición de equilibrio, en donde se muestran las distribuciones de portadores de carga y los flujos de generación y recombinación.

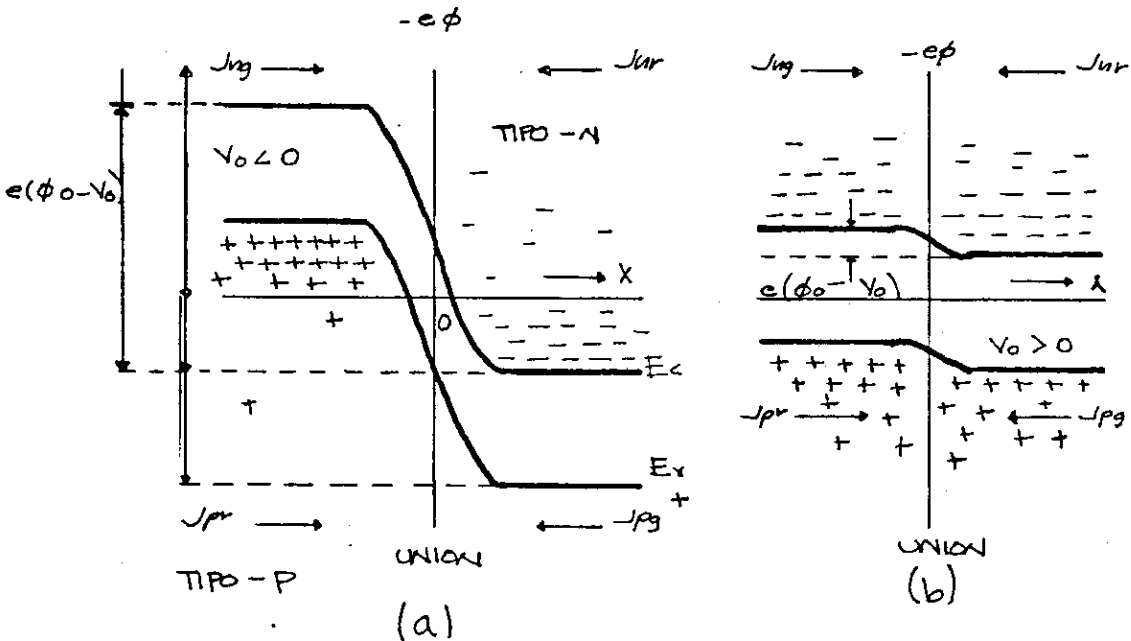


FIGURA F2-6: diagrama de potencial de una unión P-N (a) con polarización inversa, (b) con polarización directa, en donde se señala cómo se modifican los flujos de generación y recombinación.

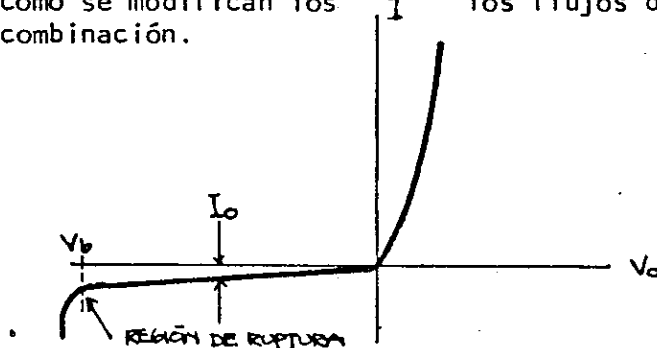


FIGURA F2-7: relación corriente-voltaje de una unión P-N, con la región de ruptura inversa.

de potencial en la unión aumenta en una cantidad igual a $-eV_0$, en donde V_0 es el voltaje aplicado. En estas condiciones es difícil para los portadores vencer la barrera por difusión y, por ende, J_{nr} y J_{pr} se hacen muy pequeños; sin embargo, los flujos J_{ng} y J_{pg} dependen sólo de la velocidad de generación térmica de pares electrón-agujero en las regiones volumétricas respectivas y, en consecuencia, no se ven reducidos. Por tanto, conforme se incrementa el voltaje de polarización, la densidad de corriente a través de la unión tiende a un valor pequeño constante $-e(J_{pg} - J_{ng})$, denominado corriente de saturación, que está limitada por el número de pares térmicamente generados que se crean en las regiones volumétricas cerca de la unión. Esta corriente depende sólo de parámetros del material y la temperatura, y no puede aumentar notablemente, no importa el alto valor que puede adquirir el voltaje. Esta situación física se ilustra en la figura F2-6(a).

Cuando se aplica un voltaje de polarización directo, la altura de la barrera de potencial se reduce y es muy fácil para los electrones mayoritarios del lado N y los agujeros mayoritarios del lado P, difundirse sobre la barrera hacia los lados opuestos respectivos en donde pueden convertirse en portadores minoritarios y, finalmente, recombinarse. En este caso, las corrientes de recombinación J_{pr} y J_{nr} pueden hacerse muy grandes en comparación con sus valores de equilibrio. No obstante, las corrientes de generación J_{ng} y J_{pg} permanecen iguales, ya que están limitadas por la velocidad de generación térmica. La estructura es un gran flujo de corriente como se ilustra en la figura F2-6(b). La estructura de la unión presenta una impedancia muy elevada en la dirección de polarización inversa y una impedancia muy pequeña en la dirección de polarización directa, por lo que se comporta como un rectificador. En la figura F2-7 se muestra la relación corriente-voltaje. En la condición de polarización inversa se originan flujos de corriente de saturación consistentes en portadores minoritarios generados de un modo térmico en las regiones volumétricas que quedan fuera, pero son adyacentes, a las capas de carga espacial de la unión y que la barrera de potencial de la unión se encarga de recoger. En la condición de polarización directa, la corriente se compone en su mayor parte de portadores mayoritarios que se difunden sobre la barrera reducida para convertirse en portadores minoritarios en la región de conductividad opuesta, al otro lado de la unión; en efecto la unión "inyecta" un gran número de portadores minoritarios hacia las regiones volumétricas adjuntas a ambos lados.

Puesto que el número de electrones que poseen la energía suficiente para sobrepasar la barrera de potencial de la unión es proporcional a

$$e^{-e(\phi_0 - V_0)/kT} = \text{cte. } e^{eV_0/kT}$$

es de esperarse que J_{nr} y J_{pr} sean proporcionales a $e^{eV_0/kT}$. Sin embargo, para $V_0 = 0$, $J_{pr} = J_{pg}$ y $J_{nr} = J_{ng}$ de donde, se tiene que

$$J_{pr} = J_{pg} e^{eV_0/kT} \quad 2-21$$

$$\text{y } J_{nr} = J_{ng} e^{eV_0/kT} \quad 2-22$$

entonces de acuerdo con la figura F2-6

$$J_p = J_{pr} - J_{pg} = J_{pg}(e^{eV_0/kT} - 1) \quad 2-23$$

$$\text{y } J_n = J_{ng} - J_{nr} = -J_{ng}(e^{eV_0/kT} - 1) \quad 2-24$$

La relación entre el voltaje aplicado y la densidad de corriente eléctrica I estará dada por,

$$I = e(J_p - J_n) = e(J_{p0} + J_{n0})(e^{eV_0/kT} - 1) = I_0(e^{eV_0/kT} - 1) \quad 2-25$$

En donde la densidad de corriente de saturación I_0 es la suma de las densidades de corriente de generación

$$I_0 = e(J_{p0} + J_{n0}) \quad 2-26$$

La relación que existe entre la corriente y el voltaje que aparece en la figura F2-7 es una gráfica de la ecuación 2-25 antes mencionada. A continuación, se procederá a sacar el resultado 2-25 por medio de un método de cálculo más riguroso que cuenta con dos ventajas: la primera, que proporciona una expresión más explícita de la corriente de saturación y, segunda, que forma la base para la técnica de cálculo que se puede aplicar a dispositivos de unión más complicados.

Relación corriente-voltaje en una junta abrupta

El procedimiento que se seguirá consiste en resolver la ecuación de continuidad para los portadores en exceso, tanto en la región N , como en la P , y aplicar las condiciones apropiadas de frontera a la unión y a todas las demás partes del sistema. Con este fin se supondrá que la unión se extiende esencialmente al infinito en las direcciones "y" y "z", de tal manera que sólo se requiere tomar en cuenta las variaciones de concentración de portadores y el flujo de corriente a lo largo de la dirección "x". También se supondrá que se ha llegado a un estado estacionario y que todo el voltaje aplicado al sistema aparece en la región de carga espacial asociada con la unión. Las caídas de voltaje y los campos eléctricos en el material que se encuentra fuera de la región de la unión son, por tanto, despreciables y el transporte en dichas regiones es puramente difusivo. En estas condiciones, las ecuaciones de continuidad para los electrones de la región N y los agujeros de la región P , tienen la forma

$$\frac{d^2(N_p - N_{p0})}{dx^2} - \frac{N_p(x) - N_{p0}}{L_n^2} = 0 \quad (\text{región } P, x < x_{0-}) \quad 2-27$$

y

$$\frac{d^2(P_n - P_{n0})}{dx^2} - \frac{P_n(x) - P_{n0}}{L_p^2} = 0 \quad (\text{región } N, x > x_{0+}) \quad 2-28$$

$$\text{en donde, } L_n = \sqrt{D_n \tau_n} \quad \text{y} \quad L_p = \sqrt{D_p \tau_p} \quad 2-29$$

Las condiciones de frontera que van a usarse indican lejos de la unión que la densidad de portadores minoritarios excedentes deben tender a cero, de donde,

$$N_p - N_{p0} = 0 \quad \text{para } x = -\infty \quad 2-30$$

y

$$P_n - P_{n0} = 0 \quad \text{para } x = +\infty \quad 2-31$$

se debe de cumplir además

$$\frac{N_p(-x_{0-})}{N_{p0}} = e^{-e(\phi_0 - V_0)/kT} \quad 2-32$$

y

$$\frac{P_n(x_{0+})}{P_{p0}} = e^{-e(\phi_0 - V_0)/kT} \quad 2-33$$

Las soluciones generales de las ecuaciones de continuidad 2-27 y 2-28 se pueden expresar como,

$$N_p - N_{p0} = A e^{x/L_n} + B e^{-x/L_n} \quad (\text{región P}) \quad 2-34$$

y

$$P_n - P_{n0} = C e^{x/L_p} + D e^{-x/L_p} \quad (\text{región N}) \quad 2-35$$

en donde A, B, C y D son constantes arbitrarias. Se puede ver claramente que, para satisfacer las condiciones de frontera 2-30 y 2-31, es necesario determinar que $B = C = 0$. Así también, puesto que cuando el voltaje aplicado V_0 es cero,

$$\frac{N_p(-x_{0-})}{N_{n0}} = e^{-e(\phi_0 - V_0)/kT} = e^{-e\phi_0/kT}$$

pero por ecuación 2-4, $\frac{e\phi_0}{kT} = \ln \frac{N_{n0}}{N_{p0}}$, entonces,

$$\frac{N_p(-x_{0-})}{N_{n0}} = \frac{N_{p0}}{N_{n0}} \quad \therefore N_p(-x_{0-}) = N_{p0} \quad 2-36$$

similarmente,

$$\frac{P_n(x_{0+})}{P_{p0}} = e^{-e(\phi_0 - V_0)/kT} = e^{-e\phi_0/kT} = e^{-\ln \frac{N_{n0}}{N_{p0}}}$$

pero por 2-1,

$$N_{n0} P_{n0} = P_{p0} N_{p0} = N_i^2 \Rightarrow N_{n0} P_{p0} = N_i^2 P_{p0} / P_{n0}$$

entonces,

$$\frac{P_n(x_{0+})}{P_{p0}} = e^{-\ln \frac{N_{n0}}{N_{p0}}} = e^{-\ln \frac{N_{n0} P_{p0}}{N_i^2}} = e^{-\ln \frac{P_{p0}}{P_{n0}}} = \frac{P_{n0}}{P_{p0}}$$

por lo tanto,

$$P_n(x_{0+}) = P_{n0} \quad 2-37$$

entonces, podemos expresar

$$e^{-e\phi_0/kT} = \frac{N_{p0}}{N_{n0}} = \frac{P_{n0}}{P_{p0}} \quad 2-38$$

Esto permite expresar las condiciones de frontera 2-32 y 2-33 como,

$$\frac{N_p(-x_{0-})}{N_{n0}} = \frac{N_{p0}}{N_{n0}} = e^{-e\phi_0/kT}$$

y entonces

$$N_p(-x_{0-}) = N_{n0} e^{-e\phi_0/kT} e^{eV_0/kT} = N_{n0} \frac{N_{p0}}{N_{n0}} e^{eV_0/kT}$$

por lo tanto,

$$N_p(-x_{0-}) = N_{p0} e^{eV_0/kT} \quad 2-39$$

similarmente,

$$P_n(x_{0+}) = P_{p0} e^{eV_0/kT} \quad 2-40$$

Si se hace que $B = C = 0$ en 2-34 y 2-35 y se aplican 2-39 y 2-40,

$$N_p - N_{p0} = A e^{x/L_N} \Rightarrow A = (N_p - N_{p0}) e^{-x/L_N} \Rightarrow$$

$$A = [N_{p0} e^{eV_0/kT} - N_{p0}] e^{-(-x_0-)/L_N} = N_{p0} (e^{eV_0/kT} - 1) e^{x_0-/L_N}$$

por lo tanto,

$$N_p - N_{p0} = N_{p0} (e^{eV_0/kT} - 1) e^{x_0-/L_N} e^{x/L_N}$$

$$N_p - N_{p0} = N_{p0} (e^{eV_0/kT} - 1) e^{(x+x_0-)/L_N} \quad 2-41$$

y similarmente,

$$P_n - P_{n0} = P_{n0} (e^{eV_0/kT} - 1) e^{-(x-x_0+)/L_P} \quad 2-42$$

Estas ecuaciones permiten ver que para una polarización inversa ($V_0 < 0$), la concentración de electrones en la región P y la concentración de agujeros en la región N son menores que el valor de equilibrio; en este caso la concentración de portadores minoritarios en exceso es negativa. Se puede afirmar lo contrario cuando el voltaje de polarización aplicado tiene un sentido directo ($V_0 > 0$). Las ecuaciones 2-41 y 2-42 se pueden emplear para encontrar las corrientes de agujeros y de electrones que fluyen dentro de las regiones de carga espacial desde el lado N y el lado P del cristal, respectivamente.

La corriente de unión se puede evaluar determinando las corrientes de difusión de los portadores minoritarios en las fronteras de la región de carga espacial, lo cual da,

$$J_n(0) \cong J_n(-x_0-) = -D_n \left(\frac{dN_p}{dx} \right)_{-x_0-} = -\frac{N_{p0} D_n}{L_N} (e^{eV_0/kT} - 1) e^{(x+x_0-)/L_N} \Big|_{x=-x_0-}$$

por lo tanto,

$$J_n(0) \cong -\frac{N_{p0} D_n}{L_N} (e^{eV_0/kT} - 1) \quad 2-43$$

y,

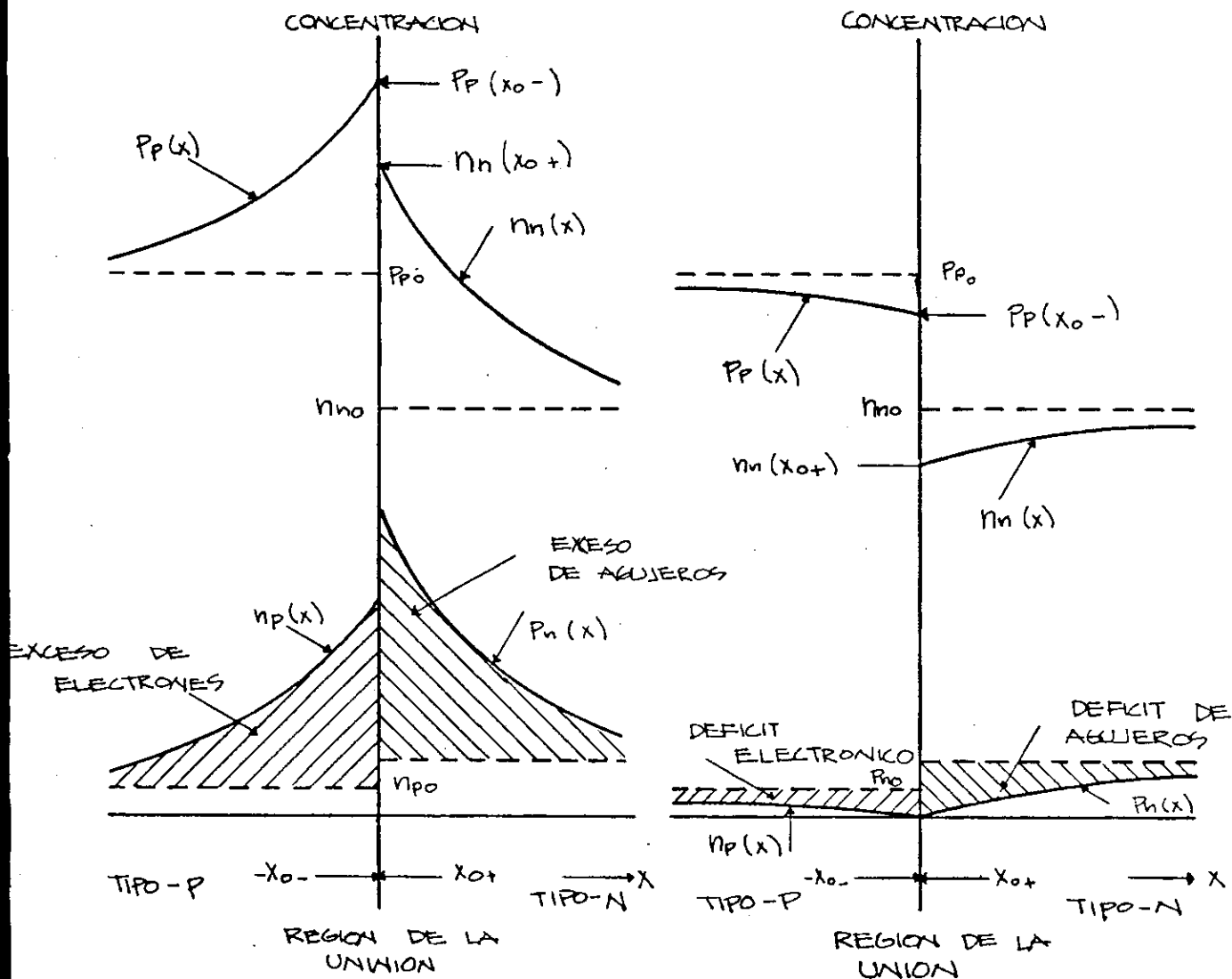
$$J_p(0) \cong J_p(x_0+) = -D_p \left(\frac{dP_n}{dx} \right)_{x_0+} = \frac{P_{n0} D_p}{L_P} (e^{eV_0/kT} - 1) \quad 2-44$$

la corriente eléctrica que cruza la unión es entonces,

$$I = e [J_p(0) - J_n(0)] = e \left(\frac{N_{p0} D_n}{L_N} + \frac{P_{n0} D_p}{L_P} \right) (e^{eV_0/kT} - 1) \quad 2-45$$

Este resultado tiene la misma forma que el que se obtuvo anteriormente como 2-25; ahora la corriente de saturación es

$$I_0 = e (J_{n0} + J_{p0}) = e \left(\frac{N_{p0} D_n}{L_N} + \frac{P_{n0} D_p}{L_P} \right) \quad 2-46$$



(a) POLARIZACIÓN DIRECTA.

(b) POLARIZACIÓN INVERSA.

FIGURA F2-8: concentraciones de electrones y agujeros en la cercanía de una unión P-N (a) con polarización directa y (b) con polarización inversa.

Las concentraciones reales de agujeros y electrones a ambos lados de la unión, tanto para la polarización directa como para la inversa se muestran en la figura F2-8 para un rectificador típico de unión P-N del tipo que se acaba de describir. De acuerdo con la ecuación anterior, se encuentra que la corriente de saturación incluye las densidades de equilibrio de portadores minoritarios N_{p0} y P_{n0} , los coeficientes de difusión D_n y D_p , así como los tiempos de vida de los portadores en exceso a través de las longitudes de difusión L_n y L_p . Esto se puede expresar en función de las densidades de portadores mayoritarios, aplicando la relación 2-1, en donde 2-46 se puede expresar como,

$$I_0 = eN_i^2 \left(\frac{D_n}{P_{p0} L_n} + \frac{D_p}{N_{n0} L_p} \right) \quad 2-47$$

Las fracciones de la corriente total de unión portada por electrones y agujeros se menciona a veces como las eficiencias de inyección de electrones y agujeros, respectivamente. De acuerdo con 2-43 y 2-44, es obvio que estas cantidades están dadas por

$$\eta_n = \frac{-eJ_n(0)}{e[J_p(0) - J_n(0)]} = \frac{\frac{eN_{p0}D_n}{L_n} (e^{eV_0/kT} - 1)}{e \left(\frac{N_{p0}D_n}{L_n} + \frac{P_{n0}D_p}{L_p} \right) (e^{eV_0/kT} - 1)} = \frac{1}{1 + \frac{L_n}{L_p} \frac{P_{n0}}{N_{p0}} \frac{D_p}{D_n}}$$

pero $N_{p0} = N_{n0} P_{p0} / P_{n0} \Rightarrow$

$$\eta_n = \frac{1}{1 + \frac{L_n}{L_p} \frac{P_{p0}}{N_{n0}}} \quad 2-48$$

y similarmente

$$\eta_p = \frac{eJ_p(0)}{e[J_p(0) - J_n(0)]} = \frac{1}{1 + \frac{bL_p}{L_n} \frac{N_{n0}}{P_{p0}}} \quad 2-49$$

donde b es la razón de las movilidades,

$$b = \frac{\mu_n}{\mu_p} = \frac{eD_n/kT}{eD_p/kT} = \frac{D_n}{D_p} \quad 2-50$$

Transistores de unión

Una estructura semiconductor que tiene un electrodo de emisor para inyectar portadores minoritarios excedentes al cristal semiconductor y un electrodo colector para recuperarlos, se conoce con el nombre de transistor y se puede fabricar para que amplifique corrientes o señales de voltaje, se usa como oscilador o dispositivo de conmutación.

En la figura F2-9 se muestra un diagrama simplificado de este dispositivo.

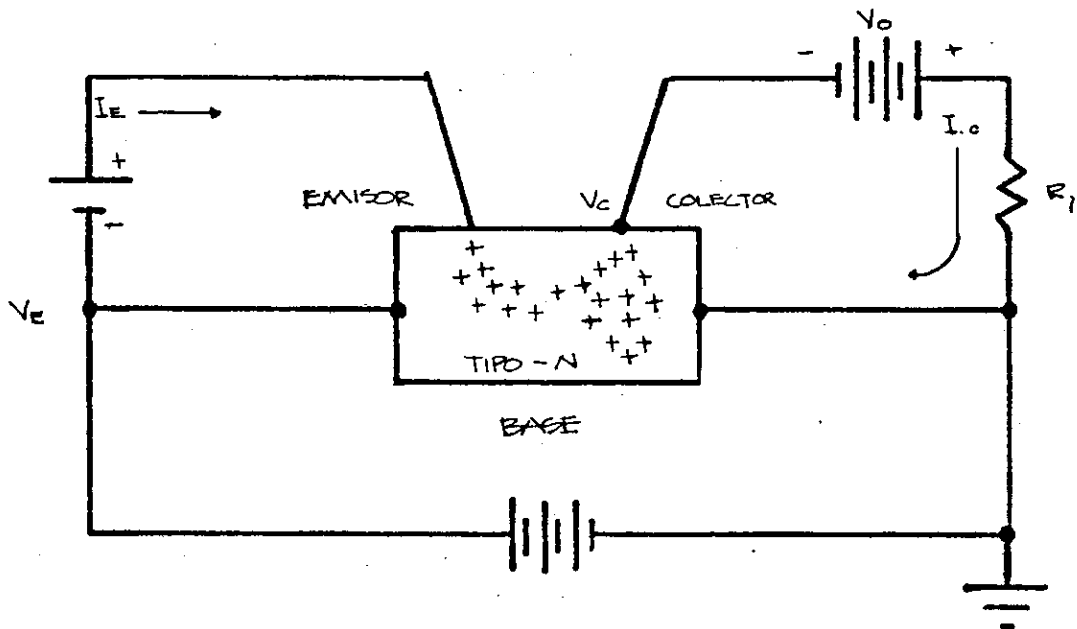


FIGURA F2-9: diagrama esquemático de un transistor de puntas de contacto y su circuito externo.

Supóngase que se hace fluir una corriente uniforme de emisor I_E en sentido directo desde el electrodo emisor de este diagrama, por medio de un voltaje de polarización directa V_E . Se inyectan agujeros al cristal semiconductor (que en este caso se supone tipo N) que se desplazan hacia el electrodo colector y son recogidos allí por medio del colector con polarización inversa, haciendo que fluya una corriente de colector I_C a través del circuito de colector y una resistencia de carga tierra.

El voltaje emisor V_E requerido para que fluya la corriente en el circuito de emisor, puede ser mucho más pequeño que el voltaje inverso V_C que se aplica al colector. Así, si casi todos los agujeros emitidos llegan al electrodo colector, la corriente de colector será casi igual a la de emisor. Por tanto, es obvio que un pequeño cambio en la corriente de emisor, producido por una pequeña entrada de potencia en el emisor, puede originar casi el mismo cambio en la corriente de colector que fluye a través de un circuito de una impedancia mucho mayor en el que se mantienen voltajes más elevados, y producir un cambio grande en la potencia que pasa a la resistencia de carga.

Esto se puede demostrar de un modo más cuantitativo expresando la corriente de emisor en función del voltaje de emisor mediante

$$I_E = I_{E0} (e^{eV_E/kT} - 1) \quad 2-51$$

en donde I_{E0} es la corriente de saturación del emisor. Del mismo modo, se puede escribir la siguiente relación de corriente-voltaje en el colector

$$I_C = I_{C0} (e^{eV_C/kT} - 1) + \alpha I_E \quad 2-52$$

en donde I_{C0} es la corriente de saturación, asociada con el colector y α representa la fracción de la corriente de emisor que se recupera subsecuentemente como corriente de portadores minoritarios en el colector. En esta ecuación, el primer término representa la componente de corriente de colector debida a la recuperación normal de portadores minoritarios que se generan térmicamente dentro del cristal semiconductor, mediante el electrodo colector con polarización inversa, en tanto que la segunda representa la componente adicional creada por los portadores minoritarios inyectados en el emisor y que después recoge el colector. En un transistor bien diseñado, el factor α es muy cercano a la unidad; pero no puede ser igual por completo ni superarla, debido a que parte de los portadores minoritarios inyectados por el emisor no podrán llegar al colector a causa de la recombinación en la parte volumétrica del cristal semiconductor o en la superficie de éste; en el funcionamiento normal del transistor, la corriente de emisor se mantiene a un valor mucho mayor que I_{E0} y el colector tiene una polarización inversa profunda dentro de la región de saturación; en estas condiciones las ecuaciones 2-51 y 2-52 pueden escribirse más o menos como sigue

$$I_E \cong I_{E0} e^{eV_E/kT} \quad (I_E \gg I_{E0}) \quad 2-53$$

$$I_C \cong I_{C0} + \alpha I_E \quad (|V_C| \gg kT/e) \quad 2-54$$

La ecuación 2-53 se puede expresar como sigue

$$V_E = \frac{kT}{e} \ln \frac{I_E}{I_{E0}} \quad 2-55$$

y, de acuerdo con esto, es evidente que una pequeña variación de la corriente de emisor dI_E se relaciona con una variación dV_E en el voltaje de emisor de acuerdo con

$$dV_E = \frac{kT}{e} \frac{dI_E}{I_E} \quad 2-56$$

El cambio correspondiente en la potencia de entrada en el circuito del emisor está dada por,

$$dP_E = d(I_E V_E) = I_E dV_E + V_E dI_E = \frac{kT}{e} \left(1 + \ln \frac{I_E}{I_{E0}}\right) dI_E \quad 2-57$$

Sin embargo, en el circuito de salida o colector se tiene que, de acuerdo con 2-54

$$dI_C = \alpha dI_E \quad 2-58$$

y, puesto que la potencia entregada a la resistencia de carga es $P_C = I_C^2 R_L$, se tiene que

$$dP_C = 2I_C R_L dI_C = (2\alpha I_{C0} R_L + 2\alpha^2 I_E R_L) dI_E \quad 2-59$$

en donde I_C y dI_C se han relacionado con I_E y dI_E por medio de 2-54 y 2-58. No obstante, si $\alpha \approx 1$ y $I_E \gg I_{C0}$, como sucede casi siempre, el primer término de 2-59 es mucho más pequeño que el segundo, y, por ende, puede desecharse, de donde, la ecuación se reduce a

$$dP_C = 2\alpha^2 I_E R_L dI_E \quad 2-60$$

Si dI_E se expresa en función de la potencia de entrada dP_E por medio de 2-57, la razón de la potencia de salida a la de entrada se expresa como

$$\frac{dP_C}{dP_E} = \frac{2\alpha^2 I_E R_L}{\frac{kT}{e} \left(1 + \ln \frac{I_E}{I_{E0}}\right)} \quad 2-61$$

Esta ecuación expresa la ganancia de potencia en c-d con señal pequeña del dispositivo. En un caso típico, se podría esperar que los valores de los diferentes parámetros fueran más o menos como sigue: $I_E \approx 10 \text{ mA}$, $I_{E0} = 10 \text{ A}$, $\alpha = 0.99$, $R_L = 1000 \text{ ohms}$, $kT/e = 1/40 \text{ V}$ a 300°K . Si se sustituyen estos valores en la ecuación anterior, se obtiene una ganancia de potencia de aproximadamente 100. La ganancia de corriente del dispositivo es la relación dI_C/dI_E . De acuerdo con 2-58, es obvio que esta cantidad es sencillamente α y que, en general, de acuerdo con 2-52, la ganancia de corriente puede expresarse como

$$\alpha = \left(\frac{\partial I_C}{\partial I_E} \right)_{V_C} \quad 2-62$$

En la figura F2-10(a) se ilustra la estructura de Haynes-Shockley,

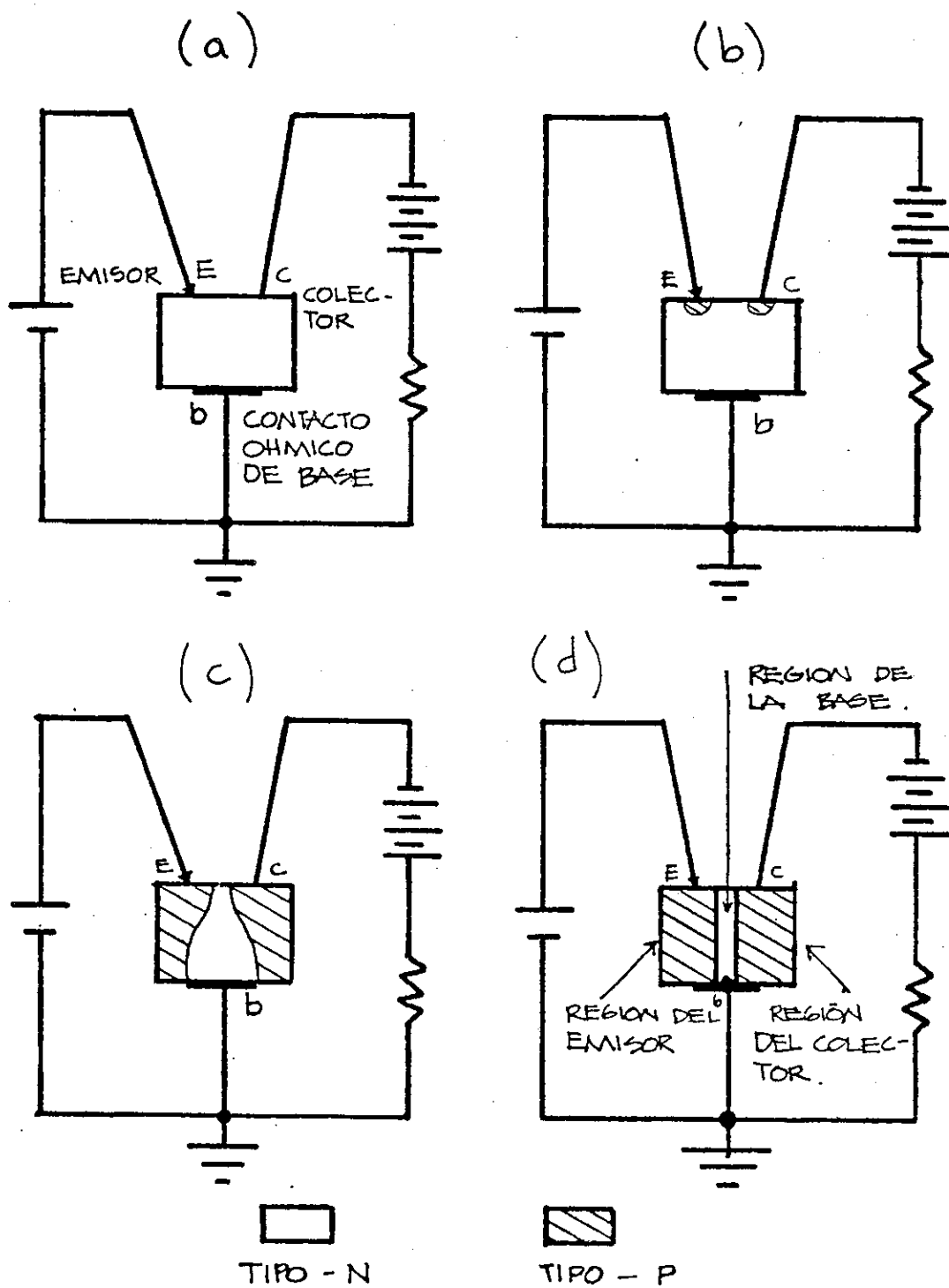


FIGURA F2-10: etapas sucesivas de la "transformación topológica" del transistor de filamentos de Haynes-Shockley para convertirlo en una estructura P-N-P planar convencional.

que se puede convertir a un transistor convencional de unión P-N-P mediante una "transformación topológica" que altera las dimensiones y las proporciones de las diferentes partes del dispositivo, pero dejan inmutables sus relaciones físicas básicas. Esto se logra tal como se ilustra en la figura F2-10, los electrodos de puntas para el contacto del emisor y el colector se sustituyen primeramente usando en su lugar dos regiones pequeñas tipo P bajo los cables de conexión que ahora se consideran como contactos óhmicos a las regiones P, como se indica en la figura F2-10(b). Es evidente que esto no altera el funcionamiento de rectificación y las uniones P-N son fundamentalmente lo mismo. A continuación se considera, que las regiones P se extienden como sucede en (c) para incluir un volumen mayor del cristal, llegando por último a la configuración convencional del transistor de unión P-N-P que aparece en (d).

Ganancia de corriente

La ganancia de corriente α puede diferir de la unidad debido a la recombinación de portadores en exceso en la región de base del transistor; también puede diferir de la unidad porque no toda la corriente de unión del emisor se debe a la inyección de agujeros en la región de base tipo N proveniente del emisor tipo P. Existe también una componente de la corriente del emisor que se origina por la inyección de electrones desde la región de la base hacia el emisor, que no produce ninguna corriente de señal en el colector, pero que fluye a través del conductor de base. En cualquier caso, la ganancia de corriente α se puede expresar como el producto de la eficiencia de inyección de portadores minoritarios del emisor por la fracción de los portadores minoritarios que, una vez que se han inyectado dentro de la región de base del dispositivo, sobreviven para ser recogidos por el colector por lo cual, si este último factor se representa por medio de f ,

$$\alpha = \eta_p f \quad 2-63$$

La ganancia de corriente de un transistor planar P-N-P como el que hemos visto, se puede calcular con métodos de análisis similares a los utilizados con el rectificador de unión. Las ecuaciones de continuidad de estado estacionario se pueden expresar como,

$$\frac{d^2(N_p - N_{p0})}{dx^2} - \frac{N_p - N_{p0}}{L_n^2} = 0 \quad (x < 0; \text{región emisor}) \quad 2-64$$

$$\frac{d^2(P_n - P_{n0})}{dx^2} - \frac{P_n - P_{n0}}{L_p^2} = 0 \quad (x > 0; \text{región base}) \quad 2-65$$

Por tanto, esto hace suponer que todas las caídas de voltaje aparecen en las uniones y, en consecuencia, no existen campos en las regiones volumétricas y, también, que las densidades de portadores inyectados son pequeñas en todas partes en comparación con las cantidades de portadores mayoritarios en equilibrio. También se supondrá que el efecto de generación y recombinación de portadores dentro de las regiones de carga espacial asociadas con las uniones, son despreciables y que la extensión de tales regiones es pequeña en comparación con el espesor de la región de base. En estas circunstancias, lo único que se requiere es aplicar condi-

ciones de frontera y evaluar las corrientes de difusión en $x = 0$ y $x = d$. En lo que respecta a las condiciones de frontera, será necesario que

$$N_p(-\infty) - N_{p0} = 0 \quad 2-66$$

es decir, que cualquier densidad de portadores minoritarios en exceso descende a cero muy lejos, dentro de la región del emisor, que

$$P_n(d) = 0 \quad 2-67$$

por lo cual, la densidad de portadores minoritarios excedentes se reduce a cero en el borde de la unión de colector con polarización inversa y, además, que en la unión del emisor

$$N_p(0) = N_{p0} e^{eV_e/kT} \quad 2-68$$

$$P_n(0) = P_{n0} e^{eV_e/kT} \quad 2-69$$

Por supuesto, se supone que la región del emisor se extiende a muchas longitudes de difusión a la izquierda de la unión del emisor.

La solución general de estas ecuaciones se puede expresar como

$$N_p - N_{p0} = A e^{x/L_n} + B e^{-x/L_n} \quad (x < 0) \quad 2-70$$

y

$$P_n - P_{n0} = C \cosh \frac{x}{L_p} + D \sinh \frac{x}{L_p} \quad (x > 0) \quad 2-71$$

De acuerdo con 2-66 se ve que $B = 0$; en este caso, las condiciones de frontera para electrones y agujeros en la unión del emisor están dadas por

$$A = N_{p0} (e^{eV_e/kT} - 1) \quad 2-72$$

y

$$C = P_{n0} (e^{eV_e/kT} - 1) \quad 2-73$$

Ahora se puede utilizar la condición de frontera de colector 2-67 para evaluar el único coeficiente restante,

$$P_n = C \cosh \frac{x}{L_p} + D \sinh \frac{x}{L_p} + P_{n0} \text{ pero } C = P_{n0} (e^{eV_e/kT} - 1) \Rightarrow$$

$$P_n(d) = P_{n0} (e^{eV_e/kT} - 1) \cosh \frac{d}{L_p} + D \sinh \frac{d}{L_p} + P_{n0} = 0$$

por lo tanto,

$$D = -P_{n0} \left[(e^{eV_e/kT} - 1) \coth \frac{d}{L_p} + \operatorname{csch} \frac{d}{L_p} \right] \quad 2-74$$

por lo cual, de acuerdo con 2-70 y 2-71,

$$N_p - N_{p0} = N_{p0} (e^{eV_e/kT} - 1) e^{x/L_n} \quad 2-75$$

y,

$$P_n - P_{n0} = P_{n0} \left[(e^{eV_e/kT} - 1) \cosh \frac{x}{L_p} - \left\{ (e^{eV_e/kT} - 1) \coth \frac{d}{L_p} + \operatorname{csch} \frac{d}{L_p} \right\} \operatorname{sech} \frac{x}{L_p} \right] \quad 2-76$$

Las corrientes de electrones y agujeros que pasan por las uniones de emisor se evalúan como se hizo en el caso del rectificador P-N, lo cual da

$$J_n(0) = -D_n \left(\frac{dN_p}{dx} \right)_0 = -\frac{N_{p0} D_n}{L_n} (e^{eV_e/kT} - 1) \quad 2-77$$

y,

$$J_p(0) = -D_p \left(\frac{dA_n}{dx} \right)_0 = \frac{P_{n0} D_p}{L_p} \left[(e^{eV_e/kT} - 1) \coth \frac{d}{L_p} + \operatorname{csch} \frac{d}{L_p} \right] \quad 2-78$$

La densidad de corriente de agujeros que pasa por el colector se determina en la misma forma y el resultado es

$$J_p(d) = -D_p \left(\frac{dP_n}{dx} \right)_d = \frac{P_{n0} D_p}{L_p} \left[\left\{ (e^{eV_e/kT} - 1) \cosh \frac{d}{L_p} + 1 \right\} \coth \frac{d}{L_p} - (e^{eV_e/kT} - 1) \operatorname{sech} \frac{d}{L_p} \right] \quad 2-79$$

La eficiencia de inyección de agujeros en la unión del emisor es la razón entre la corriente de agujeros y la corriente total que pasa por el emisor. Esto se calcula a partir de 2-77 y 2-78. Cuando se hace esto se puede suponer que en la operación normal del emisor, la corriente real de emisor es mucho mayor que la de saturación, de tal manera que $e^{eV_e/kT} \gg 1$. Los términos de las ecuaciones anteriores que no contienen el factor exponencial se pueden hacer a un lado y (recordando que $N_{n0} P_{n0} = P_{n0} N_{p0} = N_i^2$) se llega a,

$$\eta_p = \frac{e J_p(0)}{e [J_p(0) - J_n(0)]} = \frac{1}{1 + \frac{b N_{n0} L_p \tanh \frac{d}{L_p}}{P_{n0} L_n}} \quad 2-80$$

La fracción de agujeros inyectados en el emisor, que sobreviven a la recombinación para ser recogidos en el colector, en estado estacionario, es

$$f = \frac{J_p(d)}{J_p(0)} = \frac{\coth \frac{d}{L_p} \cosh \frac{d}{L_p} - \operatorname{sech} \frac{d}{L_p}}{\coth \frac{d}{L_p}} = \operatorname{sech} \frac{d}{L_p} \quad 2-81$$

También en este caso se ha hecho caso omiso de los términos que no contienen el factor exponencial $e^{eV_e/kT}$. Ahora la ganancia de corriente α se puede expresar como

$$\alpha = \eta_p f = \frac{1}{\cosh \frac{d}{L_p} + \frac{b L_p N_{n0} \operatorname{sech} \frac{d}{L_p}}{L_n P_{n0}}} \quad 2-82$$

Es evidente que para que α sea tan grande como sea posible, es necesario que la anchura de la región de base tenga un mínimo posible y que la longitud de difusión L_p (y, por ende, el tiempo de vida de los agujeros dentro de la región de base tipo N) sea también lo más grande posible; además para proporcionar la máxima eficiencia de inyección de agujeros, la densidad de impurificación del emisor P_{e0} debe ser mucho mayor que la densidad de impureza N_{n0} en la región de base del dispositivo.

Ganancia de corriente en función de la frecuencia

La variación de la ganancia de corriente en función de la frecuencia se puede entender cuantitativamente viendo la ecuación de continuidad dependiente del tiempo,

$$D_p \frac{d^2(P_n - P_{n0})}{dx^2} - \frac{P_n - P_{n0}}{\tau_p} = \frac{\partial (P_n - P_{n0})}{\partial t} \quad 2-83$$

si se supone que, $P_n(x, t) - P_{n0} = U(x) e^{i\omega t}$ 2-84

y se sustituye esto en 2-83, se obtiene la siguiente ecuación diferencial,

$$\frac{d^2 U}{dx^2} - \frac{1}{L_p^2} (1 + i\omega \tau_p) U(x) = 0 \quad 2-85$$

para $U(x)$. La solución general de esta ecuación tiene la forma

$$U(x) = A \exp\left(\frac{x}{L_p} \sqrt{1 + i\omega \tau_p}\right) + B \exp\left(-\frac{x}{L_p} \sqrt{1 + i\omega \tau_p}\right)$$

entonces,

$$U(x) = A \exp\left[\left(\frac{\beta + i\gamma}{L_p}\right)x\right] + B \exp\left[-\left(\frac{\beta + i\gamma}{L_p}\right)x\right] \quad 2-86$$

en donde, $\beta = \text{Re}(\sqrt{1 + i\omega \tau_p})$ y $\gamma = \text{Im}(\sqrt{1 + i\omega \tau_p})$ 2-87

de $\beta + i\gamma = \sqrt{1 + i\omega \tau_p}$ obtenemos, $\beta^2 + 2i\beta\gamma - \gamma^2 = 1 + i\omega \tau_p$
formamos un sistema de dos ecuaciones con dos incógnitas y obtenemos,

$$\beta^2 = \frac{1}{2} \left[1 + \sqrt{1 + \omega^2 \tau_p^2} \right] \quad 2-88$$

y

$$\gamma^2 = \frac{1}{2} \left[-1 + \sqrt{1 + \omega^2 \tau_p^2} \right] \quad 2-89$$

En las expresiones anteriores se debe escoger el signo positivo en el radical, ya que en el límite, conforme ω tiende a cero, se debe obtener

$\beta^2 = 1$ y $\gamma^2 = 0$. Si L_p representa la constante de decaimiento exponencial asociada con la solución 2-86, entonces, al usar 2-88, es evidente que

$$U(x) = A e^{x/L_p} e^{i\gamma x/L_p} + B e^{-x/L_p} e^{-i\gamma x/L_p} \quad 2-90$$

siendo,

$$\frac{1}{L_p'} = \frac{\beta}{L_p} = \frac{1}{\sqrt{2} L_p} \left[1 + \sqrt{1 + \omega^2 z_p^2} \right]^{1/2} \quad 2-91$$

en donde el valor de L_p es precisamente el valor de estado estacionario dado por 2-29. De acuerdo con 2-90 y 2-91, se ve que, a frecuencias bajas, la constante de degeneración L_p' relacionada con la ecuación de continuidad está dada por

$$L_p' \cong L_p = \sqrt{D_p z_p} \quad (\omega z_p \ll 1) \quad 2-92$$

que es precisamente el valor de estado estacionario que se obtuvo en 2-29, en tanto que para frecuencias crecientes, L_p' disminuye a partir de este valor, acercándose en el límite de altas frecuencias al valor

$$L_p' \cong L_p \sqrt{\frac{2}{\omega z_p}} = \sqrt{\frac{2 D_p}{\omega}} \quad (\omega z_p \gg 1) \quad 2-93$$

De acuerdo con esto, es evidente que en las condiciones de c-a la distancia efectiva de decaimiento que se debe utilizar para encontrar la fracción de la señal de entrada que llega al colector es L_p' y no L_p . El valor $L_p'(\omega)$ tal como lo da 2-91 debe sustituirse en 2-81, que proporciona el valor de decaimiento f . El resultado es

$$\alpha(\omega) = \eta_p f = \frac{\operatorname{sech} \frac{d}{L_p'(\omega)}}{1 + \frac{b N_{A0} L_p \tanh \frac{d}{L_p}}{p_0 L \omega}} \quad 2-94$$

Al aumentar la frecuencia se llegará a un punto en donde L_p' tiene una magnitud comparable con d y en esta región de frecuencia, el valor de f y, por ende, el de α , es notablemente menor que la unidad. La ganancia de corriente en esta región comienza a reducirse en comparación con su valor de frecuencia cero. La "frecuencia de corte α " del dispositivo, ω_0 , se puede definir como sigue

$$L_p'(\omega_0) = d \quad 2-95$$

sustituyendo este valor de L_p' en 2-91 y despejando la frecuencia, se tiene

$$\omega_0 = \frac{2\sqrt{D_p}}{d} \sqrt{\frac{D_p}{d^2} - \frac{1}{z_p}} \quad 2-96$$

Para frecuencias que sobrepasan este valor, la ganancia de corriente es mucho menor que el que se tiene a la frecuencia cero. Es obvio que para lograr altas frecuencias de corte, se requieren valores pequeños en la anchura de la base y valores grandes en los tiempos de vida en la región de la base.

Circuito de emisor a tierra

Veamos ahora brevemente, el circuito emisor a tierra y el circuito de base abierta de estos dispositivos. En el circuito que aparece en la figura F2-10 (circuito de base a tierra) se supuso que se introdujo un incremento de corriente dI_E al electrodo emisor. En la figura F2-11, el circuito de emisor conectado a tierra, se introduce una señal de corriente dI_B al cable de base. De acuerdo con la continuidad de la corriente, se debe tener siempre

$$I_E = I_C + I_B \quad 2-97$$

de donde,

$$dI_E = dI_C + dI_B \quad 2-98$$

no obstante, por definición, $dI_E = dI_C / \alpha \Rightarrow$

$$dI_C = \frac{\alpha}{1-\alpha} dI_B = \beta dI_B \quad 2-99$$

la cantidad

$$\beta = \frac{dI_C}{dI_B} = \frac{\alpha}{1-\alpha} \quad 2-100$$

es la ganancia de corriente de emisor a tierra del transistor. Evidentemente, esto demuestra que la configuración del circuito de emisor a tierra se puede emplear para proporcionar factores de amplificación de corriente mayores que la unidad. La configuración de emisor a tierra es la disposición del circuito que se usa con mayor frecuencia en amplificadores de transistor.

Circuito de base abierta

Otro efecto interesante del funcionamiento del transistor de unión, es la autoamplificación de la corriente de saturación del colector en un circuito de base abierta, ilustrado en la figura F2-12. En este caso las corrientes de emisor y colector deben ser iguales y, por tanto, a condición de que la magnitud del voltaje de batería sea mucho mayor que

kT/e , se debe tener, de acuerdo con 2-52

$$I_C = I_{C0} + \alpha I_E = I_{C0} + \alpha I_C \quad 2-101$$

de donde

$$I_C = \frac{I_{C0}}{1-\alpha} \quad 2-102$$

Por tanto, en este circuito, la corriente de colector se autoamplifica por un factor de $1/(1-\alpha)$. Este efecto proporciona una forma muy sencilla aunque aproximada para medir la ganancia de corriente en c-d de un transistor de unión. La corriente de saturación del colector I_{C0} se determina primeramente abriendo el circuito del emisor y completando el circuito de base, observando la cantidad de corriente que fluye. A continuación se abre el circuito de base y se cierra el de emisor, como se indica en F2-12, y se determina la corriente de colector autoamplificada por esta operación. Entonces, la ganancia de corriente se puede determinar a partir de 2-102.

Aunque se ha limitado el estudio a los transistores de unión P-N-P, es obvio que se puede realizar un análisis idéntico para las unidades N-P-N, llegando a ecuaciones que son iguales a las que se derivaron antes, con la excepción de que los subíndices N y P deben intercambiarse y la

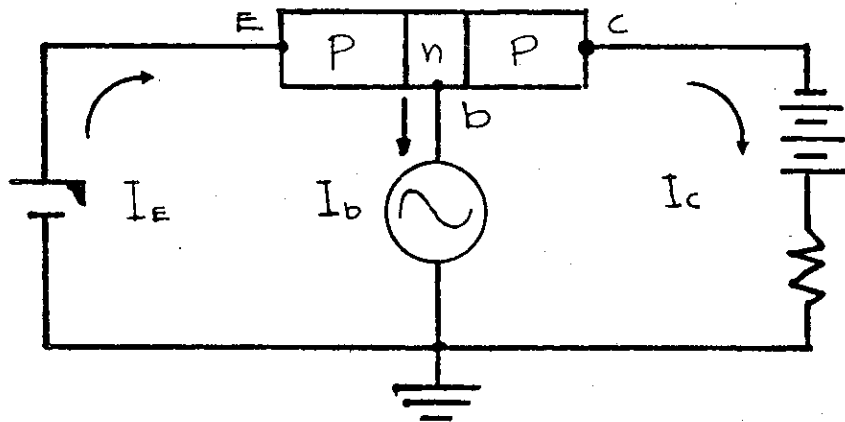


FIGURA F2-11: circuito de emisor a tierra de un transistor P-N-P.

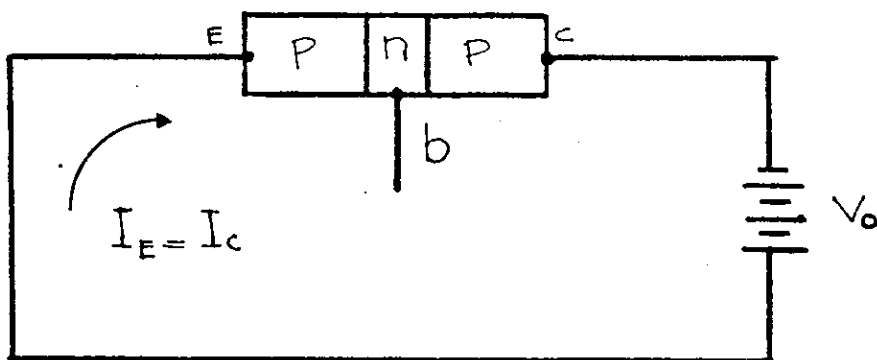


FIGURA F2-12: autoamplificación de una corriente de colector en un circuito de base abierta.

cantidad b se sustituye por $1/b$. Sin embargo, debe hacerse notar que los voltajes de polarización que deben usarse en conexión con las unidades N-P-N, tienen el signo inverso con respecto a los correspondientes a los transistores P-N-P.

3. EL TRANSISTOR DE EFECTO DE CAMPO

Consideraremos ahora otro tipo de dispositivo semiconductor, que puede ser usado para amplificación, llamado transistor de efecto de campo de unión. Lo abreviaremos como FET (field-effect transistor). La configuración física y los principios de operación, son diferentes de los del transistor de unión. El FET tiene muchas ventajas sobre el transistor de unión: tiene una muy alta resistencia de entrada, produce menor ruido, tiene una mejor estabilidad térmica y es muy poco afectado por la radiación. Pero el transistor de unión tiene una mejor respuesta en alta frecuencia.

El FET de unión se ilustra diagramáticamente en la figura F3-1. Consiste de una barra semiconductor tipo N con dos regiones tipo P, como se ilustra. (las regiones P y N pueden ser intercambiadas; la teoría de operación es la misma). Un lado de la región tipo N se llama fuente y la otra drenador. La región tipo N entre las regiones de depleción de las uniones P-N, de la fuente al drenador se llama canal. La corriente principal es la corriente del drenador I_D que va del drenador a la fuente a través del canal. Las dos regiones tipo P se llaman puertas, las cuales están eléctricamente conectadas externamente. La cantidad de impurezas de las regiones P es muy fuerte.

Hay dos uniones P-N. Están entre cada puerta y el canal tipo N. Asumamos que las dos uniones puerta-canal están polarizadas inversamente por V_{GS} , como se ilustra en F3-1. La capa de depleción se extiende más hacia el canal tipo N que hacia las puertas, debido a que la concentración de impurezas de las puertas es mucho mayor que la que hay en el canal. La región de depleción (que se ilustra sombreada en la figura F3-1) es más ancha cerca del drenador debido a la corriente de drenador.

Los portadores de carga son electrones libres. Nótese que estos son los portadores mayoritarios en el canal. Sólo hay un tipo de carga en un FET, por esta razón, se llama dispositivo unipolar. En contraste, los transistores de unión utilizan electrones libres y agujeros, por lo tanto, hay dos tipos de portadores de carga y se llama, entonces, dispositivo bipolar.

La corriente de drenador I_D es una función del voltaje aplicado V_{DS} y de la resistencia entre el drenador y la fuente. Esta resistencia está en función de la conductividad del semiconductor y de la sección transversal del área del canal. La región de depleción está casi libre de portadores de carga móviles, por lo cual no contribuye al área disponible para conducción. Por tanto, la conducción ocurre en el canal, entre las capas de depleción. Si la polarización inversa en las juntas P-N, puerta-canal, se incrementa, el ancho de la región de depleción se incrementa y el ancho del canal disminuye, aumentando la resistencia del drenador a la fuente.

Consideremos ahora que un pequeño potencial V_{GS} se aplica entre el drenador y la fuente, como se indica en la figura F3-1. La puerta se mantiene a un potencial constante V_{GS} con respecto a la fuente. Entonces, la diferencia de potencial entre la puerta y la fuente es $V_{GS} - V_{GS}$ (nótese que es negativa). Además, el drenador está a un potencial V_{DS} con res-

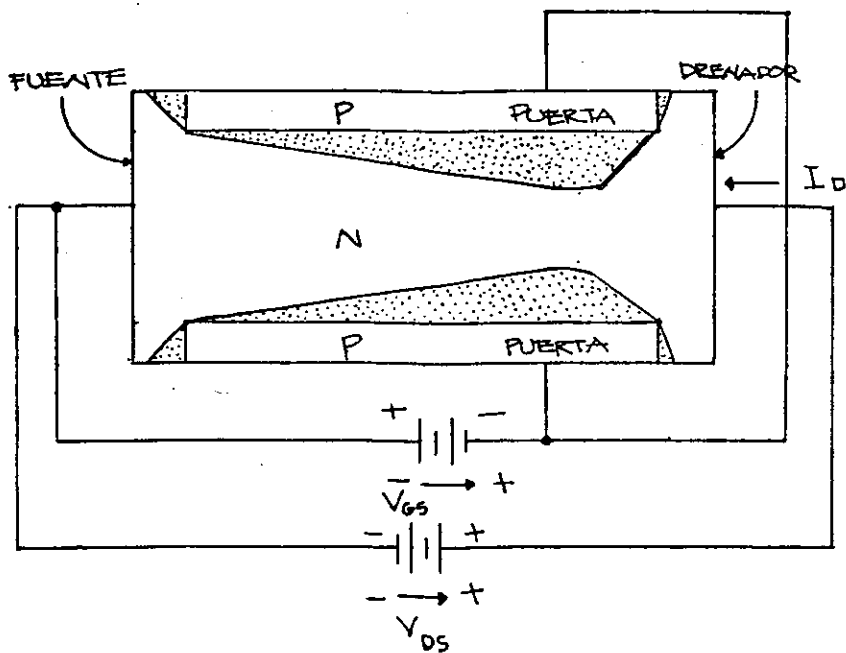


FIGURA F3-1: representación diagramática del FET de unión. Las regiones de deplexión se muestran sombreadas. Nótese que V_{GS} es un número negativo.

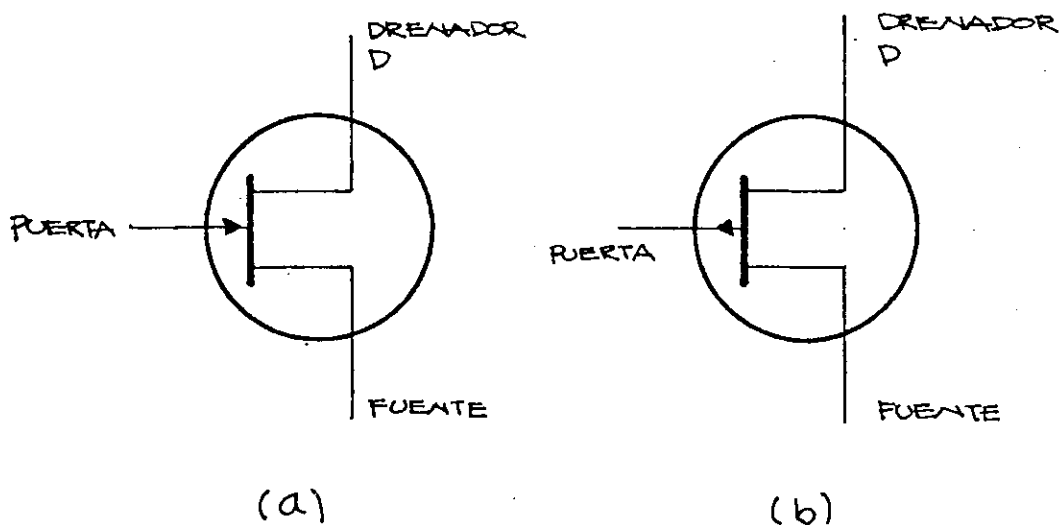


FIGURA F3-2: símbolos para el FET, (a) canal tipo N y (b) canal tipo P.

pecto a la fuente. Entonces, el potencial puerta-drenador es $V_{os} - V_{es}$. El potencial puerta-canal será mayor cerca del drenador que cerca de la fuente. Esto puede considerarse de otra manera. La corriente en el canal, ocasiona una caída de voltaje allí. El potencial entre la puerta y el canal, como una función de la posición, es $-V_{es}$ más la caída de voltaje entre la fuente y el punto en cuestión. Conforme la distancia a la fuente aumenta, la caída de voltaje se incrementa. Entonces, el potencial entre la puerta y el canal también se incrementa.

Consideremos ahora las características corriente-voltaje del FET. Asumamos que V_{es} está fijo y que V_{os} se incrementa a partir de cero. Cuando V_{os} es cero, las regiones de depleción son simétricas de acuerdo con las puertas. Al incrementarse V_{os} , resulta cierta corriente de drenador. La caída de voltaje en el canal produce que la anchura de la capa de depleción se ensanche y se sesgue, como se muestra en la figura F3-1. Conforme V_{os} se incrementa, la corriente I_D disminuye. La resistencia en el canal también se incrementa porque la anchura de la capa de depleción se ha ensanchado.

El voltaje para el cual la anchura del canal se hace muy angosta no está exactamente definida. Aún así, el punto donde, en teoría, el canal se convierte en diferencialmente angosto ó, en la práctica, cuando las curvas de V_{os} contra I_D son casi horizontales, se le llama contracción y al voltaje que produce esto se le llama voltaje de contracción y es una función de V_{es} .

Los símbolos del FET se muestran en la figura F3-2. Uno es para un FET de canal N y el otro para un FET de canal P. Nótese que los símbolos para la fuente y el drenador son los mismos. A veces, la letra D se pone cerca del drenador para clarificar la simbología.

Corriente en el FET de unión: discusión cuantitativa. Obtengamos ahora una expresión cuantitativa para la característica corriente-voltaje de un FET de unión. Para comenzar, consideremos que el voltaje drenador-fuente v_{os} es cero. La anchura del canal como función de v_{es} , el voltaje puerta-fuente, podrá ser entonces determinado. Si v_{os} es cero, entonces i_D , la corriente de drenador, será cero y la región de depleción será uniforme en la mayor parte del canal, como lo ilustra la figura F3-3. Asumamos que la concentración de impurezas varía abruptamente de A a D en la unión P-N. La anchura de la región de depleción para una junta abrupta está dada por

$$d = \sqrt{\frac{2\epsilon(V_0 - v_{es})}{eD(D/A + 1)}} \quad 3-1$$

donde ϵ = permitividad, V_0 = potencial a través de la unión P-N no polarizada, e = magnitud de carga del electrón. En el FET, $D/A \ll 1$, entonces, después de algunas manipulaciones, obtenemos para la anchura del canal (véase figura F3-3)

$$w(x) = b - 2d = b - 2\sqrt{\frac{1}{eD} 2\epsilon(V_0 - v_{es})} \quad 3-2$$

La ecuación 3-2 es válida únicamente para valores de $\omega(x)$ mayores o iguales que cero. Esto es que la anchura del canal no puede ser negativa. Aquellos valores de v_{cs} que hagan $\omega(x)$ menor que cero, no pueden ser utilizados en 3-2.

Ahora determinemos el valor de v_{cs} que produce la contracción cuando v_{os} es cero. Cuando ocurre la contracción, $\omega(x)$ debe ser cero. Entonces haciendo 3-2 igual a cero y resolviendo para v_{csp} , donde el subíndice P indica contracción (pinch-off), tenemos,

$$V_{csp} = -\frac{b^2 e D}{8 \epsilon} + V_0 \quad 3-3$$

Nótese que v_{csp} es una cantidad negativa. En general,

$$\frac{b^2 e D}{8 \epsilon} \gg V_0 \quad 3-4$$

Entonces tenemos,

$$V_{csp} = -\frac{b^2 e D}{8 \epsilon} \quad 3-5$$

Asumamos ahora que el pequeño circuito entre el drenador y la fuente se reemplaza por v_{os} . Esto producirá una corriente i_0 . Debe notarse que si v_{cs} es el voltaje de contracción v_{csp} , la anchura del canal será esencialmente cero e i_0 será también cero. En general, bajo la condición de que v_{cs} sea negativo y de que $|v_{cs}| \geq |v_{csp}|$, i_0 será muy pequeña, del orden de nano amperios.

Si v_{os} es muy pequeño, tal que

$$v_{os} \ll |v_{cs}| \quad 3-6$$

entonces la capa de deplexión no será afectada esencialmente, por el voltaje drenador-fuente. Lo sesgado y angosto de la figura F3-1 se hace aparente únicamente para grandes voltajes de drenador-fuente.

Consideremos v_{os} de tal manera que se satisfaga 3-6. Si el FET tiene "a" metros de grosor, entonces la sección de área transversal del canal será $a\omega$. Si el largo efectivo del canal tiene L metros, entonces la resistencia del canal del drenador a la fuente es

$$R = \frac{L}{\sigma a \omega} \quad 3-7$$

donde σ es la conductividad del semiconductor tipo N. Entonces, usando la ley de Ohm y sustituyendo 3-2, tenemos,

$$i_0 = v_{os} \frac{\sigma a}{L} \left[b - \sqrt{\frac{8 \epsilon (V_0 - v_{cs})}{e D}} \right] \quad 3-8$$

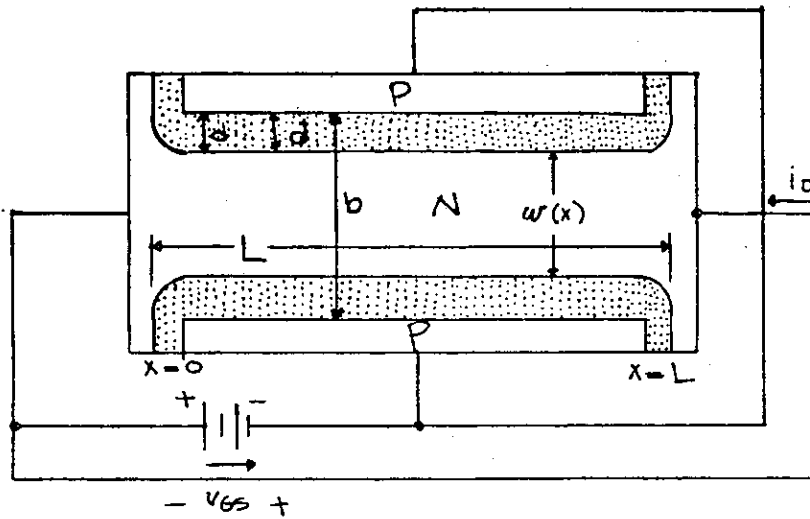


FIGURA F3-3: representación del FET de unión con $v_{gs} = 0$. Nótese que v_{gs} es negativo.

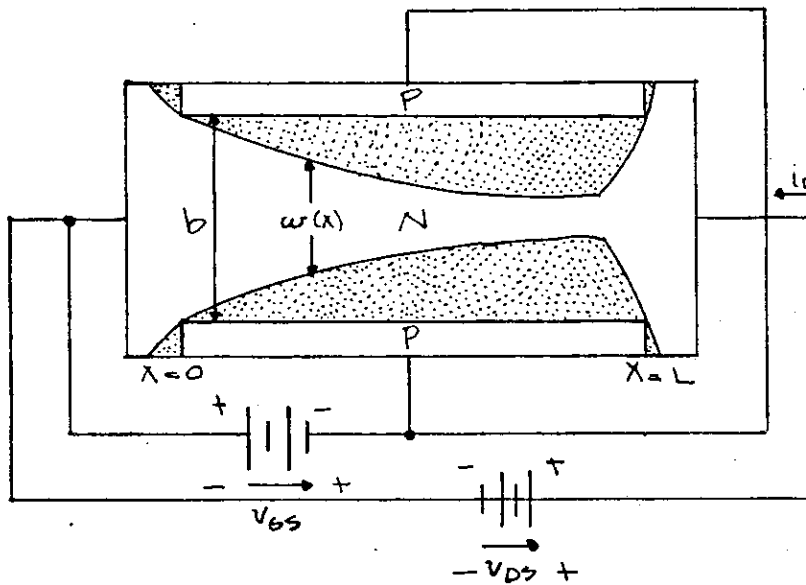


FIGURA F3-4: representación del FET de unión. Nótese que v_{gs} es negativo.

Despejando b en 3-5 obtenemos,

$$b = \sqrt{-\frac{q\epsilon}{eD} V_{gs}} \quad 3-9$$

y sustituyendo en 3-8,

$$I_0 = \frac{q a}{L} b \left[\left(1 - \sqrt{\frac{V_{gs} - V_0}{V_{gs}}} \right) \right] V_{ds} \quad 3-10$$

Consideremos ahora la característica corriente-voltaje cuando no estamos restringidos a pequeños valores de V_{ds} . Trabajaremos con la representación del FET que se ilustra en la figura F3-4. Para simplificar el análisis asumiremos que la diferencia de potencial entre el canal en $x = 0$ y la fuente es $-V_{cs}$; esto es, la caída de voltaje en el canal entre la fuente y $x = 0$ es cero. Asumiremos que la caída de voltaje en el canal entre $x = L$ y la fuente es esencialmente V_{ds} . Entonces el potencial aplicado entre la parte de la puerta en $x = L$ y el canal es $-V_{cs} + V_{ds}$. La diferencia de potencial entre la puerta y el canal es $v(x) + V_0$. Esto es, $v(x)$ es la diferencia de potencial no debida al potencial de contacto. Entonces de acuerdo con nuestros supuestos,

$$v(0) = -V_{cs} \quad 3-11$$

$$v(L) = -V_{cs} + V_{ds} \quad 3-12$$

Nótese que hemos definido $v(x)$ como una cantidad positiva. La ecuación 3-2 dará aproximadamente $\omega(x)$ en cualquier punto, reemplazando $-V_{cs}$ por $v(x)$. Entonces,

$$\omega(x) = b - \sqrt{\frac{q\epsilon [V_0 + v(x)]}{eD}} \quad 3-13$$

Para una longitud diferencial, la resistencia del canal será (véase ecuación 3-7),

$$dR = \frac{dx}{q \sigma \omega(x)} \quad 3-14$$

donde dx es ahora la longitud de la pieza en cuestión. La caída de voltaje a través de una longitud diferencial dx es igual al producto de la intensidad de campo eléctrico E_x y dx . como $\vec{E} = -\nabla V$ entonces,

$$E_x = -\frac{dv(x)}{dx} \quad 3-15$$

Entonces, la corriente a través del canal en un punto x se encuentra por la ley de Ohm y las ecuaciones 3-14 y 3-9 para dar,

$$i_0 = -\frac{q\sigma\omega(x)}{dx} \left[\frac{dV(x)}{dx} dx \right] \quad 3-16$$

Nótese que la corriente tiene un valor constante i_0 a través del canal, que no depende de x . Operando obtenemos,

$$i_0 dx = -q\sigma\omega(x) dV(x) \quad 3-17$$

sustituyendo 3-13 y 3-9, obtenemos,

$$i_0 dx = -q\sigma \sqrt{\frac{qE}{\epsilon_0}} \left[\sqrt{V_{GSF}} - \sqrt{V_0 + V(x)} \right] dV \quad 3-18$$

Integrando de $x = 0$ a $x = L$ o, equivalentemente, de $V(0)$ a $V(L)$ tenemos,

$$\int_0^L i_0 dx = -q\sigma \sqrt{\frac{qE}{\epsilon_0}} \left\{ \int_{V(0)}^{V(L)} \left[\sqrt{V_{GSF}} - \sqrt{V_0 + V(x)} \right] dV \right\} \quad 3-19$$

integrando, encontramos,

$$i_0 = -\frac{q\sigma}{L} \sqrt{\frac{qE}{\epsilon_0}} \left\{ \sqrt{V_{GSF}} V_{OS} - \frac{2}{3} \left[(V_0 + V_{OS} - V_{GS})^{3/2} - (V_0 - V_{GS})^{3/2} \right] \right\} \quad 3-20$$

Las ecuaciones 3-2 y también 3-13 y 3-20 son válidas únicamente para valores de V_{OS} y V_{GS} que no hagan negativo a $\omega(x)$. La ecuación 3-13 es válida hasta aquel valor de $V(x)$ que hace que $\omega(x)$ sea cero. Esto ocurre cuando,

$$V_{OS} - V_{GS} = -V_{GSF} \quad 3-21$$

Esto determina el máximo valor de i_0 que hace válida la ecuación 3-20. Esto corresponde a una i_0 en contracción. Arriba de la contracción, la curva de i_0 contra V_{OS} se vuelve casi horizontal y la ecuación 3-21 se puede utilizar para predecir éstos valores casi constantes. Sustituyendo 3-21 en 3-20, obtenemos la corriente en esta porción horizontal,

$$i_{0, sat.} = \frac{q\sigma}{L} \sqrt{\frac{qE}{\epsilon_0}} \left\{ \sqrt{V_{GSF}} (-V_{GSF} + V_{GS}) - \frac{2}{3} \left[(V_0 - V_{GSF})^{3/2} - (V_0 - V_{GS})^{3/2} \right] \right\} \quad 3-22$$

como i_0 no puede ser imaginaria, esta ecuación es válida sólo si V_{GS} no es más negativa que V_{GSF} . En este punto, $i_0 = 0$. Los amplificadores con FET operan usualmente en la región de esta característica, donde las curvas de i_0 contra V_{GS} son casi horizontales. Entonces la ecuación 3-22 puede usarse para predecir estas curvas. Nótese que la variación de

I_D , sat. con V_{GS} no es lineal.

Características estáticas y parámetros dinámicos del FET de unión

En general, hay tres variables que deben ser representadas: I_D , V_{GS} y V_{DS} . Nótese que la fuente (s = source) es usualmente el electrodo de referencia. Para evitar el uso de gráficas tridimensionales, se plotean gráficas dejando una variable constante. Cada curva de la gráfica es para un valor constante diferente de la tercera variable. Se especifican tres familias de curvas y las consideraremos una a una.

Las características estáticas de salida, o simplemente, características de salida, son gráficas de I_D contra V_{DS} con V_{GS} como parámetro. Un conjunto típico se muestra en la figura F3-5.

Utilicemos las características de salida para analizar un amplificador FET simple como el que se ilustra en la figura F3-6. Para hacer esto, dibujamos una línea de carga en las características de salida. La ecuación para la línea de carga está dada por

$$V_{GS} = V_{DD} - I_D R_L \quad 3-23$$

Esta es la ecuación de una línea recta con pendiente $-1/R_L$ con interceptos V_{DD} y V_{DD}/R_L .

Analicemos el amplificador de la figura F3-6 usando los siguientes valores: $V_{DD} = 25$ volts., $V_{GS} = -1.5$ volts., $R_L = 2500$ ohms y $V_{GS} = 0.5 \sin \omega t$ volts. La línea de carga se dibuja en la figura F3-5. El punto estático de operación se obtiene considerando que la señal de voltaje es $V_{GS} = 0$. Entonces se obtiene de la intersección de la línea de carga con la curva para $V_{GS} = -1.5$ volts. Entonces, $I_{DQ} = 4.9$ MA y $V_{DSQ} = 12.5$ volts.

El voltaje de señal es V_{GS} . Esto causa que el voltaje de puerta a fuente V_{GS} , varíe entre -1 y -2 volts. El punto de operación instantáneo se mueve entre los dos extremos marcados en la figura F3-5. Nótese que el voltaje de la señal de salida está desfasado 180° con el voltaje de la señal de entrada. Conforme V_{GS} se torna positivo, V_{GS} se vuelve menos negativo e I_D se incrementa. Esto incrementa la caída de voltaje en R_L y reduce V_{DS} .

La señal de salida varía entre 9 y 16 volts, mientras que la señal de entrada varía entre -1 y -2 volts. Entonces, una señal de entrada de 1 voltio produce una señal de salida de siete.

La amplificación de voltaje de este circuito es

$$A_v = - \frac{16 - 9}{2 - 1} = -7 \quad 3-24$$

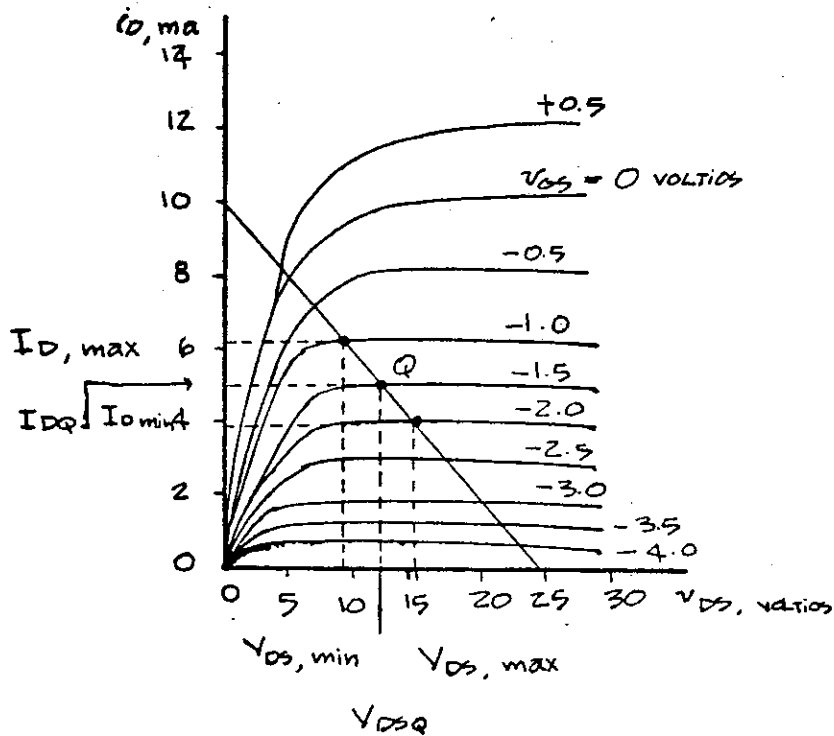


FIGURA F3-5: características de salida típicas del FET de unión. Está dibujada la línea de carga.

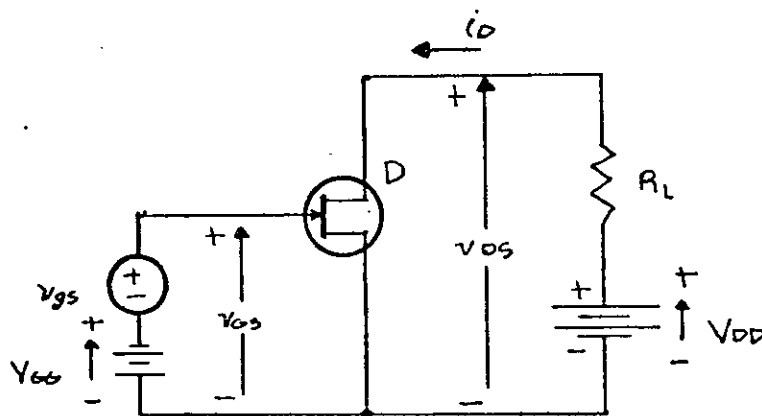


FIGURA F3-6: un amplificador FET simple. Nótese que V_{GS} es negativo. Usualmente una D se agrega para distinguir el drenador de la fuente.

El signo negativo indica el desfase en 180°. En la realidad el voltaje y la corriente de salida, no varían exactamente en forma senoidal porque las características son no lineales.

Nótese que la curva de la característica está dada para v_{gs} positivo. Cuando v_{gs} se vuelve positivo, la unión P-N puerta-canal se polariza directamente y aparece una corriente i_o . Esta corriente puede también dibujarse entre las características de salida. Los FET's de unión generalmente no operan con la unión puerta-canal, polarizada directamente puesto que la alta impedancia de entrada es una de las ventajas del FET y si la unión puerta-canal está polarizada directamente, esta ventaja se elimina.

Otro conjunto de características de salida usadas para caracterizar un FET son las características de transferencia. Estas son ploteos de i_o contra v_{gs} con v_{ds} como parámetro. Un conjunto típico de características de transferencia se muestra en la figura F3-7. Nótese que todas las curvas son muy cercanas, indicando que i_o es prácticamente independiente de v_{ds} para la mayoría de valores de v_{ds} .

El tercero de los conjuntos de características consiste de un ploteo de v_{ds} contra v_{gs} con i_o como parámetro. Se denominan características de corriente constante, y un conjunto típico se muestra en la figura F3-8. Las características son casi horizontales para la mayoría de los valores de v_{ds} , otra vez indicando que i_o es prácticamente independiente de v_{ds} para la mayoría del rango de operación.

Quando se consideran amplificadores, la mayoría de las veces estamos interesados en pequeños cambios en el voltaje o corriente de salida en respuesta a pequeños cambios en el voltaje de entrada. Las pendientes de las características estáticas en el punto estático de operación se utilizan generalmente para caracterizar estas señales. Estos parámetros dinámicos se definen a continuación.

Factor de amplificación (μ). Este factor se define como,

$$\mu = - \left. \frac{dv_{ds}}{dv_{gs}} \right|_{i_o = \text{cte.}} = - \frac{\partial v_{ds}}{\partial v_{gs}} \quad 3-25$$

Entonces, μ es el inverso de la pendiente de la característica de corriente constante y es una cantidad adimensional.

Transconductancia (g_m). La transconductancia se define como,

$$g_m = \left. \frac{di_o}{dv_{gs}} \right|_{v_{ds} = \text{cte.}} = \frac{\partial i_o}{\partial v_{gs}} \quad 3-26$$

Es la pendiente de las características de transferencia en el punto de ope

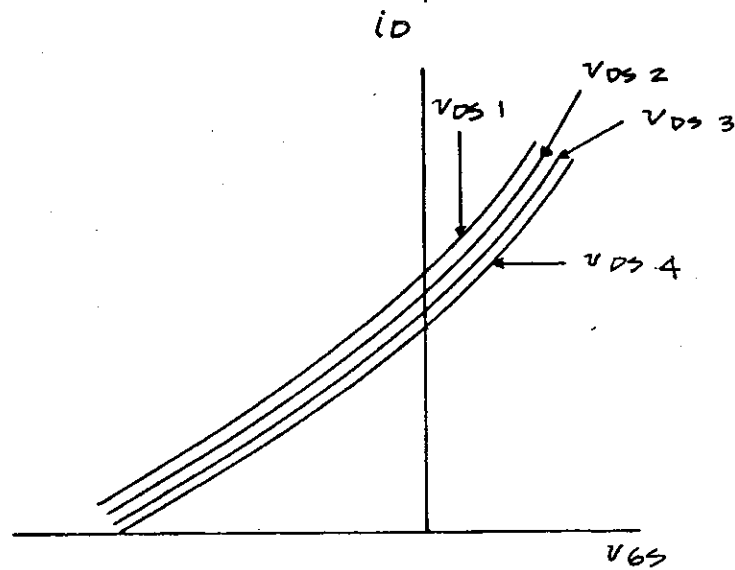


FIGURA F3-7: características de transferencia típicas para un FET.
 $V_{DS1} > V_{DS2} > V_{DS3} > V_{DS4}$

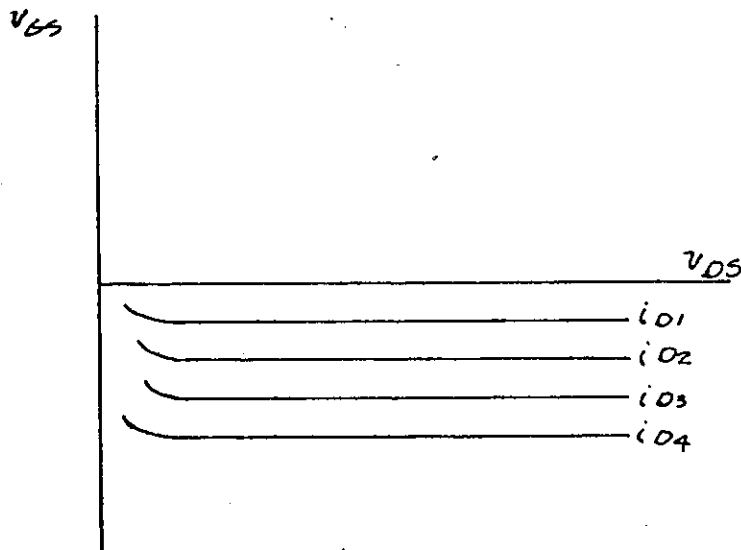


FIGURA F3-8: características de corriente constante típicas para un FET.
 $i_{D1} > i_{D2} > i_{D3} > i_{D4}$

ración. Sus unidades se dan en mhos , esto es, $1/\text{ohm}$.
Resistencia de drenador (r_d). La resistencia de drenador (o equivalentemente, la resistencia dinámica drenador fuente) se define como,

$$r_d = \left. \frac{dV_{os}}{dI_o} \right|_{V_{es} = \text{cte.}} = \frac{\partial V_{os}}{\partial I_o} \quad 3-27$$

que es el inverso de la pendiente de la característica de salida en el punto de operación. Sus unidades son ohms y algunas veces se trabaja con la conductancia de drenador g_d ,

$$g_d = 1/r_d \quad 3-28$$

Las resistencias dinámicas entre la puerta y la fuente y entre la puerta y el drenador pueden ser también determinadas. En general para una unión puerta-canal inversamente polarizada, estas pueden tener valores muy altos del orden de 10^8 ohms.

Hemos definido los parámetros dinámicos como pendientes de las características estáticas, pero podemos obtener los tres parámetros sólo de un conjunto de características. Por ejemplo, la ecuación 3-26 puede aproximarse por,

$$g_m \approx \left. \frac{\Delta I_o}{\Delta V_{es}} \right|_{V_{os} = \text{cte.}} \quad 3-29$$

El cambio en I_o correspondiente a un cambio en V_{es} para un valor constante de V_{os} puede obtenerse de las características de salida y sustituidas en 3-29 para obtener g_m .

Los tres parámetros dinámicos no son independientes. Determinemos la relación entre ellos. Si consideramos cambios diferenciales en I_o podemos escribir,

$$dI_o = \frac{\partial I_o}{\partial V_{es}} dV_{es} + \frac{\partial I_o}{\partial V_{os}} dV_{os} \quad 3-30$$

Entonces, sustituyendo 3-26 y 3-27,

$$dI_o = g_m dV_{es} + \frac{1}{r_d} dV_{os} \quad 3-31$$

Ahora asumiendo que V_{es} y V_{os} varían de tal forma que I_o permanece constante, es decir que $dI_o = 0$ tenemos,

$$g_m r_d = - \left. \frac{dV_{os}}{dV_{gs}} \right|_{i_o = cte.} \quad 3-32$$

y sustituyendo 3-25,

$$\mu = g_m r_d \quad 3-33$$

con lo cual hemos determinado la relación entre los parámetros dinámicos.

Efectos de alta frecuencia en el FET de unión

El comportamiento en alta frecuencia del FET está predominantemente afectado por las capacitancias del dispositivo. La polarización inversa de la unión P-N, puerta-canal, tiene capacitancias de transición asociadas a todo su largo. Existen simultáneamente capacitancias y resistencias a lo largo del canal. Además, existen capacitancias entre tierras y con_tactos del FET. Todos estos efectos se pueden tener en cuenta considerando que existen tres capacitores, uno, C_{gs} entre la puerta y la fuente, otro C_{gd} entre la puerta y el drenador, y el tercero, C_{ds} entre el drenador y la fuente. Estas capacitancias se denominan parásitas pues no están puestas deliberadamente y ellas, degradan la respuesta de alta frecuencia del dispositivo.

El modelo lineal de alta frecuencia del FET

Consideraremos el modelo más preciso que se muestra en la figura F3-9. Hemos incluido una terminal de sustrato (por los MOSFET). Están también las capacitancias C_{gs} , C_{gd} y C_{ds} y las resistencias r_{gs} y r_{gd} que están en paralelo con ellas. Estas resistencias, generalmente son muy grandes, del orden de 10^8 ohms en un FET, y pueden ser ignoradas, considerando circuitos abiertos.

El generador de corriente no es proporcional al voltaje de entrada sino al voltaje a través de C_a . Esto es,

$$V_a = \frac{V_{gs}}{1 + j\omega r_a C_a} \quad 3-34$$

Nótese que la corriente del generador de corriente, y por lo tanto la amplificación, cae cuando ω se incrementa. Esto es causado por dos efectos. Uno es que las reactancias de las capacitancias C_{gd} , C_{gs} y C_{ds} llegan a ser muy pequeñas conforme la frecuencia se incrementa. Si las reactancias llegan a ser suficientemente pequeñas, tienden a corto-circuitar la señal. Entonces el voltaje de entrada se hace muy pequeño. Además, la habilidad de amplificar del FET, disminuye cuando se incrementa la

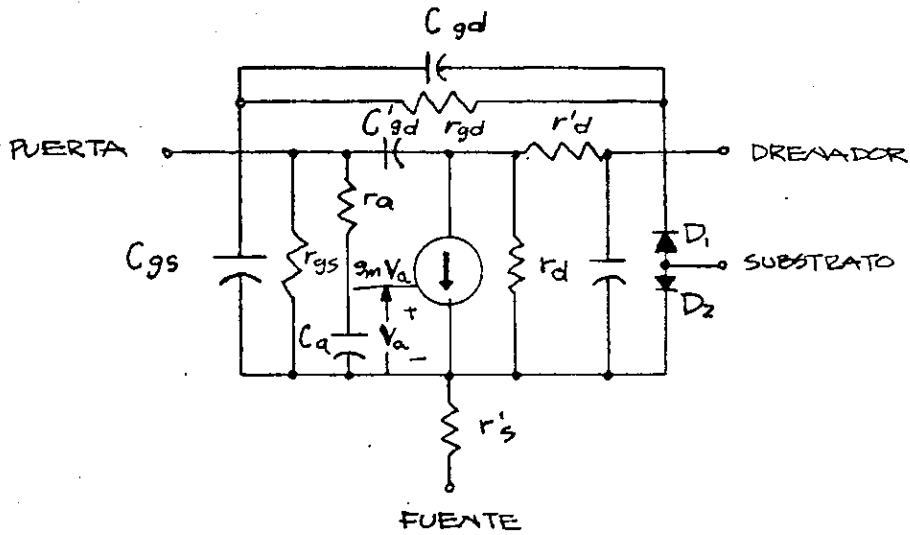


FIGURA F3-9: modelo de alta frecuencia de un FET.

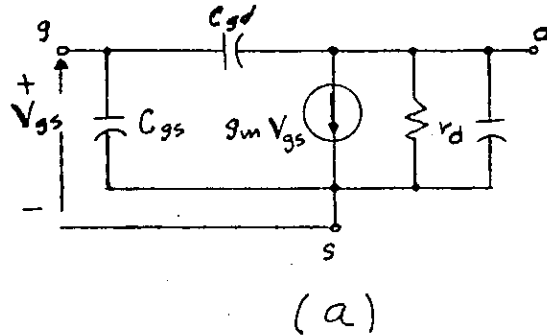
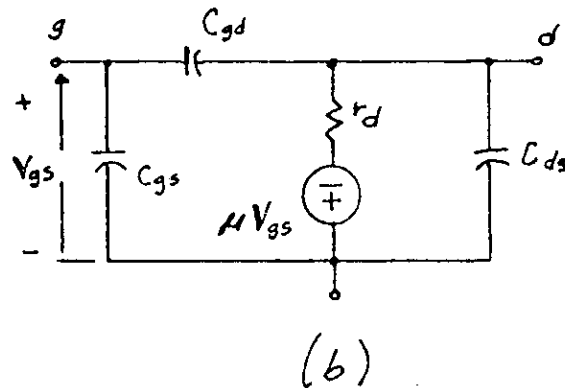


FIGURA F3-10: modelos de alta frecuencia del FET comúnmente usados; a) forma de generador de corriente y b) forma de generador de voltaje.



frecuencia. Estos efectos se toman en cuenta incluyendo r_a y C_a . Sin embargo, se ha encontrado que a las frecuencias en que los efectos de r_a y C_a se hacen sentir, la ganancia ha tenido una reducción tan substancial que el amplificador ya no es utilizable. En este caso los efectos de r_a y C_a pueden ser ignorados, considerándolos como circuitos abiertos y $g_m V_a$ puede ser reemplazado por $g_m V_{gs}$.

Los elementos r_s , r_d y C_{gd} se incluyen para hacer del modelo, un modelo más cercano a las mediciones; por ejemplo, r_s toma en cuenta alguna resistencia en serie con la fuente. Aún así, sus efectos son usualmente pequeños en el rango utilizable de frecuencias y pueden ser omitidos en el modelo lineal.

En general, para ser completamente precisos, un modelo lineal debería tener infinitos elementos, pero entonces, sería impráctico para utilizarse y las aproximaciones se hacen para obtener circuitos empleables. Los diodos D_1 y D_2 , de la figura F3-9, representan el aislamiento entre el FET y el sustrato. Estos diodos están, usualmente, inversamente polarizados y frecuentemente se omiten y se reemplazan por capacitores en el modelo. Por lo tanto un modelo lineal muy manejable es el que se muestra en la figura F3-10(a). Otra forma de este modelo se muestra en la figura F3-10(b). Aquí el generador de corriente en paralelo con r_a se reemplaza por un generador de voltaje en serie con r_a .

Amplificador FET de fuente común

Calculemos la admitancia de entrada del amplificador de fuente común mostrado en la figura F3-11(a). Usando la figura F3-11(b), obtenemos,

$$Y_i = \frac{I_1}{V_{gs}} \quad 3-35$$

$$\text{donde } I_1 = I_{gs} + I_{gd} \quad 3-36$$

De la figura F3-11(b) tenemos,

$$I_{gs} = V_{gs} j\omega C_{gs} \quad 3-37$$

$$I_{gd} = (V_{gs} - V_d) j\omega C_{gd} \quad 3-38$$

El voltaje V_d es desconocido pero podemos expresarlo en términos de V_{gs} y de la amplificación de voltaje A_v ,

$$A_v = \frac{V_d}{V_{gs}} \quad 3-39$$

$$\text{entonces, } I_{gd} = V_{gs} (1 - A_v) j\omega C_{gd} \quad 3-40$$

sustituyendo 3-37 y 3-40 en 3-36 y luego en 3-35 obtenemos,

$$Y_i = j\omega [C_{gs} + (1 - A_v) C_{gd}] \quad 3-41$$

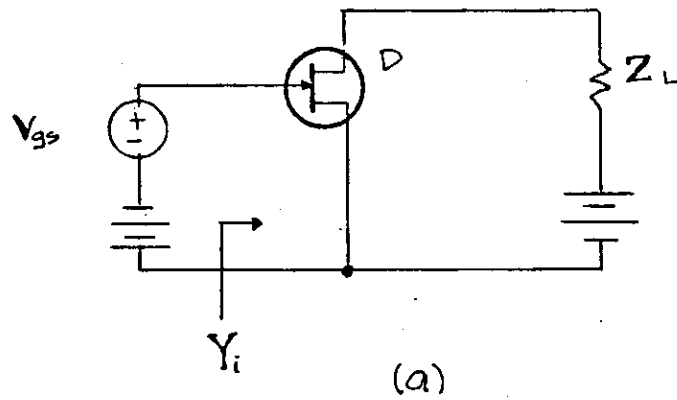
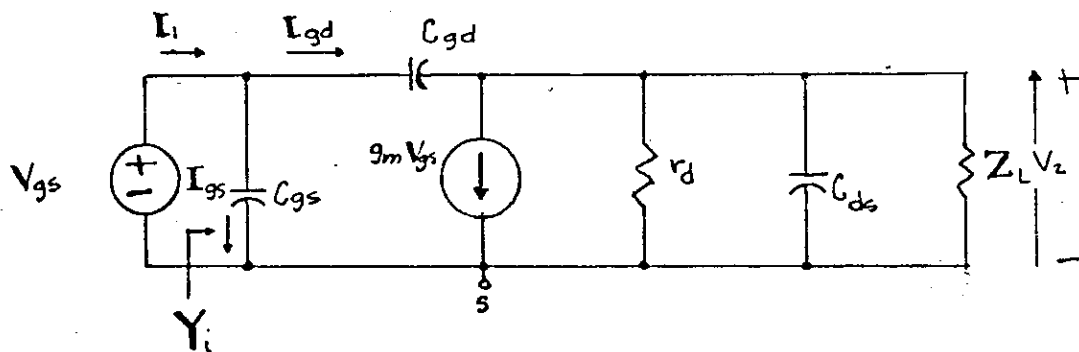


FIGURA F3-11: a) amplificador FET de fuente común, elemental y b) su modelo lineal de alta frecuencia.



La ganancia de voltaje es un número complejo. Escribámoslo como la suma de su parte real y la imaginaria,

$$A_v = A_R + j A_x \quad 3-42$$

Sustituyamos esto en 3-41 y separemos en componentes reales e imaginarias, tenemos

$$G_i = \text{Re}(Y_i) = \omega A_x C_{gd} \quad 3-43$$

$$B_i = \text{Im}(Y_i) = \omega [C_{gs} + (1 - A_R) C_{gd}] \quad 3-44$$

Las ecuaciones 3-43 y 3-44 representan la conductancia G_i en paralelo con una susceptancia $j B_i$. Esta susceptancia representa una capacitancia. Es la capacitancia de entrada C_i del amplificador, donde

$$C_i = C_{gs} + (1 - A_R) C_{gd} \quad 3-45$$

Usualmente, A_v tiene una componente real negativa cuya magnitud es mucho mayor que uno. En este caso, A_R es un gran número negativo y C_i es mucho mayor que las capacitancias que podrían esperarse de C_{gs} y C_{gd} . Este incremento en la capacitancia se denomina efecto Miller; como se desea una pequeña capacitancia de entrada, ambos, C_{gs} y C_{gd} deben ser pequeños. Es especialmente importante que C_{gd} sea pequeña pues está multiplicada por $(1 - A_R)$. En un FET, C_{gs} debe ser del orden de picofaradios.

También es deseable que la conductancia G_i sea pequeña; esto una vez más demuestra la necesidad de que C_{gd} sea pequeña. En general, A_x puede ser positiva o negativa. Si es negativa, resulta una conductancia negativa.

Para usar las ecuaciones 3-43 y 3-44, se debe conocer A_v . Un análisis de la figura F3-11(b) da,

$$A_v = \frac{V_z}{V_{gs}} = \frac{-g_m + j\omega C_{gd}}{j\omega(C_{gd} + C_{gs}) + r^{-1} + Z_i^{-1}} \quad 3-46$$

Amplificador FET de drenador común

Una de las mayores ventajas del amplificador con drenador común es que su admitancia de entrada es mucho menor que en cualquier otra configuración, especialmente cuando las capacitancias parásitas son considerables. Este amplificador también se llama seguidor de fuente y se ilustra en la figura F3-12. La admitancia de entrada se encuentra de forma similar a como se encontró en el amplificador de fuente común. Entonces,

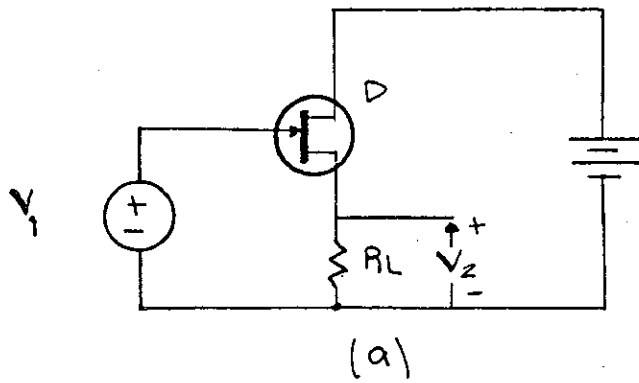
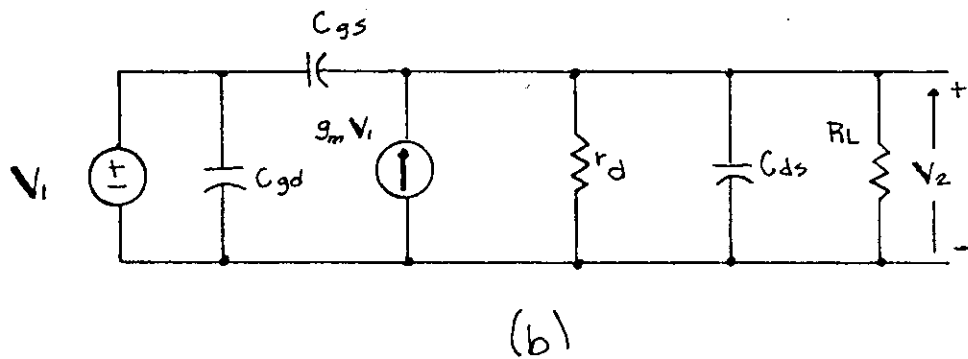


FIGURA F3-12: a) amplificador FET de drenador común y b) su modelo lineal de alta frecuencia.



$$Y_i = j\omega [C_{gd} + (1 - A_v) C_{gs}] \quad 3-47$$

donde $A_v = v_2/v_1$. La ecuación 3-47 es de la misma forma que 3-41 excepto porque C_{gd} y C_{gs} están intercambiados. Por lo tanto en el seguidor de fuente, A_v puede ser considerado muy cercano a la unidad y el factor $(1 - A_v)$ será muy pequeño. Entonces, la admitancia de entrada en 3-47 puede aproximarse como

$$Y_i \cong j\omega C_{gd} \quad 3-48$$

En general, A_v es un número complejo que varía con la frecuencia, por lo que esta aproximación es válida sólo para aquellas frecuencias donde

$A_v \cong 1$. Aún así, esta aproximación usualmente es buena para la mayor parte de frecuencias de interés.

4. DISCUSION FINAL

Apreciaciones sobre los dispositivos semiconductores

Diodos de unión. Si la temperatura de una unión P-N, llega a ser muy alta, el dispositivo operará impropriadamente, porque se generarán térmicamente muchos portadores de carga. Más aún, el exceso de temperatura, destruirá el dispositivo. La temperatura es una función de la potencia disipada en una unión P-N. La potencia instantánea disipada en una unión es igual al producto del voltaje y la corriente a través de la unión,

$$P_a = v_b i_b \quad 1$$

La temperatura depende de la potencia promedio. Si v_b e i_b varían periódicamente con período T, entonces, la potencia promedio es,

$$P_a = \frac{1}{T} \int_0^T v_b i_b dt \quad 2$$

Bajo condiciones estáticas, la potencia promedio estará dada por

$$P_a = V_{bQ} I_{bQ} \quad 3$$

La máxima potencia promedio permitida, se especifica en los manuales de dispositivos semiconductores.

El voltaje de unión puede ser expresado como una función de la corriente. Entonces, la potencia puede determinarse si se conocen las corrientes. Por esta razón, una corriente máxima siempre se especifica.

La temperatura en la unión depende no sólo de la potencia disipada en la unión, sino también de los medios que se provean para remover el calor. Los fabricantes, usualmente especifican métodos apropiados para remover este calor de los dispositivos de alta potencia.

Si el voltaje inverso excede el rompimiento, el diodo no funcionará correctamente, por lo tanto, el voltaje inverso máximo también se especifica. En un rectificador, éste voltaje inverso máximo se llama voltaje inverso de pico.

Transistores de unión. Las apreciaciones de potencia para cada unión P-N son esencialmente las mismas que para los diodos de unión. Para las condiciones de operación normal, las corrientes de emisor y colector son casi las mismas, mientras que el voltaje de colector-base es mucho mayor que el voltaje de emisor-base. Entonces, la potencia disipada en la unión de colector-base será mucho mayor que la disipada en la unión emisor-base.

Relaciones de la forma de 1, 2 y 3 pueden usarse para calcular la disipación entre el transistor cuando se utiliza el voltaje y la corriente apropiada.

La unión colector-base está inversamente polarizada en la operación normal del transistor. Este voltaje de unión no debe exceder el valor donde se produce el rompimiento de avalancha.

La corriente de colector no puede ser directamente relacionada con

la potencia disipada en la unión colector-base, pues el voltaje de colector es una función de la corriente de entrada. Aún así, una corriente máxima se especifica algunas veces. Si ésta se excede, para condiciones normales de funcionamiento, se excederán los demás parámetros.

Transistor de efecto de campo. La potencia promedio disipada en el canal del FET está dado por las ecuaciones 1, 2 y 3, donde la corriente es ahora I_D , la corriente de drenador, y el voltaje es V_{DS} , el voltaje drenador-fuente.

Si V_{DS} se hace muy grande, puede ocurrir un rompimiento por avalancha en las regiones de deplexión del FET de unión. Esto provoca un crecimiento abrupto en I_D lo que produce un funcionamiento imperfecto. Por lo tanto, se debe especificar el valor máximo de V_{DS} .

Si V_{GS} se hace muy grande en un FET de unión, puede ocurrir rompimiento por avalancha en las uniones P-N entre la puerta y el canal y entonces, el FET ya no tendrá una impedancia de entrada alta ni funcionará correctamente. Debe estar especificado, entonces, el máximo valor de V_{GS} .

5. BIBLIOGRAFIA

BIBLIOGRAFIA

- Brophy, J. James. Electrónica fundamental para científicos. Editorial Reverté, S. A.
1974
- Chirlian, Paul M. Electronic circuits: physical principles, analysis, and design. McGraw-Hill Book Company.
1971
- Dekker, A. J. Solid state physics. Prentice-Hall, Inc.
1963
- Delaney, C.F.G. Electronics for the physicist. Penguin Books Inc., Baltimore, Maryland.
1969
- Gray, Paul E.; Searle, Campbell L. Principios de electrónica. Electrónica física, modelos y circuitos electrónicos. Editorial Reverté.
1973
- Hall, H. E. Física del estado sólido. Editorial Limusa, México.
1978
- Hemenway, Henry y Caulton. Física electrónica. Editorial Limusa, México.
1976
- Mandl, Matthew. Solid-state circuit design. Users' manual. Reston Publishing Company, Inc., A Prentice-Hall Company, Reston, Virginia.
1977
- Mckelevy, John P. Física del estado sólido y de semiconductores. Editorial Limusa, México.
1976
- Millman-Halkias. Integrated electronics. International Student Edition. McGraw-Hill Kogakusha, Ltd.
1972
- Millman-Halkias. Dispositivos y circuitos electrónicos. Ediciones Pirámide, S. A., Madrid.
1977
- Reitz-Milford. Fundamentos de la teoría electromagnética. Unión Tipográfica, Editorial Hispano Americana, México.
1972
- Robinson, Vester. Manual of solid state circuit design and troubleshooting. Reston Publishing Company, Inc., Reston, Virginia. A Prentice Hall Company.
1977
- Shockley, William. Electrons and holes in semiconductors with applications to transistor electronics. D. Van Nostrand Company, Inc.
1953

A P E N D I C E S

APENDICE A

Desarrollo de las ecuaciones 1-10 y 1-11

Para un electrón en un cubo de metal, cada componente del momento es ta cuantizada. Es decir (ver figura FA-1)

$$P_x = N_x \rho \qquad P_y = N_y \rho \qquad P_z = N_z \rho \qquad \text{A-1}$$

donde $\rho \equiv h/2L$ y N_x, N_y, N_z son enteros positivos. Podemos obtener una representación gráfica conveniente, si construimos tres ejes perpendicula res entre sí, denominados P_x, P_y y P_z . Este "volumen" se llama espacio de momentos. Los únicos puntos posibles que pueden ser ocupados por un electrón en el espacio de momentums son los dados por la ecuación (A.1). Estos se indican en la figura, donde para más claridad se han representado puntos solamente en un plano, para un valor fijo de P_z (por ejemplo $P_z = 2\rho$). Por el principio de exclusión de Pauli, nunca dos electrones en un metal pueden tener los mismos cuatro números cuánticos N_x, N_y, N_z y el espín s . Por tanto, cada punto de la figura, representa dos e lectrones, uno con $s = 1/2$ y el otro con $s = -1/2$. Calculemos ahora la función de estados $N(E)$. Puesto que en el volumen ρ^3 del espacio de mo mentums, hay solamente un punto en la figura, la densidad de electrones en este espacio es de $2/\rho^3$. El módulo del momentum es $P = (P_x^2 + P_y^2 + P_z^2)^{1/2}$. El número de electrones con momentums comprendidos entre P y $P + dP$ son los que se encuentran en la capa esférica sombreada de la figura. Este número es:

$$\left(\frac{2}{\rho^3}\right)(4\pi P^2 dP) \left(\frac{1}{8}\right) = \frac{\pi P^2 dP}{(h/2L)^3} = \frac{8\pi L^3 P^2 dP}{h^3} \qquad \text{A-2}$$

Se ha introducido el factor $1/8$ en la ecuación anterior, porque sólo son posibles los valores positivos de N_x, N_y y N_z y, por tanto, debe utilizarse solamente la parte de la capa comprendida en el primer octante.

Si W es la energía (en joules), entonces $W = P^2/2m$. De donde,

$$P = (2mW)^{1/2} \qquad P dP = m dW \qquad P^2 dP = 2^{1/2} m^{3/2} W^{1/2} dW \qquad \text{A-3}$$

Si $N(W)$ es la densidad de estados (por metro cúbico), teniendo en cuenta que el volumen del metal es L^3 , se obtiene de (A.2),

$$N(W) dW = \frac{8\pi L^3 P^2 dP}{h^3} \qquad \text{A-4}$$

que da el número de electrones con momentums entre P y $P + dP$, que corresponden a energías comprendidas entre W y $W + dW$. Sustituyendo $P^2 dP$ de (A.3) en (A.4) obtenemos,

$$N(W) dW = \frac{4\pi}{h^3} (2m)^{3/2} W^{1/2} dW$$

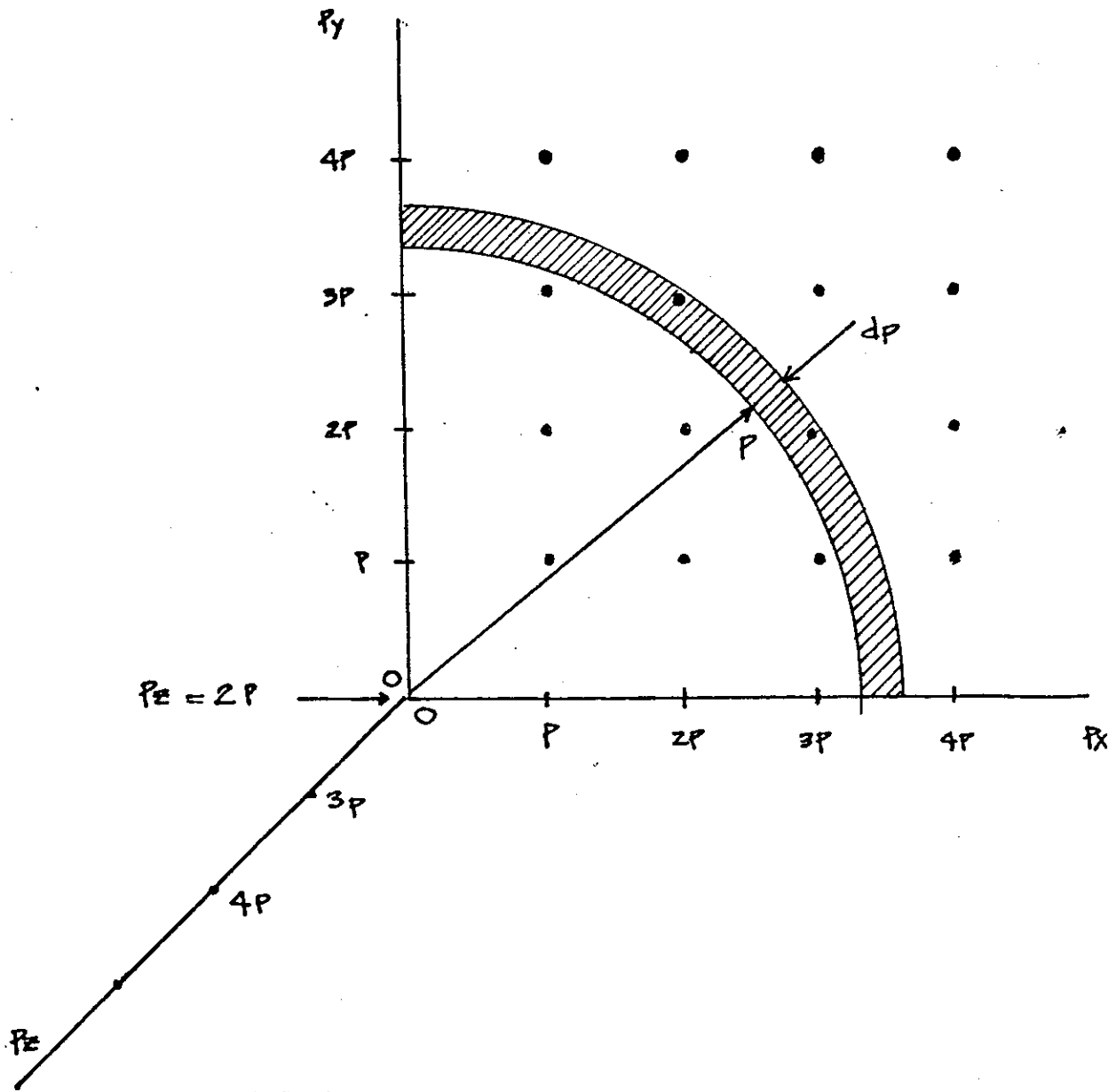


FIGURA FA-1

Si utilizamos electronvoltios en vez de joules como unidad de energía, como $W = 1.6 \times 10^{-19} E$, entonces, $N(E)$ viene dada por:

$$N(E) = \gamma E^{1/2} \quad A-5$$

donde

$$\gamma \equiv \frac{4\pi}{h^3} (2m)^{3/2} (1.6 \times 10^{-19})^{3/2}$$

En un semiconductor, la energía más baja de la banda de conducción es E_c y no cero, por lo tanto debemos modificar la ecuación (A.5), generalizándola:

$$N(E) = \gamma (E - E_c)^{1/2} \quad A-6$$

entonces,

$$N = \int_{E_c}^{\infty} \gamma (E - E_c)^{1/2} e^{-(E - E_f)/kT} dE \quad A-7$$

puesto que

$$N = \int_{E_c}^{\infty} N(E) f(E) dE = \int_{E_c}^{\infty} \frac{\gamma (E - E_c)^{1/2}}{1 + e^{(E - E_f)/kT}} dE$$

y como para $E - E_f \gg kT$ podemos despreciar el uno, caemos en la ecuación (A.7), que al multiplicarla por $e^{(E_c - E_c)/kT}$ es decir, multiplicar por uno, obtenemos:

$$N = \gamma e^{-(E_c - E_f)/kT} \int_{E_c}^{\infty} e^{-(E - E_c)/kT} (E - E_c)^{1/2} dE \quad A-8$$

Esta integral se puede evaluar al efectuar la sustitución,

$$u = \frac{E - E_c}{kT}$$

obtenemos,

$$N = \gamma e^{-(E_c - E_f)/kT} (kT)^{3/2} \int_{E_c}^{\infty} u^{1/2} e^{-u} du$$

que al consultar tablas de integrales definidas, se encuentra que el valor de la integral es $\sqrt{\pi}/2$, por lo tanto

$$N = \gamma (kT)^{3/2} \frac{\sqrt{\pi}}{2} e^{-(E_c - E_f)/kT} = N_c e^{-(E_c - E_f)/kT} \quad A-9$$

donde,

$$N_c = 2 \left(\frac{2\pi m_e kT}{h^2} \right) (1.6 \times 10^{-19})^{3/2}; \quad m_e = \text{masa efectiva del electrón}$$

Similarmente se calcula P cuyo único cambio es que en vez de N_c aparece N_v , que tiene la masa efectiva del agujero m_p , y que en vez de E_c aparece E_v :

$$P = N_v e^{-(E_f - E_v)/kT}$$

A-10

APENDICE B

Efecto Hall

En la figura FB-1 se muestran en forma esquemática los aspectos geométricos de un experimento para el efecto Hall. Se aplica un campo magnético sobre una barra portadora de corriente, en una dirección perpendicular a ésta, produciendo una deflexión de los portadores de carga hacia un lado de la barra. Es importante comprender que para una dirección dada de corriente convencional y campo magnético externo, los portadores de corriente positiva y negativa se desvían en la misma dirección. La deflexión de portadores desarrolla una carga en un lado de la barra. De donde, en la figura superior, el punto A tiene un potencial negativo respecto a A'. Al determinar la polaridad de este potencial, denominado potencial de Hall, se puede saber si los portadores de corriente son, predominantemente, agujeros o electrones.

La deflexión de los portadores de carga no puede continuar indefinidamente, debido a que el desarrollo de la carga crea un campo eléctrico transversal en el cristal. Cuando la fuerza ejercida sobre los portadores de carga, debido a este campo eléctrico transversal, llega a igualarse con la fuerza magnética, se produce el equilibrio. Haciendo que E_T , sea la magnitud del campo eléctrico transversal, la condición de equilibrio es:

$$eE_T = Bev \quad B-1$$

donde v es la velocidad promedio de desplazamiento de los agujeros o electrones en el sentido longitudinal de la barra. Puesto que la movilidad μ es por definición igual a v/E_L , en donde E_L es la magnitud del campo longitudinal, se puede eliminar la cantidad v , que no se puede medir directamente, para obtener

$$\mu = \frac{E_T}{E_L B} \quad B-2$$

El tratamiento anterior, se puede aplicar sólo cuando la densidad de un tipo de portadores es despreciable. La movilidad se puede determinar experimentalmente, al definir la relación entre el campo eléctrico transversal y el longitudinal. En la figura FB-2 se muestra una vista superior de la barra de la figura 1. Puesto que las superficies equipotenciales son perpendiculares a las líneas de campo eléctrico, los pares de puntos tales como B-B' tienen una diferencia de potencial cero entre sí. De acuerdo con la figura 2 se puede ver que $(E_T/E_L) = \tan \theta$, en donde θ es el ángulo de inclinación o de Hall de las superficies equipotenciales, de manera que:

$$\mu = \frac{\tan \theta}{B} \quad B-3$$

La medida de la magnitud del efecto de Hall se da, casi siempre, en términos de una cantidad denominada coeficiente de Hall. Partiendo de

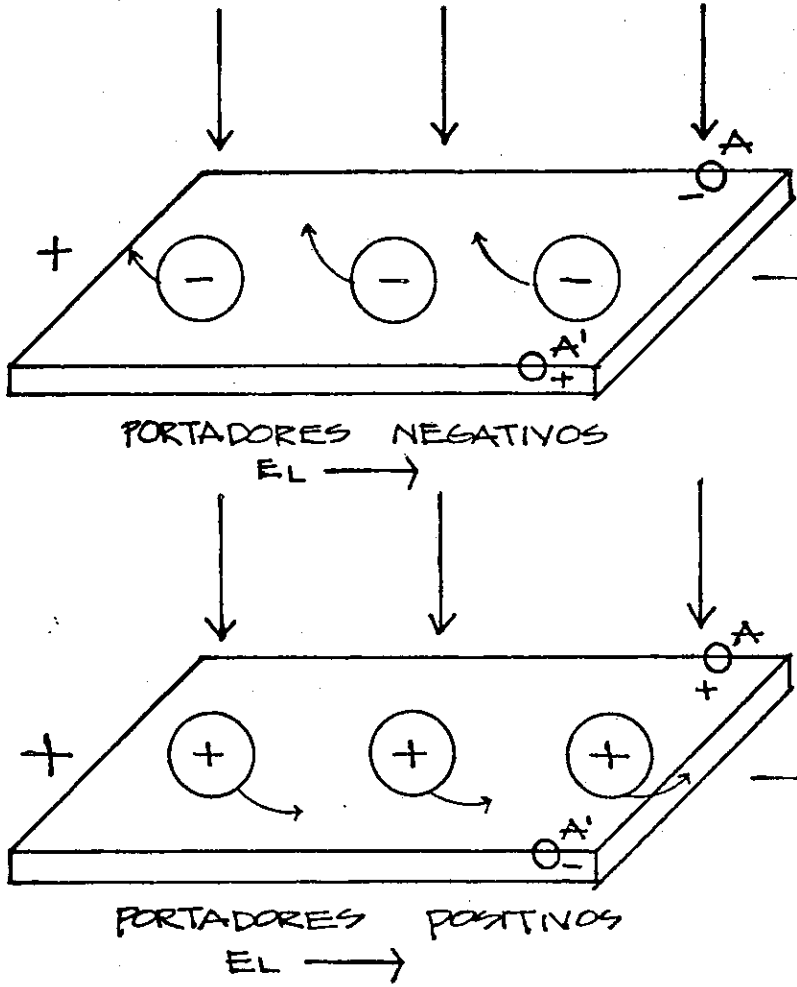
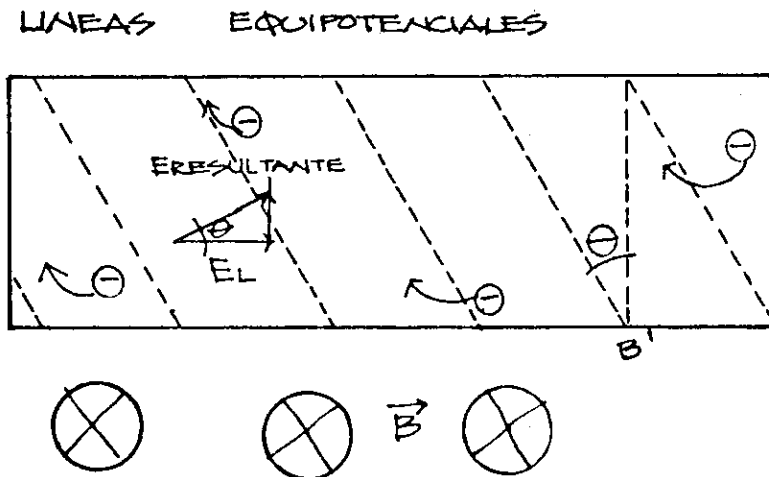


FIGURA FB-1

FIGURA FB-2



$\vec{J} = Ne \mu E_L$ y sustituyéndo (B.2), se puede expresar

$$E_T = \mu E_L B = \frac{J}{Ne} B = R_H JB$$

B-4

El coeficiente de Hall, R_H , se define como:

$$R_H \equiv \frac{E_T}{JB} = \frac{1}{Ne}$$

B-5

APENDICE C

Ecuación de Continuidad. Leyes de Fick. Relaciones de Einstein.

El propósito de la ecuación de continuidad es describir la razón de cambio, en el tiempo, de la densidad numérica de partículas en un elemento de volumen del espacio. (ver figura FC-1).

Considérese una superficie cerrada S de esa región, con objeto de calcular la velocidad de flujo de las partículas al entrar o salir del volumen rodeado por esa superficie. En la figura se muestra un elemento de área $d\vec{A}$ sobre la superficie S . Las partículas que cruzan el elemento de área $d\vec{A}$ durante el siguiente intervalo de tiempo dt , deben quedar dentro de un paralelepípedo cuya base es $d\vec{A}$ y de altura inclinada $\vec{u} dt$. El volumen del paralelepípedo que contiene a las partículas que cruzarán $d\vec{A}$ en el siguiente intervalo dt , es $\vec{u} \cdot d\vec{A} dt$. Si se multiplica este volumen por el número de partículas por unidad de volumen M que quedan cerca del elemento $d\vec{A}$, se obtiene la cantidad que cruza este elemento en el tiempo dt . En consecuencia, la cantidad de partículas que cruza $d\vec{A}$ por unidad de tiempo es $M \vec{u} \cdot d\vec{A}$. El producto $M \vec{u}$ se conoce también como densidad de flujo de las partículas que se ha denotado como \vec{J} .

Si \vec{u} está dirigida hacia el exterior, entonces, el flujo de partículas que pasa por $d\vec{A}$ contribuye a disminuir el número de partículas existentes en el interior de S . Si \vec{u} se dirige hacia el interior, el flujo que pasa por $d\vec{A}$ contribuye a un aumento en la cantidad de partículas que están en el interior de S . En general las partículas pueden fluir hacia adentro a través de las partes de la superficie S y, hacia afuera, a través de otras partes de la misma. La cantidad $M \vec{u} \cdot d\vec{A}$ es positiva cuando \vec{u} se dirige hacia el exterior de la superficie, y negativa cuando se dirige hacia el interior. En consecuencia, la suma de los términos $M \vec{u} \cdot d\vec{A}$ de cada elemento $d\vec{A}$ de la superficie cerrada, da la velocidad neta con que las partículas fluyen hacia afuera. Es decir, velocidad de flujo de salida a través de $S = \int_S M \vec{u} \cdot d\vec{A} = \int_S \vec{J} \cdot d\vec{A}$ C-1

El número de partículas dentro de S se puede expresar como $\bar{M} V$, donde \bar{M} es la densidad numérica promedio de las partículas y V el volumen encerrado por S . La velocidad con que disminuye la cantidad de partículas dentro de S , es numéricamente igual a la velocidad con que fluyen las partículas hacia el exterior a través de la superficie, en el caso especial de ninguna generación o recombinación neta de partículas dentro de V , se tiene,

$$\frac{\partial(\bar{M}V)}{\partial t} = - \int_S \vec{J} \cdot d\vec{A} \quad \text{ninguna generación o recombinación} \quad \text{C-2}$$

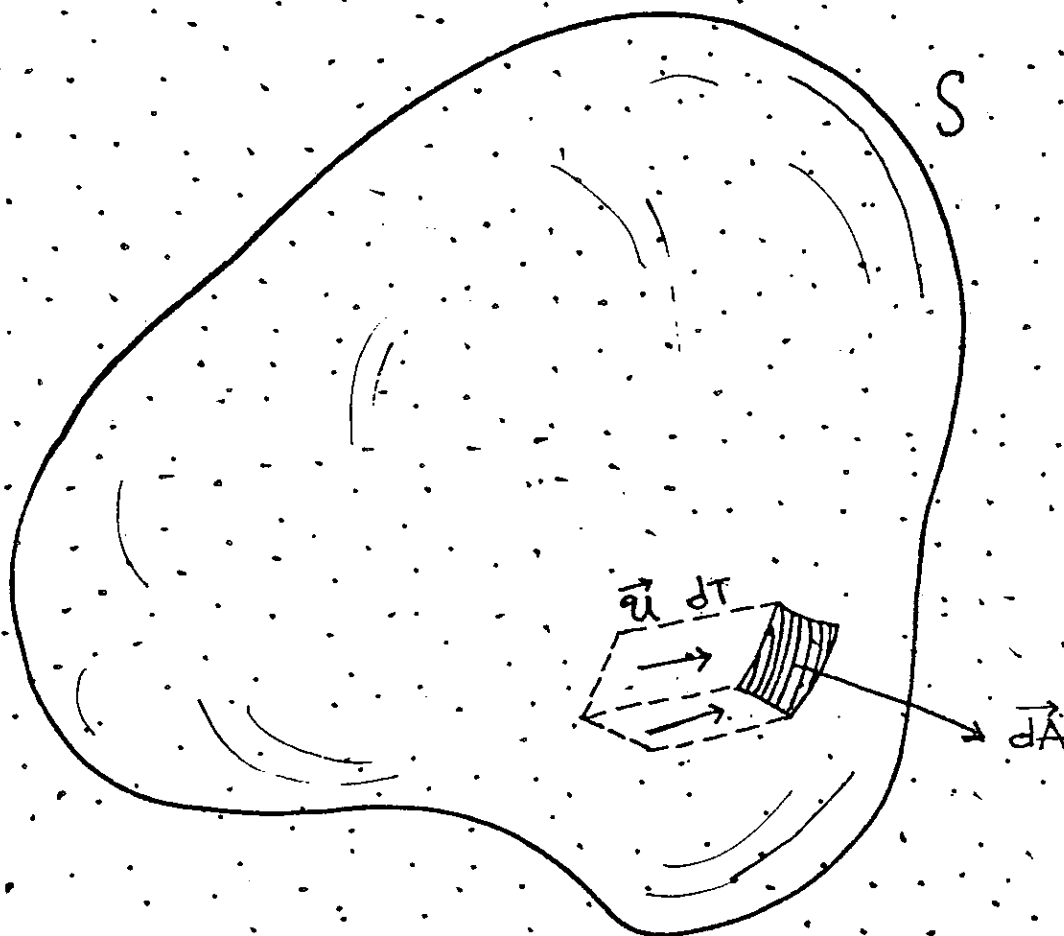


FIGURA FC-1

El signo menos de (C.2) se produce debido a que $\bar{M}V$ disminuye cuando la integral o término del flujo es positivo. Si existe una generación o recombinación neta de partículas, la velocidad neta de producción se puede expresar como $\bar{C}V$, donde \bar{C} es una función que representa la velocidad neta promedio de creación por unidad de volumen. Si \bar{C} es negativa, representa una recombinación neta. Para casos generales, la razón de cambio del número de partículas dentro de S es:

$$\frac{\partial(\bar{M}V)}{\partial t} = - \int_S \vec{\Phi} \cdot d\vec{A} + \bar{C}V \quad C-3$$

Si se divide (C.3) entre V y se hace que éste sea infinitamente pequeño, los promedios de \bar{M} y \bar{C} se pueden sustituir por sus valores M y C en un punto dado. El primer término de la derecha de (C.3) (dividido entre V) viene a ser, por definición, la divergencia de $\vec{\Phi}$. Por tanto se puede expresar lo siguiente:

$$\frac{\partial M}{\partial t} = - \nabla \cdot \vec{\Phi} + C \quad C-4$$

Esta ecuación diferencial parcial en donde se combinan conceptos geométricos y cinemáticos, se conoce como ecuación de continuidad.

Difusión es el nombre que se da al flujo neto de partículas que se mueven al azar, desde las regiones de alta concentración o densidad, hasta las de menor concentración. Se ha encontrado experimentalmente, que en la mayoría de las circunstancias, la densidad de flujo de las partículas $\vec{\Phi} (M \vec{u})$ debido a la difusión, es proporcional a la máxima derivada direccional o gradiente, de la densidad numérica de dichas partículas. Más aún, puesto que el flujo va de concentraciones altas a bajas, la dirección de la densidad del flujo es opuesta a la del gradiente, entonces, con notación vectorial,

$$\vec{\Phi} = -D \nabla M \quad C-5$$

En este caso, D es una constante de proporcionalidad conocida como constante de difusión. La ecuación (C.5) se denomina primera ley de Fick.

La ecuación (C.5), que describe la difusión, se puede combinar con la de continuidad (C.4), para obtener

$$\frac{\partial M}{\partial t} = D \nabla^2 M + C \quad C-6$$

Esta ecuación se conoce como segunda ley de Fick.

Vamos ahora a obtener una expresión para la constante de difusión D , correspondiente a partículas cargadas que se desplazan entre átomos dispersos, en función de la movilidad y la temperatura. Aunque la deducción se efectúa considerando un estado de equilibrio térmico, el resultado final, que es una relación entre ciertas constantes, tiene una validez gene

ral y se puede emplear incluso en situaciones de ausencia de equilibrio.

En este modelo, se supondrá una caja rectangular que se extiende desde $x=0$ a $x=x_0$ y que contiene una nube de partículas positivamente cargadas cuya carga es q , que están en libertad para moverse entre átomos dispersos uniformemente distribuidos, a una temperatura absoluta T . Además se tiene un campo eléctrico uniforme, aplicado externamente en la dirección negativa de x , y que tiende a hacer que las partículas cargadas "se amontonen" en el extremo de la caja correspondiente a $x=0$. Si arbitrariamente se hace que la energía potencial de una partícula cargada sea cero; cuando la partícula está en $x=0$ y que $V(x)$ sea el potencial en x debido a la suma del campo eléctrico externo y el campo eléctrico debido a la carga espacial, la energía potencial de una partícula en x se puede expresar como:

$$U = qV(x) \quad \text{C-7}$$

El número de partículas más probable, en cada estado cuántico de un sistema a la temperatura T , obedece a la estadística de Maxwell-Boltzmann y se expresa como:

$$P_{m_0(i)} = A e^{-E_i/kT} \quad \text{C-8}$$

donde E_i es la energía de un estado y A es una constante. La energía de cada partícula del sistema observado se compone tanto de la energía cinética K , como de la potencial U ,

$$E = K + U = K + qV(x) \quad \text{C-9}$$

El número total de partículas dM contenidas en un elemento de volumen Sdx se puede encontrar usando (C.8), sumando P_{m_0} a lo largo de todos los estados cuánticos disponibles para las partículas en el elemento de volumen. Esta suma equivale a sumar todas las energías cinéticas de que dispone una partícula en Sdx , debido a que la energía potencial $qV(x)$ es básicamente la misma para cada partícula de ese volumen,

$$dM = \sum_i P_{m_0} = \sum_i A e^{-[K_i + qV(x)]/kT} = e^{-qV(x)/kT} \sum_i A e^{-K_i/kT} \quad \text{C-10}$$

Si la temperatura es la misma para cada elemento de volumen, y si cada elemento se considera de igual tamaño, el factor $\sum_i A e^{-K_i/kT}$ es independiente de x . De este modo, la densidad de partículas en cada elemento es proporcional a $e^{-qV(x)/kT}$ y se puede expresar de la siguiente forma:

$$M(x) = M_0 e^{-qV(x)/kT} \quad \text{C-11}$$

donde la constante de proporcionalidad M_0 se puede interpretar como la densidad en $x=0$. Esta ecuación es conocida como relación de Boltzmann.

Por la primera ley de Fick sabemos que,

$$M u = -D \frac{\partial M}{\partial x} \quad \text{C-12}$$

(en una dimensión). Como E va en el sentido negativo del eje x , es una cantidad negativa y entonces tenemos que $q = -\mu E$. Por lo tanto,

$$-\mu E = -D \frac{\partial M}{\partial x} \quad \text{C-13}$$

Derivando (C.11),

$$\frac{\partial M}{\partial x} = \frac{M_0 q}{kT} \left(-\frac{dV(x)}{dx} \right) e^{-qV(x)/kT} = \frac{qEM}{kT} \quad \text{C-14}$$

Sustituyendo (C.14) en (C.13) y anulando los factores comunes, obtenemos,

$$D = \frac{\mu kT}{q} \quad \text{C-15}$$