
INTRODUCCION A LOS DISPOSITIVOS SEMICONDUCTORES:
PRINCIPIOS Y MODELOS



INTRODUCCIÓN A LOS DISPOSITIVOS SEMICONDUCTORES: PRINCIPIOS Y MODELOS

Pedro Julián



Editorial de la Universidad Nacional del Sur

Serie: Docencia

*A aquellos que hoy no están,
y que, con sus vidas, me mostraron el camino:*

*Mis padres, Haydée y Félix,
y Alfredo Desages*

*A quien camina conmigo hoy,
mi esposa Ana Laura;
y a nuestros pequeños hijos,
que nos siguen*

Índice general

Prefacio	VII
1. Modelos de Circuitos Eléctricos	1
1.1. Bloques constitutivos de modelos	2
1.1.1. Resistores	2
1.1.2. Capacitores	2
1.1.3. Inductores	3
1.1.4. Memristores	4
1.1.5. Fuentes independientes	5
1.1.6. Fuentes controladas	6
1.1.7. Convenciones	6
1.2. Clasificación de modelos	6
1.2.1. Modelos según la amplitud	6
1.2.2. Modelos según la frecuencia	8
1.2.3. Construcción de modelos	12
2. Introducción a los semiconductores	17
2.1. Bandas de Energía en Silicio	17
2.2. Equilibrio Térmico	21
2.3. Dopado	25
2.3.1. Dopado Tipo N	26
2.3.2. Dopado Tipo P	31
2.3.3. Compensación	33
2.4. Mecanismos de conducción	35
2.4.1. Arrastre	35
2.4.2. Difusión	40
2.4.3. Resistividad de una lámina de Silicio	44
2.5. Potenciales relativos en Silicio	46
A. Conducción: Conceptos Auxiliares	49
A.1. Tiempo de tránsito	49
A.2. Efecto Hall	49
B. Electroestática	51
C. Potenciales de contacto	53

3. Juntura Semiconductora y Diodos	55
3.1. Descripción Cualitativa	56
3.2. Electroestática de la Juntura	62
3.3. Modelo de DC	63
3.3.1. Ley de la Juntura	65
3.3.2. Solución en directa	68
3.3.3. Solución en inversa	70
3.3.4. Desviaciones del comportamiento ideal	72
3.4. Modelo Lineal Incremental	74
3.5. Modelo de AC	74
3.5.1. Capacidad en inversa	77
3.5.2. Capacidad en directa	79
3.6. Mecanismos de ruptura inversa	80
3.6.1. Efecto Túnel	80
3.6.2. Avalancha	83
4. Capacitor MOS	83
4.1. Descripción Cualitativa	84
4.2. Electroestática del capacitor MOS	84
4.2.1. Potencial de Banda Plana	85
4.2.2. Acumulación	88
4.2.3. Vaciamiento	91
4.2.4. Inversión	95
4.3. Modelo de AC	97
4.4. Otras configuraciones	98
4.4.1. Capacitor MOS sobre un sustrato P y gate P	98
4.4.2. Capacitor MOS sobre un sustrato N y gate N	100
4.4.3. Capacitor MOS sobre un sustrato N y gate P	103
5. Transistor MOS	103
5.1. Descripción cualitativa	103
5.2. El transistor NMOS	103
5.2.1. Principio básico de funcionamiento	106
5.2.2. Modelo referido al Sustrato	116
5.2.3. Modelo referido al Source	119
5.2.4. Desviaciones del comportamiento ideal	122
5.3. El transistor PMOS	122
5.3.1. Principio básico de funcionamiento	123
5.3.2. Modelo referido al Sustrato	127
5.3.3. Modelo referido al Source	130
5.3.4. Desviaciones del comportamiento ideal	131
5.4. Modelo lineal incremental	131
5.4.1. MLI referido al sustrato	136
5.4.2. MLI referido al source	138
5.5. Modelo de AC	138
5.5.1. Capacidad de gate	140
5.5.2. Capacidad de junturas	142
5.5.3. Capacidad de solapamiento	143
5.5.4. Límite de validez del modelo de AC	143

6. Transistores Bipolares	147
6.1. Descripción Cualitativa	147
6.2. Modelo de DC del transistor PNP	148
6.2.1. Región de conducción activa directa	148
6.2.2. Región de conducción activa inversa	156
6.2.3. Región de saturación y el modelo de Ebers-Moll	160
6.2.4. Modelos Simplificados	165
6.2.5. Desviaciones del comportamiento ideal	170
6.3. Modelo de DC del transistor NPN	173
6.3.1. Región de conducción activa directa	173
6.3.2. Región de conducción activa inversa	179
6.3.3. Región de saturación y el Modelo de Ebers-Moll	183
6.3.4. Modelos Simplificados	186
6.4. Modelo Lineal Incremental (MLI)	188
6.5. Modelo de AC	192
6.5.1. Capacidad de vaciamiento	192
6.5.2. Capacidad de carga de la base	192
6.5.3. Límite de validez del modelo de AC	193
A. Modelos de SPICE	197
A.1. Fuentes	197
A.1.1. Fuentes independientes	197
A.1.2. Fuentes dependientes	197
A.2. Dispositivos pasivos	197
A.2.1. Resistencias	197
A.2.2. Capacitores	198
A.2.3. Inductores	198
A.2.4. Inductores mutuos	198
A.3. Dispositivos semiconductores	198
A.3.1. Diodo	199
A.3.2. Transistores bipolares	200
A.3.3. Transistores MOS	201

Prefacio

A partir de los años setenta, la tecnología de circuitos integrados basada en transistores Complementarios Metal-Oxido-Semiconductor (CMOS) superó en utilización a la tecnología basada en transistores bipolares de juntura. Desde ese momento la tecnología CMOS se convirtió en el pilar del diseño de circuitos integrados modernos, tanto analógicos como digitales. Los contenidos curriculares de las carreras de ingeniería electrónica en los países líderes en diseño electrónico se adaptaron a la nueva tecnología. En Argentina, así como en la mayoría de los países de hispanoamérica, este cambio no se dio, manteniéndose en los cursos básicos el enfoque previo basado en transistores bipolares y componentes discretos.

Este libro se basa en más de diez años de experiencia en el dictado de las materias “Dispositivos Semiconductores” y “Análisis y Diseño de Circuitos Digitales”, materias de tercer y quinto año respectivamente, de la carrera de Ingeniería Electrónica, de la Universidad del Sur, desde el año 1999 hasta la fecha; y refleja también la experiencia adquirida a través de la enseñanza de los cursos de la Escuela Argentina de Micro-Nanoelectrónica, Tecnología y Aplicaciones (EAMTA), llevada a cabo desde el año 2006 hasta la fecha.

El objetivo central de este libro es proveer el material básico en castellano para la enseñanza inicial a nivel de grado de microelectrónica para carreras de ingeniería electrónica e ingenierías afines.

El alcance del libro comprende la descripción de los dispositivos básicos microelectrónicos, desde una perspectiva útil para un diseñador de circuitos. El libro tiene como objetivo introducir la electrónica de estado sólido, proveyendo los modelos básicos que más adelante el estudiante requiere para abordar el diseño de etapas básicas de circuitos integrados, tanto analógicas como digitales. Se describen los principios y modelos de cuatro dispositivos fundamentales: la juntura semiconductor, el capacitor MOS, el transistor MOS y el transistor bipolar de juntura. Para todos los dispositivos mencionados se describen los principios básicos de operación y se desarrollan las ecuaciones que modelan su comportamiento en corriente continua (DC), señalando claramente los rangos de funcionamiento. Estos contenidos proveen la base para el cálculo de puntos de operación (polarización) de circuitos en cursos posteriores. A continuación se desarrollan los modelos lineales incrementales, es decir, aquellos modelos destinados a representar el dispositivo en un pequeño entorno de un punto de trabajo. Estos modelos son la base para el análisis de circuitos en pequeña señal, que incluye cálculo de ganancia, sensibilidad, impedancia, etc. Por último, se derivan las expresiones de las cargas eléctricas acumuladas para cada uno de los distintos dispositivos, y se desarrollan las expresiones de las capacidades asociadas teniendo en cuenta la región de operación. La inclusión de las capacidades de los dispositivos es fundamental para el análisis y diseño frecuencial de circuitos (AC), en particular de filtros.

Hay dos características distintivas acerca del enfoque adoptado en este libro. La primera característica es la utilización de resultados experimentales para ilustrar curvas y características de los dispositivos y circuitos utilizados. De esta manera se pretende que el estudiante se familiarice con valores típicos de corrientes, tensiones y tiempos correspondientes a dispositivos reales de circuitos integrados. La segunda característica es la utilización a lo largo del libro de rutinas de MatlabTM para la visualización de ecuaciones. Esto tiene como objetivo integrar la utilización de herramientas auxiliares

de cálculo al proceso de aprendizaje del estudiante.

Organización El libro está organizado de la siguiente manera.

El Capítulo 1 incluye una breve descripción de los bloques constitutivos de circuitos (resistores, capacitores, inductores, memristores, fuentes independientes y fuentes controladas) y establece el marco conceptual dentro del cual se ubican los modelos de los dispositivos. Se introduce la noción de modelos de acuerdo a la amplitud de la señal, definiendo los modelos globales, locales y lineales incrementales, y la noción de modelos de acuerdo a la frecuencia de la señal, definiendo modelos de DC, de baja frecuencia, de media frecuencia y de alta frecuencia.

El Capítulo 2 comienza con una descripción de las propiedades electrónicas del Silicio, sus bandas de energía y la generación de portadores. Luego se describe el dopado de Silicio mediante la introducción de impurezas y se da una interpretación gráfica del resultado. Esto permite abordar los mecanismos de conducción de portadores en semiconductores: arrastre y difusión; y hallar en ambos casos la expresión analítica de la corriente resultante. Se desarrollan a continuación las leyes básicas de electrostática, y las condiciones de borde en las interfaces entre materiales. Se introduce el concepto de potencial relativo entre materiales, el cual resulta de utilidad para los desarrollos de la juntura semiconductor y del capacitor MOS. En base a los potenciales relativos, se halla la ley de la juntura que establece la diferencia de potencial entre los portadores de dos regiones de Silicio en función de los dopados respectivos. Por último, se hace una breve mención a los potenciales que surgen cuando dos materiales diferentes se ponen en contacto, lo cual resulta relevante para describir la unión entre los dispositivos semiconductores y los restantes componentes circuitales.

El Capítulo 3 introduce la juntura semiconductor. Primero se hace un análisis electrostático sin tensión aplicada y se hallan la densidad de carga, el campo eléctrico y el potencial resultante. Luego, se extiende el análisis al caso en que hay una tensión de DC aplicada y se desarrolla la expresión de la corriente de difusión, utilizando la simplificación de base corta, tanto para el caso de tensión directa como para el de tensión inversa. En base al modelo de DC hallado se deriva el modelo lineal incremental. Para el cálculo del modelo de AC se analiza primero el caso de tensión inversa y se halla la carga de vaciamiento, a partir de la cual se deduce la capacidad correspondiente. Luego, se considera el caso de tensión directa y se halla la carga almacenada, producto de la conducción de los portadores minoritarios, a partir de la cual se deduce la capacidad de difusión. Por último, se describen los mecanismos de ruptura inversa: avalancha y efecto túnel; y sus características particulares.

El Capítulo 4 analiza el capacitor MOS, formado entre un metal y un semiconductor separados por un dieléctrico. El análisis se centra en el desarrollo de la electrostática de esta estructura (utilizando polisilicio fuertemente dopado en lugar de metal) y la derivación de las cargas en juego, en función de la tensión aplicada en terminales, para cada uno de los modos de funcionamiento: acumulación, vaciamiento e inversión. Por motivos didácticos, el análisis se comienza a partir del punto de Banda Plana, donde las cargas a lo largo del capacitor son nulas. A continuación se halla el modelo de AC a partir de la expresión de carga hallada previamente. Por último, se resumen las características principales (valores de las tensiones de Banda Plana y de Umbral, y configuraciones de carga resultantes) cuando los dopados del material del gate y del sustrato cambian de tipo y valor.

El Capítulo 5 desarrolla el transistor MOS en sus dos versiones, el transistor de canal N o NMOS, y el transistor de canal P o PMOS. Ambos análisis, si bien siguen lineamientos completamente paralelos, se desarrollan en forma separada y completa. Para ambos dispositivos se siguen los mismos pasos. Para hallar el modelo de DC, se realiza un análisis electrostático donde se halla la expresión genérica de la corriente del transistor en función de la carga de inversión y su velocidad en un punto arbitrario del canal. Luego se integra esta expresión para hallar la corriente en función de las tensiones entre terminales. En este punto se plantean dos modelos de utilidad: el modelo con las tensiones referidas al sustrato, de interés particular para diseño analógico; y el modelo con las tensiones referidas al source,

de uso masivo, sobre todo en el caso de diseño digital. Las regiones de funcionamiento y sus curvas resultantes se ilustran utilizando datos experimentales medidos sobre circuitos integrados de prueba. El modelo lineal incremental se halla a partir del modelo de DC, considerando las regiones de triodo y saturación. También en este caso se ilustran los modelos resultantes de utilizar el sustrato, o el source, como terminal común. Para hallar el modelo de AC, se calculan las distintas capacidades involucradas: la capacidad del gate, las capacidades de solapamiento y las capacidades de vaciamiento de las junturas de source y drain. Para el caso de la capacidad del gate, se tienen en cuenta las diferencias en función del modo de operación del transistor. Por último, se introduce la frecuencia de transición como el límite aproximado de validez del modelo de frecuencias medias.

El Capítulo 6 desarrolla el transistor bipolar de juntura, en sus dos versiones, el transistor PNP y el transistor NPN. Al igual que en el caso del transistor MOS, ambos análisis se desarrollan en forma separada y completa. Para hallar el modelo de DC, se plantean las expresiones de los niveles de portadores minoritarios en función de las tensiones aplicadas, y se procede a hallar la expresión de las corrientes de emisor, colector y base para los tres casos relevantes: directa activa, reversa activa y saturación. A partir del caso de saturación, que se obtiene como una combinación de los casos de activa directa y activa reversa, se plantea el modelo de Ebers Moll, que captura en forma global el comportamiento del transistor bipolar de juntura. También se plantean versiones simplificadas del modelo de Ebers Moll para los casos de activa directa y saturación, de utilidad para la resolución de puntos de funcionamiento de circuitos con transistores. Las regiones de funcionamiento y sus curvas resultantes se ilustran utilizando datos experimentales. A partir del modelo de Ebers Moll, se deriva el modelo lineal incremental y se introduce el modelo híbrido- π . Para hallar el modelo de AC, se calculan dos capacidades: la capacidad de vaciamiento de la juntura base colector, que se halla en inversa; y la capacidad de difusión, producto de la acumulación de portadores minoritarios en la juntura base-emisor, que se halla en directa. De igual manera que en el caso del transistor MOS, se halla la frecuencia de transición como el límite aproximado de validez del modelo de frecuencias medias.

En la mayoría de los libros de texto, se presenta con mayor énfasis el análisis de los transistores NMOS y NPN, es decir, aquellos cuya conducción se basa en electrones, y se resumen de manera sintética las ecuaciones para los dispositivos PMOS y PNP, es decir, aquellos cuya conducción se basa en huecos. El enfoque adoptado en este libro consiste en presentar los desarrollos completos para todos los dispositivos. La explicación principal se realiza para el caso del transistor NMOS y del transistor PNP (es decir, se prioriza un dispositivo con conducción en base a electrones y otro en base a huecos) mientras que para el caso del transistor PMOS y del transistor NPN se realiza una explicación más sintética, pero manteniendo todos los desarrollos. Este enfoque se adopta a raíz de los mejores resultados observados en el aprendizaje de los estudiantes.

Agradecimientos A los colegas que proveyeron numerosas sugerencias, agregados y correcciones a lo largo de la escritura del libro, especialmente a Andrés Aymonino, Ariel Lutenberg, Félix Palumbo, Carlos Dualibe, Fernando Silveira, Alfredo Arnaud, José Lipovetzky, Santiago Sondón, Alfredo Falcón, Angel Soto, Ariel Arelovich, Martín Di Federico, Omar Lifschitz. A los alumnos de Dispositivos Semiconductores y Análisis y Diseño de Circuitos Integrados de la Universidad Nacional del Sur, en particular a Paola Ceminari, Cesar Panzita y Niria Osterman. A mi amigo Luis Bastán, Director Creativo de BDA (www.bdastyle.com), por el arte de tapa.

A mi esposa Ana Laura y a mis tres hermosos hijos, Valentina, Violeta y Baltasar. Por todo su apoyo, y por toda la atención y tiempo que este libro les quitó.

El sitio web Se ha diseñado un sitio web para el libro, que se haya disponible en la dirección de internet <http://www.gisee.uns.edu.ar/LibroDispositivos>. El sitio contiene material de soporte del libro, que incluye, entre otras cosas, ejercicios para cada capítulo, modelos de dispositivos, fe de erratas,

rutinas de Matlab y transparencias para el dictado de clases. El hecho de poner este material en web brinda la ventaja de poder actualizar y revisar este contenido periódicamente.

Capítulo 1

Modelos de Circuitos Eléctricos

Un circuito físico se define como la interconexión de dispositivos físicos, es decir, componentes reales disponibles tales como resistores, amplificadores operacionales, inductores, capacitores, cables, etc. Es necesario distinguir un “dispositivo físico” ó un “circuito físico” de su *modelo*. El *modelo* de un dispositivo físico, es una representación matemática que permite estudiar su comportamiento. De la misma manera, el *modelo* de un circuito físico es una representación matemática que permite el estudio, análisis y resolución de un conjunto de dispositivos interconectados. Hecha esta aclaración, de aquí en adelante se utilizarán las expresiones dispositivo y circuito para hacer referencia a *modelos*.

Sin excepción, cada elemento circuital será una aproximación del dispositivo físico correspondiente, con lo cual, el circuito también será una aproximación del circuito físico. Como sucede con toda aproximación, es necesario determinar, para su correcta utilización, los límites dentro de los cuales ésta produce resultados razonables. En el caso de circuitos eléctricos, ésto se traduce en determinar los rangos de amplitudes y frecuencias dentro de los cuales el modelo circuital aproxima correctamente al circuito físico. La utilización de modelos circuitales fuera de los rangos de validez del circuito físico asociado, es un error típico que se vé agravado por el uso de simuladores. Los simuladores de circuitos basan su funcionamiento en la resolución numérica de las ecuaciones resultantes de la interconexión de elementos circuitales. La solución producida por el simulador tendrá sentido siempre y cuando, para las señales utilizadas, cada uno de los elementos circuitales funcione en todo momento dentro de su región de validez. Esta verificación es algo que recae por completo en el usuario, dado que el simulador se limita a resolver las ecuaciones del circuito y no tiene información acerca de la relación entre el circuito y el circuito físico. En pocas palabras, el simulador resuelve *circuitos* (modelos) y no circuitos físicos. Por definición [1], un elemento circuital es equivalente a un elemento físico cuando ambos son indistinguibles a partir de mediciones eléctricas en sus terminales. Desafortunadamente, no siempre es posible hallar un circuito equivalente, y el ingeniero debe conformarse con un circuito *aproximadamente* equivalente en rangos definidos de funcionamiento. Es importante entender que es una tarea primordial del ingeniero producir un circuito *aproximadamente* equivalente que permita representar el circuito físico para las señales de interés.

Los circuitos eléctricos desarrollados en este libro caen en la categoría de los denominados modelos de parámetros concentrados. Un modelo puede considerarse de parámetros concentrados, cuando la dimensión del dispositivo es pequeña con respecto a las longitudes de onda de las señales aplicadas. En un modelo de parámetros concentrados las ondas electromagnéticas se propagan a través del mismo en forma instantánea, de tal manera que la corriente $i(t)$ que fluye en cualquier terminal y la tensión $v(t)$ entre cualquier par de terminales está bien definida para todo tiempo t [1]. Cada elemento puede considerarse puntual, manteniendo una relación entre la corriente a través del mismo y la tensión entre sus terminales. En otras palabras, en un circuito de parámetros concentrados, las dimensiones y formas de cada uno de sus elementos son irrelevantes. En los casos donde un modelo de parámetros

concentrados no es suficiente, por ejemplo en el caso de una línea de transmisión, donde cada sección de la misma posee un valor $i(t, x)$ de corriente y un valor $v(t, x)$ de tensión, que dependen de la posición x , es necesario recurrir a un modelo de parámetros distribuidos¹. Todos los circuitos utilizados en este libro pueden modelarse adecuadamente (siempre que las frecuencias de trabajo no sean lo suficientemente grandes) mediante modelos de parámetros concentrados.

En definitiva, el objetivo final de los modelos a utilizar es poder predecir el comportamiento eléctrico de un circuito dado. Debido a esto, será de interés producir un circuito equivalente utilizando una colección de elementos básicos. A continuación se resumen brevemente los elementos circuitales necesarios para modelar circuitos de parámetros concentrados, y luego se da una clasificación de los modelos en función de la amplitud y frecuencia de las señales utilizadas.

1.1. Bloques constitutivos de modelos

1.1.1. Resistores

Un resistor es un elemento circuital de dos terminales caracterizado por una curva en el plano tensión v versus corriente i . Matemáticamente, un resistor satisface la siguiente relación:

$$\mathcal{R}_R = \{(v, i) : f(v, i) = 0\} \quad (1.1)$$

Si el resistor consiste de una recta que pasa por el origen, es decir

$$v = R \times i \quad (1.2)$$

entonces se denomina resistor lineal. En cualquier otro caso, el resistor se dice no lineal. Si la curva $v-i$ puede expresarse como una función de la corriente (tensión), entonces el resistor se dice controlado por corriente (tensión). Una curva monótonamente creciente es controlada por tensión y corriente a la vez.

En el caso en que un resistor no lineal sea controlado por corriente, es decir $v = v(i)$, el mismo se puede representar como:

$$v(t) = R(i) \times i(t) \quad (1.3)$$

donde

$$R(i) \triangleq \frac{dv(i)}{di} \quad (1.4)$$

se denomina la resistencia incremental. El símbolo del resistor se muestra en la Fig. 1.1-a.

1.1.2. Capacitores

Un capacitor es un elemento circuital de dos terminales caracterizado por una curva en el plano tensión v versus carga q . Matemáticamente, un capacitor satisface la siguiente relación:

$$\mathcal{R}_C = \{(v, q) : f(v, q) = 0\} \quad (1.5)$$

Si el capacitor consiste de una recta que pasa por el origen, es decir

¹Un pequeño elemento metálico puede representarse con un modelo de parámetros concentrados, dado que al exponerse a una fuente de calor, presenta una temperatura bien definida. Una varilla metálica larga, expuesta a una fuente de calor en un punto particular, debe representarse con un modelo de parámetros distribuidos, dado que la temperatura depende del lugar de la varilla y del instante de tiempo considerado

$$q = C \times v \quad (1.6)$$

entonces se denomina capacitor lineal. Dado que $i(t) = dq(t)/dt$, (1.6) se puede escribir de las siguientes formas equivalentes:

$$i(t) = C \times \frac{dv(t)}{dt} \quad (1.7)$$

$$v(t) = \frac{1}{C} \int_{-\infty}^t i(\tau) d\tau \quad (1.8)$$

En cualquier otro caso, el capacitor se dice no lineal. Si la curva $v - q$ puede expresarse como una función de la tensión (carga), entonces el capacitor se dice controlado por tensión (carga). Una curva monótonamente creciente es controlada por tensión y carga a la vez.

En el caso en que un capacitor no lineal sea controlado por carga, es decir $q = q(v)$, el mismo se puede representar como:

$$i(t) = C(v) \times \frac{dv(t)}{dt} \quad (1.9)$$

donde

$$C(v) \triangleq \frac{dq(v)}{dv} \quad (1.10)$$

se denomina la capacidad incremental. El símbolo del capacitor se muestra en la Fig. 1.1-b.

1.1.3. Inductores

Un inductor es un elemento circuital de dos terminales caracterizado por una curva en el plano corriente i versus flujo ϕ . Matemáticamente, un inductor satisface la siguiente relación:

$$\mathcal{R}_L = \{(i, \phi) : f(i, \phi) = 0\} \quad (1.11)$$

Si el inductor consiste de una recta que pasa por el origen, es decir

$$\phi = L \times i \quad (1.12)$$

entonces se denomina inductor lineal. Dado que $v(t) = d\phi(t)/dt$, (1.12) se puede escribir de las siguientes formas equivalentes:

$$v(t) = L \times \frac{di(t)}{dt} \quad (1.13)$$

$$i(t) = \frac{1}{L} \int_{-\infty}^t v(\tau) d\tau \quad (1.14)$$

En cualquier otro caso, el inductor se dice no lineal. Si la curva $i - \phi$ puede expresarse como una función de la corriente (flujo), entonces el inductor se dice controlado por corriente (flujo). Una curva monótonamente creciente es controlada por flujo y corriente a la vez.

En el caso en que un inductor no lineal sea controlado por flujo, es decir $\phi = \phi(i)$, el mismo se puede representar como:

$$v(t) = L(i) \times \frac{di(t)}{dt} \quad (1.15)$$

donde

$$L(i) \triangleq \frac{d\phi(i)}{di} \quad (1.16)$$

se denomina la inductancia incremental. El símbolo del inductor se muestra en la Fig. 1.1-c.

1.1.4. Memristores

Un memristor (o resistor con memoria) es un elemento circuital de dos terminales caracterizado por una curva en el plano carga q versus flujo ϕ . Este elemento fue predicho en el año 1971 por Leon O. Chua [2], pero no fue hasta el año 2008 que se halló evidencia experimental de un dispositivo de esta naturaleza [3]. El descubrimiento fue hecho por investigadores de los laboratorios de Hewlett Packard, al analizar dispositivos de escala nanométrica. Matemáticamente, un memristor satisface la siguiente relación:

$$\mathcal{R}_M = \{(q, \phi) : f(q, \phi) = 0\} \quad (1.17)$$

Si el memristor consiste de una recta que pasa por el origen, es decir

$$\phi = M \times q \quad (1.18)$$

el dispositivo es equivalente a un resistor lineal, $v = M \times i$. Debido a esto y tal como se señala en [2], el memristor no tiene sentido en la teoría de redes lineales.

En cualquier otro caso, el memristor se dice no lineal. Si la curva $q - \phi$ puede expresarse como una función de la carga (flujo), entonces el memristor se dice controlado por carga (flujo). Una curva monótonamente creciente es controlada por flujo y carga a la vez.

En el caso en que un memristor no lineal sea controlado por carga, es decir $\phi = \phi(q)$, el mismo se puede representar como:

$$\phi(t) = M(\phi) \times q(t) \quad (1.19)$$

donde

$$M(q) \triangleq \frac{d\phi(q)}{dq} \quad (1.20)$$

se denomina la memristancia incremental. La Ec. (1.19) también se puede expresar como:

$$\int_{-\infty}^{t_0} v(\tau) d\tau = M(\phi) \times \int_{-\infty}^{t_0} i(\tau) d\tau \quad (1.21)$$

Dado que $q = \int_{-\infty}^{t_0} i(\tau) d\tau$, la memristancia para un tiempo dado t_0 es función de los valores pasados de corriente en el dispositivo:

$$M(q(i)) = M\left(\int_{-\infty}^{t_0} i(\tau) d\tau\right) \quad (1.22)$$

De aquí el nombre de memristor o resistor con memoria. El símbolo del memristor se muestra en la Fig. 1.1-d.

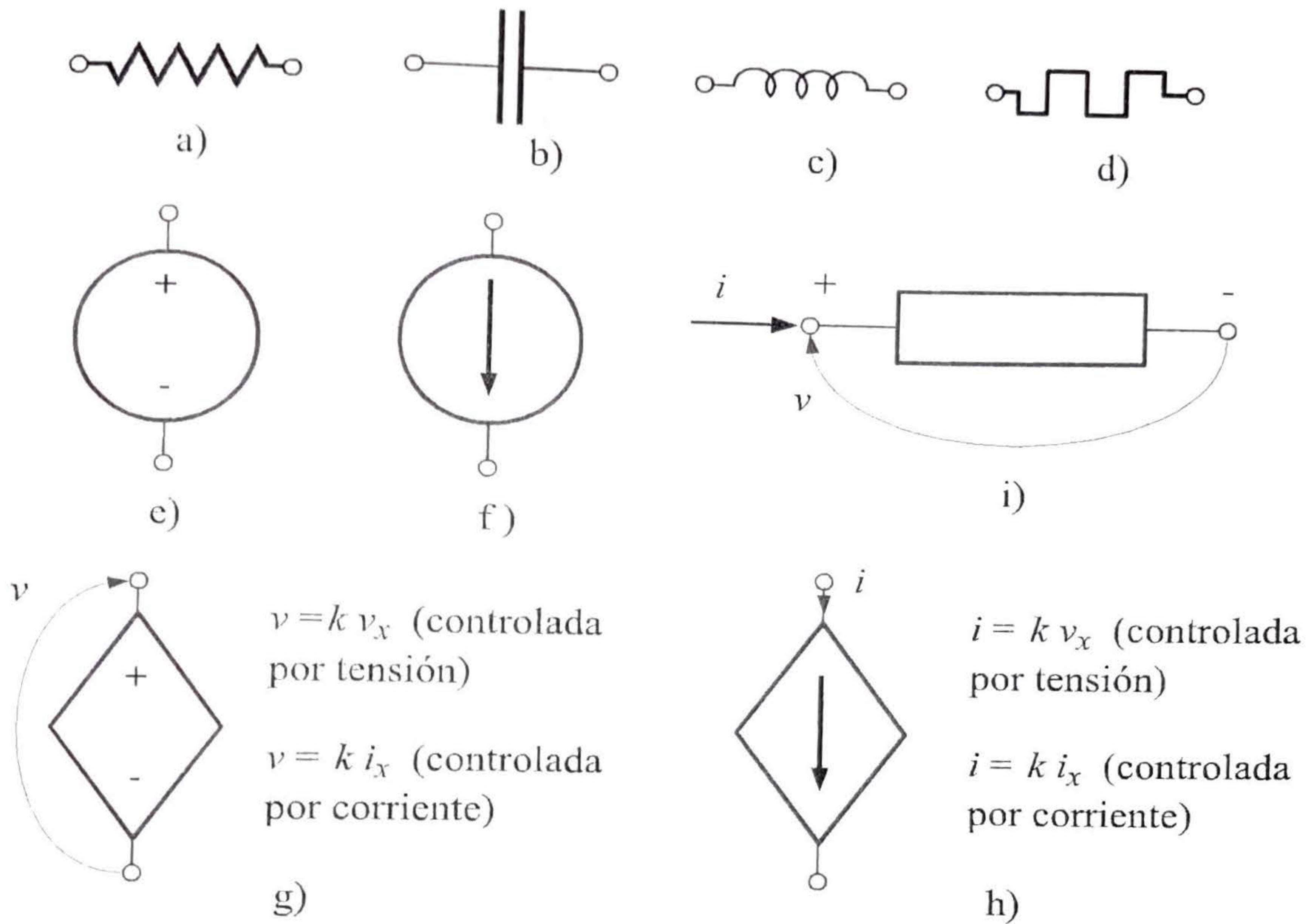


Figura 1.1: Bloques constitutivos de circuitos: a) resistor; b) capacitor; c) inductor; d) memristor; e) fuente independiente de tensión; f) fuente independiente de corriente; g) fuente de tensión controlada; h) fuente de corriente controlada; i) convenciones utilizadas de tensión y corriente.

1.1.5. Fuentes independientes

Una fuente independiente de tensión es un elemento de dos terminales, cuya tensión entre terminales $v_s(t)$ para cualquier instante de tiempo se halla definida a priori, y es independiente de la corriente. Dado que una fuente de tensión independiente se puede representar, para un dado instante de tiempo t , en el plano tensión versus corriente, se trata de un caso particular de un resistor. Es importante notar aquí que debido a la independendencia entre la tensión v_s y su corriente, la resistencia incremental de una fuente de tensión independiente verifica:

$$R = \frac{dv_s}{di} = 0 \tag{1.23}$$

Debido a ello, cuando se realizan análisis de pequeña señal de circuitos, las fuentes independientes de tensión se reemplazan por una resistencia de valor cero, es decir, un cortocircuito. El símbolo de una fuente independiente de tensión se muestra en la Fig. 1.1-e.

Una fuente independiente de corriente es un elemento de dos terminales, cuya corriente $i_s(t)$ para cualquier instante de tiempo se halla definida a priori, y es independiente de la tensión. Dado que una fuente de corriente independiente se puede representar, para un dado instante de tiempo t , mediante una curva en el plano tensión versus corriente, se trata también de un caso particular de un resistor. Es importante notar aquí que debido a la independendencia entre la corriente i_s y su tensión, la resistencia incremental de una fuente de tensión independiente verifica:

$$\frac{1}{R} = \frac{di_s}{dv} = 0 \tag{1.24}$$

Debido a ello, cuando se realizan análisis de pequeña señal de circuitos, las fuentes independientes de corriente se reemplazan por una resistencia de valor infinito, es decir, un circuito abierto. El símbolo de una fuente independiente de corriente se muestra en la Fig. 1.1-f.

1.1.6. Fuentes controladas

Una fuente controlada (dependiente) lineal es un elemento de dos terminales cuya tensión o corriente para cualquier instante de tiempo es proporcional a la tensión v_x o la corriente i_x en un elemento x de otra parte del circuito. El elemento x del cual depende la fuente controlada es denominado elemento controlador. Teniendo en cuenta las cuatro posibles combinaciones, resultan las fuentes de tensión controladas por tensión (FTCT), o controladas por corriente (FTCC), y las fuentes de corriente controladas por tensión (FCCT), o controladas por corriente (FCCC). El símbolo de una fuente controlada de tensión se muestra en la Fig. 1.1-g, y el correspondiente a una fuente de controlada de corriente se muestra en la Fig. 1.1-h.

1.1.7. Convenciones

A lo largo del libro se utilizará la convención de la Fig. 1.1-i para indicar tensiones y corrientes en un elemento. Las corrientes se indicarán mediante una flecha entrante, y las tensiones se medirán con respecto al terminal negativo del elemento $v = v^+ - v^-$.

1.2. Clasificación de modelos

1.2.1. Modelos según la amplitud

De acuerdo a la amplitud de las señales un modelo puede clasificarse como global, local o lineal incremental [4].

Un modelo global representa a un dispositivo sobre todos los rangos medibles de tensión y corriente. Un modelo local representa fielmente a un dispositivo sobre una porción determinada del rango de tensión y corriente del dispositivo. Finalmente, un modelo lineal incremental provee una aproximación a un dispositivo utilizando solamente los elementos lineales del conjunto de elementos circuitales introducidos en la Sección 1.1 (es decir, excluyendo las fuentes independientes y todos los elementos no lineales). De acuerdo a lo expuesto, un modelo lineal incremental representa solamente el comportamiento del dispositivo en un pequeño entorno de un punto de trabajo; lo suficientemente pequeño como para que la aproximación lineal sea válida. Si se agranda el entorno para cubrir un mayor rango de las variables, el modelo resultante será local, y podrá ser lineal o no lineal, dependiendo tanto del dispositivo como del rango de las variables consideradas. Si se extiende el rango de las variables a todos los valores posibles, inevitablemente aparecerán fenómenos no lineales, como pueden ser no-linealidades intrínsecas del dispositivo y/o fenómenos de ruptura. Debido a esto, un modelo global de un dispositivo electrónico es necesariamente no lineal.

Ejemplo 1 La Fig. 1.2-a muestra el símbolo de un diodo semiconductor y su modelo global a través de la curva de corriente i_D en función de la tensión en sus terminales v_D . La tensión $V_{ON} \approx 700\text{mV}$ es la tensión a partir de la cual el diodo comienza a conducir una corriente significativa, mientras que V_{BR} es la tensión de ruptura inversa, que dependiendo del tipo de diodo puede variar entre algunas decenas o centenas de voltios. La Fig. 1.2-b muestra un circuito equivalente aproximado de tipo global, construido con diodos ideales y fuentes de tensión, que es capaz de reproducir el comportamiento del dispositivo para tensiones de amplitud grande, tanto positivas como negativas. Un diodo ideal presenta corriente nula para $-V_{BR} \leq v_D \leq V_{ON}$, con $v_D = V_{ON}$ si $i_D > 0$, y $v_D = -V_{BR}$ si $i_D < 0$. Este

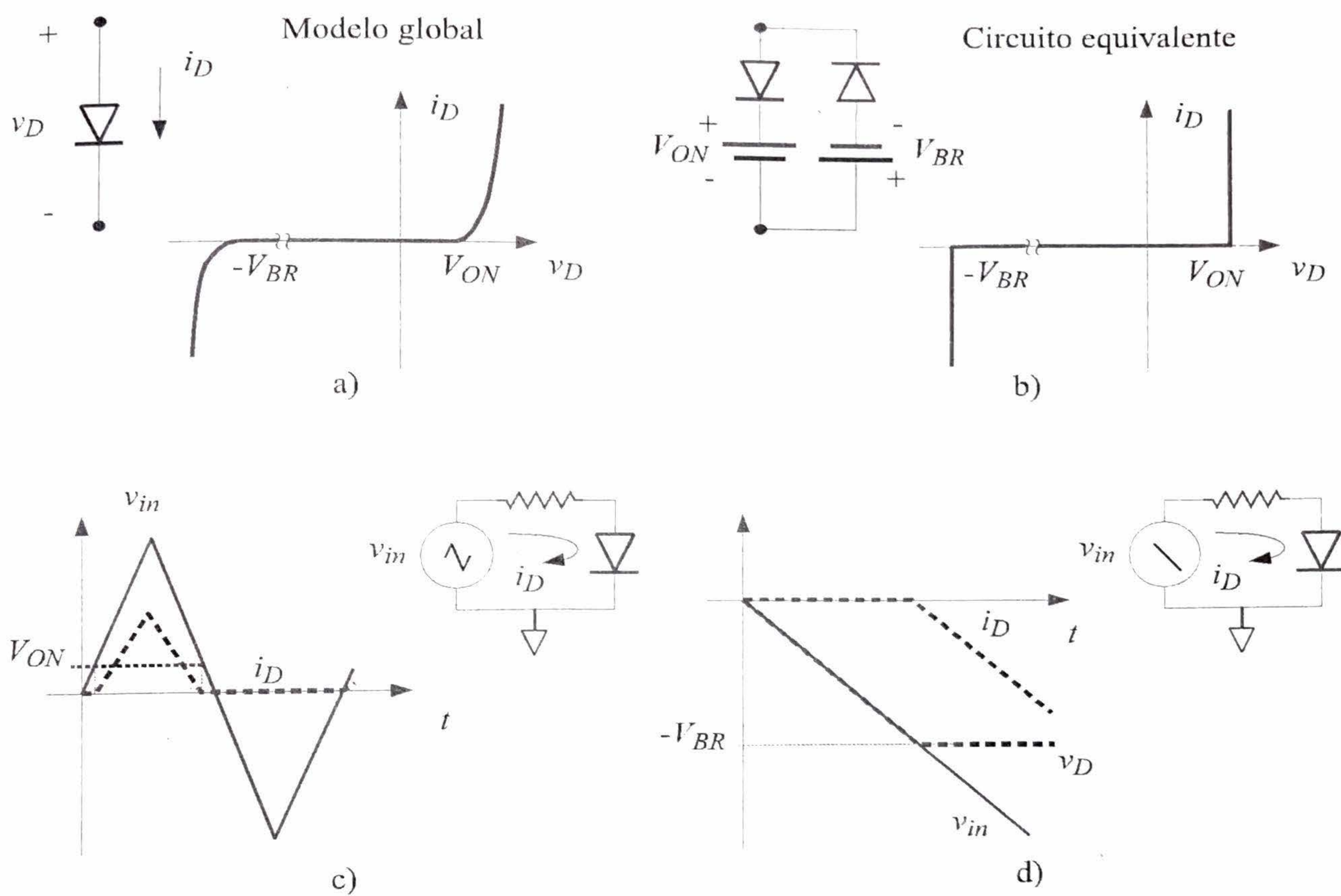


Figura 1.2: Modelo global de un diodo: a) símbolo y curva corriente-tensión que define el modelo global; b) circuito equivalente para aproximar el modelo global utilizando diodos ideales; c) respuesta temporal del diodo ante una entrada positiva y negativa (sin alcanzar V_{BR}); e) respuesta temporal del diodo ante una entrada negativa.

circuito se puede utilizar para determinar la respuesta del diodo ante una entrada de valores positivos y negativos, como se muestra en la Fig. 1.2-c, y también reproduce fielmente el comportamiento para tensiones negativas cercanas a la tensión de ruptura inversa, como se muestra en el circuito de la Fig. 1.2-d.

Ejemplo 2 La Fig. 1.3-a muestra el modelo local de un diodo y el circuito equivalente que consiste en una fuente de tensión y una resistencia serie. Este modelo es válido siempre y cuando sea $v_D > V_{ON}$. La Fig. 1.3-c muestra la respuesta temporal de la corriente en un circuito utilizando este modelo, cuando se aplica una señal triangular de tensión.

Ejemplo 3 La Fig. 1.3-b muestra el modelo lineal incremental de un diodo y el circuito equivalente que consiste en una resistencia. Este modelo es válido en un entorno del punto de trabajo definido por el par (V_D, I_D) y relaciona las variables de pequeña señal i_d y v_d . La Fig. 1.3-d muestra la respuesta temporal de un circuito utilizando este modelo. La señal de entrada v_{IN} tiene una componente continua de valor V_{IN} y una componente de señal v_{in} que es la que se aplica al modelo lineal incremental.

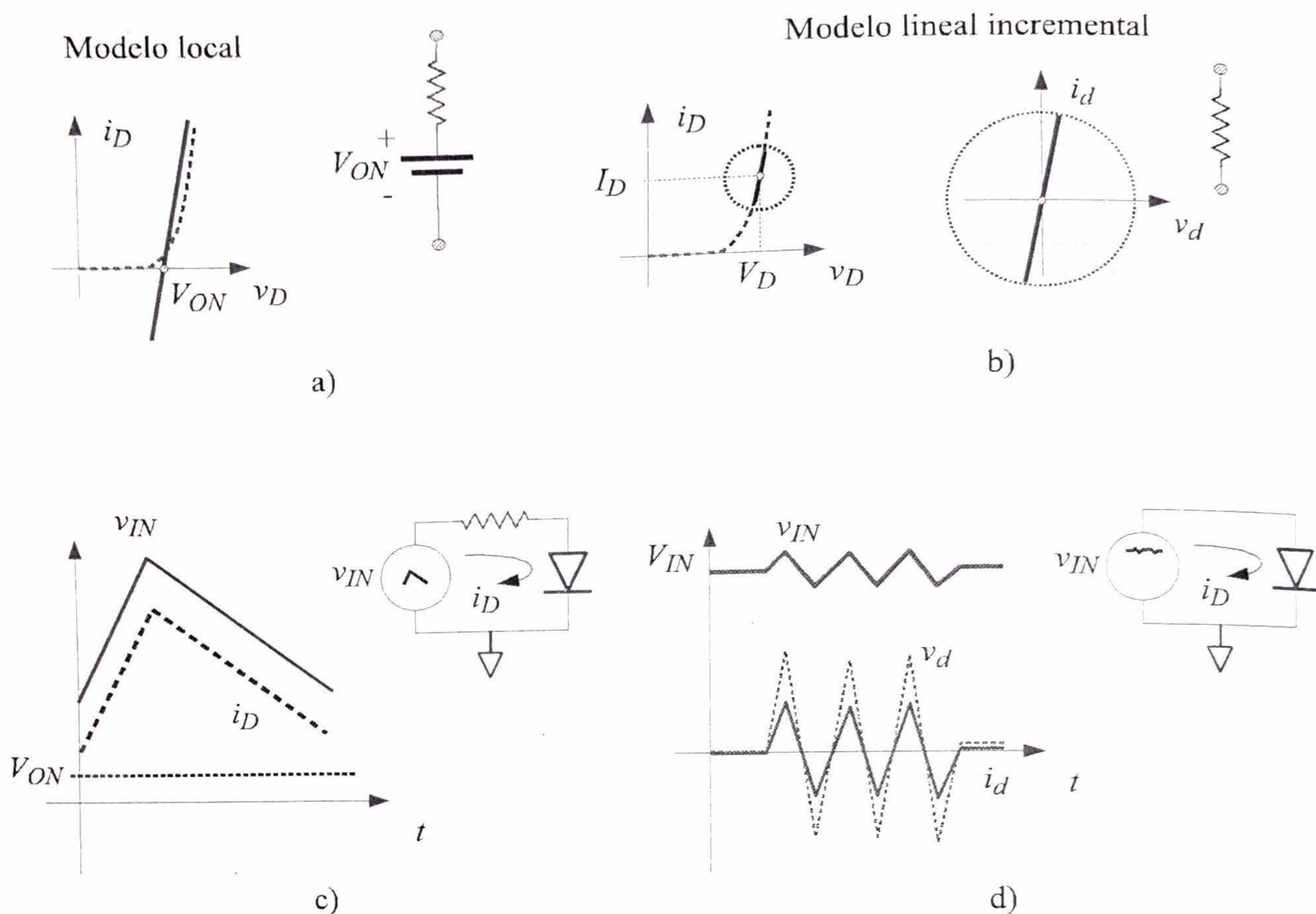


Figura 1.3: Modelos locales y lineales incrementales de un diodo: a) modelo local y circuito equivalente; b) modelo lineal incremental y circuito equivalente; c) respuesta temporal utilizando el modelo local; e) respuesta temporal utilizando el modelo lineal incremental.

1.2.2. Modelos según la frecuencia

De acuerdo a la frecuencia de las señales un modelo puede clasificarse como de corriente continua (DC) o de corriente alterna (AC). Un modelo de DC solo posee elementos resistivos, fuentes independientes y dependientes de DC; es decir, no posee capacitores, inductores o memristores (no lineales). Los modelos de DC en general son capaces de modelar desde DC ($f = 0\text{Hz}$) hasta alguna frecuencia relativamente baja. Los modelos de AC contienen capacitores, inductores, memristores, y/o fuentes independientes o dependientes de AC. Su comportamiento presenta variaciones dependiendo de la frecuencia de las señales utilizadas. Son capaces de modelar desde frecuencias muy bajas hasta frecuencias muy altas. En virtud del amplio rango de frecuencias utilizadas en la electrónica, que vá desde algunos Hz o fracciones de Hz hasta GHz, usualmente se divide el espectro frecuencial en frecuencias bajas, frecuencias medias y frecuencias altas. Las bajas frecuencias comprenden en general el rango de audio, desde algunos Hz hasta algunas decenas de KHz. Las frecuencias medias comprenden desde algunas decenas de KHz hasta algunas decenas de MHz. Por último, altas frecuencias comprenden frecuencias superiores a las centenas de MHz. Los modelos que se desarrollan en este libro, tanto para diodos como para transistores son válidos desde DC hasta frecuencias medias.

Ejemplo 4 La Fig. 1.4-a ilustra una resistencia de Silicio policristalino en un proceso moderno de

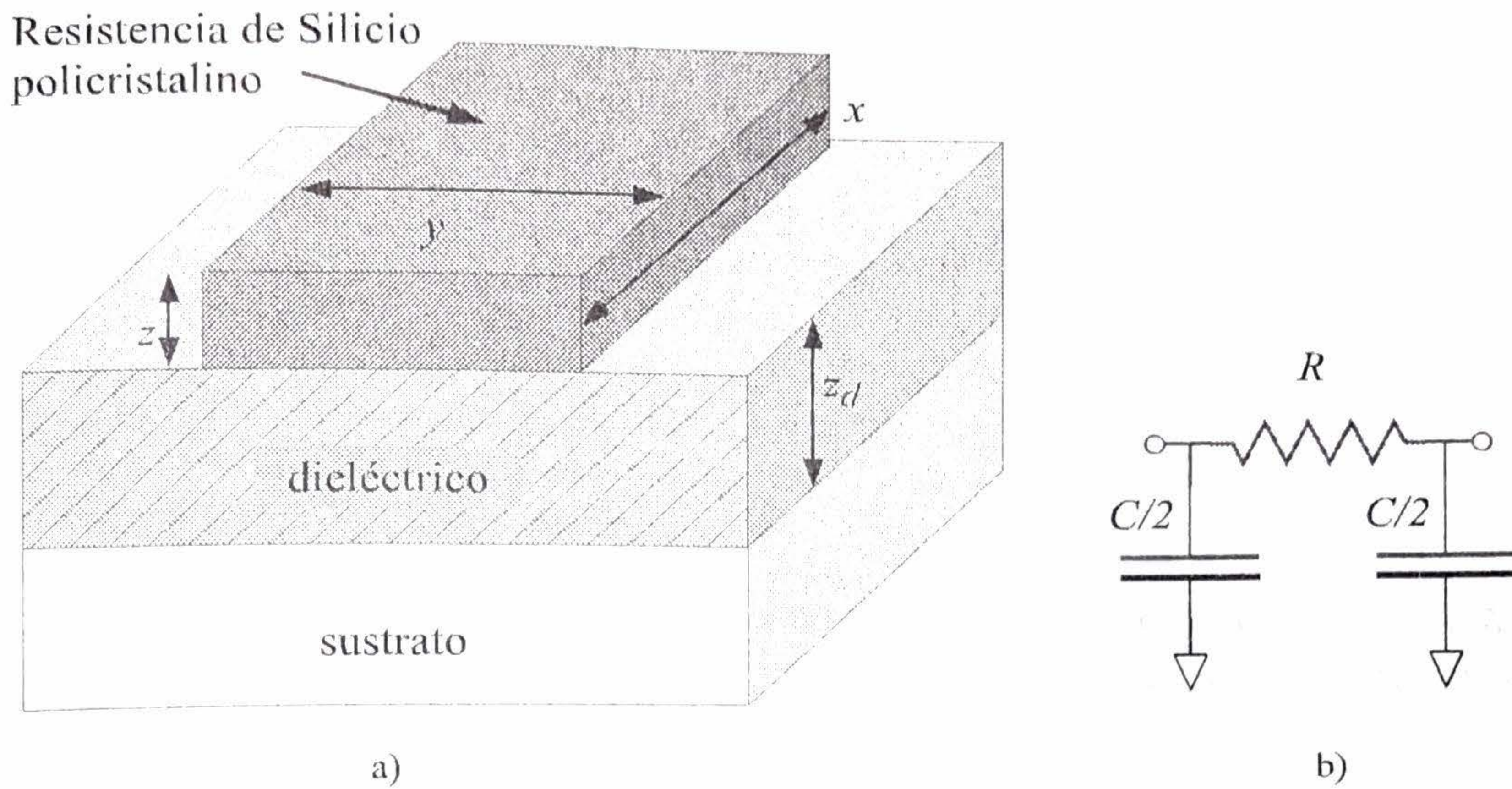


Figura 1.4: a) Diagrama de una resistencia realizada con Silicio policristalino en un proceso de circuitos integrados; b) Circuito equivalente.

circuitos integrados. Para DC y frecuencias bajas, es suficiente con considerar la resistencia dada por

$$R = \rho \times \frac{x}{yz} \quad (1.25)$$

donde x es la longitud, y el ancho, y z la profundidad, tal como se ilustra en la Fig. 1.4. Si $y = 0,5\mu m$, $x = 1500\mu m$, $z = 0,5\mu m$, y considerando que la resistividad del Silicio policristalino es $\rho = 1,17 \times 10^{-5}\Omega m$, la resistencia del mismo es igual a:

$$R = \frac{1,17 \times 10^{-5}\Omega m \times 1500\mu m}{0,5\mu m \times 0,5\mu m} = 70,2K\Omega \quad (1.26)$$

El resistor se aísla del sustrato de Silicio mediante un dieléctrico de espesor z_d , típicamente dióxido de Silicio (SiO_2) el cual posee una constante dieléctrica $\epsilon_{SiO_2} = 3,9 \times \epsilon_0$. Debido a esto, el resistor presenta una capacidad de placas paralelas con respecto al sustrato, dada por:

$$C = \frac{\epsilon_{SiO_2}xy}{z_d} \quad (1.27)$$

Si $z_d = 0,5\mu m$ y teniendo en cuenta que $\epsilon_{SiO_2} = 3,9 \times 8,854 \times 10^{-12}F/m$, esta capacidad resulta igual a:

$$C = \frac{3,9 \times 8,854 \times 10^{-12}F/m \times 1500\mu m \times 0,5\mu m}{0,5\mu m} = 51,8fF \quad (1.28)$$

La resistencia debe entonces modelarse incluyendo la capacidad relativa al sustrato. Una opción es distribuir la capacidad total en dos partes iguales, a ambos extremos de la resistencia, como se muestra en la Fig. 1.4-b.

La frecuencia

$$f_c = \frac{1}{2\pi RC} = \frac{1}{2\pi \times 70,2K\Omega \times 51,8fF} = 43,7MHz \quad (1.29)$$

y la constante de tiempo asociada:

$$\tau_c = R \times C = 70,2K\Omega \times 51,8fF = 3,6ns \quad (1.30)$$

determinan el punto donde la resistencia es igual a la impedancia capacitiva, es decir cuando:

$$Z_C = \frac{1}{2\pi f_c} = 70,2K\Omega \quad (1.31)$$

Por lo tanto, para frecuencias mucho menores a f_c (o tiempos mucho mayores que τ_c), por ejemplo $f \leq f_c/10$ la resistencia efectivamente se comporta como tal. Sin embargo, a partir de $f = f_c/10$ es necesario incluir el efecto de la capacidad. Para valores mayores a $10f_c$ puede considerarse que la impedancia equivalente es puramente capacitiva, siendo el efecto resistivo despreciable.

Ejemplo 5 Un capacitor electrolítico de Tantalio puede representarse mediante el circuito aproximado de la Fig. 1.5-a. La capacidad nominal C_N representa la capacidad nominal del dispositivo; la resistencia equivalente serie ESR representa las pérdidas en el dieléctrico y la resistencia óhmica del electrolito y los terminales; la inductancia equivalente serie ESL representa la inductancia causada por los arrollamientos de los dos terminales (ánodo y cátodo). La fuente de corriente $i(v, T)$ representa una corriente de pérdida de DC que depende de la tensión aplicada en terminales tal como se ilustra en la Fig. 1.5-b; esta fuente de corriente también depende de la temperatura. Por último, la capacidad C_p representa capacidades parásitas del encapsulado. La 1.5-c muestra la característica frecuencial de la impedancia del capacitor, para un capacitor KAL0009-N de $C_N = 100\mu F$ y 63V [5]. A bajas frecuencias (desde DC hasta 10KHz) la impedancia es de la forma

$$Z_C = \frac{1}{j2\pi f C_N} \quad (1.32)$$

A partir de allí, entre 50KHz y 200KHz se puede ver una parte plana donde el comportamiento está dominado por la resistencia equivalente ESR, y por lo tanto la curva es independiente de la frecuencia. Para frecuencias mayores, el comportamiento de la impedancia está dominado por la inductancia, y es de la forma:

$$Z_C = j2\pi f ESL \quad (1.33)$$

Para frecuencias mucho mayores, las capacidades parásitas C_p terminan cortocircuitando el dispositivo y reduciendo su impedancia a cero (fuera del gráfico).

Ejemplo 6 Un cable de metal en un proceso moderno de circuitos integrados debe representarse por distintos modelos de acuerdo a la frecuencia de operación. Si se considera un cable de Aluminio ($\rho = 2,7 \times 10^{-2}\Omega\mu m$) con la misma geometría que la resistencia de Polisilicio de la Fig. 1.4-a, con dimensiones $x = 5000\mu m$, $y = 5\mu m$ y $z = 0,3\mu m$, la resistencia del mismo es:

$$R = \frac{2,7 \times 10^{-2}\Omega\mu m \times 5000\mu m}{5\mu m \times 0,3\mu m} = 90\Omega \quad (1.34)$$

El fabricante especifica una capacidad de la forma:

$$C = 28 \frac{aF}{\mu m^2} \times y \times x + 43 \frac{aF}{\mu m} \times x \quad (1.35)$$

De manera que la capacidad total de la línea es:

$$C = 0,915pF \quad (1.36)$$

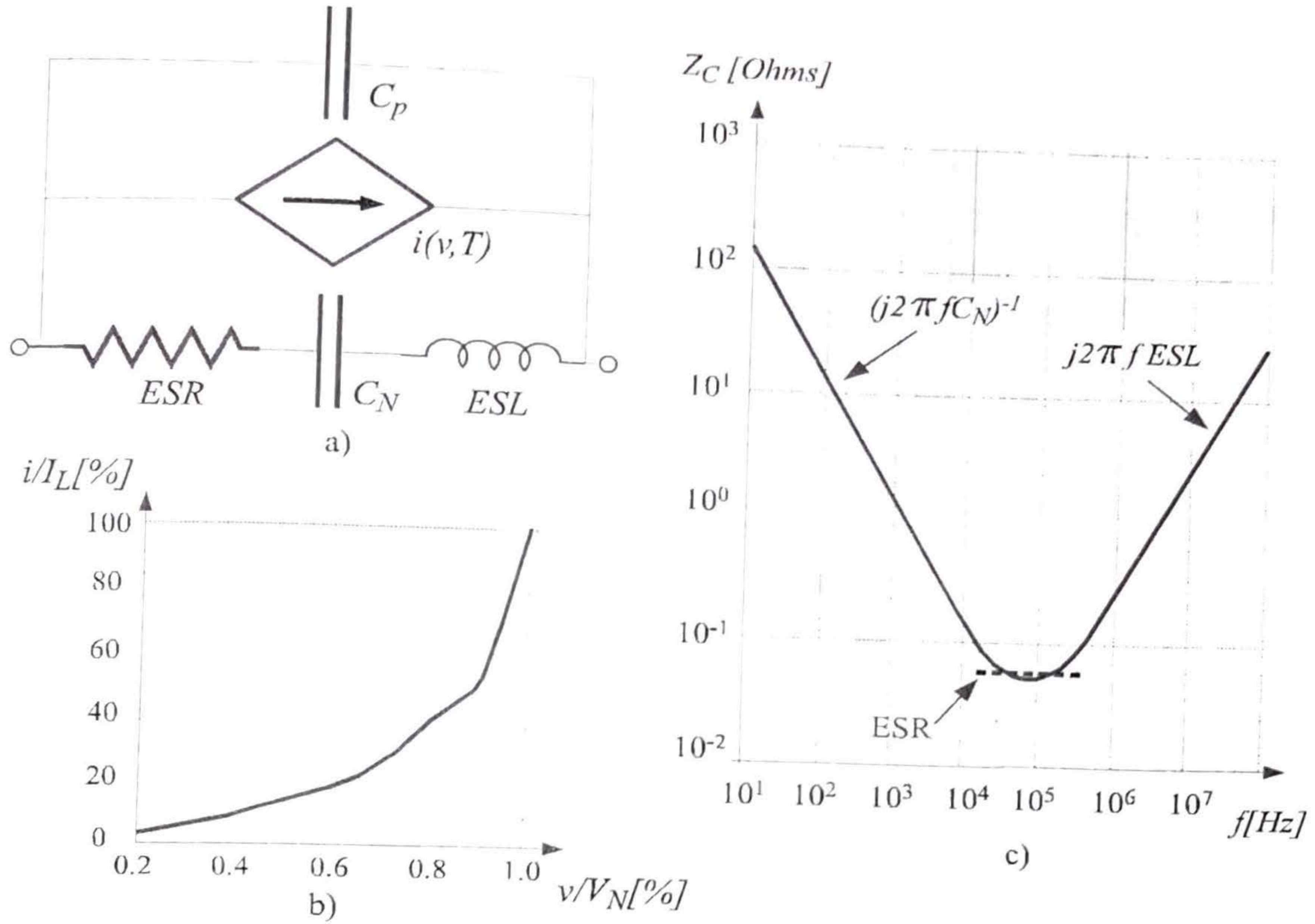


Figura 1.5: Modelo de un capacitor electrolítico de Tantalio: a) circuito equivalente; b) característica de la fuente de corriente $i(v, T)$, donde $I_L = 4,14\mu A$ y $V_N = 63V$; c) impedancia versus frecuencia.

El mismo modelo de la Fig. 1.4-b puede aplicarse inicialmente en este caso. La frecuencia donde la impedancia capacitiva es igual a la resistencia es:

$$f_c = \frac{1}{2\pi RC} = \frac{1}{2\pi \times 90 \times 0,915 \times 10^{-12}} = 1,93GHz \quad (1.37)$$

que corresponde a una constante de tiempo $\tau = 90 \times 0,915 \times 10^{-12} = 82ps$.

Para frecuencias mucho menores que f_c ($f < 193Mhz$) la resistencia es despreciable frente a la impedancia capacitiva. La combinación serie de la resistencia y el capacitor puede representarse sólo con el capacitor. Por lo tanto, desde el terminal de entrada (izquierda) del cable, la impedancia equivalente está dada por la de ambos capacitores en paralelo. En este caso, el modelo de la Fig. 1.6-a es suficiente para modelar el cable.

Para señales cuyas frecuencias sean mayores que f_c , la resistencia es significativa y puede ser necesario el empleo de una red RC tal como se muestra en las Figs. 1.6-b ó c.

La velocidad de propagación de la señal electromagnética en el cable (en caso que el dieléctrico es SiO_2) es de $15,2cm/ns$, con lo cual el tiempo de propagación a lo largo del cable es:

$$t_e = \frac{0,5cm}{15,2cm/ns} = 32,9ps \quad (1.38)$$

Para señales muy rápidas (con constantes de tiempo menores a $2,5 \times t_e = 82,50ps$ [6]) puede ser necesario utilizar un modelo de línea de transmisión como el que se muestra en la Fig. 1.6-d.

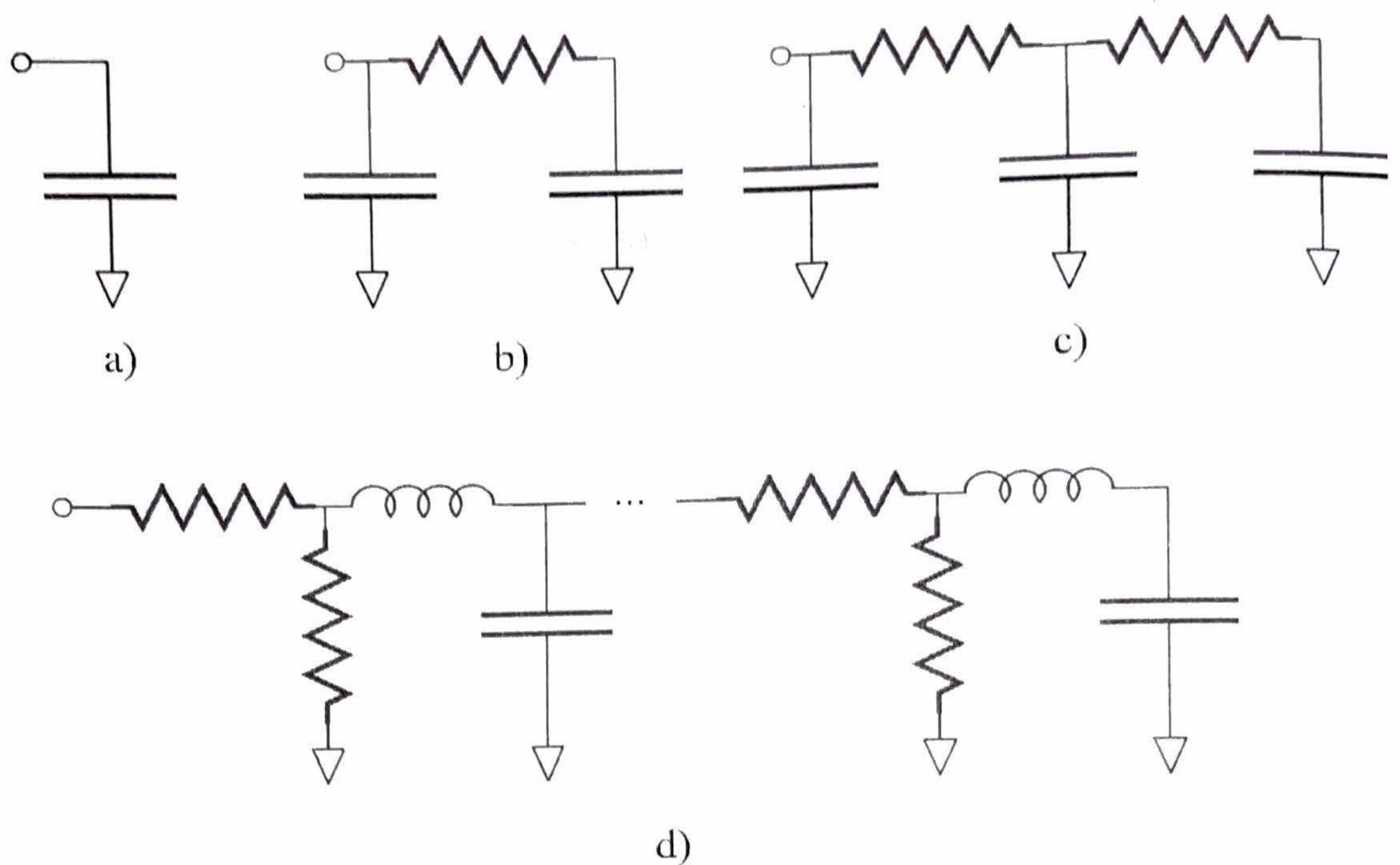


Figura 1.6: Modelo de un cable en un proceso avanzado de circuitos integrados: a) modelo capacitivo; b) modelo RC; c) modelo RC; d) línea de transmisión.

1.2.3. Construcción de modelos

La Fig. 1.7 ilustra la clasificación de modelos de acuerdo a la amplitud y frecuencia de las señales bajo consideración [4].

Comenzando por el modelo de mayor complejidad, es decir el modelo global de AC, la remoción de todos los capacitores y los inductores lleva al modelo global de DC. Por otro lado, la remoción de algunos elementos no lineales lleva al modelo local de AC. Desde este último, la remoción de todos los capacitores e inductores lleva al modelo local de DC, y la remoción de todos los elementos no lineales lleva al modelo lineal incremental de AC. La remoción de todos los inductores y capacitores de este último lleva al modelo incremental de DC.

A lo largo del libro, se construirán modelos de DC y se obtendrán los modelos de AC incorporando las capacidades e inductancias según corresponda. Los modelos a utilizar pertenecerán a la clase de dispositivos cuasi-resistivos, que se definen de la siguiente manera.

Definición 1 *Un dispositivo se dice cuasi-resistivo si los mecanismos de almacenamiento de energía del dispositivo pueden tenerse en cuenta añadiendo inductores lineales y capacitores lineales a un modelo de DC apropiadamente elegido.*

El siguiente ejemplo ilustra la metodología descripta.

Ejemplo 7 *Un transistor MOS de tres terminales² tal como se muestra en la Fig. 1.8-a, tiene como variables principales la tensión de gate v_G , la tensión de drain v_D , y la corriente de drain i_D . El gate está aislado mediante un material dieléctrico, por lo cual en general resulta $i_G = 0$.*

²El cuarto terminal, denominado sustrato, se haya conectado al terminal de source.

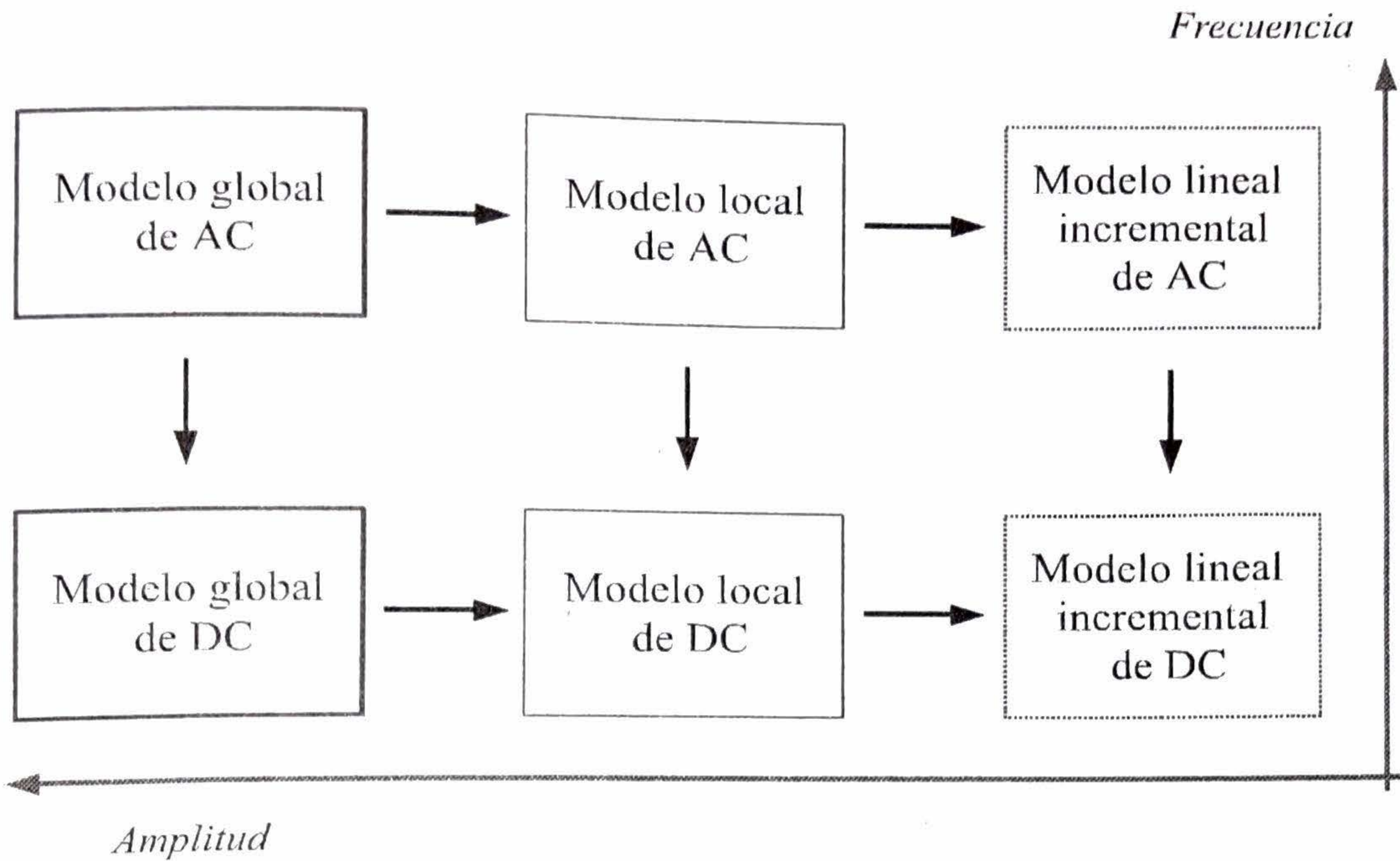


Figura 1.7: Modelos clasificados de acuerdo a la amplitud y frecuencia de las señales.

Un modelo local de DC válido para tensiones positivas de gate y drain consiste en un circuito abierto entre gate y source, y una FCCT no lineal

$$i_D(v_G, v_D) = f(v_G, v_D) \quad (1.39)$$

entre los terminales de drain y source, tal como se muestra en la Fig. 1.8-b. Si se considera el punto de trabajo (V_G, V_D) indicado en la curva característica de la Fig. 1.8-b, ampliado en la Fig. 1.8-c, el transistor puede modelarse por medio de una FCCT lineal de valor

$$i_d = g_m v_g \quad (1.40)$$

y una resistencia de valor r en paralelo. Este modelo lineal incremental (LI) de DC sólo es válido en un pequeño entorno del punto de trabajo indicado.

El modelo de la Fig. 1.8-b no es global dado que no modela valores negativos de v_D , ni valores grandes de v_D donde la corriente de i_D se incrementa en forma pronunciada producto del campo eléctrico elevado que se produce en el dispositivo. Si bien es raramente utilizado, la Fig. 1.9-a ilustra un modelo global de DC que tiene en cuenta estos aspectos. El diodo en inversa entre drain y sustrato modela la corriente en inversa, la fuente de corriente I_h representa la corriente de impacto de ionización que surge a altos valores de v_D y el transistor bipolar es el causante del incremento de la corriente i_D [7]. La Fig. 1.9-b muestra el circuito equivalente resultante de introducir un diodo y una FCCT no lineal

$$i_{BR}(v_G, v_D) = f_{BR}(v_G, v_D) \quad (1.41)$$

cuya característica completa de salida es la de la Fig. 1.9-c.

Ejemplo 8 Para extender el modelo local de DC del transistor MOS del ejemplo anterior a un modelo local de AC, es necesario incluir las capacidades del dispositivo como se ilustra en la Fig. 1.10-a. La capacidad más importante de un transistor MOS es la capacidad que presenta el gate. Para valores pequeños de v_G , más específicamente, para $v_G < V_T$, donde V_T es la tensión de umbral del transistor,

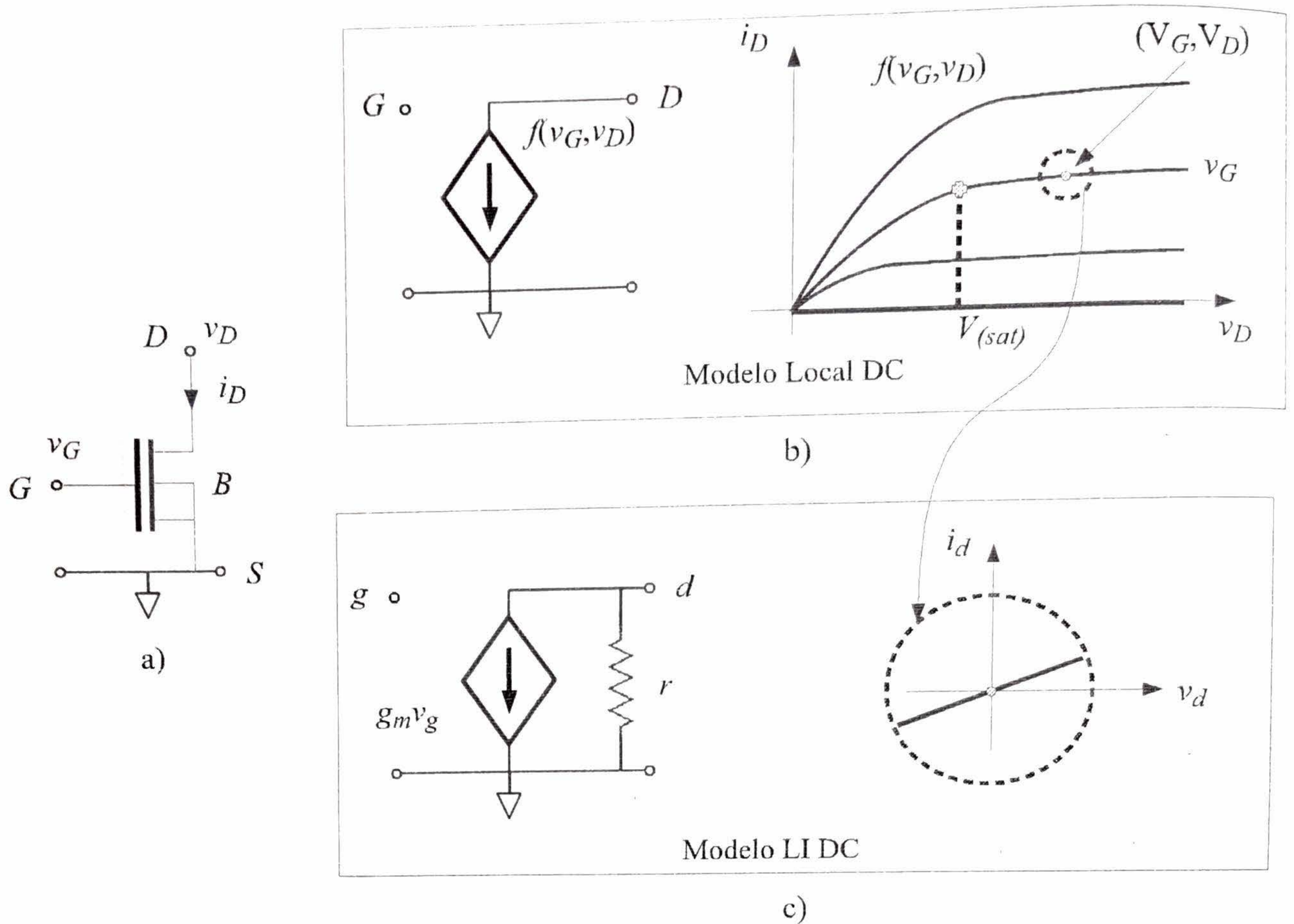


Figura 1.8: Modelos de DC de un transistor MOS: a) transistor MOS; b) modelo local de DC; c) modelo lineal incremental de DC.

esta capacidad se desarrolla entre el gate y el sustrato, es no lineal y depende de v_G como se muestra en la Fig. 1.10-d. Cuando $v_G > V_T$ esta capacidad alcanza un valor constante, y se puede reemplazar por dos capacitores: uno de ellos conectado entre gate y source, llamado C_{GS} , y otro conectado entre gate y drain, llamado C_{GD} . Para valores pequeños de la tensión de drain v_D , estas capacidades son iguales $C_{GD} = C_{GS} = C_{ox}/2$, donde C_{ox} es la capacidad del dieléctrico del transistor. Si se aumenta la tensión v_D , la capacidad C_{GS} aumenta hasta lograr un valor máximo $C_{GS} = 2C_{ox}/3$ mientras que la capacidad C_{GD} se reduce hasta cero, como se muestra en la Fig. 1.10-e. Esta variación ocurre mientras la tensión de salida varía en el intervalo $0 \leq v_D \leq V_{(sat)}$. Para valores mayores de v_D , ambas capacidades permanecen constantes. Por último, existe otra capacidad entre el terminal de drain y el terminal de referencia que disminuye con la tensión aplicada como se muestra en la Fig. 1.10-f.

Si se toma el punto de trabajo (V_G, V_D) indicado en la curva característica de la Fig. 1.8-b, y se reemplazan los valores de capacidad por los correspondientes a los valores $v_G = V_G$ y $v_D = V_D$, se obtiene el modelo LI de AC de la Fig. 1.10-b. Este modelo es apto para frecuencias medias de algunos centenares de MHz. Para valores mayores de frecuencia, es necesario incluir una resistencia adicional R_g entre el terminal de gate y source, y una constante dependiente de la frecuencia para la FCCT que modele el desfase entre la tensión de gate y la corriente de drain, de la forma

$$g_m(f) = g_m \times \frac{1}{1 + jf/f_c} \quad (1.42)$$

donde f_c es la frecuencia de corte. Este modelo puede utilizarse hasta algunos GHz.

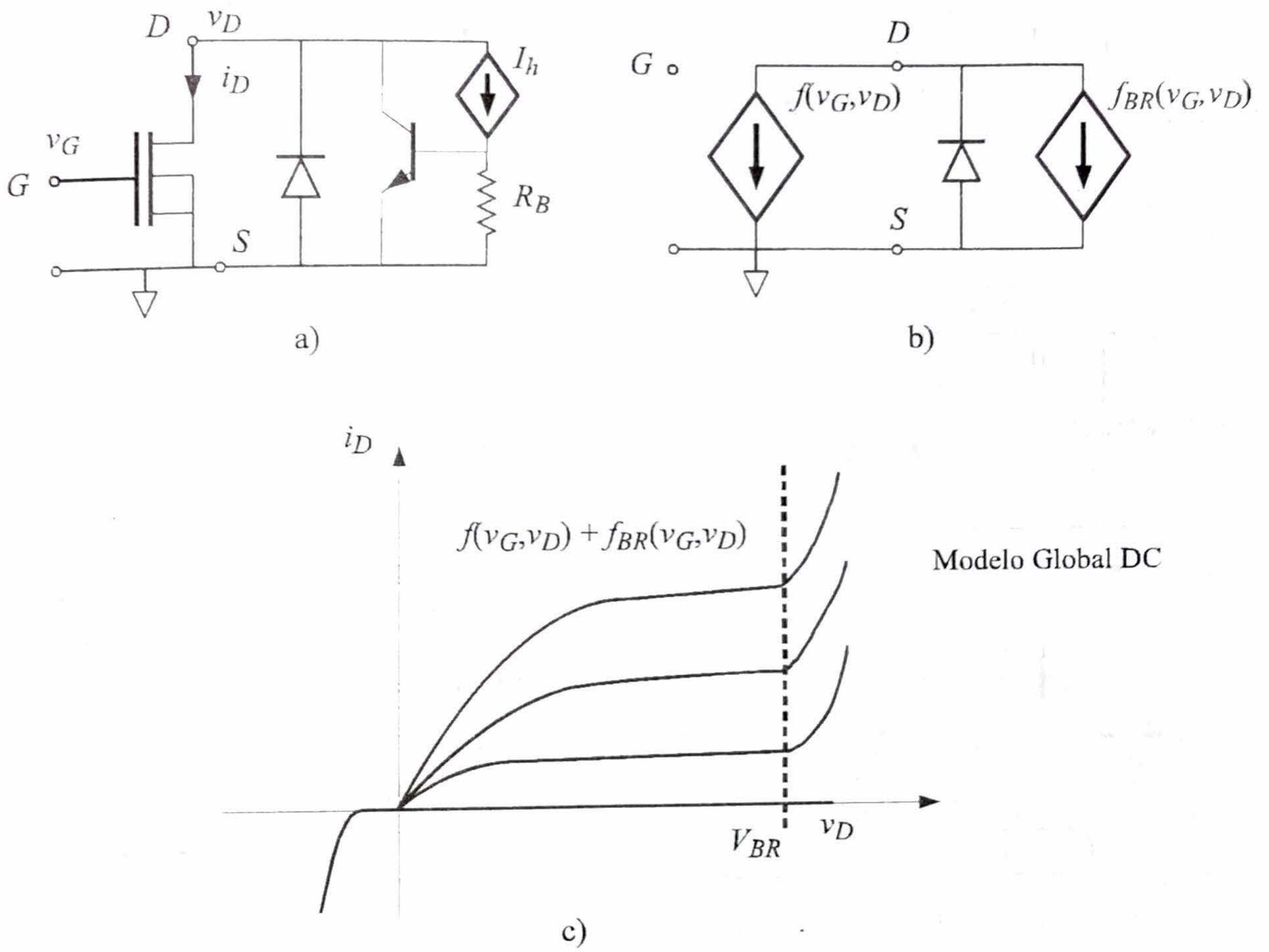


Figura 1.9: Modelo global de un transistor MOS: a) esquemático; b) modelo equivalente ; c) curvas características de DC.

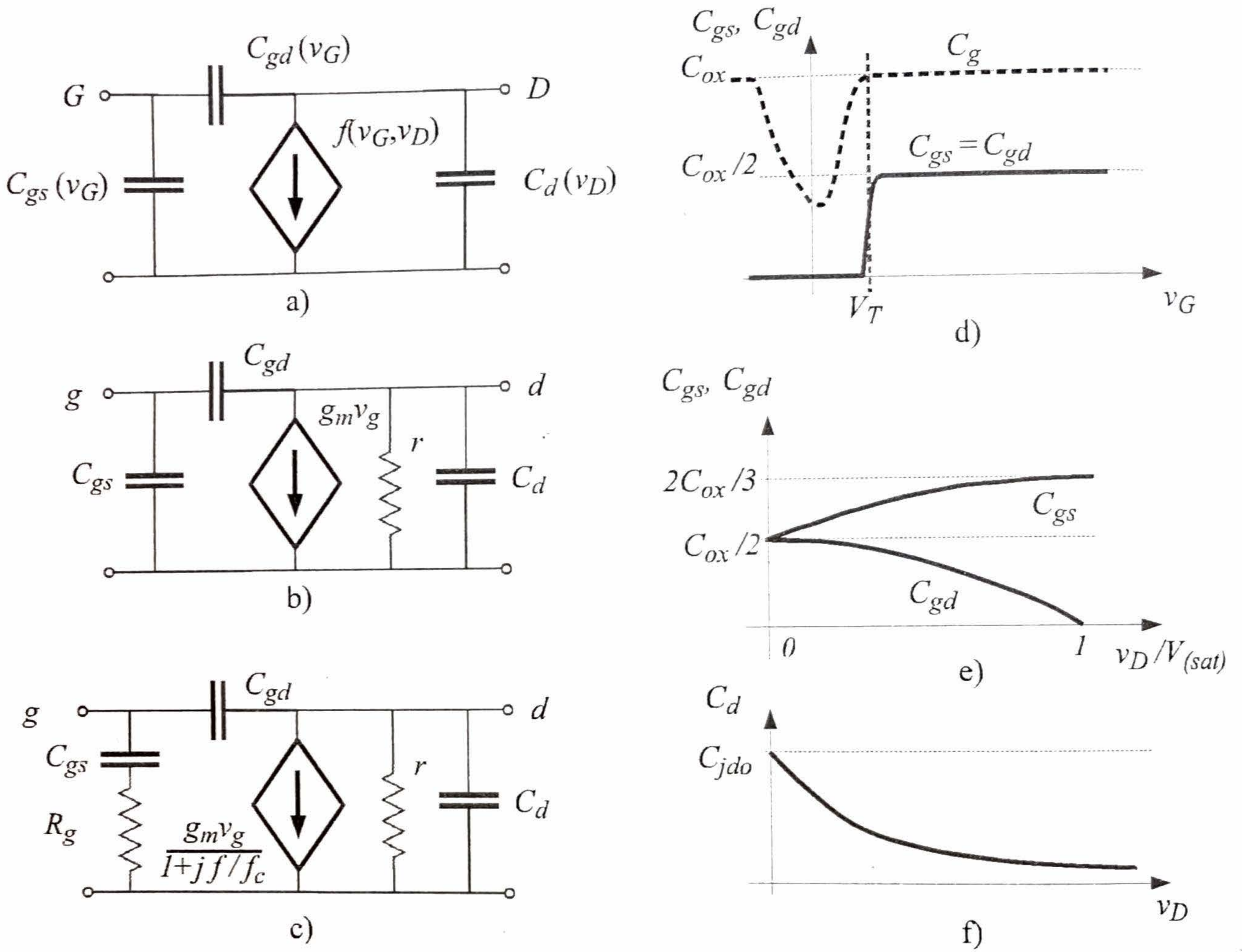


Figura 1.10: Modelos de AC de un transistor MOS: a) modelo local de AC; b) modelo lineal incremental de AC; c) modelo lineal incremental para alta frecuencia; d) capacidad de gate con respecto a la tensión v_G ; e) capacidades C_{gs} y C_{gd} con respecto a la tensión normalizada de salida $v_D/V_{(sat)}$; f) capacidad del drain C_d con respecto a la tensión de salida v_D .

Capítulo 2

Introducción a los semiconductores

2.1. Bandas de Energía en Silicio

El Silicio es el elemento número 14 en la tabla periódica de los elementos y pertenece al grupo IV (junto con el Carbón y el Germanio, entre otros). Es un elemento que posee en estado aislado catorce protones y catorce electrones. De acuerdo a la mecánica cuántica, un sistema o partícula que se halla confinado espacialmente solo puede tener ciertos valores particulares –o discretos– de energía¹. Estos valores discretos son llamados niveles de energía.

En el caso de una partícula eléctrica, se define la energía potencial como el trabajo que se debe realizar para mover una carga (en presencia de otras) desde un punto de referencia hasta su localización final. En general, se elige por convención el nivel de referencia de energía potencial (es decir, energía cero) en infinito. Si se considera una carga positiva en el origen ($+q$), la energía potencial de un electrón situado a una distancia r , está dada por:

$$\text{Energía} = \frac{1}{4\pi\epsilon_0} \frac{-q \times q}{r} \quad (2.1)$$

En el caso del Silicio –en ausencia de energía térmica, es decir cuando $T = 0K$ – los electrones se distribuyen de la siguiente manera: dos de ellos en el primer nivel de energía, ocho en el segundo nivel, y cuatro más en el tercer nivel como se ilustra en la Fig. 2.1. En consecuencia, los electrones poseen energía potencial negativa, y cuanto más cerca se encuentran del núcleo del átomo, menor energía poseen.

La mecánica cuántica representa al electrón confinado en un átomo mediante un modelo de onda estacionaria que rodea al núcleo. Esta onda define la probabilidad de hallar al electrón en una determinada posición, y su longitud de onda determina la energía del electrón. En términos geométricos, esta onda puede alinearse en un cierto número (bien definido) de direcciones posibles (propiedad relacionada con el momento angular); es decir que para un cierto nivel de energía (o lo que es equivalente, para una cierta longitud de onda) un electrón puede tener diferentes estados. La orientación y longitud de la onda estacionaria definen entonces un estado posible. Las superficies de igual valor –de probabilidad– de las ondas estacionarias se denominan orbitales. La Fig. 2.2 ilustra la forma de los tres primeros tipos de orbitales. De acuerdo a las leyes de la mecánica cuántica², cada estado orbital solo puede albergar dos electrones con diferentes momentos angulares³, como máximo. Cada orbital se nota con el número de nivel de energía seguido del tipo de orbital. El Silicio posee dos electrones en el primer orbital $1s$, dos en el orbital $2s$, seis en los orbitales tipo⁴ $2p$, dos en el orbital $3s$ y dos más en los orbitales de tipo

¹En la mecánica clásica, las partículas pueden poseer cualquier valor de energía.

²Más específicamente, de acuerdo al principio de exclusión de Pauli.

³Esta propiedad se conoce como *spin*.

⁴Existen tres orbitales tipo $2p$ de acuerdo a la orientación en los ejes x , y ó z ; lo mismo sucede con los orbitales $3p$.

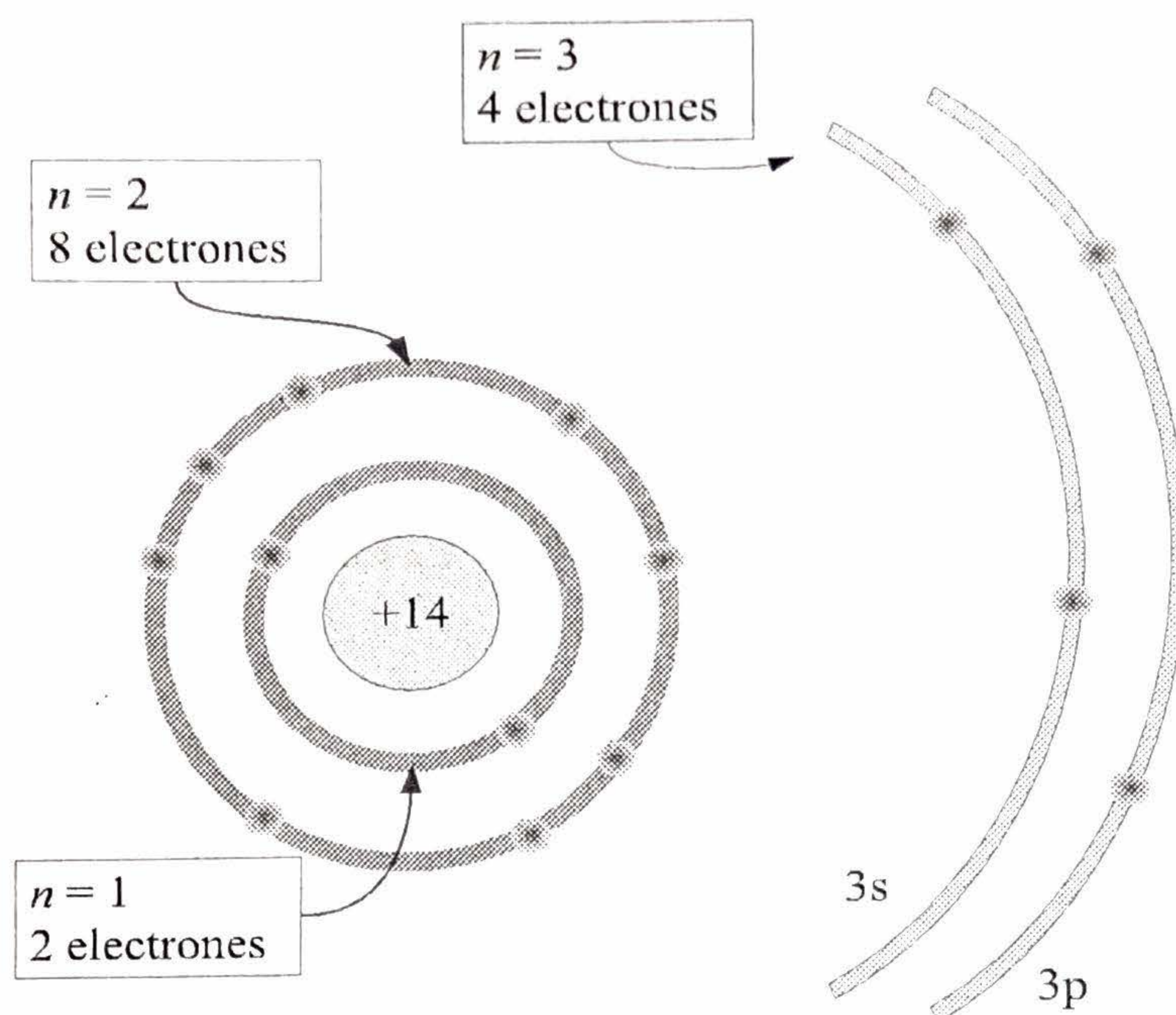


Figura 2.1: Niveles de energía ($n = 1, 2, 3$) y electrones en Silicio.

$3p$. Este ordenamiento se denomina configuración electrónica y se nota: $1s^2 2s^2 2p^6 3s^2 3p^2$. El último nivel de energía ($n = 3$) tiene ocho estados posibles y cuatro electrones. Este nivel es llamado nivel de valencia y posee los electrones con mayor energía. La cantidad de electrones de valencia (cuatro para el Silicio) determina cómo un elemento se relaciona e interactúa con otros.

Si se considera un conjunto de N átomos aislados de Silicio, en los niveles más altos de energía potencial hay $8N$ estados posibles y solo $4N$ de ellos están ocupados. Si se imagina un experimento con una cantidad lo suficientemente grande de átomos, en el cual la distancia entre átomos se fuese reduciendo, gradualmente los $6N$ estados s y los $2N$ estados p irían convergiendo a $4N$ estados vacíos y $2N + 2N$ estados ocupados. Cuando N es lo suficientemente grande, se puede considerar que en lugar de niveles discretos hay dos bandas continuas de energía. De la Fig. 2.3 se puede observar que la banda inferior (completa con electrones) alcanza un mínimo energético para un valor particular de espaciamiento. Este valor es la separación atómica de una red de Silicio en estado cristalino ($2,35\text{\AA}$, donde $1\text{\AA} = 0,1\text{nm}$). La banda inferior se denomina banda de valencia y la superior se denomina banda de conducción. También es importante notar que entre estas dos bandas de energía existe una banda intermedia donde no hay estados disponibles; esta banda se denomina banda prohibida.

El nivel inferior de la banda prohibida se denomina E_v y el nivel superior se denomina E_c . La diferencia entre ambos se denomina E_G y es la energía requerida por un electrón para pasar de la banda de valencia a la banda de conducción. Esta diferencia de energía es por lo tanto un indicador de la conducción del material. En los materiales aislantes E_G es grande; por ejemplo, para dióxido de Silicio: $E_G(\text{SiO}_2) = 8\text{eV}$; y para diamante: $E_G(\text{diamante}) = 5\text{eV}$. En los materiales conductores E_G es muy pequeño o negativo (en cuyo caso las bandas de conducción y valencia se hallan solapadas). En los materiales semiconductores los valores de E_G son intermedios; por ejemplo, en Silicio: $E_G(\text{Si}) = 1,12\text{eV}$; en Germanio: $E_G(\text{Ge}) = 0,66\text{eV}$; en Arseniuro de Galio: $E_G(\text{GaAs}) = 1,42\text{eV}$. La Fig. 2.4 ilustra las diferencias entre materiales de los tres tipos.

El nivel de energía para el cual la probabilidad de ocupación es 0,5 se denomina "Nivel de Fermi"

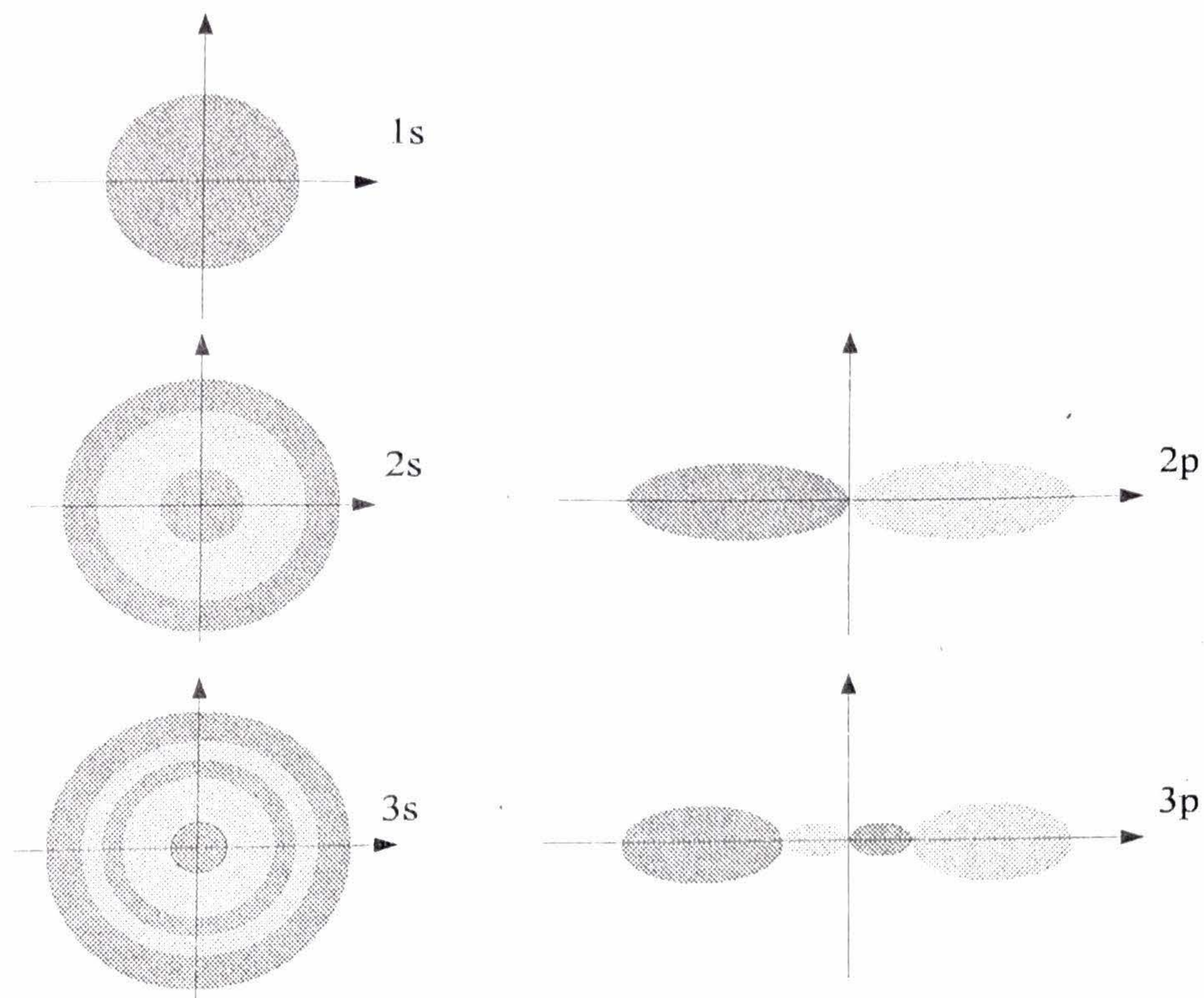


Figura 2.2: Sección transversal de los tres primeros tipos de orbitales.

y se nota E_F ⁵. Este nivel de energía depende de las masas efectivas de los electrones y en el caso de Silicio sin impurezas se halla aproximadamente en la mitad de la banda prohibida. La banda de conducción tiene un nivel mínimo de energía potencial E_c y un nivel máximo de energía potencial E_{sup} . La banda de valencia, a su vez, tiene un nivel mínimo de energía potencial E_{inf} y un nivel máximo de energía potencial E_v . Las bandas de energía potencial menores que E_{inf} se hallan más cercanas y más firmemente vinculadas al núcleo. Cuanto más cercano está un electrón al núcleo, más energía es necesaria para llevarlo, por ejemplo, a la banda de conducción. Los electrones por debajo de la banda de valencia están sólidamente vinculados a la red cristalina y no pueden moverse. Los electrones en banda de conducción pueden moverse. Es posible que haya electrones con más energía que E_{sup} y en tal caso, la diferencia entre la energía total del electrón E_e y la energía potencial E_{sup} es energía cinética, lo que indica que el electrón está en movimiento con cierta velocidad. También es importante destacar lo siguiente: cuando la cantidad de átomos N es lo suficientemente grande, las bandas de energía resultantes son una propiedad de la red cristalina entera y no de cada átomo aislado. En otras palabras, es la red cristalina la que posee bandas de energía. Una consecuencia directa de ello es que cada electrón en la banda de conducción ya no tiene una relación con el átomo que originalmente "lo contribuyó" sino que pertenece a la red y puede trasladarse a lo largo de ella.

Como se ha supuesto que la temperatura es lo suficientemente baja, no se producen cambios de estado en los electrones de la banda de valencia. Por lo tanto, cada átomo de Silicio tiene cuatro electrones en su banda de valencia y cuatro estados vacíos en su banda de conducción. Al combinarse en una red cristalina, cada átomo de Silicio se asocia con otros cuatro átomos vecinos produciendo lo que se denomina un enlace covalente. En esta situación, cada átomo de Silicio comparte un electrón de

⁵La estadística de Fermi-Dirac, en honor a sus pioneros, Enrico Fermi y Paul Dirac, describe la energía de partículas simples en un conjunto de partículas idénticas que obedecen el Principio de Exclusión de Pauli, como es el caso de los electrones.

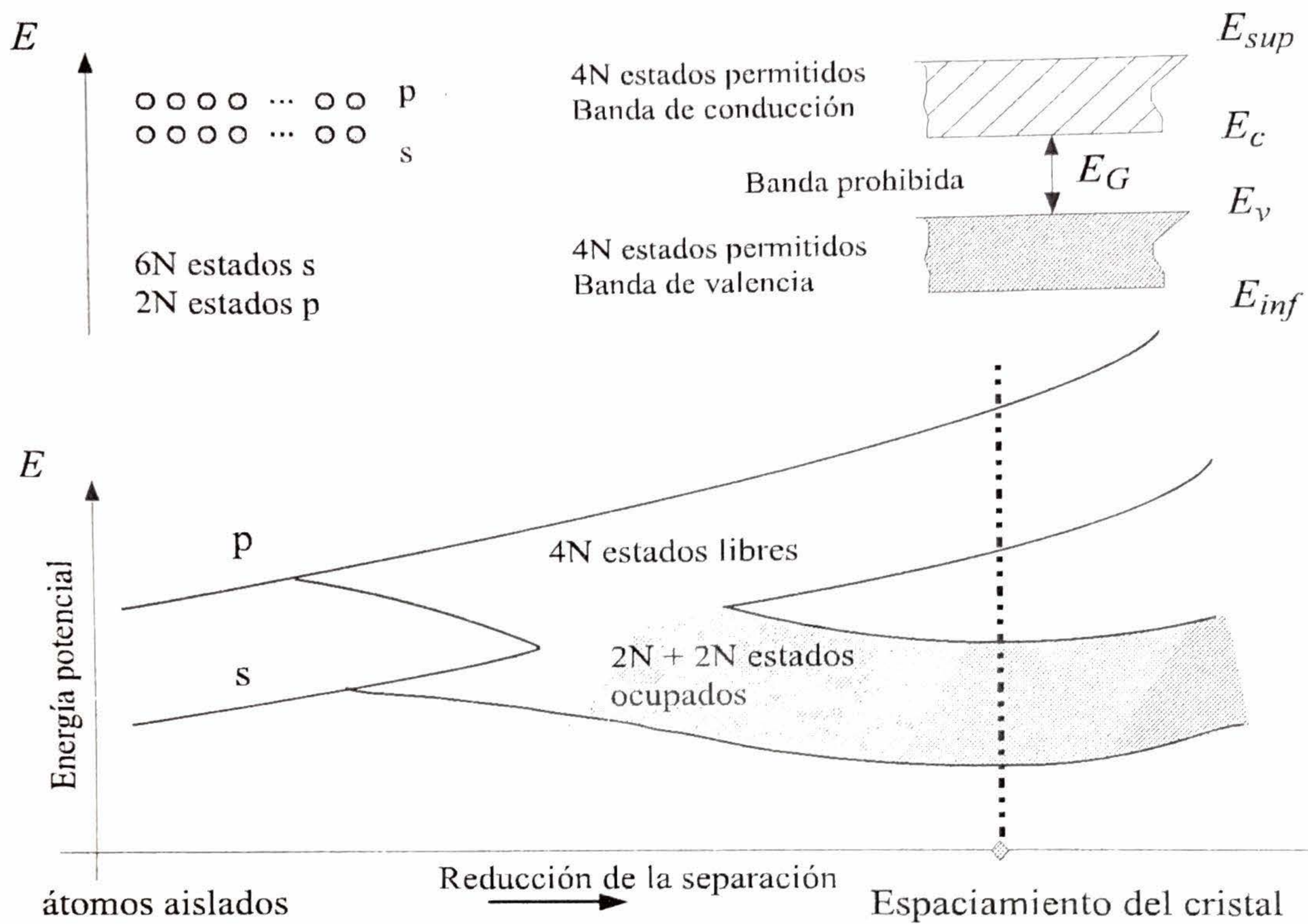


Figura 2.3: Bandas de energía y distribución de estados.

su banda de valencia con un átomo vecino y completa así sus ocho estados produciendo una estructura sólida conocida como cristal. Si se permite que la temperatura aumente, algunos electrones de la banda de valencia pueden adquirir energía (térmica) suficiente como para pasar a la banda de conducción. Esto produce dos efectos: Por un lado, aparece un electrón en la banda de conducción, es decir, un electrón con capacidad de trasladarse, que a partir de ahora se denominará portador negativo; por otro lado, en la banda de valencia se observa la ausencia de un electrón y la consiguiente aparición de una carga positiva. La ausencia del electrón es equivalente a la presencia de una partícula con carga positiva denominada hueco. El hueco puede trasladarse, al moverse los electrones en la banda de valencia, de manera análoga al movimiento de una burbuja de aire en agua. Estos dos tipos de portadores (negativos debido a los electrones en banda de conducción y positivos debido a los huecos en banda de valencia) son los responsables de la conducción de corriente.

Como ya se ha dicho, los átomos de Silicio están separados unos de otros $2,35\text{\AA}$. El cubo unitario de Silicio, definido como aquel cuya repetición produce la estructura cristalina, se ilustra en la Fig. 2.5. El mismo tiene $5,43\text{\AA}$ de lado y contiene un átomo por cada vértice, un átomo por cada lado, y cuatro átomos en su interior. El volumen del cubo unitario de Silicio es:

$$V = (5,43 \times 10^{-10} m)^3 = 1,6 \times 10^{-22} cm^3 \quad (2.2)$$

Cada cubo unitario de Silicio posee cuatro átomos propios, seis átomos compartidos de a pares con cubos vecinos, y ocho átomos compartidos de a ocho con cubos vecinos. Teniendo en cuenta que los átomos compartidos deben considerarse parcialmente, teniendo en cuenta su aporte al cubo unitario, la cantidad de átomos en un cubo unitario es $N = 4 + 6/2 + 8/8 = 8$.

De aquí surge la densidad atómica del Silicio, que es:

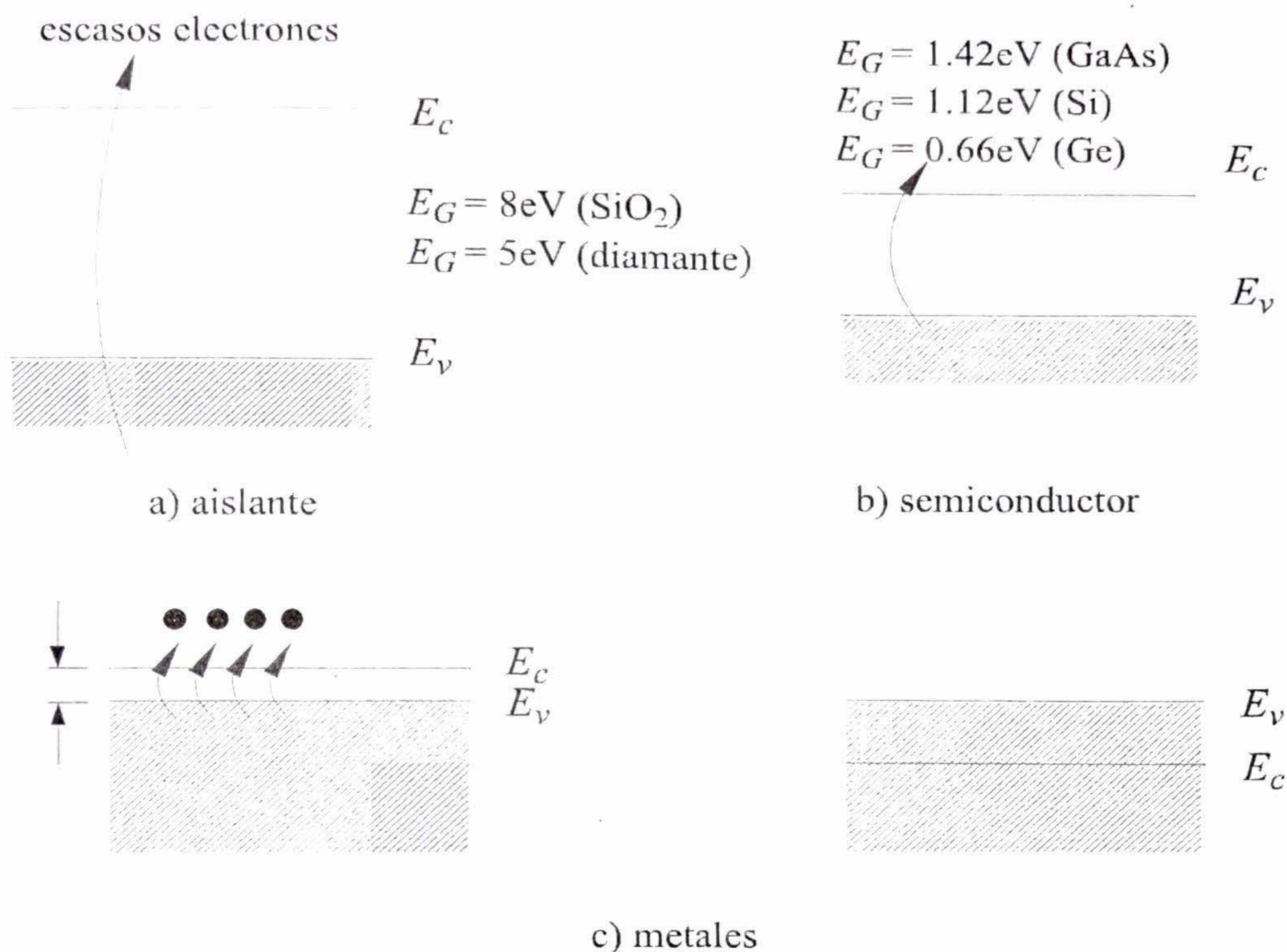


Figura 2.4: Bandas de energía: a) aislantes; b) semiconductores; c) metales.

$$d = \frac{8}{1,6 \times 10^{-22} \text{cm}^3} = 5 \times 10^{22} \text{cm}^{-3} \quad (2.3)$$

2.2. Equilibrio Térmico

Si bien la estructura cristalina del Silicio es tridimensional, es conveniente visualizar la red en dos dimensiones. La Fig. 2.6 muestra el estado de una red cristalina ideal a temperatura $T = 0K$. Esta representación se conoce comúnmente como “modelo de enlace”. La red cristalina se dice ideal cuando no hay ninguna impureza; en tal caso, al Silicio se lo denomina intrínseco. Los átomos de silicio están fijos a la red y son iones con una carga positiva $+4q$ dado que han contribuido cuatro electrones (con una carga neta de $-4q$) a sus vecinos⁶. De esta manera, se ve que la estructura es eléctricamente neutra (la carga neta es igual a cero), lo cual es natural dado que los átomos individuales son eléctricamente neutros. A esta temperatura ($T = 0K$), se puede observar que no hay portadores disponibles para conducción, con lo cual el material es un aislante.

Cuando la temperatura aumenta ($T > 0K$) la energía térmica produce la rotura de algunos enlaces covalentes. Cuando esto sucede, un electrón en banda de valencia obtiene la energía suficiente para pasar a la banda de conducción y moverse, y a su vez deja un hueco en la banda de valencia que también posee la capacidad de moverse (ver Fig. 2.7). En esta situación, se dice que se ha generado un par hueco-electrón. La tasa de generación térmica de portadores por unidad de volumen es una función exponencial de la temperatura [8], se nota $G_{th}(T)$ y sus unidades son cantidad de portadores por unidad de volumen por segundo: $1/\text{cm}^3\text{s}$. Los portadores también pueden generarse por efecto óptico

⁶ $q = 1,6 \times 10^{-19}$ Coulombs ([C]) es la carga del electrón.

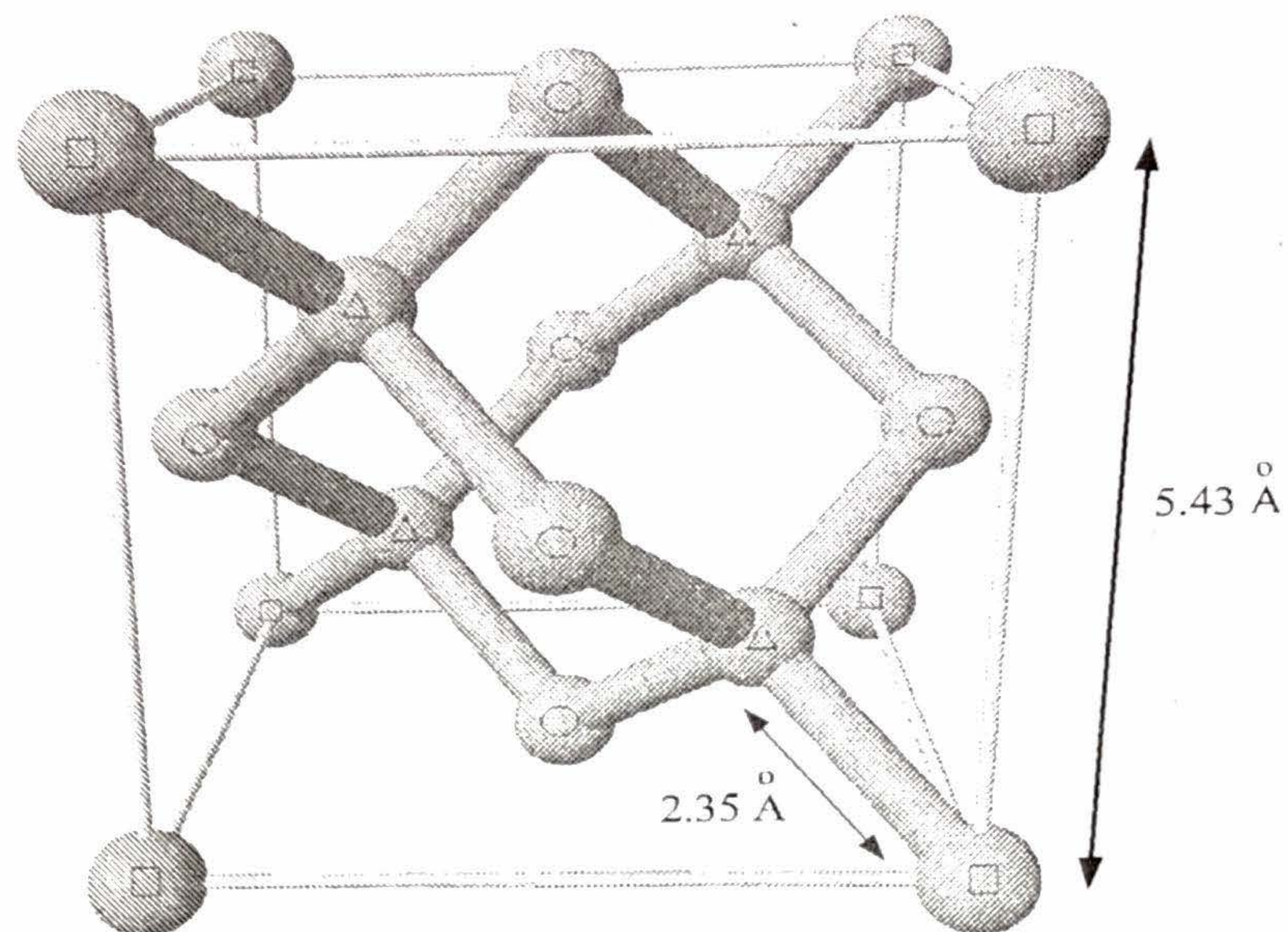


Figura 2.5: Estructura atómica del Silicio: los cuadrados indican los átomos sobre los vértices; los círculos indican los átomos sobre las caras; los triángulos indican los átomos interiores.

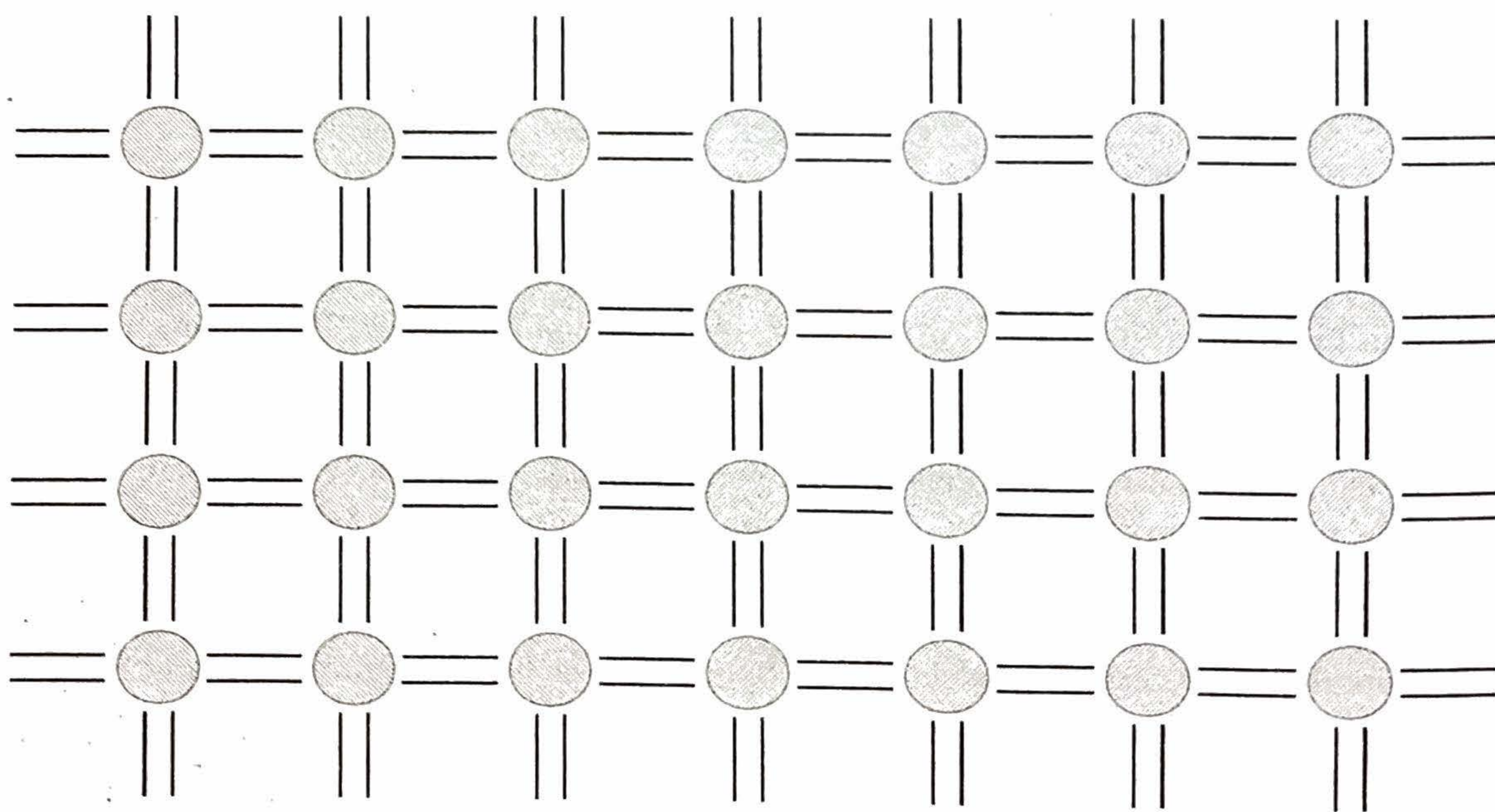


Figura 2.6: Representación bidimensional de la estructura cristalina del Silicio.

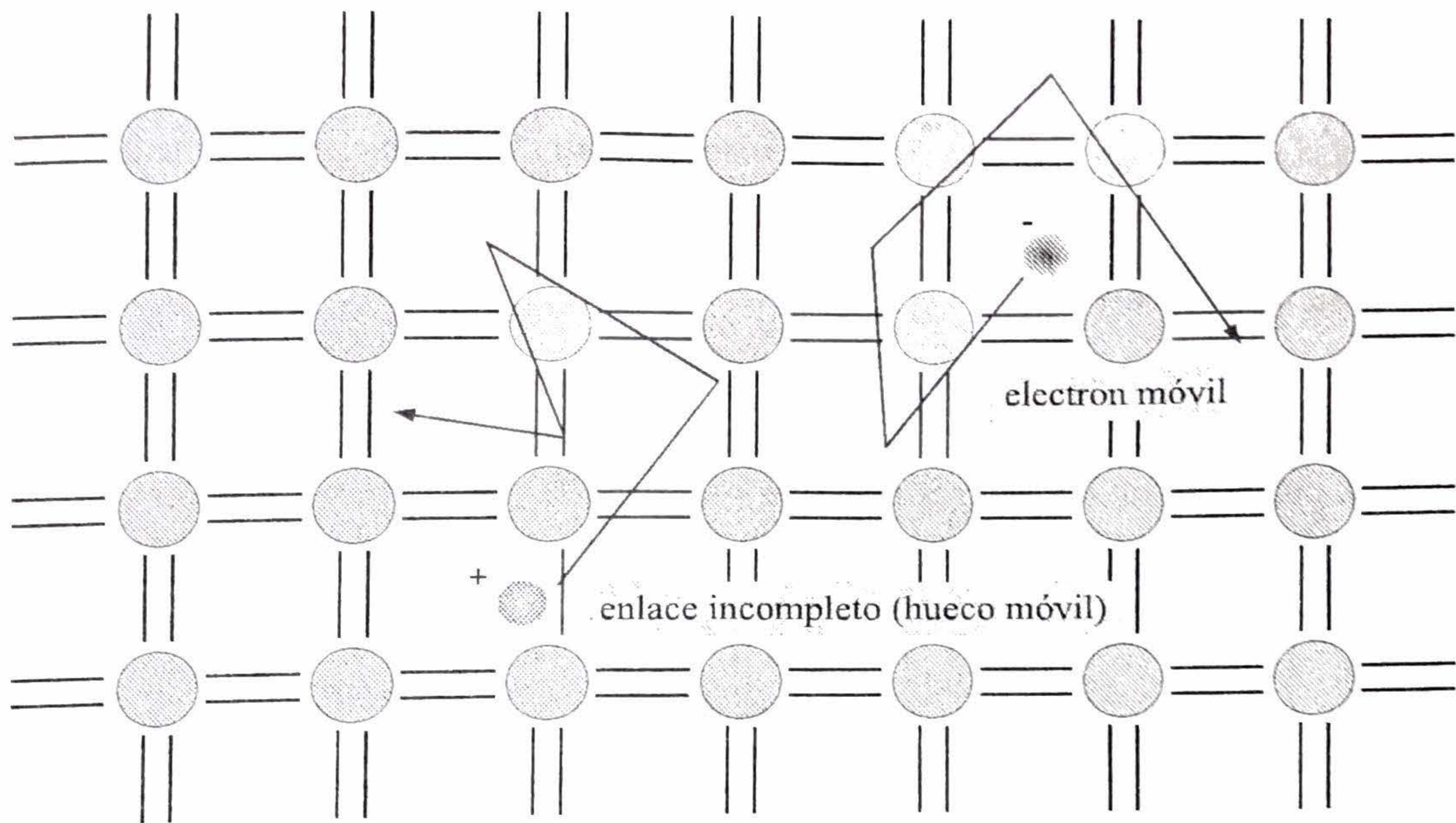


Figura 2.7: Representación bidimensional de la estructura cristalina del Silicio, indicando los portadores móviles para $T > 0K$.

si se ilumina el material con una fuente de luz apropiada. Si un fotón impacta contra la estructura puede proveer la energía suficiente para romper un enlace covalente y generar un par hueco-electrón. La tasa de generación por efecto óptico se nota G_{op} y es esencialmente independiente de la temperatura. La tasa total de generación de portadores es entonces:

$$G = G_{th}(T) + G_{op} \quad (2.4)$$

Por otro lado, cuando un electrón de la banda de conducción y un hueco de la banda de valencia se encuentran, se recombinan recomponiendo un enlace covalente. Este mecanismo compensa y balancea la producción de portadores. La recombinación de portadores puede clasificarse de acuerdo al tipo de energía liberada:

- Recombinación térmica: La recombinación se traduce en una vibración de la estructura cristalina o fonón, es decir calor. Este es el tipo de recombinación predominante en Silicio.
- Recombinación óptica: La recombinación produce un fotón. Este mecanismo es casi inexistente en Silicio, pero muy significativo en Arseniuro de Galio. Este mecanismo es la base del funcionamiento de los diodos emisores de luz o LEDs ("light emitting diodes").

Para que la recombinación ocurra se necesita un electrón y un hueco. Por lo tanto si una población de portadores es alta y la otra es baja, la tasa de recombinación será baja. Por el contrario, si ambas poblaciones son altas, la tasa de recombinación será alta. En efecto, la tasa de recombinación R resulta proporcional al producto de las poblaciones de portadores positivos y negativos:

$$R = k \times n \times p \quad (2.5)$$

donde n es la cantidad de portadores negativos, p es la cantidad de portadores positivos, y k es una constante de proporcionalidad.

Si se mantiene la temperatura constante hasta alcanzar un estado estable o de equilibrio, la tasa de generación de portadores debe ser igual a la tasa de recombinación de portadores, es decir:

$$k \times n \times p = G_{th}(T) + G_{op} \quad (2.6)$$

Un caso importante es el de equilibrio térmico, definido por la inexistencia de intercambio de energía del sistema con el medio exterior. En términos prácticos, esto se da ante la ausencia prolongada de estímulos externos (excitación óptica, o campos eléctricos aplicados), a una temperatura constante. En este caso especial, se obtiene⁷:

$$n \times p = \frac{G_{th}(T) + G_{op}}{k} \quad (2.7)$$

Dado que la población de portadores positivos es igual a la población de portadores negativos, se define la concentración intrínseca de portadores en equilibrio térmico n_i de la siguiente manera:

$$n_i \triangleq n = p \quad (2.8)$$

De (2.7) se deduce directamente que esta concentración satisface:

$$\boxed{n_i^2 = \frac{G_{th}(T) + G_{op}}{k}} \quad (2.9)$$

La Ec. (2.9) es conocida como la Ley de Acción de Masas. La cantidad de portadores excitados a la banda de conducción para una cierta temperatura T puede obtenerse utilizando mecánica cuántica [8] y está dada por la siguiente fórmula:

$$n_i = \left[2 \left(\frac{2\pi m_e kT}{h^2} \right)^{3/2} \right] e^{-E_G/2kT} \quad (2.10)$$

donde

T = temperatura en grados Kelvin

k = constante de Boltzmann: $1,38 \times 10^{-23} J/K$ ó $8,617 \times 10^{-5} eV/K$

m_e = masa del electrón: $9,1 \times 10^{-31} kg$

h = constante de Planck: $6,626 \times 10^{-34} joule \times s$

E_G = diferencia de energías entre los niveles de conducción y valencia

Por ejemplo, en el caso de Silicio a temperatura ambiente ($T = 300K$), la cantidad de portadores en la banda de conducción es:

$$n_i = 1 \times 10^{10} cm^{-3} \quad (2.11)$$

En el caso de Germanio la concentración intrínseca (a temperatura ambiente) es $n_i = 7,2 \times 10^{13} cm^{-3}$, mientras que para Arseniuro de Galio la concentración intrínseca (a temperatura ambiente) es $n_i = 3 \times 10^7 cm^{-3}$.

Nótese de (2.3) y (2.11) que en Silicio, a temperatura ambiente, hay un portador por cada 5×10^{12} átomos de Silicio, lo cual evidencia una concentración de portadores notablemente baja en términos relativos.

⁷Desde aquí en adelante, n y p indicarán las densidades de portadores en equilibrio térmico.

2.3. Dopado

Como ya se ha visto, la cantidad de portadores existentes en Silicio para conducción, a temperatura ambiente, es baja. Para modificar la conducción, se contamina o dopa el Silicio con átomos de otros elementos que se denominan impurezas. Antes de examinar en detalle el efecto del dopado, es conveniente analizar la distribución de portadores en Silicio intrínseco. En cada nivel de energía, hay una cierta cantidad de estados disponibles que pueden ser ocupados. La densidad de estados⁸ disponibles en las bandas de conducción y valencia están dados por las siguientes ecuaciones [8]:

$$g_c(E) = \frac{m_n \sqrt{2m_n(E - E_c)}}{\pi^2 h^3} \quad E \geq E_c \quad (2.12)$$

$$g_v(E) = \frac{m_p \sqrt{2m_p(E_v - E)}}{\pi^2 h^3} \quad E \leq E_v \quad (2.13)$$

y se ilustran en la Fig. 2.8.

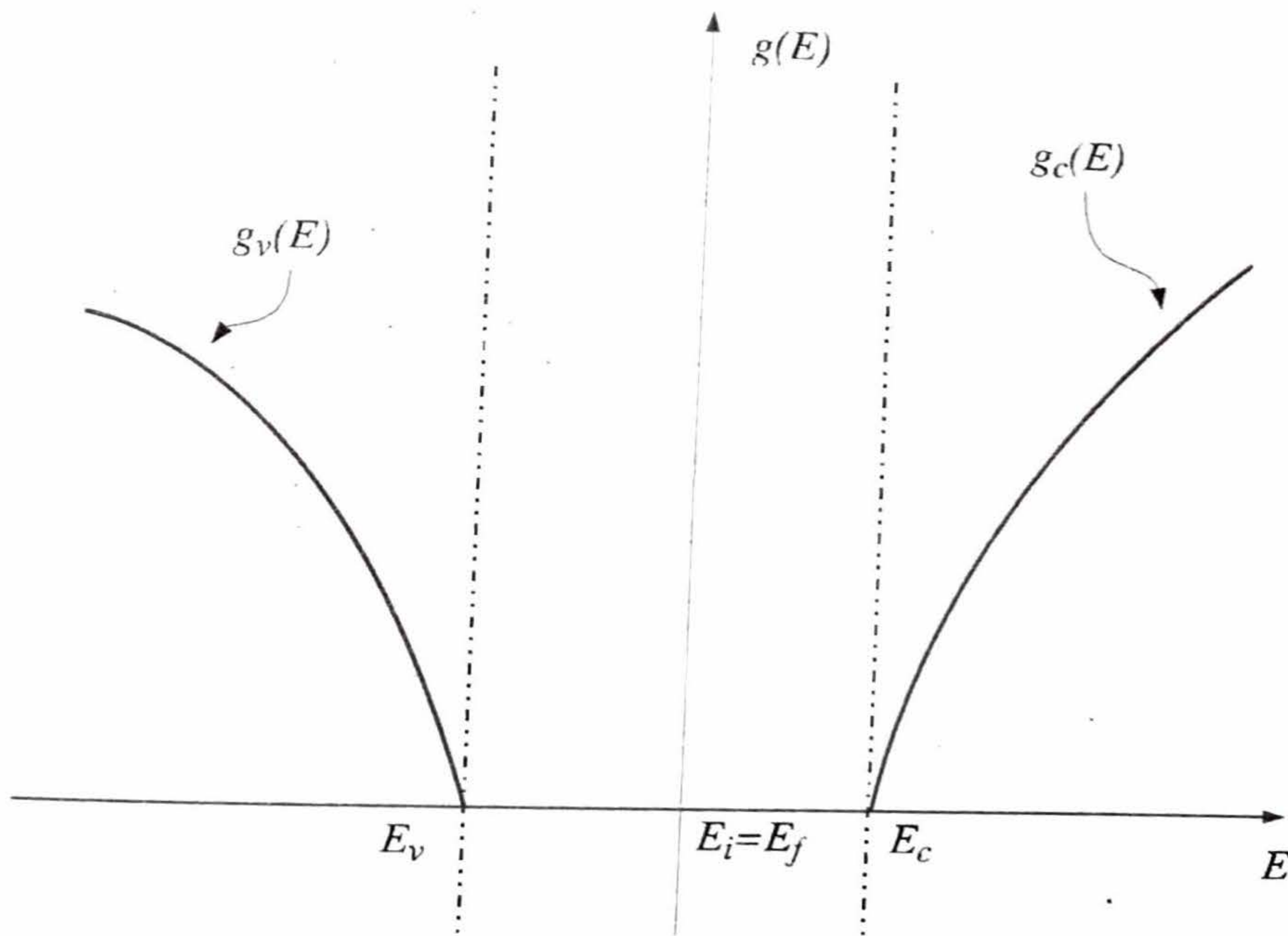


Figura 2.8: Funciones de densidad de estados de electrones en la banda de conducción, g_c , y de estados de huecos en la banda de valencia, g_v .

Esta distribución de estados implica que para energías mayores a E_c hay cada vez más estados disponibles para electrones, mientras que para energías menores a E_v hay cada vez más estados disponibles para huecos. La cantidad de portadores, en banda de conducción o valencia, dependen del producto de estados disponibles y de la probabilidad que los mismos estén ocupados. Como se discutió en la sección anterior, la probabilidad de ocupación de un estado se describe mediante la función de distribución de Fermi. Por ejemplo, la probabilidad de que un estado de energía E esté ocupado por un electrón es la siguiente:

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}} \quad (2.14)$$

donde E_F es el nivel de energía de Fermi.

⁸Esto es la cantidad de estados por unidad de volumen y unidad de energía.

La Fig. 2.9 muestra la distribución de Fermi (para electrones) a varias temperaturas. Se puede ver que a $T = 0K$ la probabilidad de existencia de un electrón en la banda de conducción es 0, mientras que la probabilidad en la banda de valencia es 1. A su vez, la probabilidad de que un estado de energía E esté ocupado por un hueco es:

$$h(E) = 1 - \frac{1}{1 + e^{(E-E_F)/kT}} \quad (2.15)$$

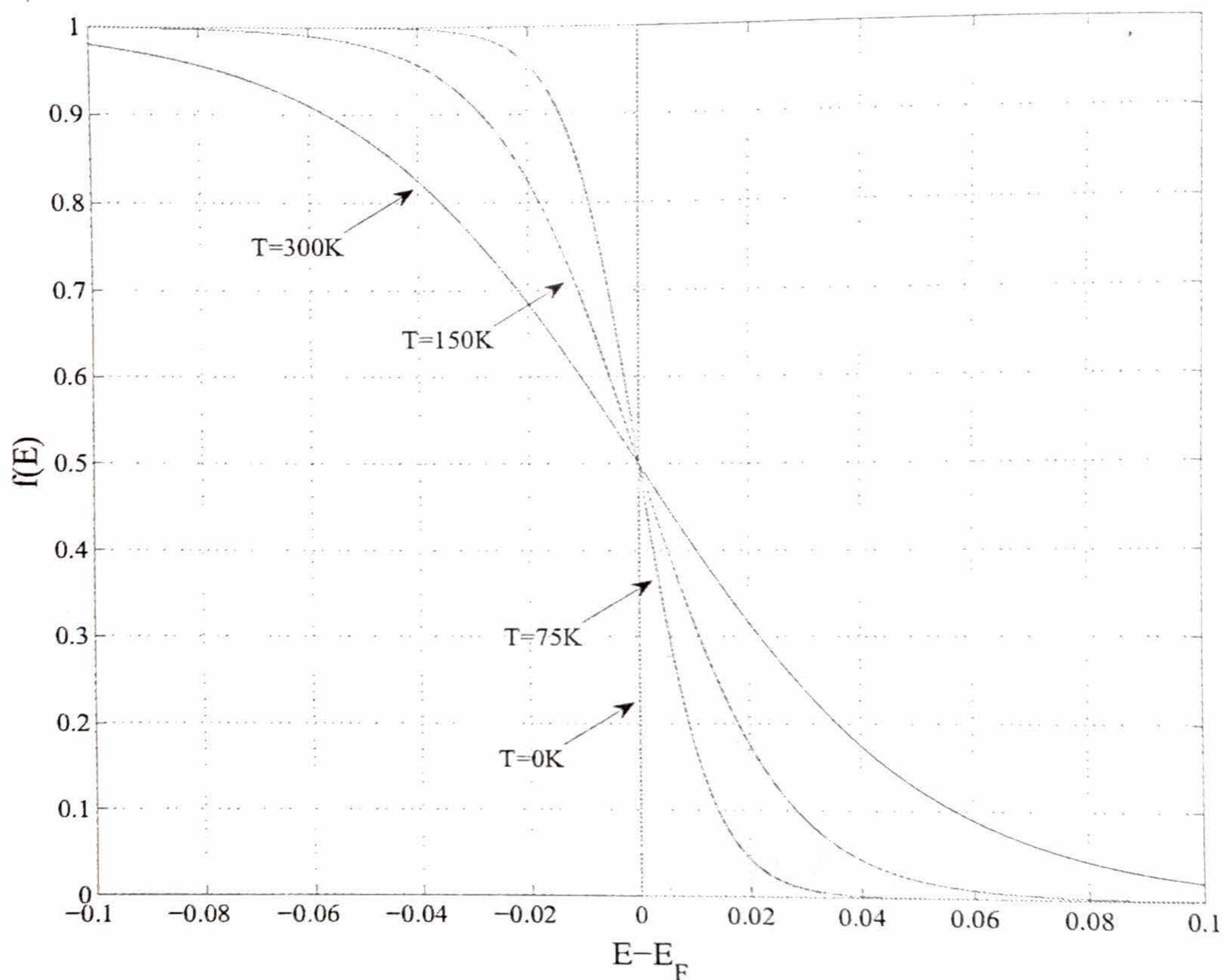


Figura 2.9: Función de distribución de Fermi de electrones $f(E)$ para varias temperaturas: $T = 0K$, $T = 75K$, $T = 150K$ y $T = 300K$.

La Fig. 2.10 muestra la densidad de estados, la función de Fermi y la cantidad de portadores para Silicio intrínseco.

2.3.1. Dopado Tipo N

Una manera de modificar la conductividad del Silicio consiste en aumentar la cantidad de electrones en la banda de conducción, a costa de la reducción de huecos en la banda de valencia, mediante la introducción de átomos, llamados impurezas, que “donan” electrones. Para ello, se introducen en el Silicio átomos del grupo V, como por ejemplo: Arsénico (As), Fósforo (P) o Antimonio (Sb). Los átomos del grupo V tienen en su banda de valencia cinco electrones. Dada la compatibilidad geométrica del Arsénico y del Fósforo con el Silicio, el átomo donador comparte cuatro de sus cinco electrones con átomos vecinos de Silicio, y el quinto queda con un vínculo energético débil. Un aumento de la temperatura hace que este electrón tenga la energía necesaria para acceder a la banda de conducción y pueda desplazarse por la red; esto deja al átomo de impureza ionizado positivamente. La Fig. 2.11-a

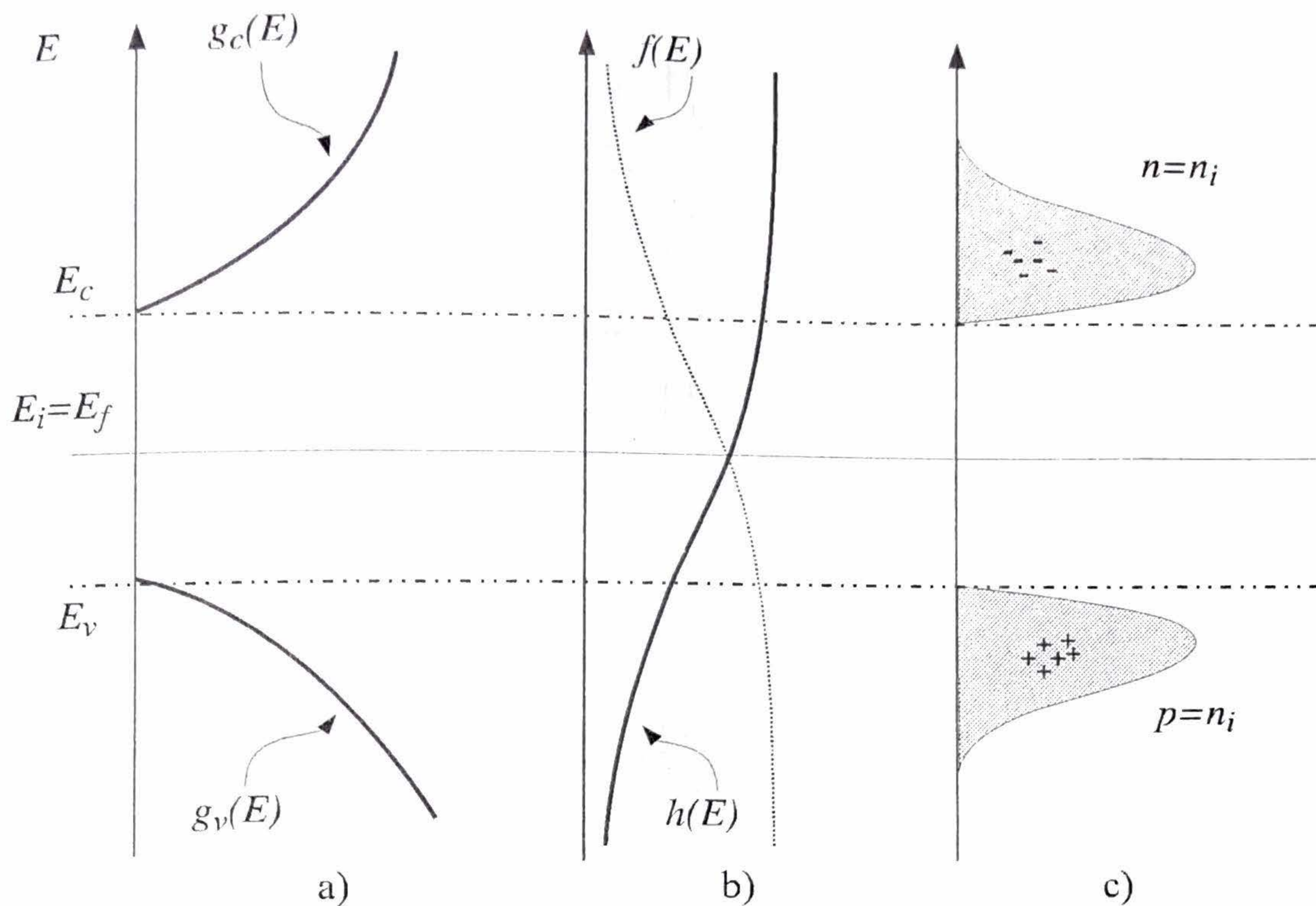


Figura 2.10: Portadores en Silicio intrínseco: a) densidad de estados de electrones $g_c(E)$ y huecos $g_v(E)$; b) distribución de Fermi de electrones $f(E)$ y huecos $h(E)$; c) portadores negativos (electrones) n y positivos (huecos) p .

ilustra la inserción de la impureza donadora en la red cristalina, y el portador negativo que resulta a temperatura ambiente. En términos energéticos, el quinto electrón de la impureza tiene un nivel de energía E_{imp} muy cercano al nivel de energía E_c (ver Fig. 2.11-b), de tal manera que se requiere muy poca energía para que este electrón pase a la banda de conducción⁹.

Para analizar como se modifican las concentraciones de portadores con la introducción de N_d [impurezas/cm³] impurezas donadoras, es necesario, en primer lugar, analizar el balance de cargas en el material. El Silicio es eléctricamente neutro, al igual que las impurezas introducidas en el material. Cada impureza introducida en el material, como ya se mencionó, provee un electrón a los portadores negativos del material. En consecuencia, cada impureza queda ionizada, es decir, con una carga neta positiva $+q$. El balance de carga dicta entonces, que la densidad neta de carga ρ , que es igual a la suma de las cargas de los portadores positivos, de los portadores negativos y de los iones donadores, debe ser nula:

$$\rho = 0 = q(-n + p + N_d) \quad (2.16)$$

Es útil recordar aquí, que cuando $N_d = 0$ (en el caso de Silicio intrínseco), resulta $n = p = n_i$. Sin embargo, al introducir impurezas, la cantidad n de portadores negativos, y la cantidad p de portadores positivos, difieren con respecto a n_i . Si N_d es pequeño con respecto a n_i , su efecto en (2.16) puede desprejarse. Si por el contrario, N_d es mucho mayor que n_i , habrá un incremento de la carga de iones positivos dado que cada uno de los átomos de impurezas queda ionizado positivamente. Debido a esto, los portadores negativos n deberán aumentar con respecto a su valor en Silicio intrínseco,

⁹A temperatura ambiente todas las impurezas resultan ionizadas.

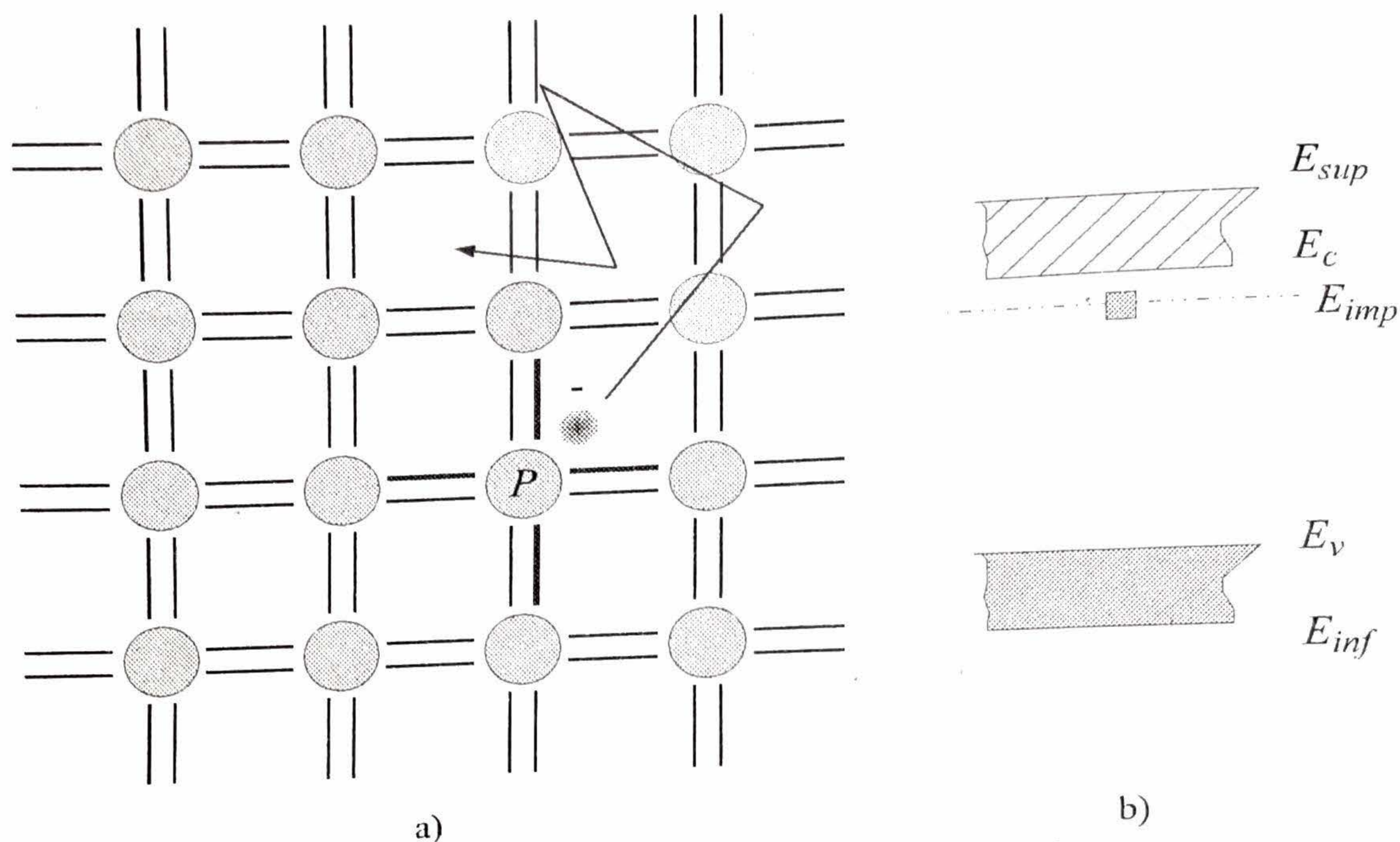


Figura 2.11: Dopado tipo N: a) inserción de un átomo donador en la grilla cristalina y su portador negativo a temperatura ambiente; b) nivel de energía del portador de la impureza donadora E_{imp} .

n_i , de manera de equilibrar la carga neta. Intuitivamente se puede ver que el efecto es aumentar los portadores negativos. Cuánto aumentan los portadores negativos y qué sucede con los portadores positivos? Dado que la población estable de portadores surge del equilibrio entre la recombinación y la generación, establecido por la Ley de Acción de Masas, deben cumplirse simultáneamente (2.16) y

$$n \times p = n_i^2 \quad (2.17)$$

La Ec. (2.17) es válida mientras que las impurezas y el Silicio no interactúen fuertemente, lo que sucede para dopados menores a $10^{19}/cm^3$; en esta situación el Silicio se dice "no-degenerado". Para dopados mayores, el Silicio se dice "degenerado" y el producto $n \times p$ resulta mayor que n_i^2 , y exhibe una dependencia no-lineal con respecto al dopado [9].

Si se sustituye $p = n_i^2/n$ en (2.16), resulta:

$$0 = q \left(-n + \frac{n_i^2}{n} + N_d \right) \quad (2.18)$$

La solución a esta ecuación es:

$$n = \frac{N_d + \sqrt{N_d^2 + 4n_i^2}}{2} = \frac{N_d}{2} + \frac{N_d}{2} \sqrt{1 + \frac{4n_i^2}{N_d^2}} \quad (2.19)$$

El cristal siempre¹⁰ se dopa de manera que $N_d \gg n_i$, razón por la cual, (2.19) puede simplificarse de la siguiente manera:

¹⁰Si la concentración de impurezas es menor que la concentración intrínseca, los niveles de portadores no resultan alterados.

y la población de huecos resulta:

$$n \approx N_d \tag{2.20}$$

$$p \approx n_i^2 / N_d \tag{2.21}$$

Se ve entonces que la población de electrones portadores aumenta en “un portador por impureza”, mientras que la población de huecos se reduce drásticamente. Es por esto que, en materiales dopados con impurezas donadoras, los portadores negativos son portadores mayoritarios, y los portadores positivos son minoritarios. El Silicio dopado con impurezas donadoras se denomina Silicio tipo N. En Silicio tipo N, la conducción se produce por la circulación de los portadores negativos.

El dopado tipo N produce un corrimiento del nivel de Fermi E_f hacia la banda de conducción, que se coloca por encima del nivel E_i , como se muestra en la Fig. 2.12-b. La Fig. 2.12-a muestra la densidad de estados, que permanece sin cambios. En la Fig. 2.12-c se muestran los portadores, y allí se puede apreciar el incremento de los portadores negativos con respecto a los portadores positivos.

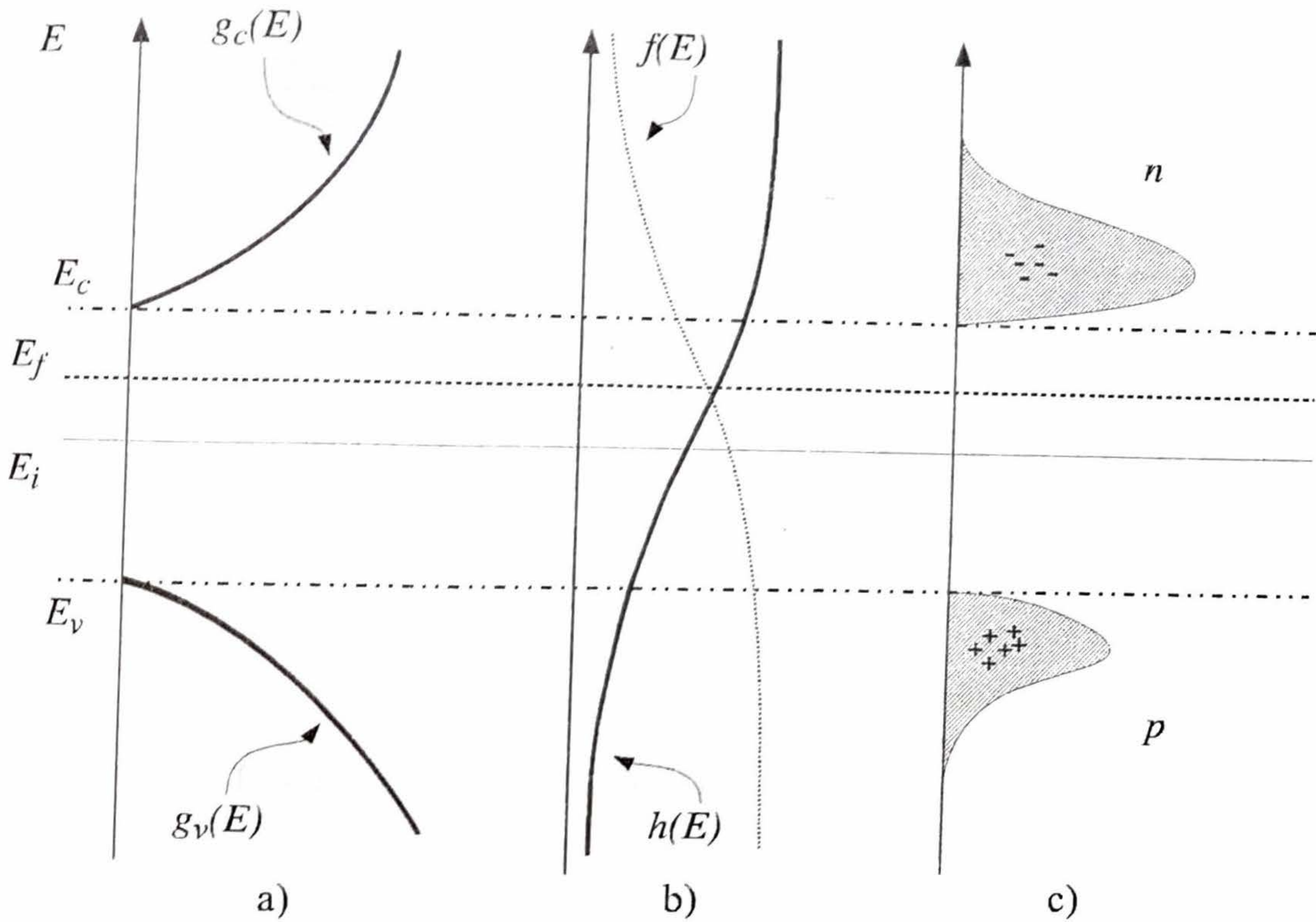


Figura 2.12: Portadores en Silicio dopado tipo N: a) densidad de estados de electrones $g_c(E)$ y huecos $g_v(E)$; b) distribución de Fermi de electrones $f(E)$ y huecos $h(E)$; c) portadores negativos (electrones) n y positivos (huecos) p .

La solución (2.19) puede visualizarse de manera gráfica, si se dibujan las dos ecuaciones que rigen el equilibrio; es decir, $n = p + N_d$, debido al balance de cargas, y $n = n_i^2/p$, debido a la Ley de Acción de Masas. La Fig. 2.13 muestra ambas ecuaciones sobre el par de ejes $n - p$. El punto de intersección entre ambas curvas es el punto de equilibrio de concentración de portadores.

Ejemplo 9 Una muestra de Silicio se dopa con Fósforo utilizando $N_d = 1 \times 10^{15}$. Cuáles son las concentraciones resultantes de portadores ?

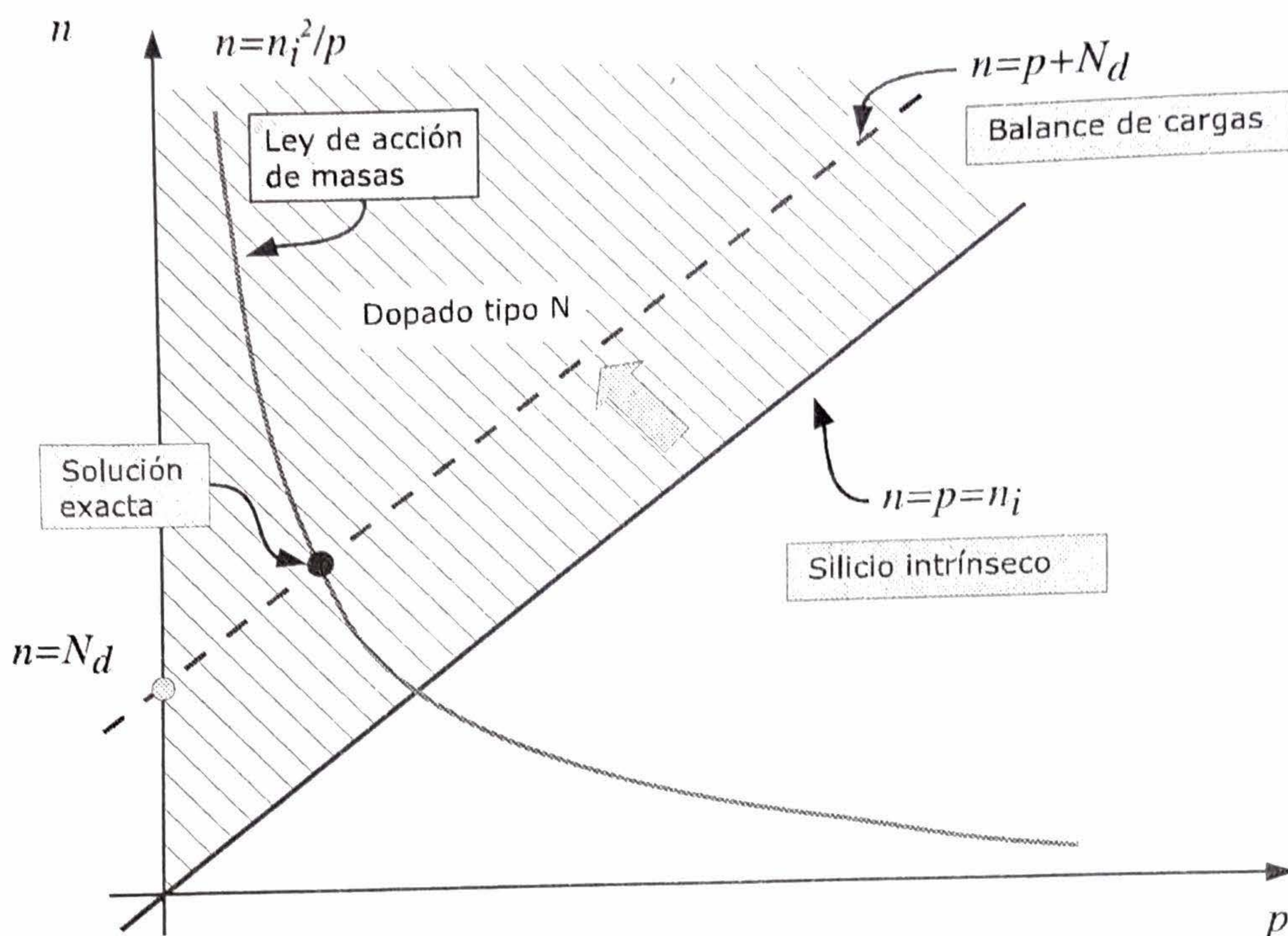


Figura 2.13: Visualización gráfica del dopado tipo N.

Dado que se verifica $N_d \gg n_i$, es posible utilizar la aproximación (2.20), con lo cual la población de portadores negativos es:

$$n = N_d = 1 \times 10^{15} \quad (2.22)$$

y la población de huecos es:

$$p = n_i^2 / N_d = 10^{20} / 10^{15} = 10^5 \quad (2.23)$$

Nótese que la población de huecos se ha reducido a 100.000 portadores, mientras que la de electrones ha aumentado 5 órdenes de magnitud.

El lector puede corroborar que el cálculo exacto de portadores es $n = 1,00000000010 \times 10^{15}$ y $p = 9,9999999990 \times 10^4$, lo cual confirma la aproximación realizada. Una rutina de Matlab, que produce el cálculo de las concentraciones, se lista a continuación:

```
format long;
ni=1e10; Nd=1e15; a=-1; b=Nd; c=ni^2 ;
raices=roots([a b c]);
raiz=max(raices);
n=raiz
p=ni^2/raiz
```

Por último, y en relación con las concentraciones típicamente utilizadas para el dopado de Silicio, el límite inferior es aproximadamente 10^{13} cm^{-3} por los motivos expuestos, y el límite superior ronda 10^{19} cm^{-3} y está limitado por la *solubilidad* del material donador en Silicio. La solubilidad determina en qué momento las impurezas dejan de entremezclarse con el Silicio para empezar a formar dominios entre sí. Aquí debe recordarse que para que las impurezas contribuyan al aumento de los portadores,

deben asociarse con átomos de Silicio. En el momento en que las impurezas comienzan a asociarse entre sí, la contribución efectiva a la generación de portadores no varía¹¹.

2.3.2. Dopado Tipo P

La otra manera de incrementar la conductividad del Silicio es aumentar el número de huecos en la banda de valencia, a costa de una reducción de los electrones de banda de conducción, mediante la introducción de átomos de impurezas que "aceptan" un electrón. Para ello, se introducen en el Silicio átomos del grupo III, por ejemplo: Boro (B), Galio (Ga), Indio (In) o Aluminio (Al); siendo Boro el más utilizado. Los átomos del grupo III tienen en su banda de valencia tres electrones. Dada la compatibilidad geométrica del Boro con el Silicio, el átomo aceptor comparte sus tres electrones con los átomos vecinos, produciendo un hueco en la red (debido a la ausencia de un cuarto electrón) que queda con un vínculo energético débil. Un aumento de la temperatura¹² hace que un electrón de la banda de valencia tenga la energía necesaria para salir de ella y unirse a un átomo aceptor, el cual queda ionizado negativamente. El hueco resultante en banda de valencia puede desplazarse por la red. La Fig. 2.14-a ilustra la inserción de la impureza aceptor en la red cristalina, y el portador positivo que resulta a temperatura ambiente. En términos energéticos, el hueco de la impureza tiene un nivel de energía E_{imp} muy cercano al nivel de energía E_v (ver Fig. 2.14-b), de tal manera que se requiere muy poca energía para que un electrón de banda de valencia se recombine con él, y se genere un hueco que puede trasladarse por la banda de valencia.

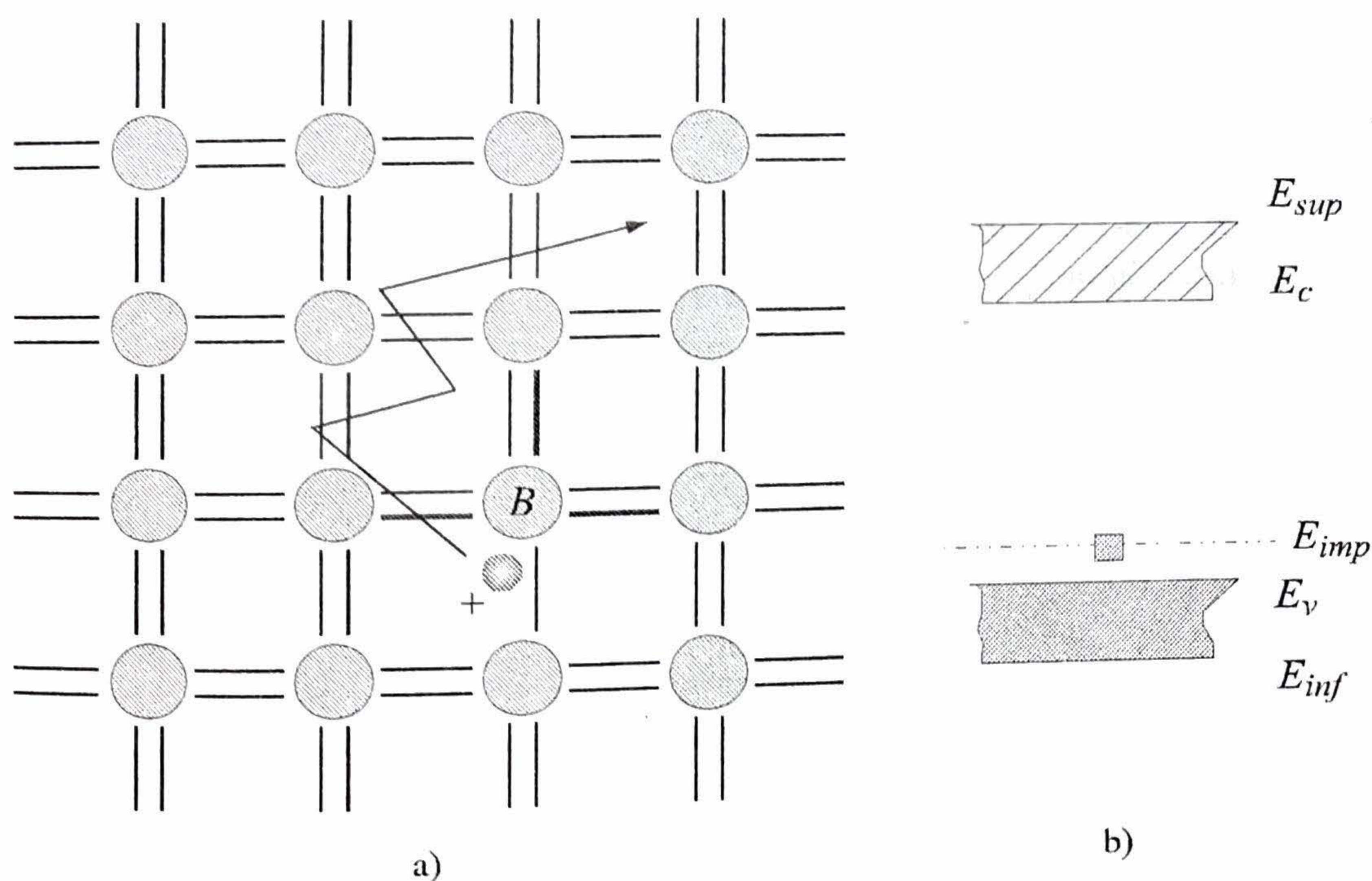


Figura 2.14: Dopado tipo P: a) inserción de un átomo aceptor en la grilla cristalina y su portador positivo a temperatura ambiente; b) nivel de energía del portador de la impureza E_{imp} .

Para analizar como se modifican las concentraciones de portadores con la introducción de N_a

¹¹El fenómeno es similar a la disolución de azúcar en agua. A partir del momento que se alcanza el límite de solubilidad, el agregado de azúcar no se disuelve en el agua, y el agua cesa de endulzarse.

¹²A temperatura ambiente todas las impurezas resultan ionizadas.

impurezas aceptoras, es necesario realizar el balance de cargas en el material. El Silicio es eléctricamente neutro, al igual que las impurezas introducidas. Cada impureza introducida en el material provee un hueco a los portadores positivos del material. En consecuencia, cada impureza queda ionizada, es decir, con una carga neta negativa $-q$. El balance de carga dicta entonces, que la densidad neta de carga ρ , que es igual a la suma de las cargas de los portadores positivos, de los portadores negativos y de los iones aceptores, debe ser nula:

$$\rho = 0 = q(-n + p - N_a) \quad (2.24)$$

De aquí se puede ver que si N_a es mucho mayor que n_i , entonces p deberá aumentar con respecto a su valor en equilibrio térmico, n_i , de manera de equilibrar la carga neta. Intuitivamente se puede ver que el efecto es aumentar los portadores positivos. Para calcular la población de portadores positivos se deben aplicar simultáneamente (2.24) y la Ley de Acción de Masas:

$$n \times p = n_i^2 \quad (2.25)$$

Sustituyendo $n = n_i^2/p$ en (2.24) resulta:

$$0 = q \left(p - \frac{n_i^2}{p} - N_a \right) \quad (2.26)$$

La solución a esta ecuación es:

$$p = \frac{N_a + \sqrt{N_a^2 + 4n_i^2}}{2} = \frac{N_a}{2} + \frac{N_a}{2} \sqrt{1 + \frac{4n_i^2}{N_a^2}} \quad (2.27)$$

El cristal siempre se dopa de manera que $N_a \gg n_i$ dado que si la concentración de impurezas es menor que la concentración intrínseca, los niveles de portadores no se alteran. Teniendo en cuenta esto, (2.27) puede simplificarse de la siguiente manera:

$$\boxed{p \approx N_a} \quad (2.28)$$

En este caso, la población de electrones resulta:

$$\boxed{n \approx n_i^2/N_a} \quad (2.29)$$

Se ve entonces que la población de huecos portadores aumenta en "un portador por impureza", mientras que la población de electrones se reduce drásticamente. Es por esto que en materiales dopados con impurezas aceptoras, los portadores positivos son portadores mayoritarios, y los portadores negativos son minoritarios. El Silicio dopado con impurezas aceptoras se denomina Silicio tipo P. En Silicio tipo P, la conducción se produce por la circulación de los portadores positivos.

El dopado tipo P produce un corrimiento del nivel de Fermi hacia la banda de valencia, como se muestra en la Fig. 2.15.

Como en el caso anterior, la solución (2.27) puede verse de manera más intuitiva si se grafican las dos ecuaciones que rigen el equilibrio, es decir, $n = p - N_a$, por el balance de cargas, y $n = n_i^2/p$, por la Ley de Acción de Masas. La Fig. 2.16 muestra ambas ecuaciones sobre el par de ejes $n - p$. El punto intersección entre ambas curvas es el punto de equilibrio de concentración de portadores.

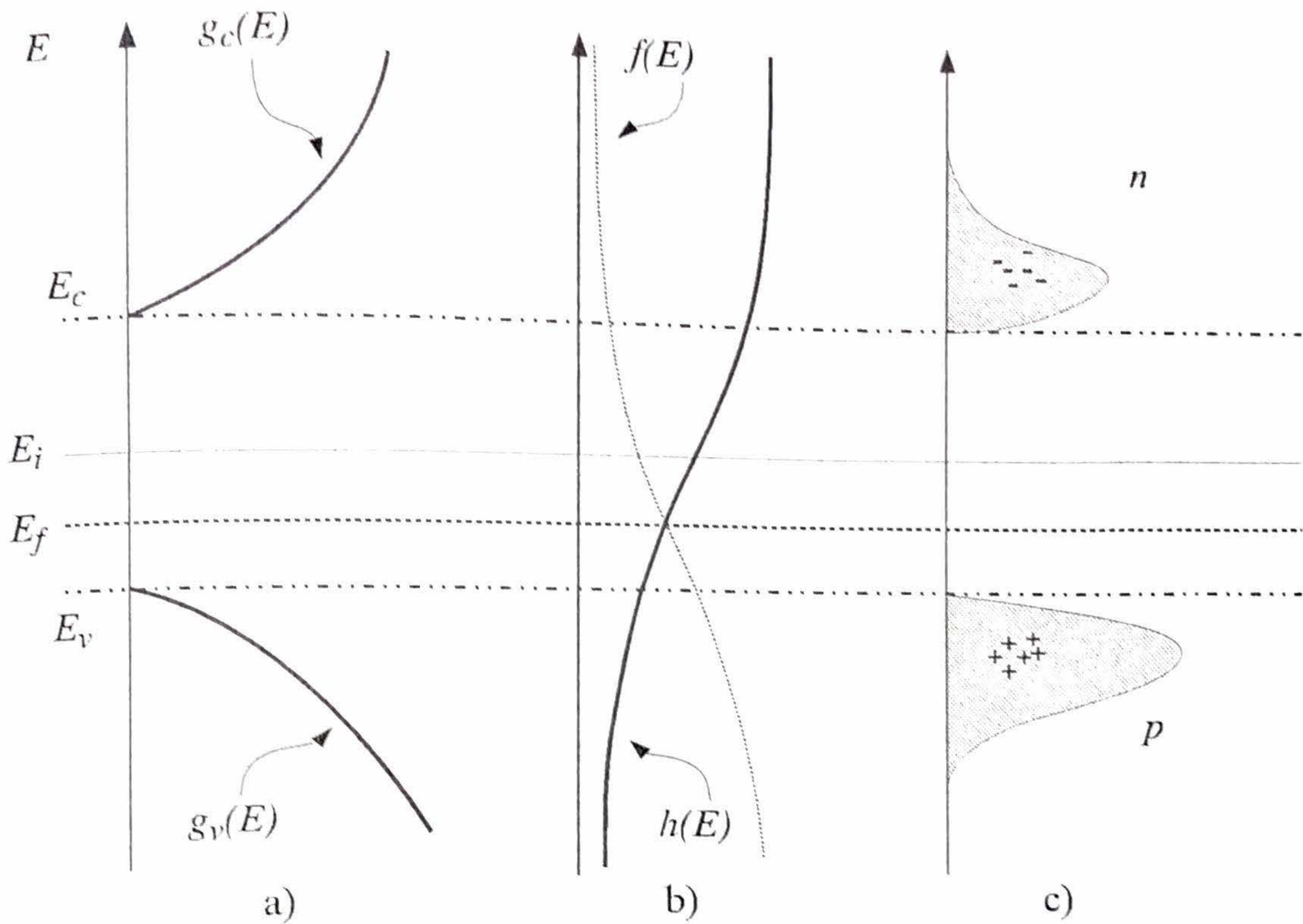


Figura 2.15: Portadores en Silicio dopado tipo P: a) densidad de estados de electrones $g_c(E)$ y huecos $g_v(E)$; b) distribución de Fermi de electrones $f(E)$ y huecos $h(E)$; c) portadores negativos (electrones) n y positivos (huecos) p .

2.3.3. Compensación

Por último, es posible combinar los dos tipos de dopados en lo que se denomina *compensación*. Este proceso se utiliza habitualmente en la secuencia de procesamiento de obleas de Silicio, cuando se desea dopar una zona con el tipo de dopado contrario al existente en la oblea, o en una región particular. La compensación consiste en someter al Silicio a un dopado tipo P (ó N), seguido por un dopado tipo N (ó P, respectivamente). El análisis es similar a los realizados anteriormente. En este caso particular, la Ley de Acción de Masas seguirá siendo válida, y la conservación de cargas requiere:

$$\begin{aligned} \rho = 0 &= q(-n + p - N_a + N_d) \\ &= q(-n + p + (N_d - N_a)) \end{aligned} \quad (2.30)$$

De aquí se puede ver que (2.30) es similar a (2.16) y (2.24), con la diferencia que ahora quien define el tipo de dopado resultante es el término $N_d - N_a$. En efecto, en el caso que $N_d - N_a > 0$, el material tendrá un dopado efectivo tipo N, de magnitud $\tilde{N}_d = N_d - N_a$. Si, por el contrario, $N_d - N_a < 0$, entonces el material tendrá un dopado efectivo tipo P, de magnitud $\tilde{N}_a = N_a - N_d$. Con estas consideraciones, es posible aplicar los desarrollos de las secciones anteriores para arribar a las cantidades totales de portadores. En efecto, si $N_d - N_a > 0$, entonces

$$n = \frac{(N_d - N_a)}{2} + \frac{(N_d - N_a)}{2} \sqrt{1 + \frac{4n_i^2}{(N_d - N_a)^2}} \quad (2.31)$$

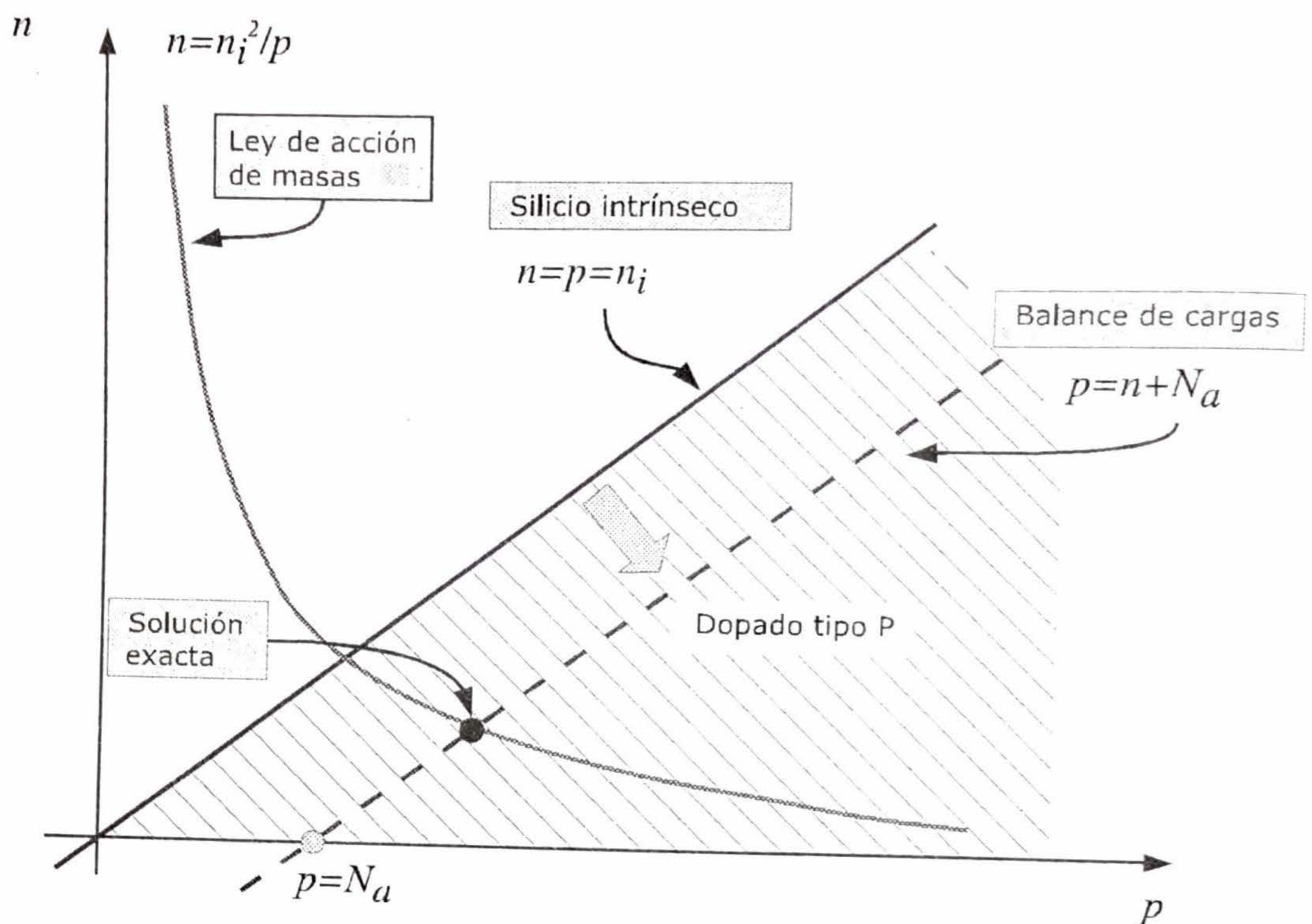


Figura 2.16: Visualización gráfica del dopado tipo P.

Si $N_d \gg N_a$, y considerando que ambos dopados siempre son mayores que la concentración intrínseca, es decir, $N_d, N_a \gg n_i$, (2.31) se simplifica a:

$$n \approx N_d - N_a \tag{2.32}$$

En este caso, la población de huecos resulta:

$$p \approx \frac{n_i^2}{N_d - N_a} \tag{2.33}$$

Por el contrario, cuando $N_d - N_a < 0$, resulta:

$$p = \frac{(N_a - N_d)}{2} + \frac{(N_a - N_d)}{2} \sqrt{1 + \frac{4n_i^2}{(N_a - N_d)^2}} \tag{2.34}$$

Si $N_a \gg N_d \gg n_i$, entonces (2.34) se simplifica a:

$$p \approx N_a - N_d \tag{2.35}$$

En este caso, la población de electrones es:

$$n \approx \frac{n_i^2}{N_a - N_d} \tag{2.36}$$

La Fig. 2.17 ilustra las posibles situaciones, según sea $N_a \gg N_d$, o $N_d \gg N_a$, lo cual equivale a trasladar la recta de balance de carga hacia la derecha o la izquierda, respectivamente.

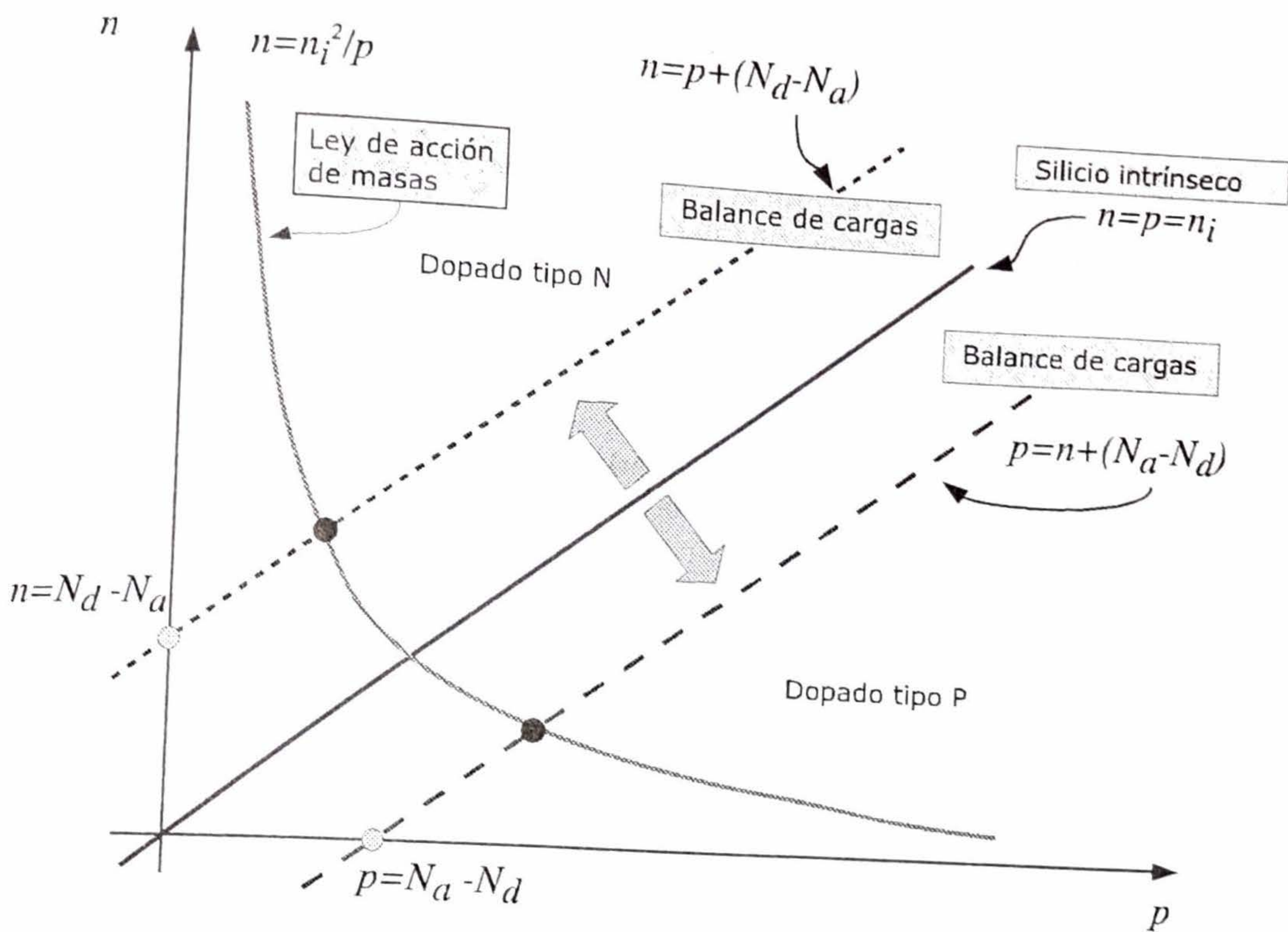


Figura 2.17: Visualización gráfica de la compensación.

2.4. Mecanismos de conducción

La conducción de corriente en un semiconductor puede producirse por dos mecanismos. El primer mecanismo ocurre cuando un campo eléctrico aplicado produce una fuerza de *arrastre* sobre los portadores, forzando el movimiento. El segundo mecanismo ocurre cuando la concentración de portadores es mayor en un lugar que en otro, de manera que los portadores (movilizados aleatoriamente por acción de la temperatura) *difunden* de la zona de mayor concentración a la de menor concentración. En general, ambos mecanismos se encuentran en todos los dispositivos, aunque salvo casos particulares¹³, en la mayoría de los casos es posible representar el movimiento de portadores por uno solo de estos mecanismos, lo que simplifica significativamente los desarrollos.

2.4.1. Arrastre

Si se considera una pieza de Silicio en equilibrio térmico (sin la acción de un campo eléctrico o fuentes de energía óptica) se verá a los electrones moviéndose por acción térmica a una velocidad promedio de $v_{th} = 1 \times 10^7 \text{ cm/s}$, y colisionando cada $\tau_c = 0,1 \text{ ps}$. Es decir, los electrones se agitan por efecto de la temperatura y se aceleran, cambiando de rumbo de manera aleatoria al interactuar con la red cristalina (fenómeno conocido como "scattering"). Considerando la velocidad y el tiempo de colisión, un electrón recorre, entre colisiones, tramos de una longitud promedio:

$$\lambda = v_{th} \times \tau_c \approx 1 \times 10^7 \text{ cm/s} \times 10^{-13} \text{ s} = 10 \text{ nm} \quad (2.37)$$

Esta distancia se conoce como "camino libre promedio". A pesar de ello, el movimiento neto de la masa de electrones es nulo, con lo cual la corriente neta es nula. Un razonamiento análogo es aplicable

¹³Un ejemplo es el transistor MOS, descrito en el Cap. 5, que en un entorno de la tensión de umbral conduce por acción de ambos mecanismos

a los huecos.

Si se considera ahora un campo eléctrico E actuando sobre una pieza de silicio, los electrones estarán bajo la acción de una fuerza electrostática $F = -qE$, y los huecos bajo una fuerza $F = +qE$. Debido a ello, se producirá un desplazamiento neto de la masa de electrones en la dirección opuesta al campo eléctrico, y un desplazamiento neto de la masa de huecos en la dirección del campo eléctrico; esto evidencia la aparición de una corriente neta. La velocidad de arrastre de los portadores (tanto electrones como huecos) se define como:

$$v^a = \frac{\overline{\Delta x}}{\Delta t} \quad (2.38)$$

donde $\overline{\Delta x}$ es el desplazamiento promedio de los portadores en el intervalo Δt .

Experimentalmente se verifica que la relación entre el campo eléctrico y la velocidad de los portadores satisface:

$$\boxed{v_n^a = -\mu_n E} \quad (2.39)$$

y

$$\boxed{v_h^a = +\mu_p E} \quad (2.40)$$

donde μ_n y μ_p son denominadas, respectivamente, las movilidades de electrones y huecos. Nótese aquí que un campo eléctrico constante aplicado a un electrón aislado –libre en el espacio– produce una aceleración constante $F = -qE$, y por lo tanto un crecimiento lineal de la velocidad con respecto al tiempo. Por el contrario, en un cristal, la interacción con la red hace que los portadores se frenen periódicamente, razón por la cual la velocidad, ante un campo eléctrico constante, resulta constante.

Las movilidades de huecos y electrones no son constantes, sino que varían de acuerdo al dopaje de la muestra de Silicio. Intuitivamente puede razonarse que cuanto mayor sea la concentración de impurezas en el material, mayor será la probabilidad de que un electrón/hueco interactúe con la red cristalina produciendo colisiones. Es por ello que el dopado neto de la muestra ocasiona una reducción de la movilidad de los portadores (ver Fig. 2.18).

La movilidad de huecos y electrones en Silicio puede aproximarse por la siguiente fórmula empírica:

$$\mu = \mu_{min} + \frac{\mu_o}{1 + (N/N_{ref})^\alpha} \quad (2.41)$$

donde (considerando $T = 300K$) para el caso de electrones: $N_{ref} = 1,3 \times 10^{17} cm^{-3}$, $\mu_{min} = 92 cm^2/Vs$, $\mu_o = 1268 cm^2/Vs$ y $\alpha = 0,91$; y para el caso de huecos: $N_{ref} = 2,35 \times 10^{17} cm^{-3}$, $\mu_{min} = 54,3 cm^2/Vs$, $\mu_o = 406,9 cm^2/Vs$ y $\alpha = 0,88$.

Nótese que la movilidad de los electrones es aproximadamente tres veces mayor que la de los huecos. En otras palabras, ante un mismo campo eléctrico, los electrones se desplazan tres veces más rápido que los huecos. Para dopados típicos alrededor de $5 \times 10^{16} cm^{-3}$,

$$\mu_n = 1000 cm^2/Vs, \quad \mu_p = 400 cm^2/Vs \quad (2.42)$$

Ejemplo 10 *El siguiente código puede utilizarse para dibujar en Matlab las curvas de movilidad versus dopado:*

```
% Movilidad de electrones
N=logspace(13,19,50);
% Parametros
Nref=1.3e17;
umin=92;
```


2.4. Mecanismos de conducción

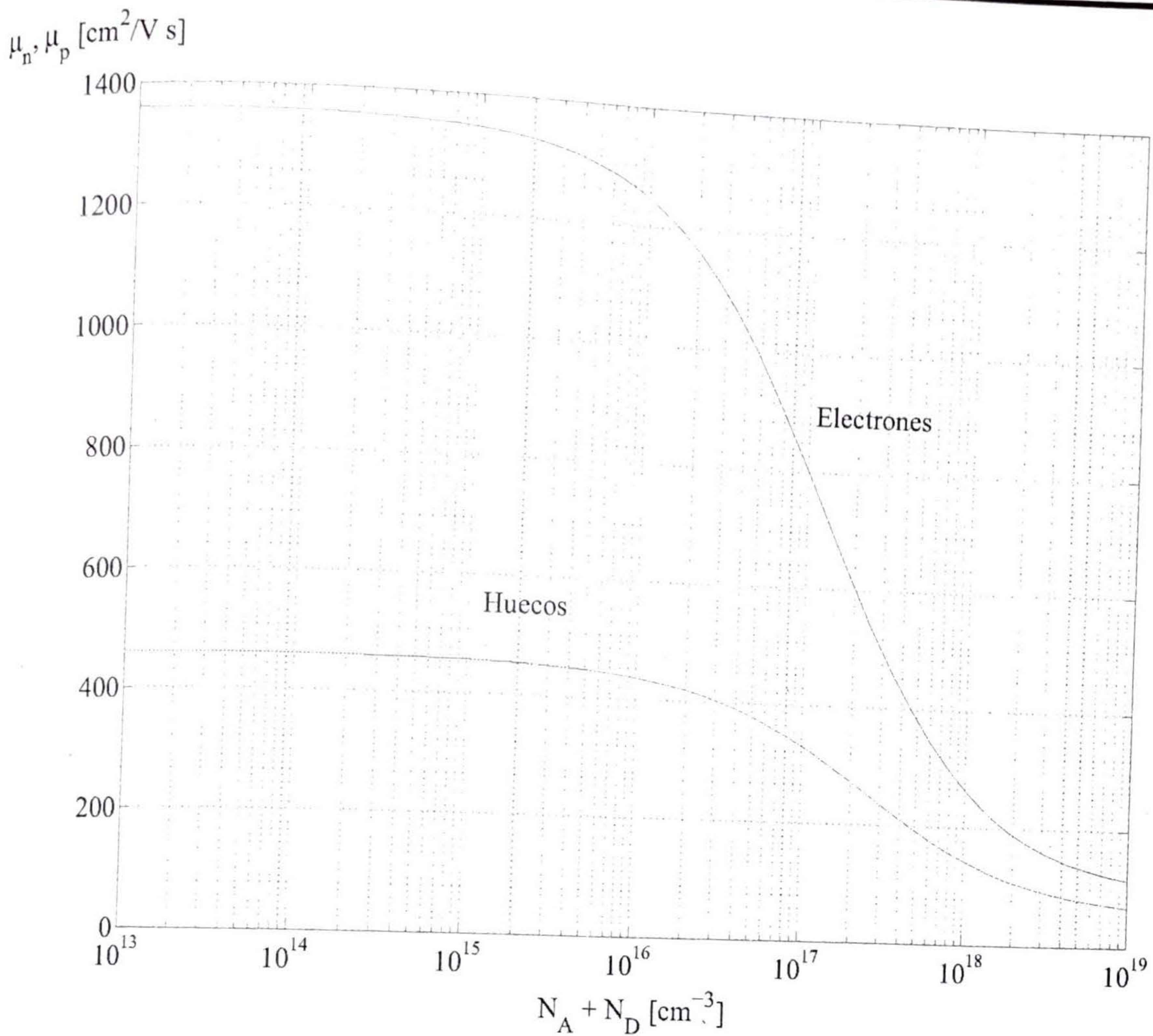


Figura 2.18: Movilidad de huecos y electrones en Silicio en función del dopado neto.

```

uo=1268;
alfa=0.91;
un=umin+uo./(1+(N./Nref).^alfa);
% Movilidad de huecos
% Parametros
Nref=2.35e17;
umin=54.3;
uo=406.9;
alfa=0.88;
up=umin+uo./(1+(N./Nref).^alfa);
semilogx(N,un);
semilogx(N,up);

```

Otro efecto en la conducción sucede con el aumento de la temperatura, el cual produce mayores efectos de colisión, reduciendo la movilidad. Para bajos dopados, la disminución de movilidad con respecto a la temperatura sigue una ley aproximadamente polinomial: $\mu_n \approx T^{-2,3}$ y $\mu_p \approx T^{-2,3}$. Para dopados mayores, la movilidad de los portadores sigue disminuyendo con la temperatura, pero a una tasa menor debido a los efectos de las colisiones con las impurezas ionizadas. Estos efectos pueden

modelarse, utilizando una corrección-función de la temperatura-para cada uno de los parámetros de (2.41), de acuerdo a la siguiente fórmula:

$$[\text{parámetro}] = [\text{parámetro}]_{300} \left(\frac{T}{300} \right)^\eta \quad (2.43)$$

donde [parámetro] representa uno cualquiera de los siguientes parámetros: μ_{min} , μ_o , N_{ref} ó α ; $[\text{parámetro}]_{300}$ es el valor del parámetro a $T = 300K$, T es la temperatura en Kelvin y η es una constante que depende del parámetro específico. En el caso particular de Silicio, para el parámetro N_{ref} , debe considerarse $\eta = 2,4$; para el parámetro μ_{min} , debe considerarse $\eta = -0,57$; para el parámetro μ_o , en el caso de electrones, debe considerarse $\eta = -2,33$, y en el caso de huecos, debe considerarse $\eta = -2,23$; para el parámetro α , debe considerarse $\eta = -0,146$.

Por último, es preciso señalar que, conforme el campo eléctrico aumenta, la velocidad de los portadores comienza a experimentar un efecto de saturación. Este efecto se nota para campos eléctricos cercanos a $10^4 V/cm$, cuando la velocidad de saturación se aproxima a $v_{(sat)} = 10^7 cm/s$, como se muestra en la Fig. 2.19. Esta situación es típica en dispositivos VLSI de dimensiones sub-micrométricas debido a los mayores campos eléctricos¹⁴.

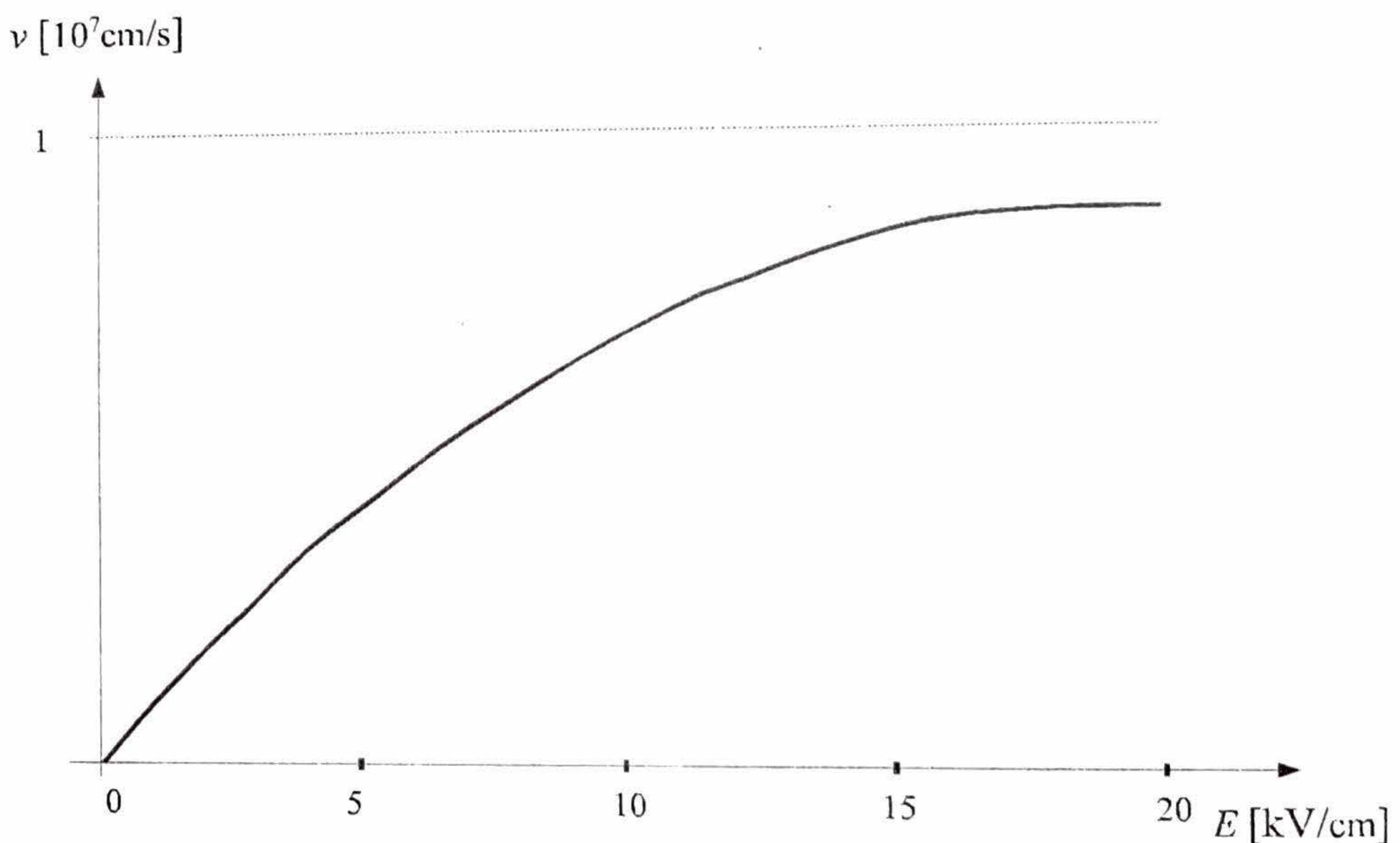


Figura 2.19: Velocidad de arrastre de electrones en función del campo eléctrico aplicado.

Densidad de corriente de arrastre

Para calcular la densidad de corriente de arrastre, se debe analizar la variación de la carga con respecto al tiempo ($i = dQ/dt$), ante la presencia de un campo eléctrico. Para ello, se considerará un volumen dado de portadores y se calculará cuál es la variación de carga por unidad de tiempo, a través de un plano de referencia. Se asumirá un campo eléctrico positivo, es decir, un campo eléctrico que mueve cargas positivas hacia $x > 0$ y cargas negativas hacia $x < 0$. En el caso de huecos, en un intervalo de tiempo Δt el volumen de carga existente es:

¹⁴El campo eléctrico se origina cuando un potencial dado se aplica sobre un material de cierta longitud. En los procesos sub-micrométricos, las dimensiones de los dispositivos se reducen agresivamente. Las tensiones de funcionamiento también se reducen, para disminuir los campos eléctricos, pero en menor relación.

$$\Delta V = Av_p^a \Delta t \quad (2.44)$$

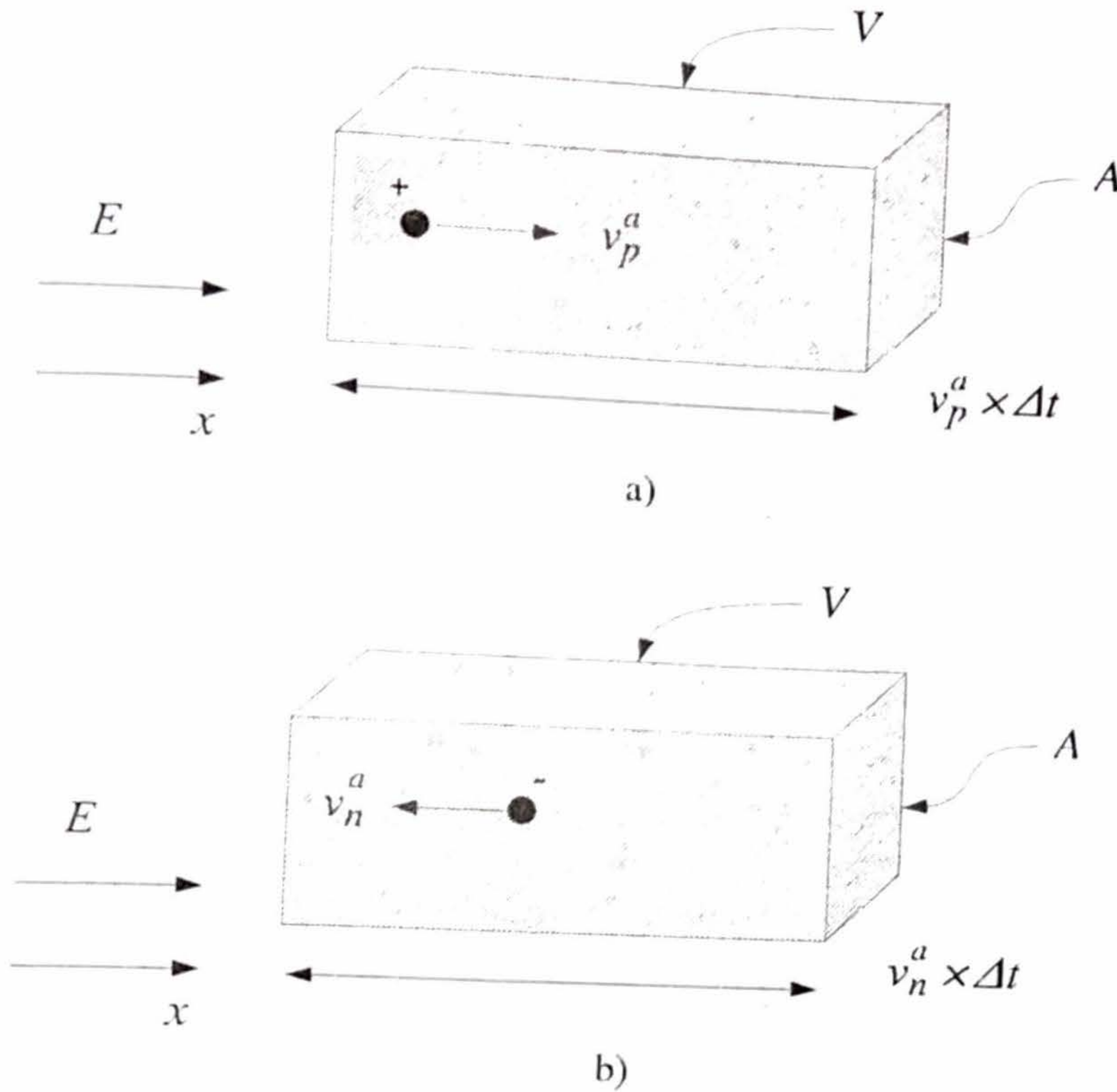


Figura 2.20: Volumen de referencia y conducción de portadores por arrastre: a) huecos; b) electrones.

El número de portadores positivos ΔN_p , que cruzan el plano de referencia, es igual a la concentración de portadores p , multiplicada por el volumen ΔV (ver Fig. 2.20-a):

$$\Delta N_p = p \Delta V \quad (2.45)$$

La carga asociada a estos portadores es:

$$\Delta Q = q \Delta N_p = qp \Delta V \quad (2.46)$$

La densidad de corriente de arrastre se calcula como el número de portadores que atraviesan el plano de referencia, por unidad de área y de tiempo:

$$j_p^a = \frac{\Delta Q}{A \Delta t} = \frac{qpAv_p^a \Delta t}{A \Delta t} = qp v_p^a \quad (2.47)$$

Si se recuerda que $v_p^a = \mu_p E$, luego:

$$\boxed{j_p^a = qp \mu_p E} \quad (2.48)$$

De aquí se ve que la corriente de arrastre de huecos es proporcional a la densidad de portadores, a la movilidad y al campo eléctrico. Debe notarse en este caso, que un campo eléctrico positivo $E > 0$, produce una velocidad de huecos positiva v_p^a (es decir, huecos moviéndose de izquierda a derecha), y por lo tanto, una corriente de arrastre $j_p^a > 0$.

Para el caso de los electrones, la carga en el volumen (ver Fig. 2.20-b) es:

$$\Delta Q = -qn\Delta V \quad (2.49)$$

con lo cual, la corriente de arrastre de electrones es:

$$j_n^a = \frac{\Delta Q}{A\Delta t} = \frac{-qnAv_n^a\Delta t}{A\Delta t} = -qnv_n^a \quad (2.50)$$

Si se recuerda que $v_n^a = -\mu_n E$, luego:

$$\boxed{j_n^a = qn\mu_n E} \quad (2.51)$$

A diferencia del caso anterior, un campo eléctrico positivo $E > 0$ produce una velocidad de electrones negativa $v_n^a < 0$ (es decir, electrones moviéndose de derecha a izquierda), y por lo tanto, una corriente de arrastre $j_n^a > 0$.

Por último, es importante notar tanto en (2.48) como en (2.51), que la corriente es proporcional al producto del número de portadores y la movilidad. El aumento del dopado hace que el número de portadores aumente y que la movilidad disminuya. Por ende, estos dos factores tienen efectos opuestos sobre la conducción. Cuál es el efecto predominante? El aumento de portadores es directamente proporcional a la introducción de impurezas, y varía aproximadamente seis órdenes de magnitud ($10^{13} - 10^{19}$); la disminución de movilidad es de catorce veces para los electrones (de $1400\text{cm}^2/\text{Vs}$ a $100\text{cm}^2/\text{Vs}$) y de diez veces para los huecos (de $500\text{cm}^2/\text{Vs}$ a $50\text{cm}^2/\text{Vs}$), para la variación mencionada de dopaje. Es claro que el efecto del aumento de portadores es significativamente mayor que la disminución de movilidad. Por ende, el efecto neto del dopado es el aumento de la conducción de corriente, debido al aumento de la cantidad de portadores.

2.4.2. Difusión

Tanto los electrones como los huecos en Silicio están en movimiento producto de la energía térmica. Como se vió, esto provoca un movimiento aleatorio de ambos tipos de portadores. Si la concentración de portadores es homogénea, entonces el movimiento aleatorio de portadores no produce una corriente neta, y la misma es nula.

A continuación se verá que, si hay diferencias en las concentraciones de portadores, el movimiento aleatorio produce un reordenamiento que tiende a equilibrar las concentraciones, produciendo un movimiento de portadores, desde las regiones de mayor concentración hacia las de menor concentración¹⁵.

Para el análisis se considera una concentración con un gradiente no nulo, y un plano de referencia a través del cual pasarán los portadores. Se considera, por simplicidad, que el movimiento de portadores sucede solo en el eje x . A cada lado del plano de referencia, se analiza una región de sección transversal A y de longitud Δx , tal como se ilustra en la Fig. 2.21. En esta figura, el volumen de cada región resultante a ambos lados del plano de referencia, es el producto de la longitud a lo largo del eje x y la sección transversal A . La altura de la curva define la concentración conforme varía x . Si se considera que $\Delta x = \lambda$, luego de un tiempo τ_c , todos los portadores contenidos en el volumen $A\lambda$ se moverán, por efecto térmico, una distancia λ . De esta manera, todos los portadores iniciales habrán abandonado el volumen luego de transcurrido el tiempo τ_c . Dado que el movimiento térmico es aleatorio, es natural suponer que los portadores no tienen preferencia por alguna dirección; debido a ello, una mitad migrará hacia el volumen de la derecha, y la otra hacia el volumen de la izquierda. El plano de referencia verá pasar, entonces, la mitad de los portadores del volumen a su izquierda, $P_i/2$, hacia la derecha, y la mitad de los portadores del volumen a su derecha, $P_d/2$, hacia la izquierda. Resulta claro entonces que el flujo

¹⁵El fenómeno es equivalente a la difusión de un gas en el aire, o de tinta en el agua.

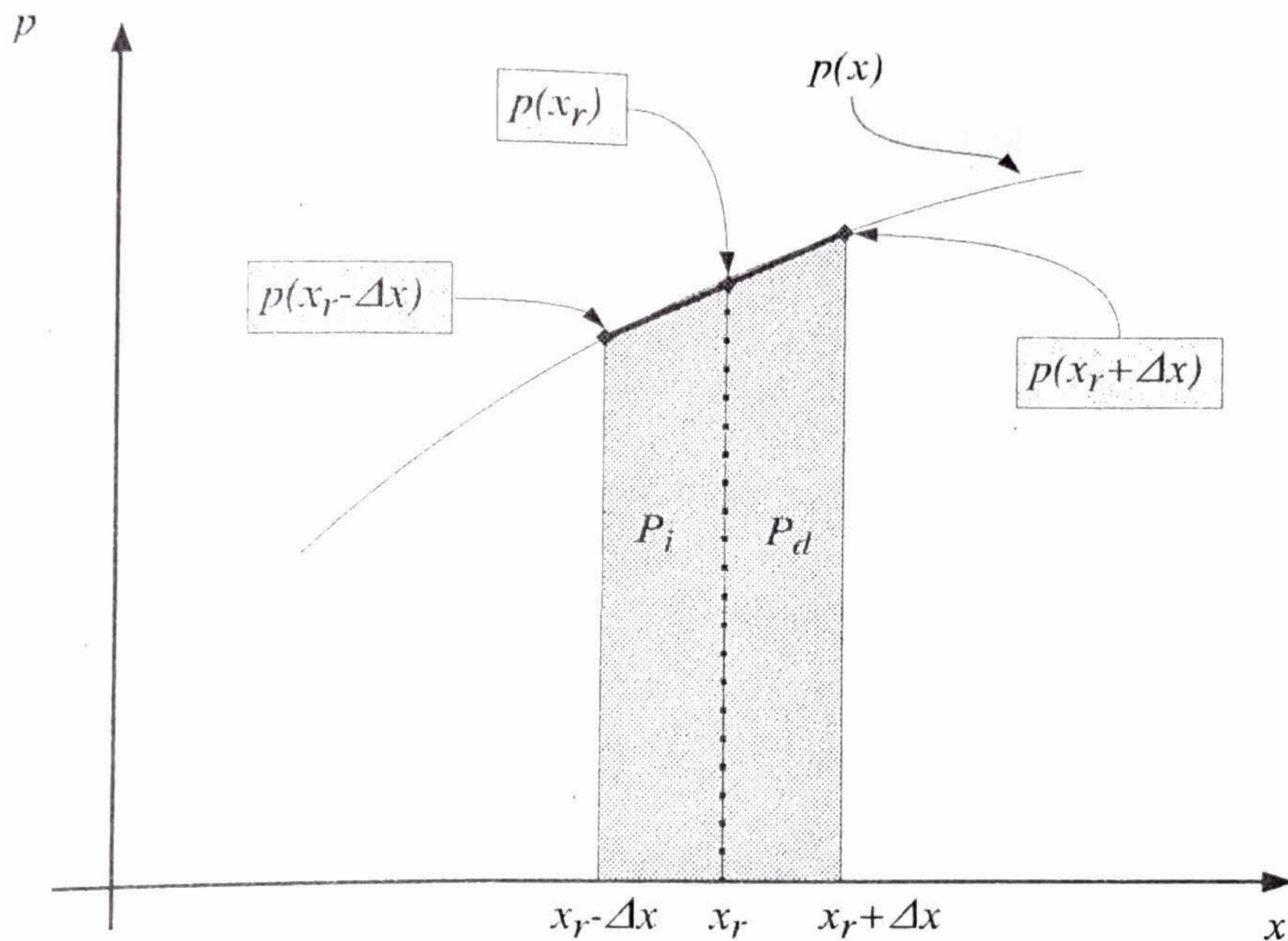


Figura 2.21: Difusión de portadores con un gradiente no nulo de concentración.

neto de portadores, de izquierda a derecha, F , es:

$$F = \frac{P_i}{2} - \frac{P_d}{2} \quad (2.52)$$

Por lo tanto, el flujo de portadores ocurrirá del volumen con mayor concentración de portadores hacia el volumen con menor concentración de portadores. Por ejemplo, en el caso que la concentración a la izquierda, P_i , sea menor que la concentración a la derecha, P_d , como sucede en la Fig. 2.21, el flujo satisfará:

$$F = \frac{P_i - P_d}{2} < 0 \quad (2.53)$$

Esta es la base conceptual para entender el desplazamiento de portadores por difusión. A continuación, se halla la relación matemática entre el flujo y la concentración de portadores.

Densidad de corriente de difusión

Se analiza primero el caso en que los portadores son huecos. En primer lugar, se aproxima la concentración de portadores, en un entorno de la ubicación x_r del plano de referencia, por medio de la expansión de Taylor:

$$p(x_r + \Delta x) = p(x_r) + \left. \frac{dp}{dx} \right|_{x=x_r} \Delta x \quad (2.54)$$

Es claro que, valores de Δx positivos corresponden a puntos a la derecha de x_r , y valores de Δx negativos corresponden a puntos a la izquierda de x_r .

Si se supone que la expansión lineal aproxima con precisión el perfil de portadores (lo cual es cierto, dado que el intervalo en consideración, λ , es pequeño), los portadores en el intervalo $[x_r - \Delta x, x_r]$, pueden aproximarse como¹⁶:

¹⁶El lector puede verificar que (2.55) corresponde al área del trapecio de la izquierda en la Fig. 2.21.

$$P_i = A\lambda \frac{p(x_r - \Delta x) + p(x_r)}{2} \quad (2.55)$$

Si se reemplaza la expansión de Taylor (2.54), considerando $\Delta x = -\lambda$, en (2.55), resulta:

$$P_i = A\lambda \frac{1}{2} \left(2p(x_r) - \left. \frac{dp}{dx} \right|_{x=x_r} \lambda \right) \quad (2.56)$$

De manera similar, los portadores a la derecha de x_r pueden aproximarse como:

$$P_d = A\lambda \frac{p(x_r + \Delta x) + p(x_r)}{2} \quad (2.57)$$

que luego de reemplazar (2.54), considerando $\Delta x = \lambda$, resulta:

$$P_d = A\lambda \frac{1}{2} \left(2p(x_r) + \left. \frac{dp}{dx} \right|_{x=x_r} \lambda \right) \quad (2.58)$$

De acuerdo al razonamiento realizado, la densidad de corriente puede calcularse como:

$$J_p = \frac{q}{A\tau_c} \left(\frac{1}{2}P_i - \frac{1}{2}P_d \right) \quad (2.59)$$

Luego de reemplazar las expresiones de P_i y P_d en (2.59), se obtiene:

$$J_p = \frac{q}{2A\tau_c} A\lambda \frac{1}{2} \left(2p(x_r) - \left. \frac{dp}{dx} \right|_{x=x_r} \lambda - 2p(x_r) - \left. \frac{dp}{dx} \right|_{x=x_r} \lambda \right) \quad (2.60)$$

que puede reducirse a:

$$J_p = -qD_p \left. \frac{dp}{dx} \right|_{x=x_r} \quad (2.61)$$

donde $D_p = \frac{\lambda^2}{2\tau_c}$, es el coeficiente de difusividad de huecos.

La Fig. 2.22 ilustra las corrientes resultantes de huecos para un perfil de portadores genérico. En la zona previa al máximo, la concentración de huecos aumenta, y por ende, su gradiente es positivo. En esta zona, de acuerdo a (2.61), la corriente será negativa, producto que los huecos se mueven de derecha a izquierda. En la zona posterior al máximo, la concentración disminuye, y el gradiente es negativo. De acuerdo a (2.61), la corriente es positiva, producto que los huecos se mueven de izquierda a derecha.

En el caso en que los portadores son electrones, la aproximación de Taylor al perfil de concentración resulta:

$$n(x_r + \Delta x) = n(x_r) + \left. \frac{dn}{dx} \right|_{x=x_r} \Delta x \quad (2.62)$$

Los portadores en el intervalo $[x_r - \Delta x, x_r]$, pueden aproximarse como:

$$P_i = A\lambda \frac{n(x_r - \Delta x) + n(x_r)}{2} \quad (2.63)$$

Si se reemplaza (2.62) en (2.63), considerando $\Delta x = \lambda$, resulta:

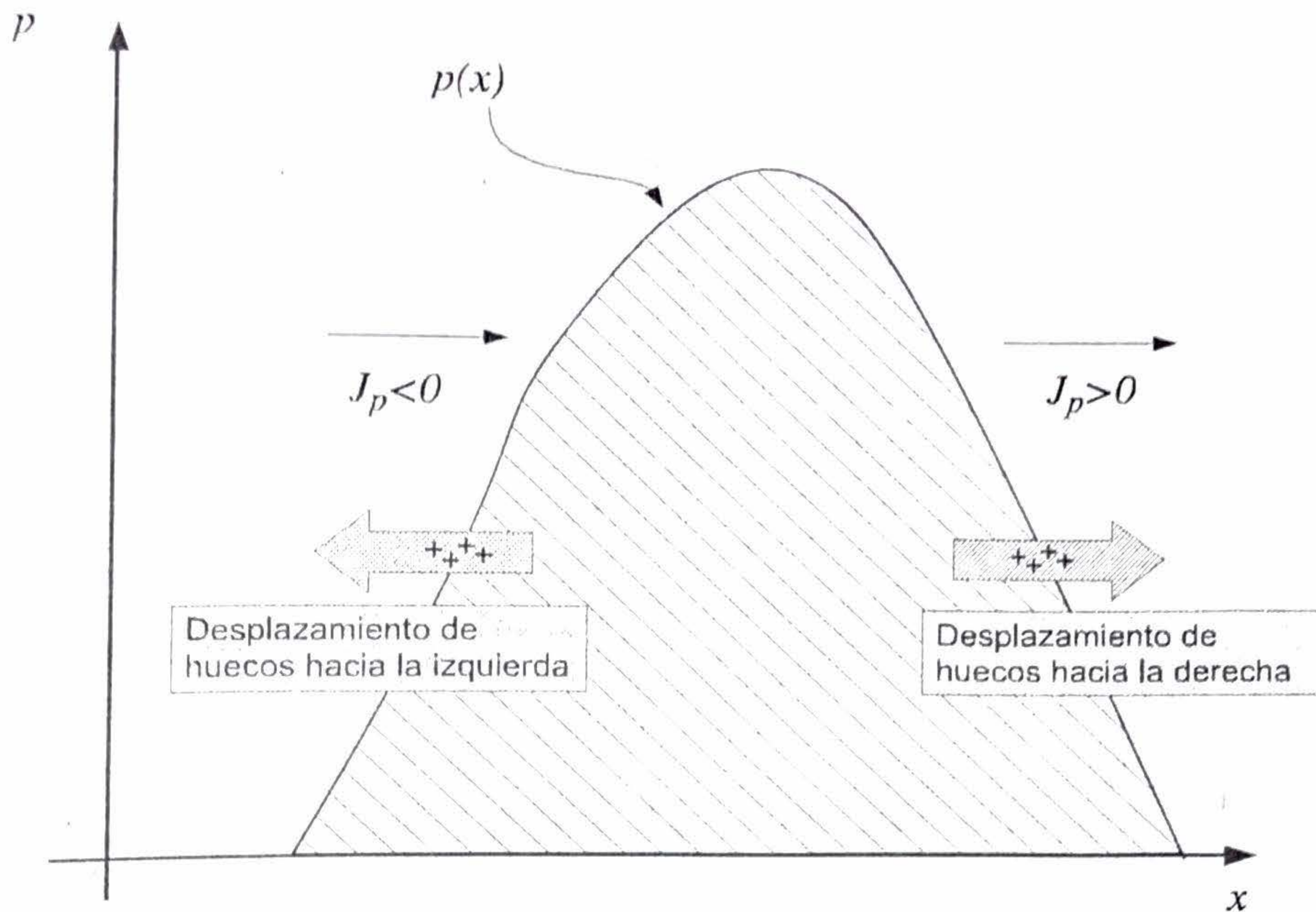


Figura 2.22: Difusión de huecos y corrientes resultantes.

$$P_i = A\lambda \frac{1}{2} \left(n(x_r) - \left. \frac{dn}{dx} \right|_{x=x_r} \lambda \right) \quad (2.64)$$

De manera similar, los portadores a la derecha de x_r pueden aproximarse como:

$$P_d = A\lambda \frac{n(x_r + \Delta x) + n(x_r)}{2} \quad (2.65)$$

que luego de reemplazar (2.62), considerando $\Delta x = \lambda$, resulta:

$$P_d = A\lambda \frac{1}{2} \left(n(x_r) + \left. \frac{dn}{dx} \right|_{x=x_r} \lambda \right) \quad (2.66)$$

De acuerdo al razonamiento realizado, la densidad de corriente puede calcularse como:

$$J_n = -\frac{q}{A\tau_c} \left(\frac{1}{2}P_i - \frac{1}{2}P_d \right) \quad (2.67)$$

Luego de reemplazar las expresiones de P_i y P_d en (2.67), se obtiene:

$$J_n = -\frac{q}{2A\tau_c} A\lambda \frac{1}{2} \left(n(x_r) - \left. \frac{dn}{dx} \right|_{x=x_r} \lambda - n(x_r) - \left. \frac{dn}{dx} \right|_{x=x_r} \lambda \right) \quad (2.68)$$

que puede reducirse a:

$$J_n = qD_n \left. \frac{dn}{dx} \right|_{x=x_r} \quad (2.69)$$

donde $D_n = \lambda^2/2\tau_c$, es el coeficiente de difusividad de electrones.

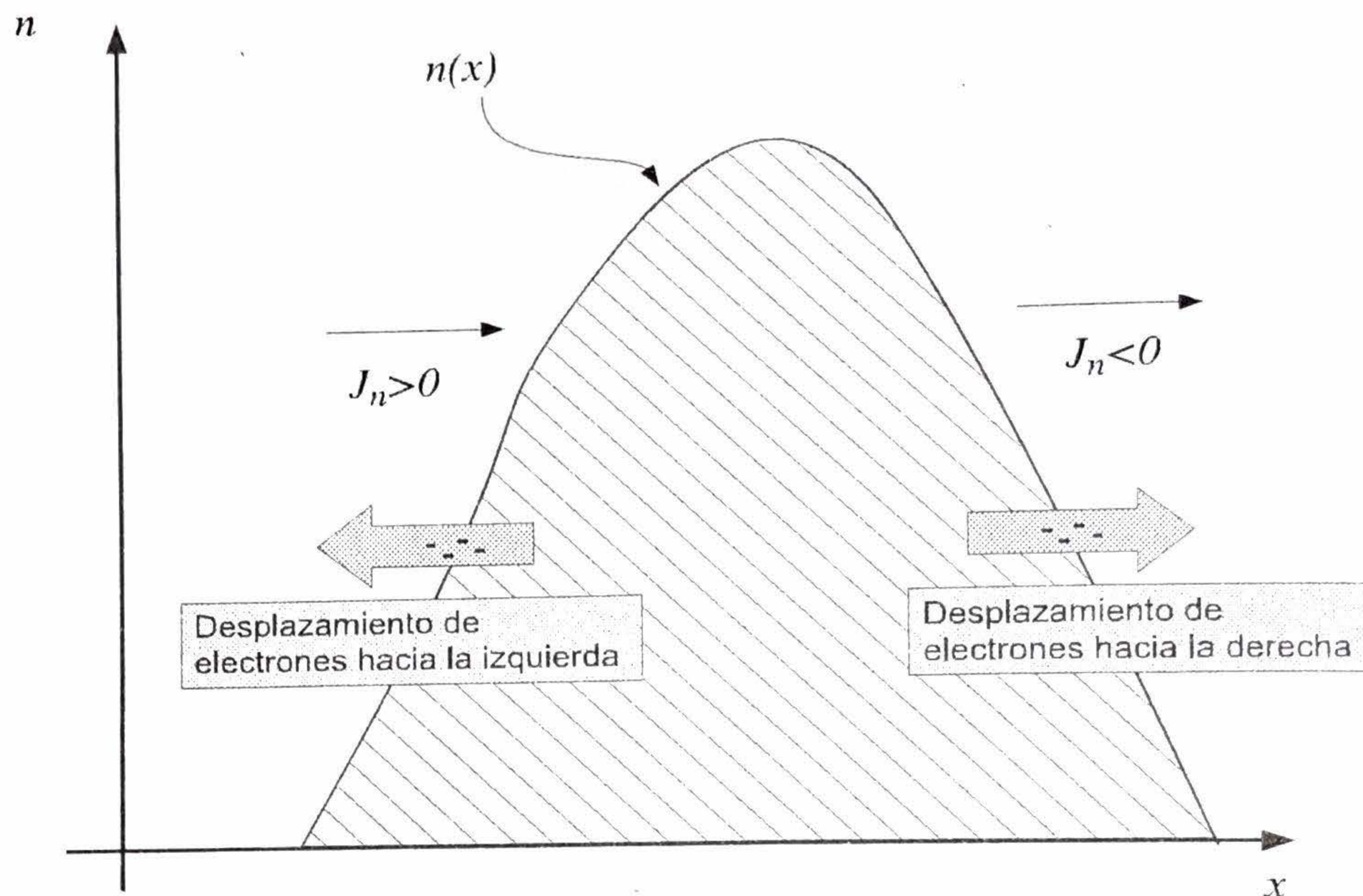


Figura 2.23: Difusión de electrones y corrientes resultantes.

La Fig. 2.23 ilustra las corrientes resultantes de electrones para un perfil de portadores genérico. En la zona previa al máximo, la concentración de electrones aumenta, y por ende, su gradiente es positivo. En esta zona, de acuerdo a (2.69), la corriente es positiva, producto que los electrones se mueven de derecha a izquierda. En la zona posterior al máximo, la concentración disminuye, y el gradiente es negativo. De acuerdo a (2.69), la corriente es negativa, producto que los electrones se mueven de izquierda a derecha.

2.4.3. Resistividad de una lámina de Silicio

Esta sección ejemplifica la conducción por arrastre, mediante el cálculo de la resistividad de una lámina de Silicio, de longitud L , ancho W , y profundidad t . Se supone que la misma tiene ambos tipos de portadores, tanto electrones como huecos. Debido a la ausencia de estímulos externos, la concentración de portadores en la muestra es constante, y no hay corriente de difusión. La corriente total será de arrastre, producto del campo eléctrico aplicado, y tendrá dos componentes:

$$j^a = j_p^a + j_n^a = qp\mu_p E + qn\mu_n E \quad (2.70)$$

Dado que la densidad de carga neta en la lámina es nula, el campo eléctrico será constante. La tensión aplicada a la muestra V , satisfará $V = E \times L$. Por lo tanto, (2.70) puede reescribirse como:

$$j^a = q(p\mu_p + n\mu_n)V/L \quad (2.71)$$

Teniendo en cuenta que $j^a = I/(Wt)$, (2.71) puede escribirse como:

$$\frac{V}{I} = \frac{L}{Wtq(p\mu_p + n\mu_n)} \quad (2.72)$$

En esta última ecuación, puede reconocerse la resistividad de la lámina:

$$\rho \triangleq \frac{1}{q(p\mu_p + n\mu_n)} \quad (2.73)$$

2.4. Mecanismos de conducción

En el diseño de circuitos integrados, es común dar el valor de resistencia de una lámina, cuya longitud es igual a su ancho¹⁷, es decir, $L = W$. En este caso, se define la resistencia superficial, o resistencia por cuadrado, de la siguiente manera:

$$R_{\square} \triangleq \frac{1}{tq(p\mu_p + n\mu_n)} \quad (2.74)$$

Ejemplo 11 Una muestra de Silicio, de dimensiones $L = 10\mu m$, $W = 2\mu m$, $t = 1\mu m$, se dopa con una concentración de Arsénico $N_d = 10^{15} cm^{-3}$. Determine la resistencia, resistividad, y resistencia por cuadrado de la lámina. Si se aplica una tensión $V = 1V$ sobre la misma: cuál es la corriente y el tiempo de tránsito de los portadores ?

Para este dopado, utilizando la rutina del Ejemplo 10, se obtiene $\mu_n = 1345 cm^2/Vs$. La cantidad de portadores negativos es $n = N_d = 10^{15} cm^{-3}$, y los portadores positivos son despreciables. La resistividad es entonces:

$$\begin{aligned} \rho &= \frac{1}{1345 \frac{cm^2}{Vs} \times 10^{15} \frac{1}{cm^3} \times 1,6 \times 10^{-19} C} \\ &= \frac{1}{1345 \times 10^{15} \times 1,6 \times 10^{-19}} \Omega cm \\ &= \frac{1}{0,2152} \Omega cm \\ &= 4,64 \Omega cm \end{aligned} \quad (2.75)$$

La resistencia por cuadrado es:

$$R_{\square} = \frac{\rho}{t} = \frac{4,64 \Omega \times 10,000 \mu m}{1 \mu m} = 46,4 K\Omega \quad (2.76)$$

y la resistencia de la lámina es:

$$R = 4,64 \Omega \frac{10,000 \mu m \times 10 \mu m}{2 \mu m \times 1 \mu m} = 232 K\Omega \quad (2.77)$$

Si se aplica una tensión $V = 1V$, la corriente es $I = 1V/232K\Omega = 4,31\mu A$. Para calcular el tiempo de tránsito es necesario calcular la carga total en la lámina, que es igual a:

$$\begin{aligned} Q &= W \times L \times t \times q \times N_d = 10 \mu m \times 2 \mu m \times 1 \mu m \times 10^{15} cm^{-3} \times q \\ &= 2 \times 10^5 \times 1,6 \times 10^{-19} C = 3,2 \times 10^{-14} C \end{aligned} \quad (2.78)$$

El tiempo de tránsito es:

$$\tau = \frac{3,2 \times 10^{-14} C}{4,31 \mu A} = 7,42 ns \quad (2.79)$$

¹⁷Se trata de una resistencia de forma cuadrada, vista desde arriba.

2.5. Potenciales relativos en Silicio

En esta sección se introduce un potencial de referencia que resulta conveniente para futuros análisis. Para ello, se analiza la corriente total de portadores en un trozo de Silicio, *sin potenciales externos aplicados*. Si se comienza analizando el caso de los electrones, la corriente total debe ser nula, es decir:

$$\begin{aligned} J_n &= qn\mu_n E + qD_n \frac{dn}{dx} = 0 \\ &= qn\mu_n \frac{-d\phi}{dx} + qD_n \frac{dn}{dx} = 0 \end{aligned} \quad (2.80)$$

De (2.80), surge una relación entre potencial y densidad de concentración de portadores:

$$d\phi = \frac{D_n}{\mu_n} \left(\frac{dn}{n} \right) = U_T \left(\frac{dn}{n} \right) \quad (2.81)$$

donde U_T es la tensión térmica, que se define como:

$$\boxed{U_T \triangleq kT/q} \quad (2.82)$$

y posee un valor de $25mV$ a temperatura ambiente; y

$$\boxed{D_n = \mu_n U_T} \quad (2.83)$$

es llamada, junto a su ecuación análoga para el caso P, relación de Einstein [8].

Si se consideran dos lugares del material, x_a y x_b , con distintas concentraciones, $n_a = n(x_a)$ y $n_b = n(x_b)$, y se integra (2.81) para hallar la diferencia de potencial, se llega a:

$$\int_{x_a}^{x_b} d\phi = U_T \int_{n_a}^{n_b} \frac{dn}{n} \quad (2.84)$$

La solución a (2.84) es:

$$\phi(x_b) - \phi(x_a) = U_T \times \ln \frac{n_b}{n_a} \quad (2.85)$$

El potencial es una magnitud relativa, por lo tanto, lo que resulta relevante es la diferencia de potencial entre dos puntos, siendo arbitrario el valor de referencia que se tome para ambos. Un valor conveniente como referencia es el de Silicio intrínseco. Para definir este valor de referencia, se asume que:

$$\phi(x) = 0 \text{ si } n(x) = n_i \quad (2.86)$$

Si se reemplaza (2.86) en (2.85), resulta:

$$\phi(x_b) = U_T \times \ln \frac{n_b}{10^{10}} \quad (2.87)$$

Considerando que:

$$U_T \times \ln \left(\frac{n_b}{10^{10}} \right) = \frac{U_T}{\log(e)} \times \log \left(\frac{n_b}{10^{10}} \right) \quad (2.88)$$

la Ec. (2.87) resulta en la "Regla de los $60mV$ " [10]:

$$\boxed{\phi(x_b) = 60mV \times \log \left(\frac{n_b}{10^{10}} \right)} \quad (2.89)$$

2.5. Potenciales relativos en Silicio

En esta sección se introduce un potencial de referencia que resulta conveniente para futuros análisis. Para ello, se analiza la corriente total de portadores en un trozo de Silicio, *sin potenciales externos aplicados*. Si se comienza analizando el caso de los electrones, la corriente total debe ser nula, es decir:

$$\begin{aligned} J_n &= qn\mu_n E + qD_n \frac{dn}{dx} = 0 \\ &= qn\mu_n \frac{-d\phi}{dx} + qD_n \frac{dn}{dx} = 0 \end{aligned} \quad (2.80)$$

De (2.80), surge una relación entre potencial y densidad de concentración de portadores:

$$d\phi = \frac{D_n}{\mu_n} \left(\frac{dn}{n} \right) = U_T \left(\frac{dn}{n} \right) \quad (2.81)$$

donde U_T es la tensión térmica, que se define como:

$$\boxed{U_T \triangleq kT/q} \quad (2.82)$$

y posee un valor de $25mV$ a temperatura ambiente; y

$$\boxed{D_n = \mu_n U_T} \quad (2.83)$$

es llamada, junto a su ecuación análoga para el caso P, relación de Einstein [8].

Si se consideran dos lugares del material, x_a y x_b , con distintas concentraciones, $n_a = n(x_a)$ y $n_b = n(x_b)$, y se integra (2.81) para hallar la diferencia de potencial, se llega a:

$$\int_{x_a}^{x_b} d\phi = U_T \int_{n_a}^{n_b} \frac{dn}{n} \quad (2.84)$$

La solución a (2.84) es:

$$\phi(x_b) - \phi(x_a) = U_T \times \ln \frac{n_b}{n_a} \quad (2.85)$$

El potencial es una magnitud relativa, por lo tanto, lo que resulta relevante es la diferencia de potencial entre dos puntos, siendo arbitrario el valor de referencia que se tome para ambos. Un valor conveniente como referencia es el de Silicio intrínseco. Para definir este valor de referencia, se asume que:

$$\phi(x) = 0 \text{ si } n(x) = n_i \quad (2.86)$$

Si se reemplaza (2.86) en (2.85), resulta:

$$\phi(x_b) = U_T \times \ln \frac{n_b}{10^{10}} \quad (2.87)$$

Considerando que:

$$U_T \times \ln \left(\frac{n_b}{10^{10}} \right) = \frac{U_T}{\log(e)} \times \log \left(\frac{n_b}{10^{10}} \right) \quad (2.88)$$

la Ec. (2.87) resulta en la "Regla de los $60mV$ " [10]:

$$\boxed{\phi(x_b) = 60mV \times \log \left(\frac{n_b}{10^{10}} \right)} \quad (2.89)$$

2.5. Potenciales relativos en Silicio

Esta regla permite calcular fácilmente el potencial de una especie de Silicio dopado, con respecto a Silicio intrínseco: por cada orden de magnitud de dopado con respecto a Silicio intrínseco, se producen $60mV$ adicionales de tensión. Por ejemplo, si x_b está dopado con $N_d = 10^{15}$, entonces, $n = 10^{15}$, $\log(10^{15}/10^{10}) = 5$, y la diferencia de potencial resultante es de $5 \times 60mV = 300mV$. Por el contrario, si x_b está dopado con $N_a = 10^{15}$, entonces, $n = 10^5$, $\log(10^5/10^{10}) = -5$, y la diferencia de potencial resultante es de $-5 \times 60mV = -300mV$.

Para el caso de la corriente de huecos, se puede realizar el mismo desarrollo, partiendo de:

$$\begin{aligned} J_p &= qp\mu_p E - qD_p \frac{dp}{dx} = 0 \\ &= qp\mu_p \frac{-d\phi}{dx} - qD_p \frac{dp}{dx} = 0 \end{aligned} \quad (2.90)$$

de donde surge la relación:

$$d\phi = \frac{D_p}{\mu_p} \left(\frac{dn}{n} \right) = U_T \left(\frac{dn}{n} \right) \quad (2.91)$$

con

$$\boxed{D_p = \mu_p U_T} \quad (2.92)$$

por la relación de Einstein. También se puede operar directamente (2.89); de hecho, dado que $n_b \times p_b = n_i^2$, por la Ley de Acción de Masas, (2.89) puede escribirse como:

$$\phi(x_b) = -60mV \times \log\left(\frac{p_b}{10^{10}}\right) \quad (2.93)$$

Repetiendo el ejemplo anterior, si x_b está dopado con $N_d = 10^{15}$, entonces, $p = 10^5$, $\log(10^5/10^{10}) = -5$, y la diferencia de potencial resultante es de $5 \times 60mV = 300mV$. Por el contrario, si x_b está dopado con $N_a = 10^{15}$, entonces, $p = 10^{15}$, $\log(10^{15}/10^{10}) = 5$, y la diferencia de potencial resultante es de $-5 \times 60mV = -300mV$.

Cual es el significado de la aparición de una diferencia de potencial entre Silicio con diferentes niveles de dopado? Es importante notar que esta diferencia de potencial surge como consecuencia del requisito impuesto de que la corriente neta sea nula, tal como indican (2.80) y (2.90), que se repiten por comodidad:

$$\begin{aligned} J_n &= qn\mu_n E + qD_n \frac{dn}{dx} = 0 \\ J_p &= qp\mu_p E - qD_p \frac{dp}{dx} = 0 \end{aligned}$$

Es decir, la diferencia de potencial es necesaria para que la corriente de difusión de electrones sea cancelada por una corriente de arrastre de electrones, y para que la corriente de difusión de huecos sea cancelada por una corriente de arrastre de huecos. Las corrientes de difusión se originan naturalmente por la diferencia de concentraciones. Nótese entonces de (2.80), que en el caso en que haya una región con dopado tipo N, como se ilustra en la Fig. 2.24, la misma tendrá potencial positivo con respecto a Silicio intrínseco. Este potencial positivo producirá un flujo de electrones hacia el silicio dopado, que cancelará la corriente de difusión de electrones del silicio dopado al intrínseco (producto de la mayor concentración de electrones en esa zona).

Por el contrario, la región con dopado tipo P tendrá potencial negativo con respecto a Silicio intrínseco. Este potencial negativo produce un flujo de huecos hacia el silicio dopado, que cancela la corriente de difusión de huecos del silicio dopado al intrínseco (producto de la mayor concentración de huecos en esa zona). En la Fig. 2.24, las flechas con portadores indican el flujo de portadores (no la dirección de la corriente) resultante.

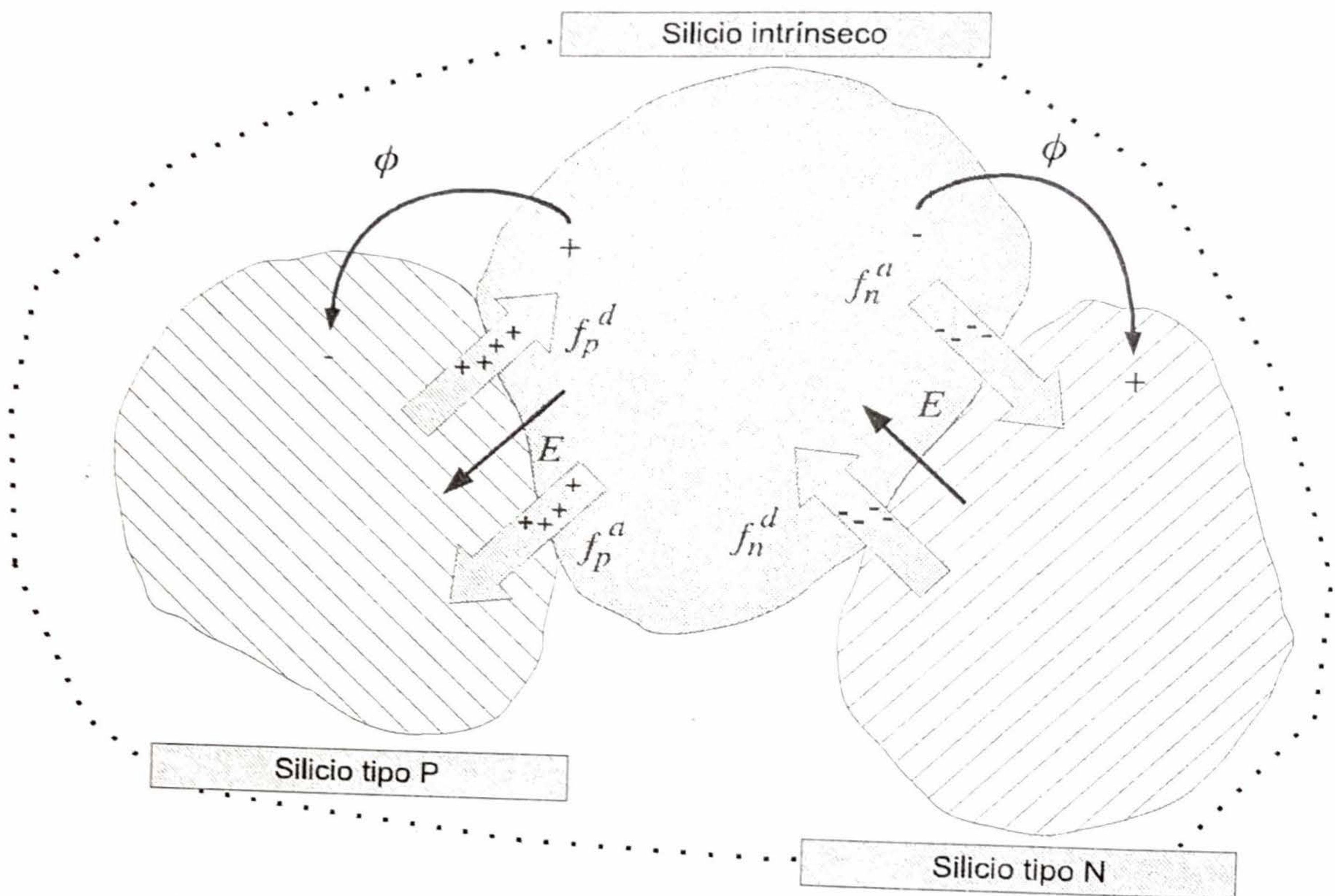


Figura 2.24: Dos regiones con diferentes dopados y los flujos de portadores resultantes, con respecto a Silicio intrínseco: f_n^d y f_n^a son los flujos de difusión y arrastre, respectivamente, de electrones; f_p^d y f_p^a son los flujos de difusión y arrastre, respectivamente, de huecos.

Apéndice A

Conducción: Conceptos Auxiliares

A.1. Tiempo de tránsito

La noción de tiempo de tránsito en un semiconductor es frecuentemente utilizada, razón por la cual se introduce en esta sección.

Se supone una pieza semiconductor de ciertas dimensiones, con una carga total de electrones¹ igual a Q , y una corriente producida por electrones que se desplazan de un extremo al otro del material, empleando para ello un tiempo τ . Tal como se describiera en la Sección 2.4.1, los electrones ingresan a la pieza semiconductor a una cierta tasa (cantidad de electrones por unidad de área y por segundo), y egresan a la misma tasa. Esto implica que, luego de transcurrido un tiempo τ , todos los electrones dentro de la pieza semiconductor, cuya carga es Q , habrán salido de ella. En consecuencia, la corriente I que está circulando por la pieza semiconductor es:

$$I = \frac{|Q|}{\tau} \quad (\text{A.1})$$

El valor

$$\tau \triangleq \frac{|Q|}{I} \quad (\text{A.2})$$

se define como tiempo de tránsito, y puede interpretarse como el tiempo promedio que le toma a un electrón recorrer la pieza semiconductor.

A.2. Efecto Hall

Una de las maneras de medir concentración de portadores en forma directa, es a través del Efecto Hall. La Fig. A.1 ilustra el esquema de medición, donde se dispone una muestra de Silicio, a la cual se le hace circular una densidad de corriente J_x , en el sentido del eje x . A su vez, se produce un campo magnético de intensidad B , en el sentido del eje z .

El campo magnético provoca una fuerza de Lorentz, en el sentido del eje y , dada por:

$$F = q(\mathbf{v} \times \mathbf{B}) \quad (\text{A.3})$$

Esta fuerza hace que los portadores positivos y negativos se acumulen en la cara inferior de la lámina, dejando una carga de signo opuesto en la cara superior de la lámina. En el eje y no hay circulación de corriente por tratarse de un circuito abierto, con lo cual, la acumulación de cargas produce un campo eléctrico E_y , y un potencial asociado V_y . Este potencial puede ser medido para

¹El caso de conducción por huecos sigue los mismos lineamientos.

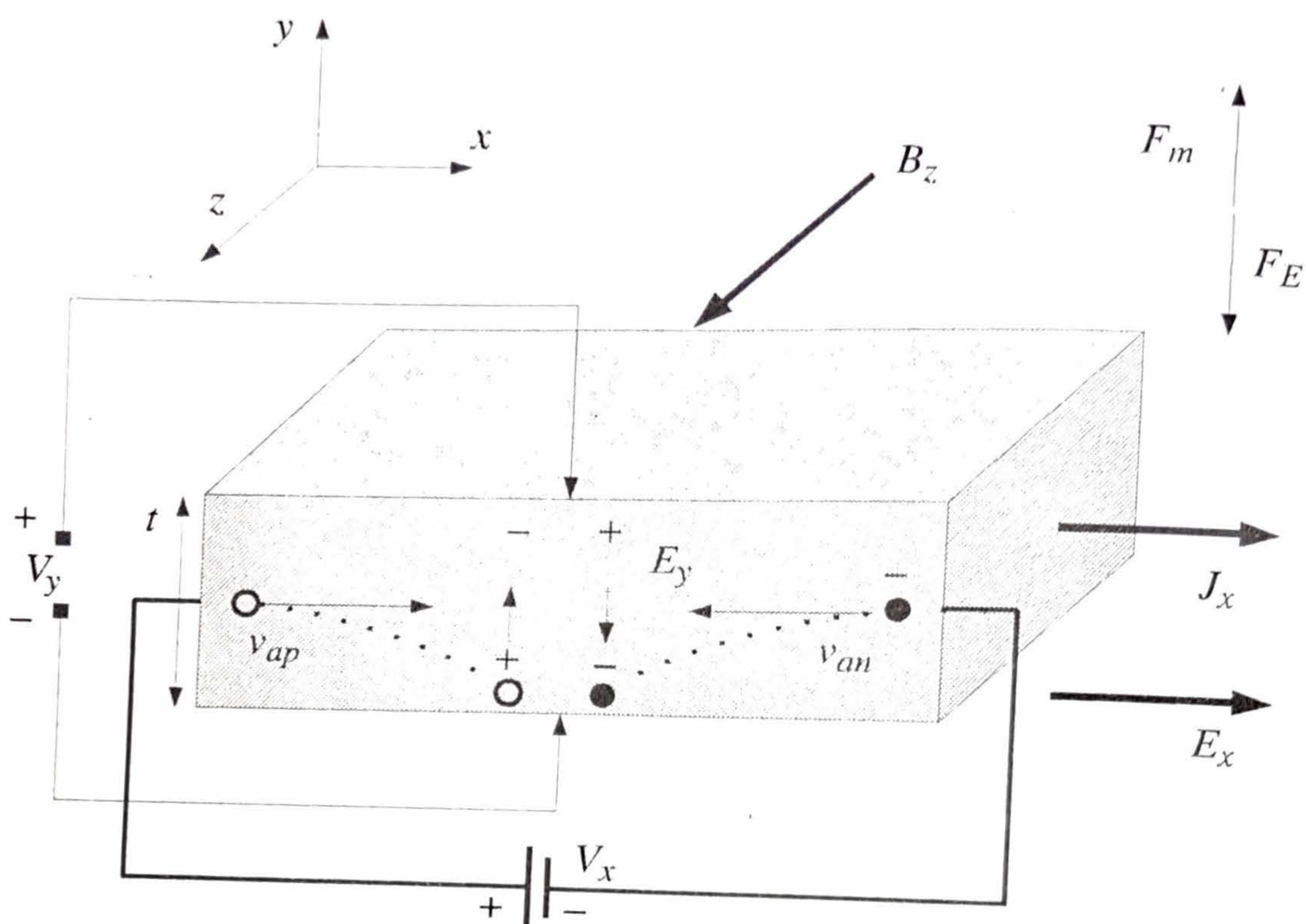


Figura A.1: Esquema de medición de concentraciones utilizando Efecto Hall.

determinar la concentración de portadores. Para ver esto, se supone, sin pérdida de generalidad, que la lámina está dopada con donadores N_d , de manera que la conducción es por electrones². Los electrones se trasladan de derecha a izquierda con una velocidad v_n^a , producto de la aplicación del potencial V_x . El valor de v_n^a surge de la expresión de la corriente:

$$J_x = -qnv_n^a \tag{A.4}$$

De esta ecuación, se obtiene $v_n^a = -J_x/(qn)$. Reemplazando esta expresión en (A.3) se obtiene el valor de la fuerza ejercida sobre los electrones por el campo magnético:

$$F_m = -\frac{J_x B_z}{n} \tag{A.5}$$

Esta fuerza, que impulsa a los electrones hacia el fondo de la lámina, debe ser igual a la fuerza del campo eléctrico E_y , resultante de la acumulación de huecos en la parte superior y electrones en la parte inferior, que está dada por:

$$F_E = -qE_y = -qV_y/t \tag{A.6}$$

Igualando las fuerzas (A.5) y (A.6) se puede despejar la expresión de la concentración de electrones:

$$n = \frac{J_x B_z t}{qV_y} \tag{A.7}$$

Nótese de (A.7) que, si se conoce el campo magnético, es posible calcular n , a partir de la medición de la tensión V_y y de la corriente J_x . Además, el signo de V_y permite establecer el tipo de dopado existente en el material.

²El análisis para el caso de impurezas aceptoras se basa en la conducción de los huecos y sigue los mismos lineamientos.

Apéndice B

Electroestática

El desarrollo de las ecuaciones de electroestática para los distintos dispositivos a analizar, se basa en la ley de Gauss:

$$\nabla \cdot (\varepsilon E) = \rho \quad (\text{B.1})$$

y en la definición de potencial

$$\nabla \phi = -E \quad (\text{B.2})$$

La combinación de ambas da lugar a la denominada "Ecuación de Poisson":

$$\nabla \cdot (-\varepsilon \nabla \phi) = -\varepsilon \nabla^2 \phi = \rho \quad (\text{B.3})$$

Estas ecuaciones deben utilizarse para el análisis de la estructura tridimensional del Silicio, aunque para un análisis de primer orden, resulta suficiente un análisis uni-dimensional, en cuyo caso, las ecuaciones anteriores se reducen a:

$$\begin{cases} \frac{d(\varepsilon E)}{dx} = \rho \\ \frac{d\phi}{dx} = -E \end{cases} \quad (\text{B.4})$$

y

$$\varepsilon \frac{d^2 \phi}{dx^2} = -\rho \quad (\text{B.5})$$

Además de las ecuaciones mencionadas, es necesario tener en cuenta las condiciones de borde. En primer lugar, en todos los casos, la densidad de carga estará localizada en un lugar acotado del espacio, con lo cual, en virtud de la relación entre carga y campo eléctrico (B.4), el campo eléctrico se desarrollará en el interior de esa región, y será nulo fuera. Es importante recordar que las líneas de campo eléctrico, comienzan en cargas positivas, y terminan en cargas negativas.

En la interfaz entre dos materiales con distintas constantes dieléctricas ε_1 y ε_2 (ver Fig. B.1-a), en virtud de (B.4), se verifica:

$$\int_{-\xi}^{\xi} d(\varepsilon E) = \int_{-\xi}^{\xi} \rho dx \quad (\text{B.6})$$

de donde resulta:

$$\varepsilon_2 E_2 - \varepsilon_1 E_1 = Q \quad (\text{B.7})$$

donde Q es la carga pelicular en la interfaz (que puede existir, o no). En caso que no haya carga en la interfaz, (B.7) se reduce a:

$$\varepsilon_2 E_2 = \varepsilon_1 E_1 \quad (B.8)$$

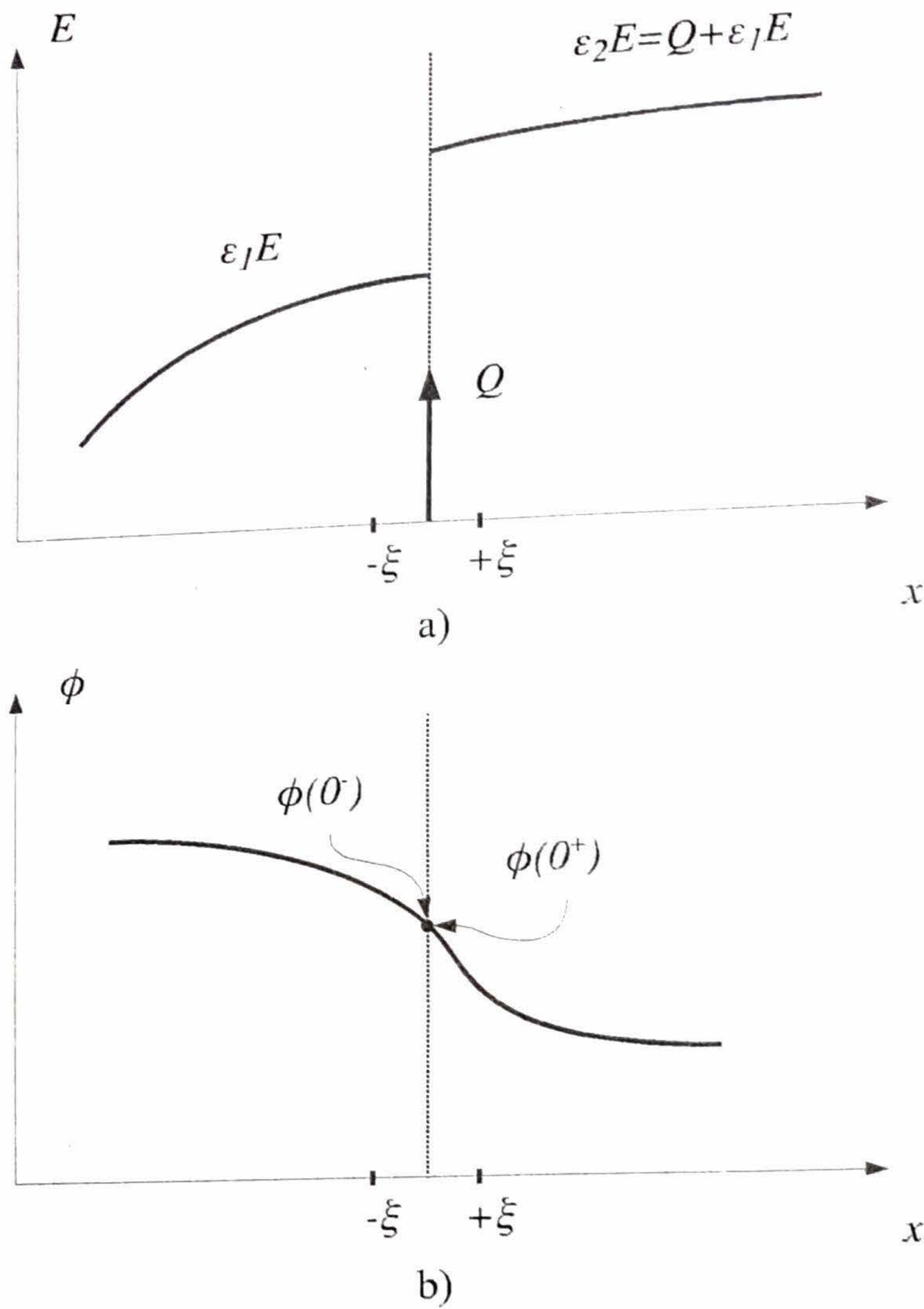


Figura B.1: Condiciones de borde para: a) campo eléctrico; b) potencial

En relación al potencial, se deduce de (B.4) que una variación abrupta del potencial provoca un campo eléctrico elevado, y por ende, una fuerza electrostática elevada sobre el material. Un potencial discontinuo implicaría un campo eléctrico infinito, y por ende, una fuerza infinita actuando sobre la interfaz del material. Esto descarta la discontinuidad del potencial, por lo cual en las interfaces entre dos materiales siempre se cumplirá (ver Fig. B.1-b):

$$\phi(0^-) = \phi(0^+) \quad (B.9)$$

Apéndice C

Potenciales de contacto

La aplicación de un potencial externo sobre un material como el Silicio, hace necesario la introducción de contactos metálicos. Toda vez que dos materiales diferentes se contactan, aparece entre ellos una diferencia de potencial, producto de la tendencia al equilibrio de las diferentes concentraciones de portadores. En el caso de los contactos con metal, se asumirá que los mismos son *óhmicos*, lo que significa que la diferencia de tensión resultante solo depende de los materiales, y no de la corriente que circula a través de ellos. La Tabla C.1 muestra los potenciales relativos de contacto de diversos metales con respecto a Silicio intrínseco.

En el caso que haya m materiales diferentes conectados en serie, como se muestra en la Fig. C.1, es sencillo ver que el potencial neto es la diferencia de potencial entre los dos materiales extremos. De hecho, de acuerdo a la Fig. C.1,

$$\begin{aligned}\phi_T &= \phi_{1,2} + \phi_{2,3} + \cdots + \phi_{m-1,m} \\ &= \phi_1 - \phi_2 + \phi_2 - \phi_3 + \cdots + \phi_{m-1} - \phi_m \\ &= \phi_1 - \phi_m \\ &= \phi_{1,m}\end{aligned}\tag{C.1}$$

En el caso que se utilice un voltímetro para medir el potencial ϕ_T , el primer y último material son iguales, dado que son los terminales del voltímetro, y por lo tanto, resulta:

$$\phi_T = 0\tag{C.2}$$

Si entre un par de materiales cualquiera de la secuencia serie, se intercala una fuente de tensión V_S (ver Fig. C.2), de acuerdo a (C.1) y (C.2), es sencillo ver que:

Cuadro C.1: Potenciales de contacto de varios materiales con respecto a Silicio intrínseco.

Material	Potencial de contacto ϕ [V]
Ag	-0.40
Cu	+0.00
Ni	+0.15
Al	+0.60
Au	-0.30
Mg	+1.35

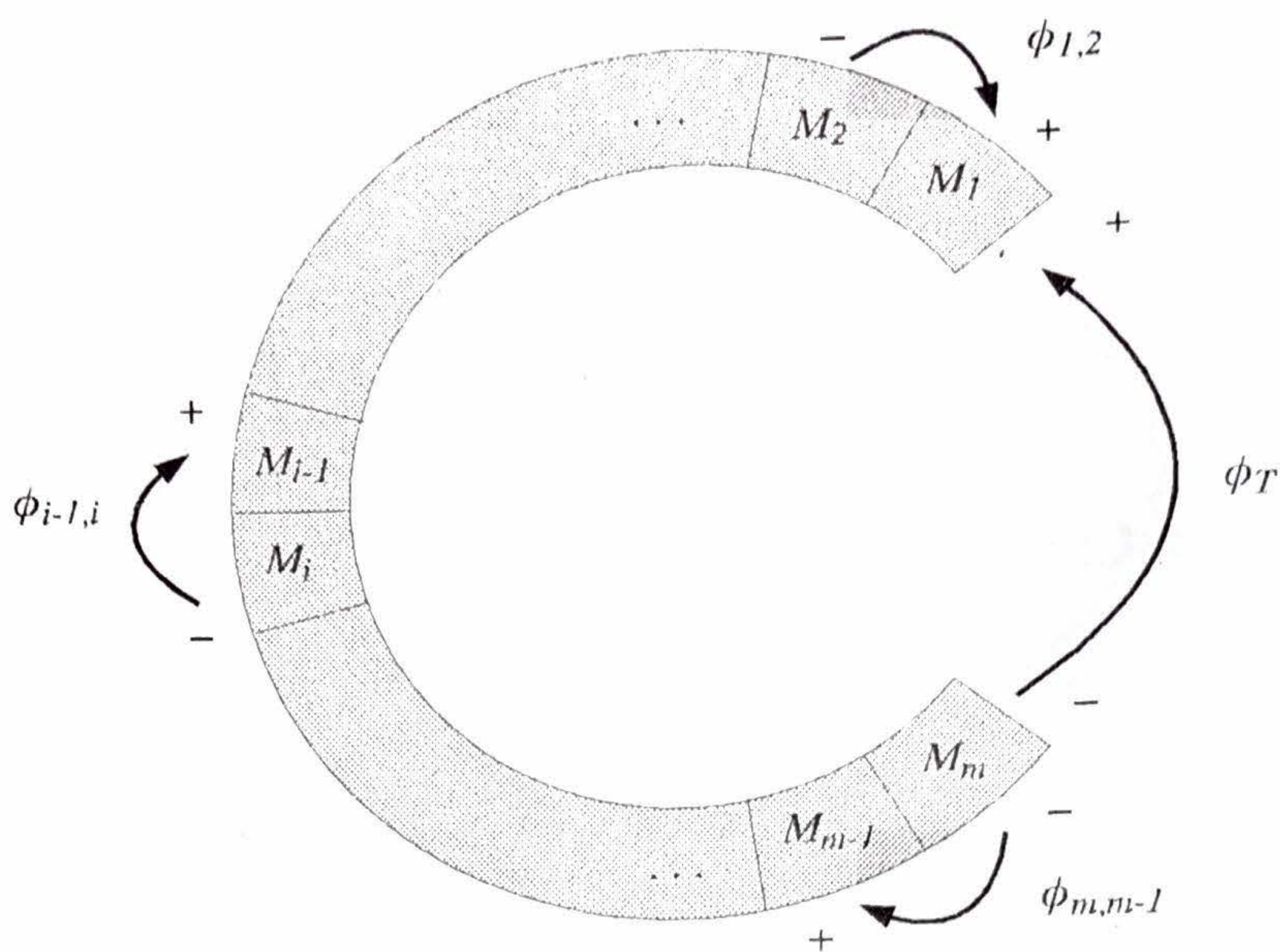


Figura C.1: Conexión serie de m materiales y el potencial de contacto resultante

$$\phi_T = V_S \tag{C.3}$$

De lo expuesto, se puede ver que un voltímetro no permite medir los potenciales de contacto de los materiales: ello explica porqué, en general, no son tenidos en cuenta al analizar circuitos. Para el análisis de ciertos dispositivos, como se verá más adelante, es fundamental considerar su existencia.

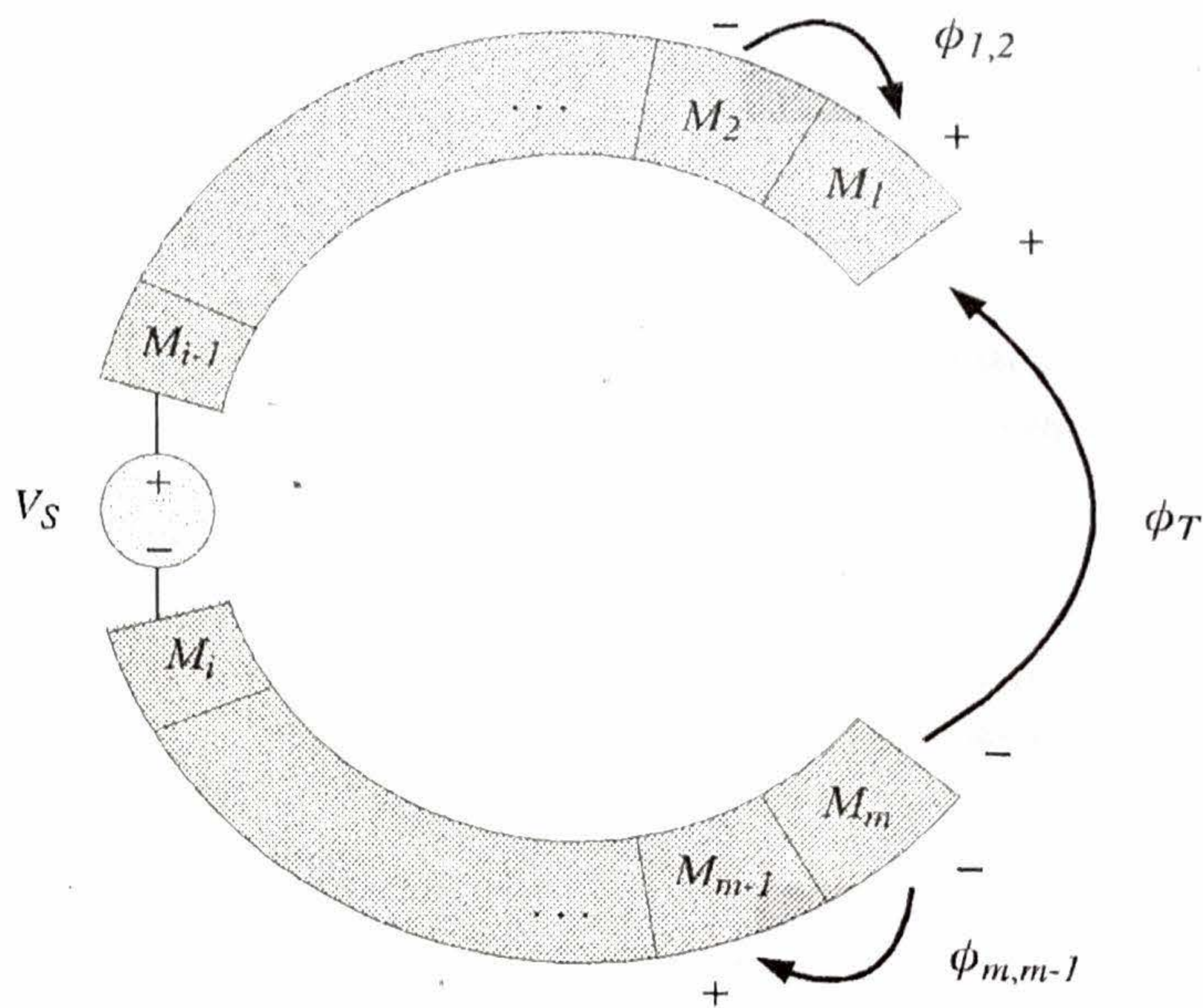


Figura C.2: Conexión serie de materiales, agregando una fuente de potencial, y el potencial de contacto resultante

Capítulo 3

Juntura Semiconductora y Diodos

3.1. Descripción Cualitativa

La juntura P-N es un elemento fundamental de los dispositivos semiconductores y constituye el bloque constructivo de los diodos de juntura, los transistores bipolares y los rectificadores controlados de Silicio (SCR), compuestos por diacs, tiristores y triacs. Por otro lado, aparecen también como elementos parásitos en todo dispositivo que se realice sobre Silicio, como capacitores MOS, transistores MOS, y aún resistores y capacitores.

Desde un punto de vista cualitativo, cuando dos piezas de Silicio, una con dopado P y la otra con dopado N, se ponen en contacto, la diferencia de concentraciones produce un flujo de electrones desde el material dopado N hacia el P, y un flujo de huecos desde el material dopado P hacia el N, tal como se ilustra en la Fig. 3.1. Los electrones que migran del material N, dejan tras de sí átomos con carga neta positiva, es decir, *iones*, los cuales están fijos a la estructura cristalina y no pueden moverse. En forma similar, los huecos que migran del material P, dejan tras de sí *iones* negativos (fijos también a la estructura cristalina). Esta migración de portadores ocurre durante un lapso de tiempo transitorio y a medida que sucede, las cargas netas de los iones van desarrollando un campo eléctrico que se opone a la corriente de difusión. Debido a este proceso se forma una zona sin portadores, que se denomina “zona de vaciamiento”, a ambos lados de la juntura –lugar exacto de unión entre el material dopado P y el dopado N. A medida que esta zona crece, es mayor la carga iónica y, por ende, el campo eléctrico resultante. Cuando la zona de vaciamiento tiene la carga iónica suficiente para que el campo eléctrico cancele la corriente de difusión inicial, el proceso se detiene y se alcanza un equilibrio. La zona de vaciamiento presenta entonces un potencial neto, que es positivo en el material N con respecto al material P, y está dado por:

$$\phi_B \triangleq \phi_{n,p} = \phi_n - \phi_p \quad (3.1)$$

Si se aplica una tensión externa, positiva en el lado N (cátodo) con respecto al lado P (ánodo), el efecto neto será el aumento del campo eléctrico de la juntura. Este campo eléctrico externo favorece el movimiento de huecos de la zona N hacia la P, y de electrones de la zona P hacia la N; es decir, sólo favorece el movimiento de los portadores minoritarios. Dado que los portadores minoritarios son despreciables, la corriente neta es una corriente de fuga que puede considerarse nula. Los portadores mayoritarios no pueden difundir de un lado al otro debido al campo eléctrico en la zona de vaciamiento. En esta situación, se dice que la juntura está en inversa y no permite el paso de corriente.

Si se aplica una tensión externa, positiva en el ánodo con respecto al cátodo, el campo eléctrico de la juntura se reduce y se produce un desbalance de las corrientes de difusión y arrastre. La disminución del campo eléctrico produce una disminución de la corriente de arrastre, y se observa una corriente neta circulando por el dispositivo, dada por la –ahora mayor– corriente de difusión. En este caso, hay

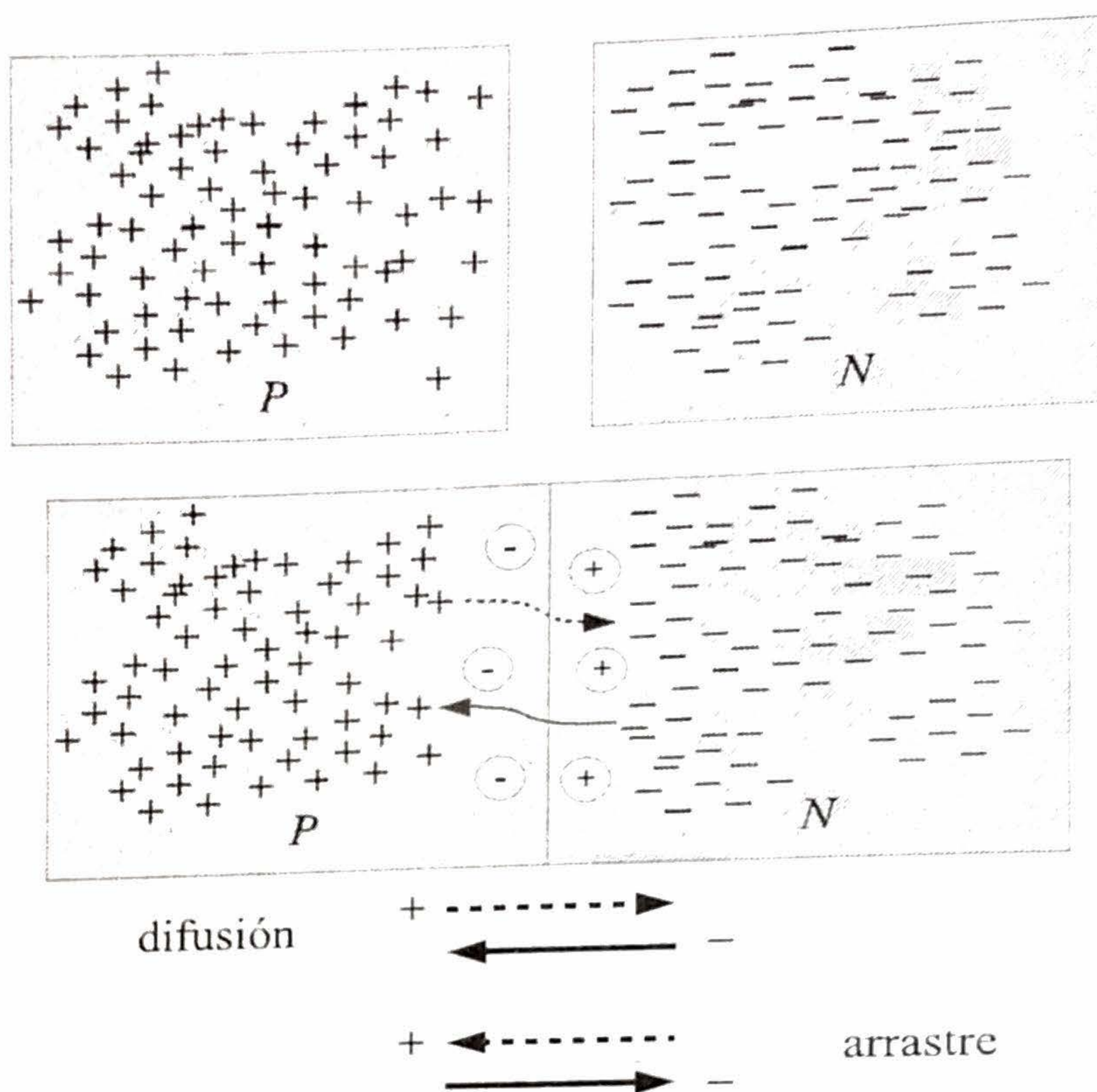


Figura 3.1: Unión entre Silicio tipo P y Silicio tipo N. La difusión de huecos y electrones produce una corriente del material P al N, y el arrastre producto del campo eléctrico produce una corriente en sentido inverso.

una corriente de difusión neta de huecos del lado P hacia el N, y de electrones del lado N hacia el P. En esta situación, se dice que la juntura está en directa y permite la conducción de corriente.

En definitiva, el diodo modelado en DC permite la conducción de corriente en un sentido, y la bloquea en el sentido contrario.

En las siguientes secciones se desarrolla el modelo del diodo de acuerdo a los lineamientos del Capítulo 1. En la Sección 3.2 se desarrolla la electrostática de la juntura; en la Sección 3.3 se obtiene el modelo de conducción en DC; en la Sección 3.4 se desarrolla el modelo lineal incremental; y en la Sección 3.5 se desarrollan las capacidades resultantes del modelo de AC. Finalmente, en la Sección 3.6 se explican los mecanismos de ruptura que son responsables de la conducción del diodo ante tensiones inversas grandes.

3.2. Electroestática de la Juntura

Para comenzar el análisis, se considerarán dos partes de Silicio: una dopada tipo P, y otra dopada tipo N, tal como se ilustra en la Fig. 3.2.

Tal como se ha descrito en la sección anterior, luego de un período transitorio, se forma una zona de vaciamiento con iones negativos del lado del material P, e iones positivos del lado del material N. Quedan determinadas así, cuatro regiones diferentes, que se ilustran en la Fig. 3.2: a) el sustrato tipo P; b) la zona de vaciamiento en el material P; c) la zona de vaciamiento en el material N; d) el sustrato tipo N. En cada una de estas regiones, es posible hacer un balance de cargas.

El sustrato tipo P está dopado con N_a aceptores, y tiene n_P y p_P portadores, de manera que en esta región, la carga neta es igual a:

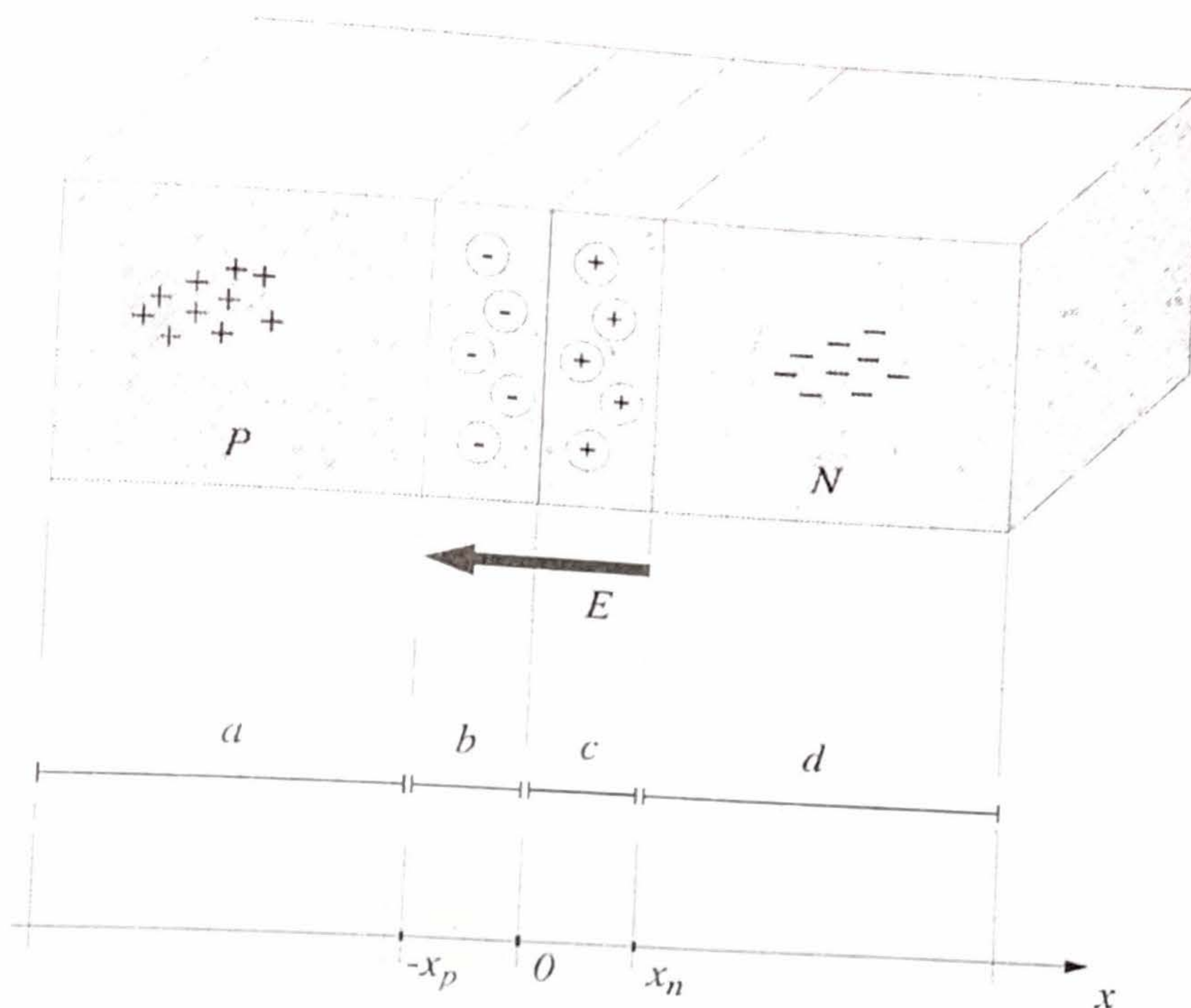


Figura 3.2: Juntura semiconductor y sus regiones: a) el sustrato tipo P; b) la zona de vaciamiento en el material P; c) la zona de vaciamiento en el material N; d) el sustrato tipo N.

$$\rho = q(-N_a - n_P + p_P) = 0 \quad (3.2)$$

La zona de vaciamiento en el material P tiene una carga neta, producto de la presencia de los iones negativos, igual a:

$$\rho = q(-N_a - n_P + p_P) \quad (3.3)$$

La zona de vaciamiento en el material N tiene una carga neta, producto de la presencia de los iones positivos, igual a:

$$\rho = q(N_d - n_N + p_N) \quad (3.4)$$

El sustrato tipo N está dopado con N_d donadores, y tiene n_N y p_N portadores de manera que en esta región, la carga neta es igual a:

$$\rho = q(N_d - n_N + p_N) = 0 \quad (3.5)$$

En este punto, resulta conveniente utilizar la "aproximación de vaciamiento", la cual consiste en suponer que en la región de vaciamiento no hay portadores -ni minoritarios, ni mayoritarios. Esta aproximación es razonable, dado que los portadores mayoritarios en estas regiones migran masivamente atravesando la juntura, y los minoritarios son removidos por el campo eléctrico. La aproximación produce un error numéricamente despreciable, a la vez que permite desarrollar las ecuaciones electroestáticas de una manera sencilla.

Utilizando esta aproximación, la densidad de carga en la zona de vaciamiento, en el material P, resulta:

$$\rho = -qN_a \quad (3.6)$$

mientras que, en la zona de vaciamiento del material N, resulta:

$$\rho = qN_d \quad (3.7)$$

La densidad de carga a lo largo de la juntura se ilustra en la Fig. 3.3.

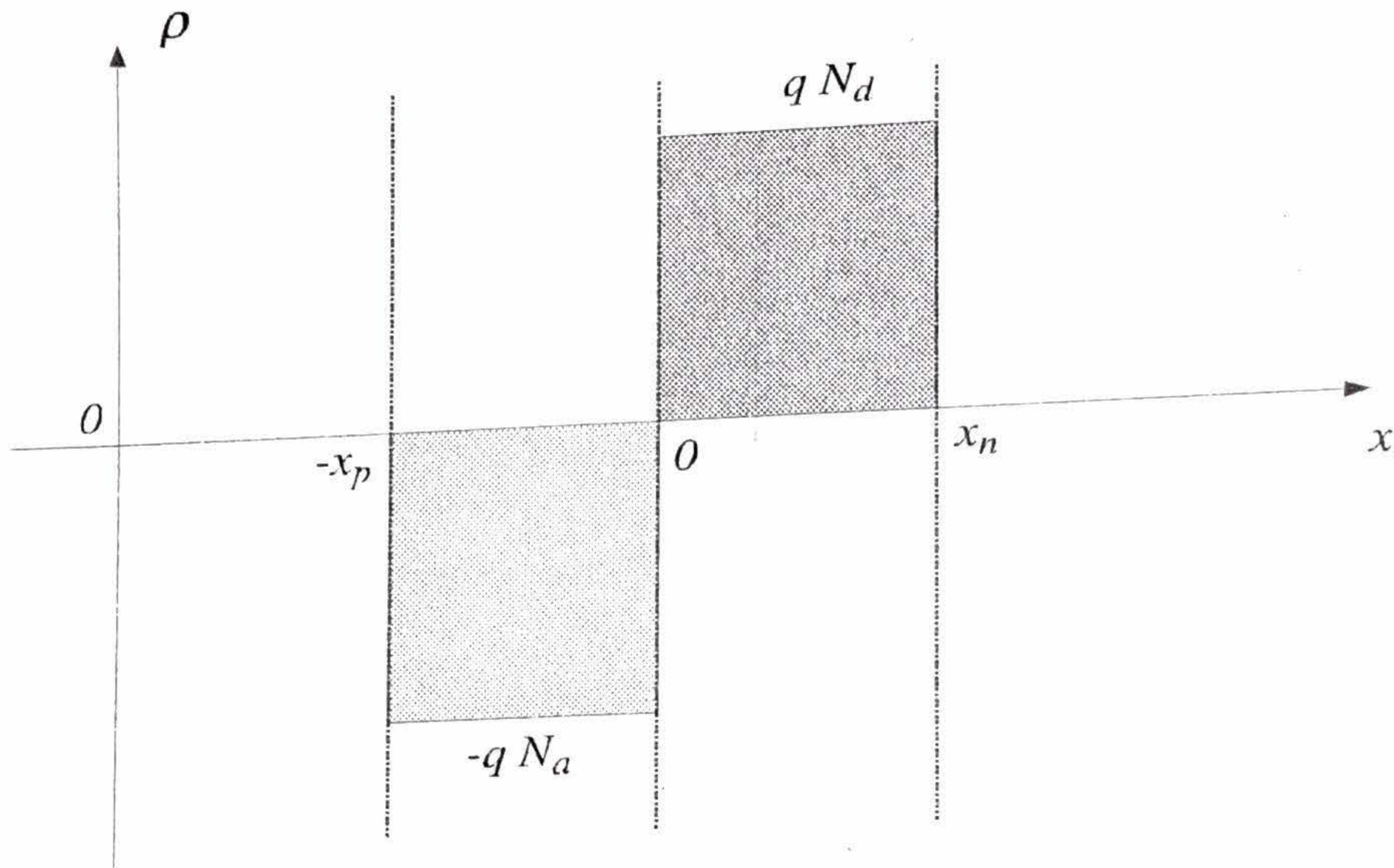


Figura 3.3: Densidad de carga a lo largo de la juntura.

Tanto en el sustrato P, como en el N, la densidad de cargas es nula, por ende, el campo eléctrico es nulo.

Dentro de la región de vaciamiento del material P, el campo eléctrico puede hallarse utilizando (B.4):

$$E(x) = \int_{-x_p}^x -\frac{qN_a}{\epsilon_{Si}} d\xi = -\frac{qN_a \xi}{\epsilon_{Si}} \Big|_{-x_p}^x \quad (3.8)$$

de donde resulta:

$$E(x) = -\frac{qN_a}{\epsilon_{Si}}(x + x_p) \quad (3.9)$$

para $-x_p \leq x \leq 0$.

De la misma manera, dentro de la región de vaciamiento del material N, el campo eléctrico está dado por:

$$E(x) = E(0) + \int_0^x \frac{qN_d}{\epsilon_{Si}} d\xi = E(0) + \frac{qN_d}{\epsilon_{Si}} \xi \Big|_0^x = E(0) + \frac{qN_d}{\epsilon_{Si}} x \quad (3.10)$$

El valor de $E(0)$ puede averiguarse evaluando (3.10) en $x = x_n$, donde el campo debe valer $E = 0$;

$$E(x_n) = E(0) + \frac{qN_d}{\epsilon_{Si}} x_n = 0 \quad (3.11)$$

por lo tanto,

$$E(0) = -\frac{qN_d}{\epsilon_{Si}} x_n \quad (3.12)$$

Reemplazando (3.12) en (3.10), se obtiene la expresión del campo eléctrico:

$$E(x) = \frac{qN_d}{\epsilon_{Si}}(x - x_n) \quad (3.13)$$

para $0 \leq x \leq x_n$.

El campo eléctrico, a lo largo de la juntura, se ilustra en la Fig. 3.4.

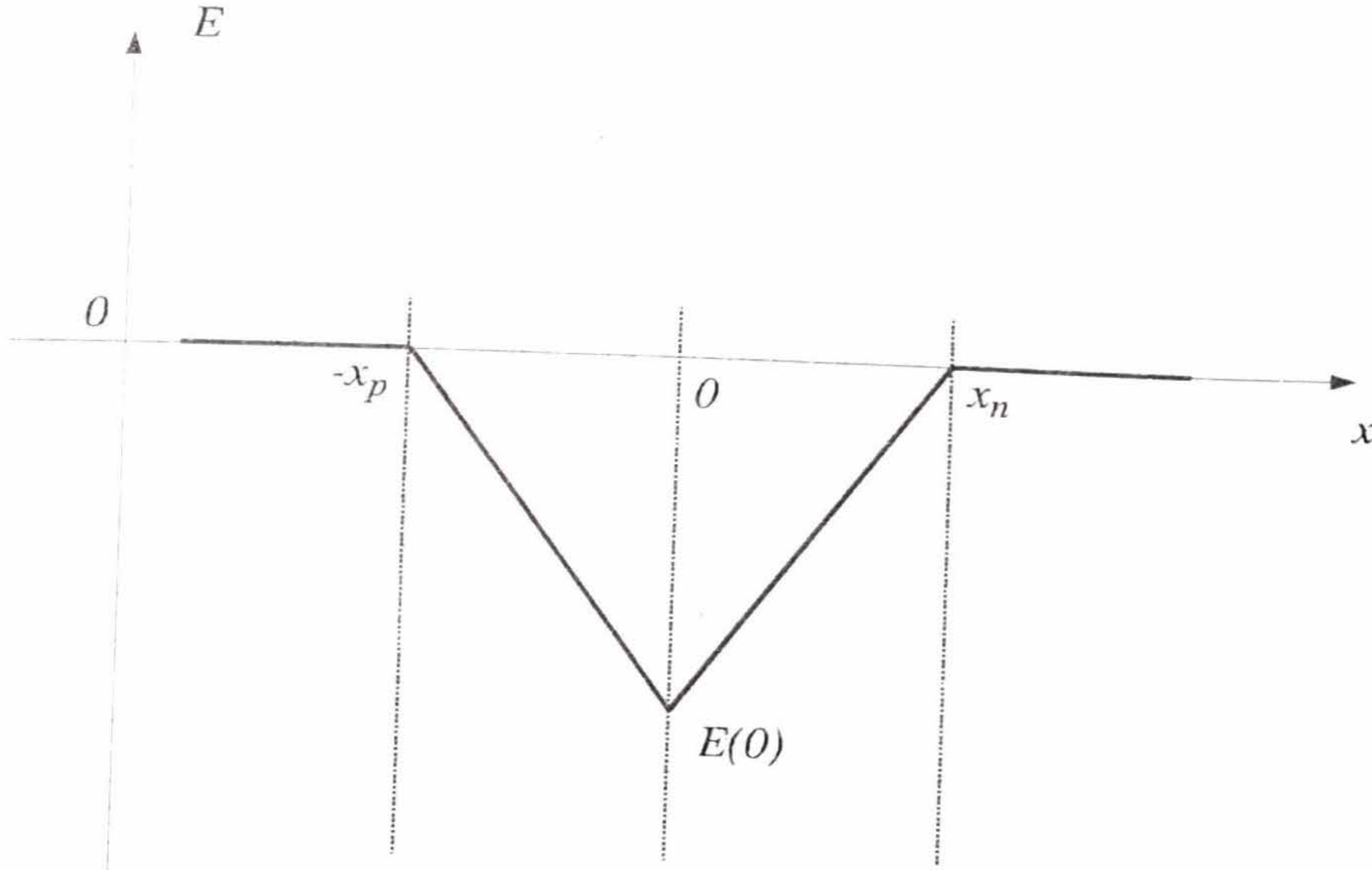


Figura 3.4: Campo eléctrico a lo largo de la juntura.

La continuidad del campo eléctrico (dado que no hay cambio de material), implica que las expresiones halladas para las regiones de vaciamiento, P y N, deben ser iguales en $x = 0$, es decir:

$$E(0) = -\frac{qN_a}{\epsilon_{Si}}(0 + x_p) = \frac{qN_d}{\epsilon_{Si}}(0 - x_n) \quad (3.14)$$

de donde surge que:

$$N_a x_p = N_d x_n \quad (3.15)$$

La Ec. (3.15) pone de manifiesto la igualdad de las cargas, a uno y otro lado de la juntura. De aquí, también se deduce que la zona de vaciamiento tiene menor profundidad de penetración en el lado de mayor concentración de impurezas. Esto es equivalente a decir que la zona de vaciamiento se desarrolla, mayormente, en la región de menor dopado.

La solución del potencial, a lo largo de la estructura, requiere de la integración del campo eléctrico, de acuerdo a la definición de potencial (B.4).

En la zona de vaciamiento del lado P, el potencial se halla de la siguiente manera:

$$\begin{aligned} \phi &= \phi_0 + \int_{-x_p}^x \frac{qN_a}{\epsilon_{Si}}(\xi + x_p)d\xi \\ &= \phi_0 + \frac{qN_a}{\epsilon_{Si}} \left(\frac{1}{2}\xi^2 + x_p\xi \right) \Big|_{-x_p}^x \\ &= \phi_0 + \frac{qN_a}{\epsilon_{Si}} \frac{1}{2} (x^2 + 2x_p x + x_p^2) \end{aligned} \quad (3.16)$$

Si se evalúa esta expresión en $x = 0$, se obtiene el potencial en la unión de ambos materiales:

$$\phi(0) = \phi_n - \frac{1}{2} \frac{qN_d}{\epsilon_{Si}} x_n^2 \quad (3.22)$$

Dado que el potencial debe ser continuo, (3.18) y (3.22), deben ser iguales, es decir:

$$\phi_p + \frac{1}{2} \frac{qN_a}{\epsilon_{Si}} x_p^2 = \phi_n - \frac{1}{2} \frac{qN_d}{\epsilon_{Si}} x_n^2 \quad (3.23)$$

De esta ecuación, es posible hallar el valor de x_p y x_n . Reagrupando términos, se llega a:

$$\frac{qN_a}{\epsilon_{Si}} x_p^2 + \frac{qN_d}{\epsilon_{Si}} x_n^2 = 2\phi_B \quad (3.24)$$

donde $\phi_B = \phi_n - \phi_p$. Reemplazando x_p , de (3.15), en función de x_n , resulta:

$$2\epsilon_{Si}\phi_B = qN_a \left(\frac{N_d x_n}{N_a} \right)^2 + qN_d x_n^2 \quad (3.25)$$

$$= qN_d \left(1 + \frac{N_d}{N_a} \right) x_n^2 \quad (3.26)$$

De aquí surge el valor de x_n :

$$x_n = \sqrt{\frac{2\epsilon_{Si}N_a\phi_B}{qN_d(N_a + N_d)}} \quad (3.27)$$

y utilizando (3.15), se puede hallar también el valor de x_p :

$$x_p = \sqrt{\frac{2\epsilon_{Si}N_d\phi_B}{qN_a(N_a + N_d)}} \quad (3.28)$$

En base a (3.27) y (3.28), se puede calcular el valor total de la zona de vaciamiento:

$$x_B = x_p + x_n = \sqrt{\frac{2\epsilon_{Si}\phi_B}{q} \left(\frac{1}{N_d} + \frac{1}{N_a} \right)} \quad (3.29)$$

En el caso que se aumente el dopado de ambas regiones, manteniendo la relación entre ellas, es decir, manteniendo $N_a/N_d = cte$, se puede ver de (3.27)-(3.29), que las zonas de vaciamiento individuales se reducen en forma inversamente proporcional a la raíz cuadrada del dopado, es decir:

$$x_p, x_n \propto \frac{1}{\sqrt{N_a + N_d}} \quad (3.30)$$

Ejemplo 12 Un diodo semiconductor está dopado con $N_a = 10^{15}$ y $N_d = 10^{16}$. Halle la longitud de las zonas de vaciamiento en cada uno de los materiales, y la total.

En primer lugar, por la regla de los 60mV, resulta $\phi_n = 60mV \times \log_{10}(10^6) = 360mV$, $\phi_p = -60mV \times \log_{10}(10^5) = -300mV$ y $\phi_B = 360mV - (-300mV) = 660mV$. Luego, las longitudes de las zonas de vaciamiento se pueden calcular como:

$$\begin{aligned}
 x_n &= \sqrt{\frac{2 \times 11,8 \times 8,854 \times 10^{-14} \times 10^{15} \times 660 \times 10^{-3}}{1,6 \times 10^{-19} \times 10^{16} \times (10^{16} + 10^{15})}} \\
 &= 8,852 \times 10^{-6} \text{ cm} = 88,52 \text{ nm}
 \end{aligned}
 \tag{3.31}$$

$$\begin{aligned}
 x_p &= \sqrt{\frac{2 \times 11,8 \times 8,854 \times 10^{-14} \times 10^{16} \times 660 \times 10^{-3}}{1,6 \times 10^{-19} \times 10^{15} \times (10^{16} + 10^{15})}} \\
 &= 88,52 \times 10^{-6} \text{ cm} = 0,8852 \mu\text{m}
 \end{aligned}
 \tag{3.32}$$

$$x_B = 0,8852 \mu\text{m} + 0,08852 \mu\text{m} = 0,974 \mu\text{m}
 \tag{3.33}$$

3.3. Modelo de DC

Para el análisis de la conducción de la juntura, se considerará el circuito de la Fig. 3.6, donde se ha dispuesto una fuente de tensión V_A . La ecuación de mallas indica:

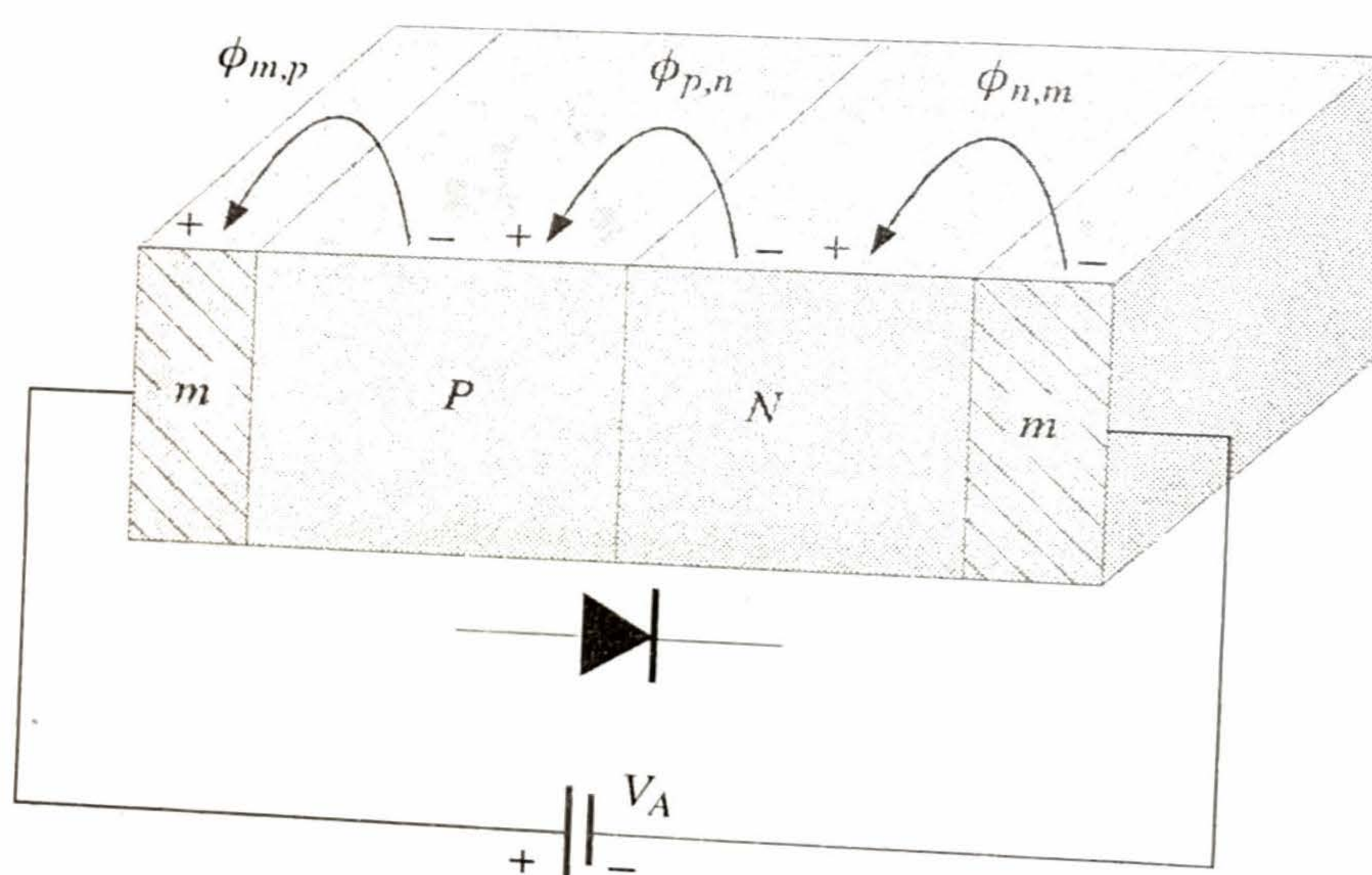


Figura 3.6: Juntura semiconductor con contactos metálicos y símbolo eléctrico del diodo.

$$\begin{aligned}
 V_A &= \phi_{m,p} + \phi_{p,n} + \phi_{n,m} \\
 &= (\phi_{m,p} - \phi_{m,n}) + \phi_{p,n}
 \end{aligned}
 \tag{3.34}$$

De aquí puede verse que cuando la tensión externa V_A es nula, resulta:

$$-\phi_{p,n} = \phi_B = \phi_{m,p} - \phi_{m,n}
 \tag{3.35}$$

Las tensiones resultantes a lo largo del dispositivo, para este caso, se muestran en la Fig. 3.7.

Si ahora se introduce un valor distinto de cero para la fuente de tensión V_A , la ecuación de mallas (3.34) es:

$$V_A = \phi_{m,p} - \phi_{m,n} - \phi_J \quad (3.36)$$

donde ϕ_J es el valor del potencial de juntura, que ahora es diferente de ϕ_B , debido a la aplicación del potencial externo. Utilizando (3.35), la ecuación anterior se puede reescribir como:

$$\phi_J = \phi_B - V_A \quad (3.37)$$

La situación se ilustra en la Fig. 3.8 para los casos $V_A > 0$ y $V_A < 0$, respectivamente.

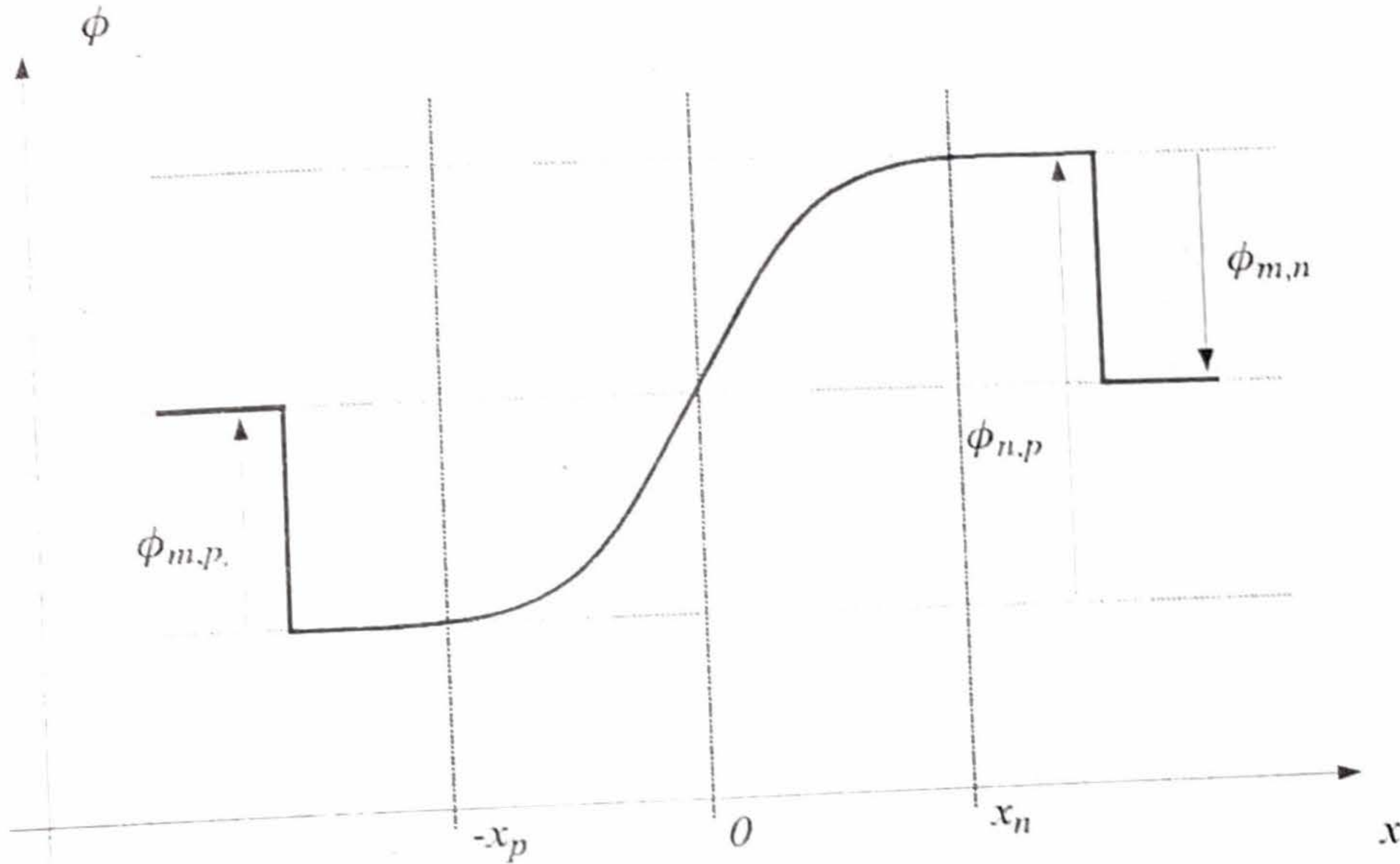


Figura 3.7: Potenciales a lo largo de los contactos metálicos y la juntura, con $V_A = 0$.

3.3.1. Ley de la Juntura

Previo al cálculo de las corrientes en una juntura semiconductor, resulta de utilidad derivar la relación existente entre los portadores minoritarios y mayoritarios, a ambos lados de la zona de vaciamiento.

Los potenciales, en función de las concentraciones de electrones, son:

$$\phi_p = U_T \ln \left(\frac{n_p}{n_i} \right) \quad (3.38)$$

$$\phi_n = U_T \ln \left(\frac{n_n}{n_i} \right) \quad (3.39)$$

donde n_p es la concentración de electrones en el material P, y n_n es la concentración de electrones en el material N. Restando ambas expresiones, resulta:

$$\begin{aligned} \phi_B &= \phi_n - \phi_p = U_T \left(\ln \left(\frac{n_n}{n_i} \right) - \ln \left(\frac{n_p}{n_i} \right) \right) \\ &= U_T \ln \left(\frac{n_n n_i}{n_i n_p} \right) = U_T \ln \left(\frac{n_n}{n_p} \right) \end{aligned} \quad (3.40)$$

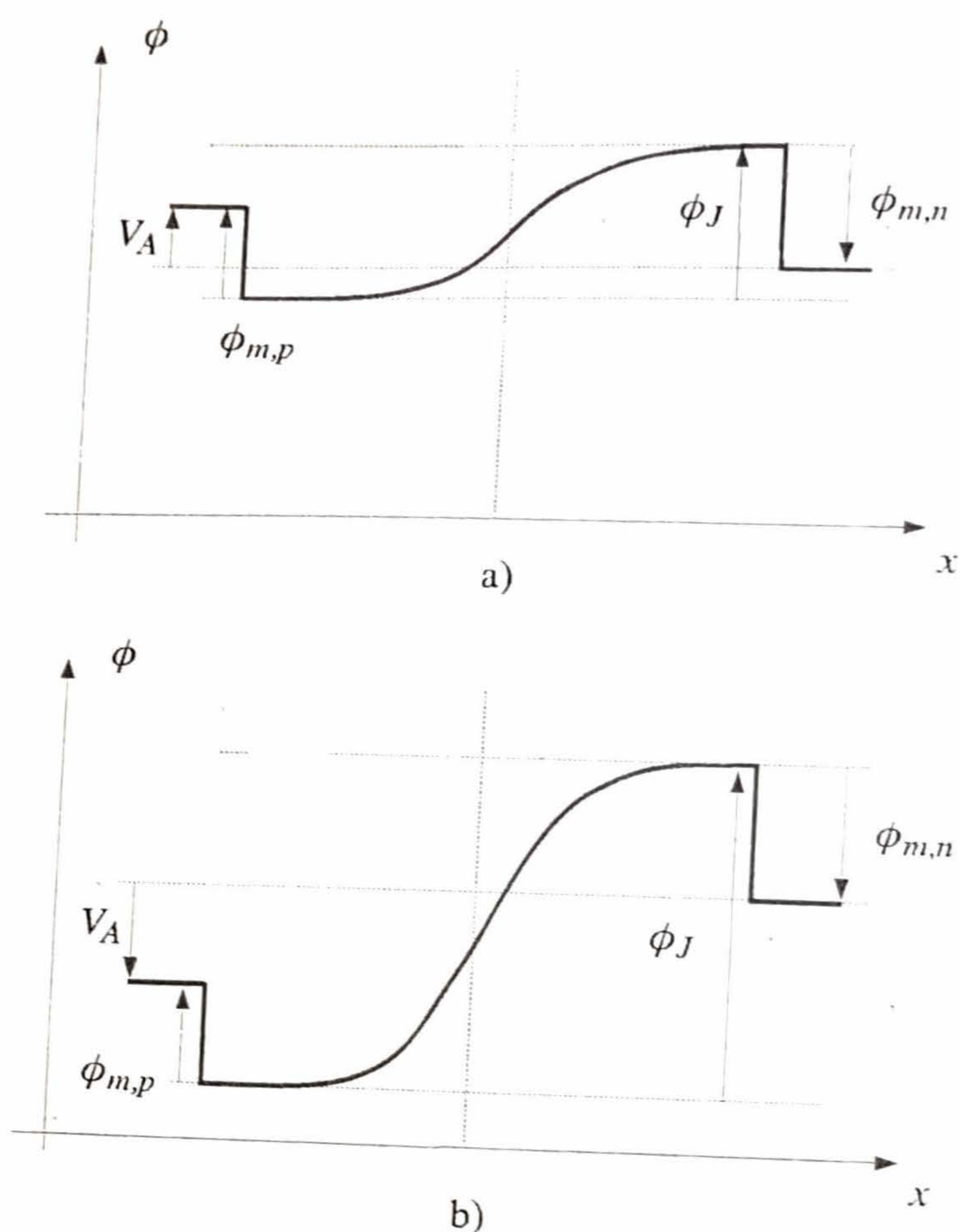


Figura 3.8: Potenciales a lo largo de los contactos metálicos y la juntura, con un potencial externo aplicado: a) $V_A > 0$; b) $V_A < 0$.

Dado que $n_N = N_d$, los portadores minoritarios del lado P pueden escribirse como:

$$n_P = N_d e^{-\phi_B/U_T} \quad (3.41)$$

De manera similar, los potenciales, en función de las concentraciones de huecos, son:

$$\phi_P = -U_T \ln \left(\frac{p_P}{n_i} \right) \quad (3.42)$$

$$\phi_N = -U_T \ln \left(\frac{p_N}{n_i} \right) \quad (3.43)$$

donde p_P es la concentración de huecos en el material P, y p_N es la concentración de huecos en el material N. Restando ambas expresiones, resulta:

$$\phi_B = \phi_N - \phi_P = -U_T \ln \left(\frac{p_N}{p_P} \right) \quad (3.44)$$

Dado que $p_P = N_a$, los portadores minoritarios del lado N, es decir p_N , pueden escribirse como:

$$p_N = N_a e^{-\phi_B/U_T} \quad (3.45)$$

3.3. Modelo de DC

En síntesis, la concentración de portadores minoritarios en un lado de la juntura se puede hallar multiplicando la concentración de portadores mayoritarios al otro lado de la juntura por el factor $e^{-\phi_B/U_T}$. Las expresiones (3.41) y (3.45), se conocen como la Ley de la Juntura, y se ilustran gráficamente en la Fig. 3.9.

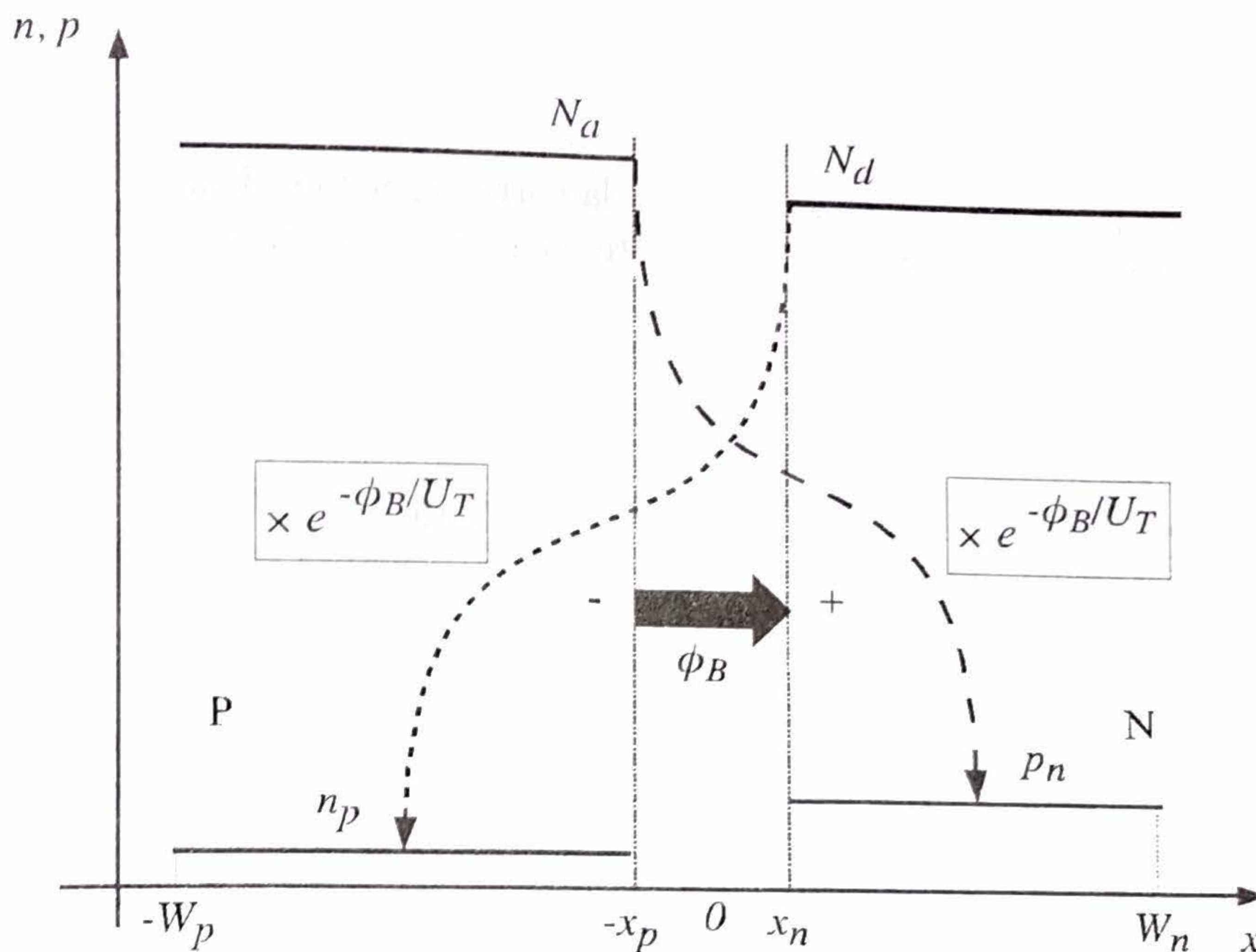


Figura 3.9: Ilustración de la Ley de la Juntura: Relación entre portadores mayoritarios y minoritarios, en función de una diferencia de potencial ϕ_B .

3.3.2. Solución en directa

Si se considera que $V_A > 0$, la juntura está en modo de conducción directa. De (3.37), se puede ver que el efecto de esta fuente de tensión positiva es disminuir el potencial intrínseco de la juntura, permitiendo la conducción de corriente por difusión. Esta corriente sucede por difusión de huecos del lado P hacia el lado N, y por difusión de electrones del lado N hacia el lado P. Dada la gran cantidad de huecos en el lado P, y de electrones en el lado N, los niveles de portadores mayoritarios no se ven alterados por la difusión de portadores, por lo que se asumen constantes. Por el contrario, los portadores minoritarios se incrementan notablemente por la inyección de portadores, provenientes del otro lado de la zona de vaciamiento. Debido a esto, se calculará la corriente utilizando la expresión de los portadores minoritarios.

Se harán dos suposiciones importantes. En primer lugar, se supondrá que la longitud del material es lo suficientemente pequeña, de manera tal que los portadores minoritarios atraviesan el sustrato es lo suficientemente pequeña, de manera tal que los portadores minoritarios atraviesan el sustrato correspondiente sin recombinarse, hasta alcanzar el contacto metálico. De esta manera, si no hay recombinación, el perfil de concentración es lineal, su derivada es constante, y por lo tanto, el flujo de portadores también es constante. Este caso se denomina solución corta, y simplifica el cálculo de la corriente¹.

La segunda suposición tiene que ver con la corriente. La circulación de corriente, producto de la aplicación de un potencial externo, indica que la suposición de equilibrio térmico *no es válida*; debido a esto, no sería posible relacionar los niveles de concentración con el potencial en la juntura, de acuerdo

¹En el caso general, el perfil de concentración estacionario resulta exponencial [8].

a (3.41) y (3.45). De hecho, las relaciones entre concentraciones y potenciales, que dan lugar a la Regla de los $60mV$, descrita en el Cap. 2, se sustenta en el hecho que las corrientes de electrones:

$$J_n = J_n^a + J_n^d \quad (3.46)$$

y de huecos:

$$J_p = J_p^a + J_p^d \quad (3.47)$$

son ambas nulas.

La segunda suposición que se hace, entonces, es que la corriente neta es lo suficientemente pequeña con respecto a las componentes individuales de corriente en equilibrio: J_n^a , J_n^d , J_p^a y J_p^d ; de manera tal que, el material puede considerarse en *equilibrio térmico*. Esto es válido mientras que los niveles de portadores minoritarios a ambos lados de la zona de vaciamiento sean mucho menores que los niveles de portadores mayoritarios, es decir:

$$n_p(-x_p) \ll N_a \quad p_n(x_n) \ll N_d \quad (3.48)$$

La condición (3.48), se denomina condición de "baja inyección de portadores".

Con las suposiciones realizadas, el perfil de carga de la juntura es como se muestra en la Fig. 3.10. Nótese que es preciso introducir un quiebre en la escala para graficar N_a y N_d en el gráfico, dado que sus valores son varios órdenes de magnitud superiores a los niveles de los portadores minoritarios.

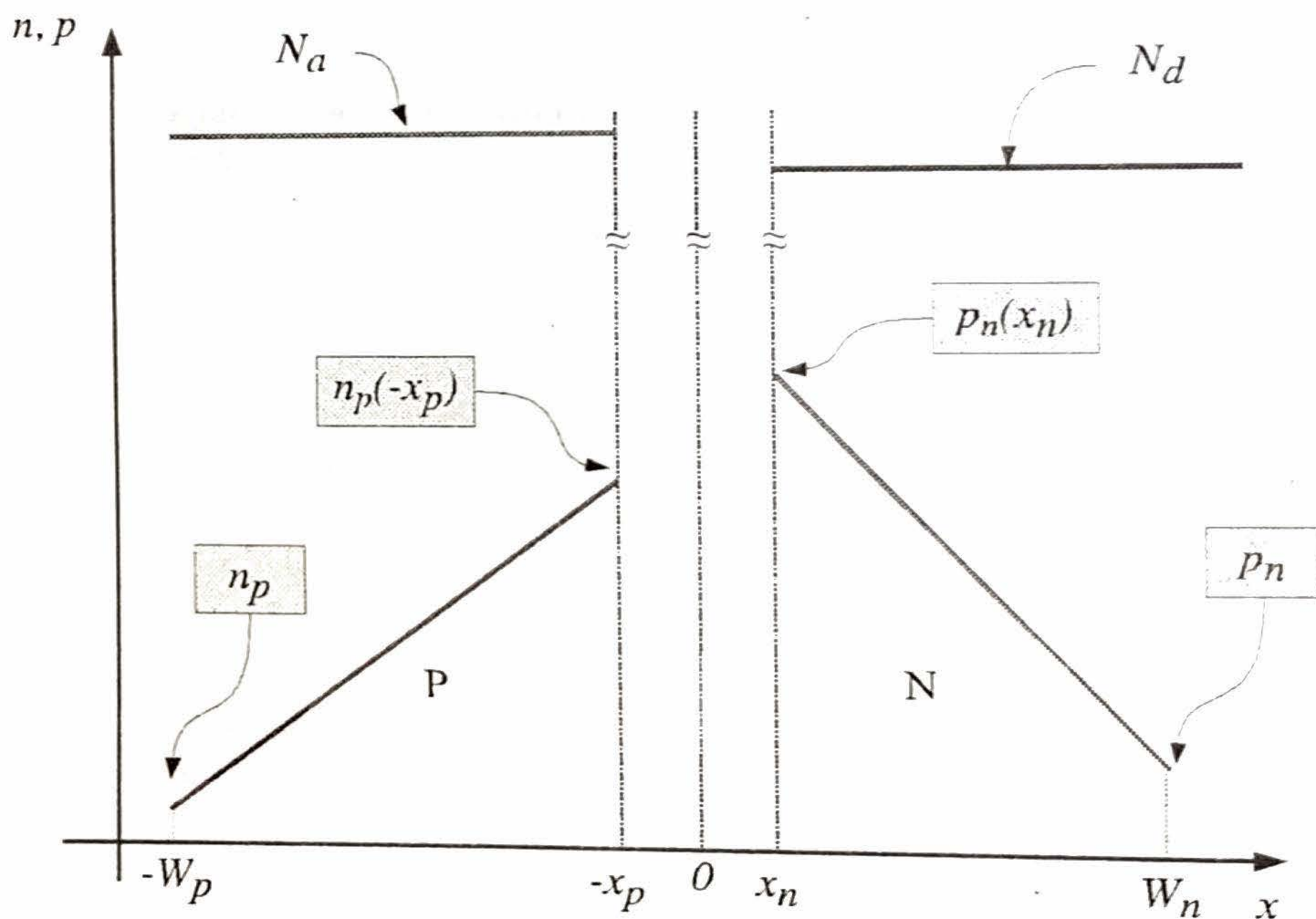


Figura 3.10: Niveles de portadores a ambos lados de la juntura ante un potencial externo positivo.

La corriente de difusión de portadores negativos, en el lado P, se puede hallar como:

$$\begin{aligned} J_n^d &= qD_n \frac{dn}{dx} \\ &= qD_n \frac{n_p(-x_p) - n_p(-W_p)}{-x_p + W_p} \end{aligned} \quad (3.49)$$

Teniendo en cuenta que, por (3.41):

$$n_P(-x_p) = N_d e^{-\phi_J/U_T}$$

y también que $n_P(-W_p) = n_P$, y $x_p \ll W_p$, (3.49) resulta en:

$$J_n^d = qD_n \frac{N_d e^{-\phi_J/U_T} - n_P}{W_p} \quad (3.51)$$

Dado que

$$N_d e^{-\phi_J/U_T} = N_d e^{-\phi_B/U_T} e^{V_A/U_T} = n_P e^{V_A/U_T} \quad (3.52)$$

donde $n_P = N_d e^{-\phi_B/U_T}$ surge de la aplicación de la Ley de la Juntura, sin potencial externo aplicado, la Ec. (3.51) puede escribirse como:

$$J_n^d = \frac{qD_n n_P}{W_p} (e^{V_A/U_T} - 1) \quad (3.53)$$

La corriente de difusión de portadores positivos, en el lado N, se puede hallar como:

$$\begin{aligned} J_p^d &= -qD_p \frac{dp}{dx} \\ &= -qD_p \frac{p_N(W_n) - p_N(x_n)}{W_n - x_n} \end{aligned} \quad (3.54)$$

Teniendo en cuenta que, por (3.45):

$$p_N(x_n) = N_a e^{-\phi_J/U_T}, \quad (3.55)$$

y también que $p_N(W_n) = p_N$, y $x_n \ll W_n$, (3.54) resulta en:

$$J_p^d = -qD_p \frac{p_N - N_a e^{-\phi_J/U_T}}{W_n} \quad (3.56)$$

Dado que

$$N_a e^{-\phi_J/U_T} = N_a e^{-\phi_B/U_T} e^{V_A/U_T} = p_N e^{V_A/U_T} \quad (3.57)$$

donde $p_N = N_a e^{-\phi_B/U_T}$ surge de la aplicación de la Ley de la Juntura, sin potencial externo aplicado, la Ec. (3.56) puede escribirse como:

$$J_p^d = \frac{qD_p p_N}{W_n} (e^{V_A/U_T} - 1) \quad (3.58)$$

Es claro que la corriente total es la suma de las corrientes de difusión de huecos y electrones, es decir:

$$J_D = J_p^d + J_n^d \quad (3.59)$$

Combinando (3.53) y (3.58), e incluyendo el área A del diodo, se obtiene la corriente del diodo:

$$\boxed{i_D = I_s (e^{v_D/U_T} - 1)} \quad (3.60)$$

donde v_D es la tensión aplicada en terminales del diodo, y

$$I_s = qA \left(\frac{D_p p_N}{W_n} + \frac{D_n n_P}{W_p} \right) \quad (3.61)$$

La Fig. 3.11 muestra la curva característica del diodo, que refleja la característica exponencial de la Ec. (3.60).

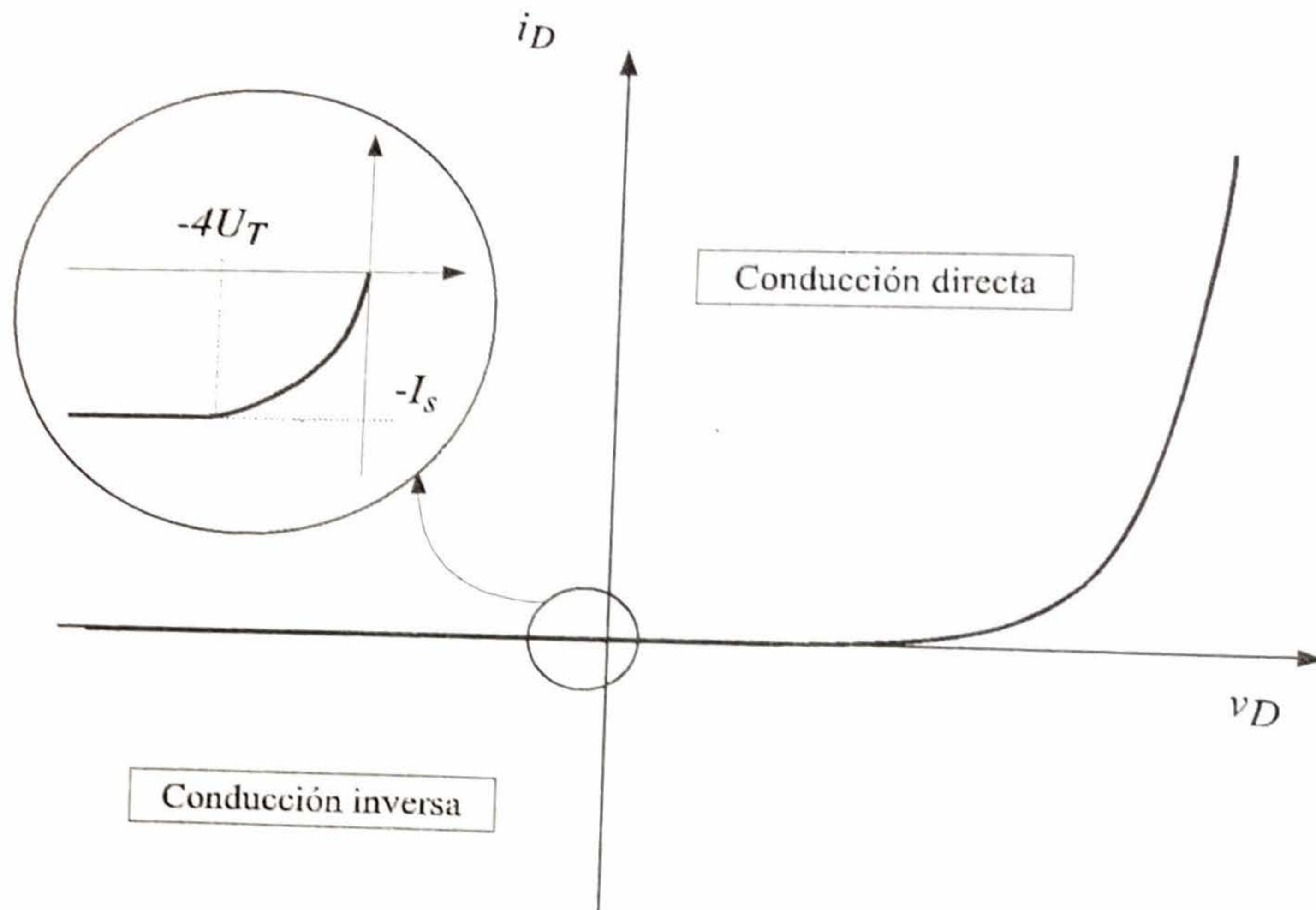


Figura 3.11: Curva i_D vs. v_D del diodo, mostrando la zona de conducción directa, y la zona de conducción inversa.

3.3.3. Solución en inversa

Si la tensión de la fuente de alimentación es negativa, es decir $V_A < 0$, entonces la juntura se dice en modo de conducción inversa. En el desarrollo realizado en la sección anterior, la tensión de alimentación V_A se consideró positiva, pero el desarrollo sigue siendo válido si se considera $V_A < 0$. De hecho, en el modo de conducción inversa, la corriente de la juntura es despreciable, con lo cual la juntura está más cerca de las condiciones de equilibrio térmico.

En el caso de tensión inversa, los niveles de portadores, a ambos lados de la zona de vaciamiento, disminuyen con respecto a los valores de equilibrio térmico. En el lado P del material, se verifica:

$$\begin{aligned} n_P(-x_p) &= N_d e^{-\phi_B/U_T} e^{V_A/U_T} \\ &= n_P e^{V_A/U_T} \end{aligned} \quad (3.62)$$

Dado que V_A es negativa, si $V_A < -4U_T$, es decir $V_A < -100\text{mV}$, luego $e^{V_A/U_T} \approx 0$ y (3.62) se reduce a:

$$n_P(-x_p) = 0 \quad (3.63)$$

De la misma manera, en el lado N del material, se verifica:

$$\begin{aligned}
 p_N(x_n) &= N_a e^{-\phi_B/U_T} e^{V_A/U_T} \\
 &= p_N e^{V_A/U_T}
 \end{aligned}
 \tag{3.64}$$

Para $V_A < -4U_T$, (3.64) se reduce a:

$$p_N(x_n) = 0 \tag{3.65}$$

En este caso, el perfil de portadores es como se ilustra en la Fig. 3.12.

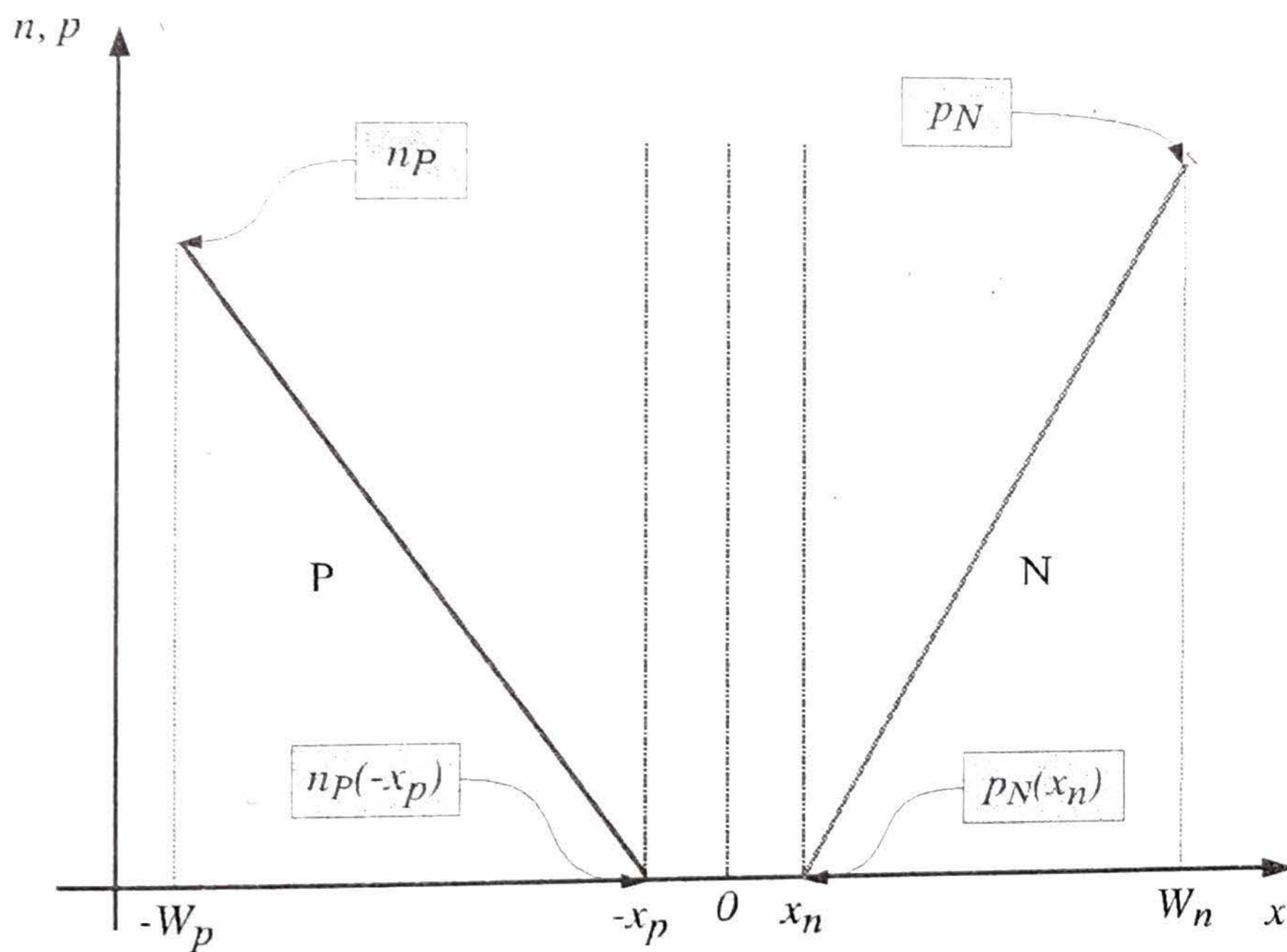


Figura 3.12: Niveles de portadores minoritarios, a ambos lados de la juntura, ante un potencial externo negativo.

De aquí puede verse que el aumento del campo eléctrico produce un flujo de electrones (minoritarios) desde el lado P hacia el lado N, y un flujo de huecos (minoritarios) desde el lado N hacia el lado P. A diferencia del caso de conducción en directa, aquí, una pequeña tensión inversa de unos 100mV es suficiente para disminuir los portadores a ambos lados de la zona de vaciamiento *prácticamente a cero*. A partir de allí, la corriente permanece constante, dado que los niveles de portadores no pueden caer por debajo de cero. La expresión analítica de la corriente es igual a la hallada para el caso de conducción directa, y está dada por (3.60). Considerando $V_A < -100mV$, resulta:

$$I_D = -I_s \tag{3.66}$$

que es la denominada corriente de fuga de la juntura. El detalle en la Fig. 3.11 ilustra la zona de conducción inversa.

Nótese como la zona de vaciamiento, que se obtiene reemplazando ϕ_B por ϕ_J , en (3.27)-(3.29), aumenta ante la aplicación de una tensión inversa ($V_A < 0$):

$$x_n(V_A) = \sqrt{\frac{2\varepsilon_{Si}N_a(\phi_B - V_A)}{qN_d(N_a + N_d)}} \quad (3.67)$$

$$x_p(V_A) = \sqrt{\frac{2\varepsilon_{Si}N_d(\phi_B - V_A)}{qN_a(N_a + N_d)}} \quad (3.68)$$

$$x_B(V_A) = x_p(V_A) + x_n(V_A) = \sqrt{\frac{2\varepsilon_{Si}(\phi_B - V_A)}{q} \left(\frac{1}{N_d} + \frac{1}{N_a} \right)} \quad (3.69)$$

Estas tres ecuaciones pueden reescribirse de la siguiente manera:

$$x_n(V_A) = \sqrt{\frac{2\varepsilon_{Si}N_a\phi_B}{qN_d(N_a + N_d)} \left(1 - \frac{V_A}{\phi_B}\right)} = x_n(0) \sqrt{\left(1 - \frac{V_A}{\phi_B}\right)} \quad (3.70)$$

$$x_p(V_A) = \sqrt{\frac{2\varepsilon_{Si}N_d\phi_B}{qN_a(N_a + N_d)} \left(1 - \frac{V_A}{\phi_B}\right)} = x_p(0) \sqrt{\left(1 - \frac{V_A}{\phi_B}\right)} \quad (3.71)$$

$$\begin{aligned} x_B(V_A) &= \sqrt{\frac{2\varepsilon_{Si}\phi_B}{q} \left(\frac{1}{N_d} + \frac{1}{N_a} \right) \left(1 - \frac{V_A}{\phi_B}\right)} \\ &= x_B(0) \sqrt{\left(1 - \frac{V_A}{\phi_B}\right)} \end{aligned} \quad (3.72)$$

Dado que la concentración de impurezas es constante, al aumentar la zona de vaciamiento aumentan tanto la concentración de cargas, a ambos lados de la juntura, como el campo eléctrico, cuyo valor máximo es ahora:

$$E_{max} = E(0) = -q \frac{N_d}{\varepsilon_{Si}} x_n(V_A) = -q \frac{N_a}{\varepsilon_{Si}} x_p(V_A) \quad (3.73)$$

La situación se ilustra, en la Fig. 3.13, para tres tensiones aplicadas diferentes. Por las mismas razones, en conducción directa, el aumento de la tensión directa ($V_A > 0$, con $V_A < \phi_B$) produce una disminución tanto de la zona de vaciamiento, como del campo eléctrico resultante.

3.3.4. Desviaciones del comportamiento ideal

Los desarrollos realizados en las secciones previas, aún cuando fueron realizados bajo la suposición de longitud corta, permiten visualizar, de una manera sencilla, la conducción de corriente en un diodo, y arribar a un modelo de utilidad para el análisis de circuitos. Sin embargo, es importante tener en cuenta que un dispositivo real presenta otros efectos, cuyas consecuencias pueden necesitar de un análisis más detallado, en función de la aplicación concreta. Entre ellos, se pueden citar [8]: efectos de superficie, generación y recombinación de portadores en la zona de vaciamiento, conducción de portadores por efecto túnel, alta inyección de portadores y efectos debidos a la resistencia serie.

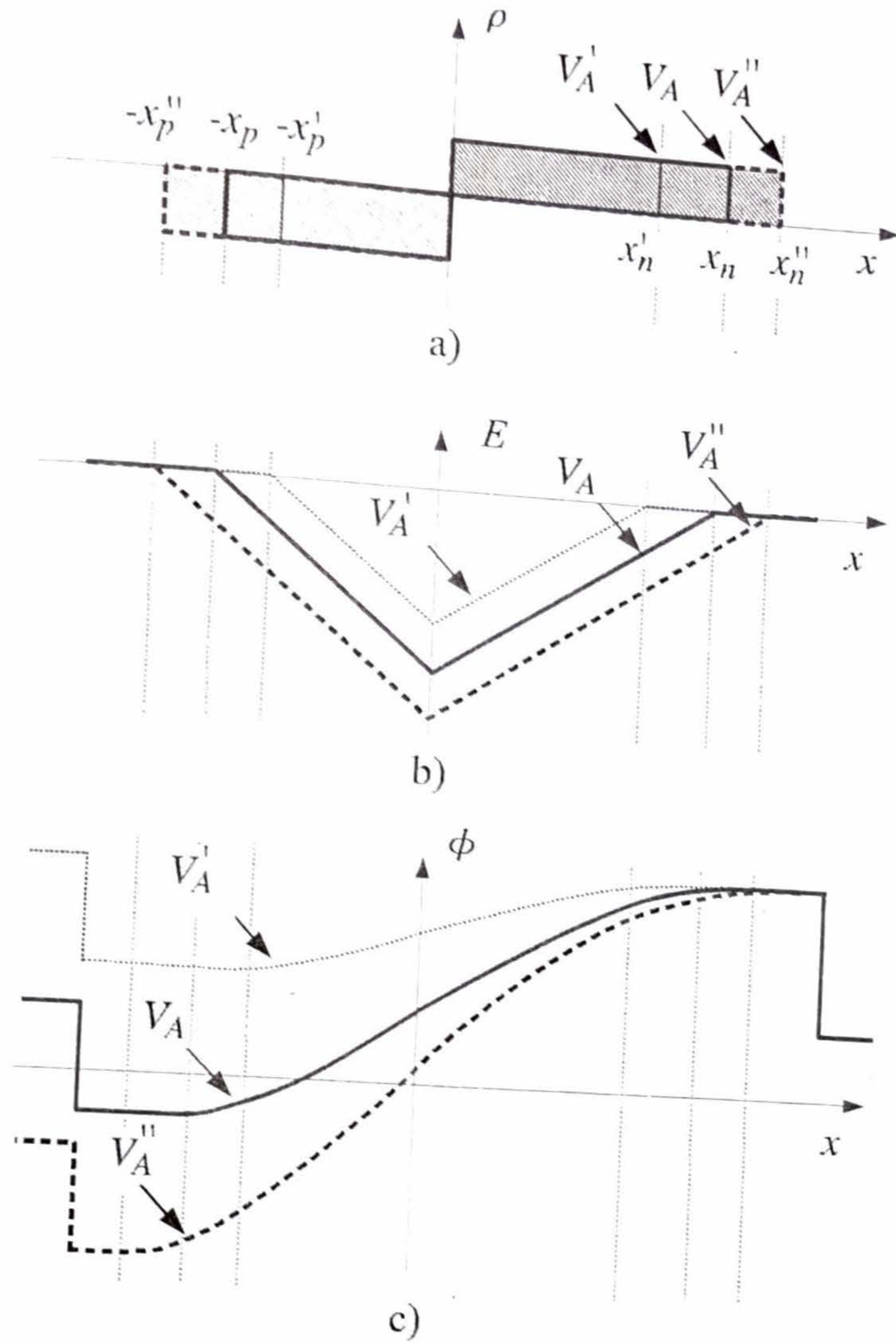


Figura 3.13: Situación de la juntura ante tres tensiones aplicadas, $V_A' > V_A > V_A''$: a) densidades de carga; b) campos eléctricos resultantes; c) potenciales.

Efectos de la temperatura

De acuerdo a (3.61), I_s es proporcional a n_i^2 . Tal como se puede apreciar de (2.10), n_i^2 depende de la temperatura a través de dos términos, de los cuales, el de mayor peso es el término exponencial [8], de manera que:

$$I_s \sim T^{(3/2)} \exp \frac{-E_g}{kT} \approx \exp \frac{-E_g}{kT} \quad (3.74)$$

Una expansión lineal de (3.74), con respecto a la temperatura, resulta en:

$$\begin{aligned} \exp \frac{-E_g}{kT} &= \exp \frac{-E_g}{kT_a} + \frac{E_g}{kT_a^2} \exp \frac{-E_g}{kT_a} \Delta T \\ &\sim I_s|_{T_a} \left(1 + \frac{E_g}{kT_a^2} \Delta T \right) \approx I_s|_{T_a} (1 + \Delta T/7) \end{aligned} \quad (3.75)$$

De acuerdo a esta ecuación, a temperatura ambiente ($T_a = 300K$), la corriente I_s , en el caso de Silicio, se duplica ante un incremento de $7^\circ C$ a temperatura ambiente. La corriente real en un diodo,

ante un cambio de temperatura, varía duplicándose cada $10^\circ C$ de incremento de la temperatura (para valores cercanos a la temperatura ambiente).

Por otro lado, para una corriente constante, la tensión del diodo puede expresarse de la siguiente manera:

$$v_D = U_T \ln \left(\frac{i_D}{I_s} \right) \quad (3.76)$$

Considerando una expansión lineal con respecto a I_s , y teniendo en cuenta la variación de I_s con la temperatura:

$$\begin{aligned} v_D &= v_D \Big|_{T_a} - \frac{U_T}{I_s} \Delta I_s \\ &= v_D \Big|_{T_a} - \frac{U_T}{I_s} \times \frac{I_s \Delta T}{7} = v_D \Big|_{T_a} - 3,5 mV/^\circ C \times \Delta T \end{aligned} \quad (3.77)$$

En diodos reales, la disminución es de $2mV$ por cada grado de temperatura.

3.4. Modelo Lineal Incremental

Para la obtención del modelo lineal incremental (MLI) del diodo, es necesario suponer que el mismo está operando en un punto de funcionamiento, definido por un par de valores de tensión y corriente continua, (V_D, I_D) . Se supone, entonces, que sobre el diodo hay una tensión aplicada, de la forma:

$$v_D = V_D + v_d \quad (3.78)$$

y la corriente resultante será:

$$i_D = I_D + i_d \quad (3.79)$$

El MLI describe la relación entre v_d e i_d .

Para ello, se toma la expresión de la corriente del diodo, dada por (3.60), y se realiza una expansión de Taylor, en un entorno del punto de funcionamiento (V_D, I_D) (ver Fig. 3.14):

$$i_D = I_D + \frac{\partial i_D}{\partial v_D} \Big|_{V_D} v_d + \frac{1}{2} \frac{\partial^2 i_D}{\partial v_D^2} \Big|_{V_D} v_d^2 + \dots \quad (3.80)$$

Para valores pequeños de v_d , los términos de orden dos y superiores pueden despreciarse y la corriente puede expresarse como:

$$i_d = g_d v_d \quad (3.81)$$

donde g_d es la conductancia del MLI, que se define como:

$$g_d \triangleq \frac{\partial i_D}{\partial v_D} \Big|_{I_D} \quad (3.82)$$

Si se tiene en cuenta la ecuación del diodo (3.60), la expresión de la conductancia es:

$$g_d = \frac{1}{U_T} I_s e^{V_D/U_T} \quad (3.83)$$

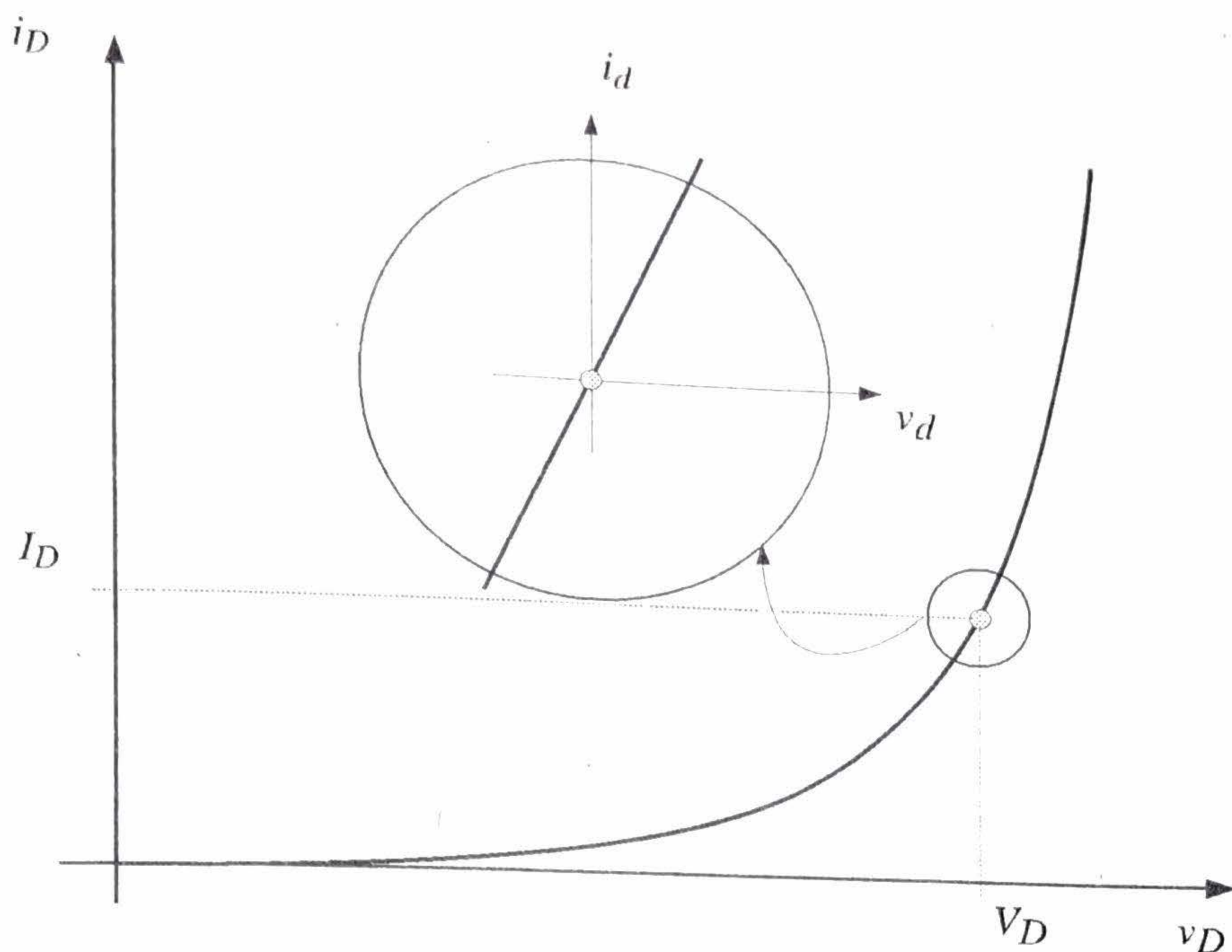


Figura 3.14: Curva característica del diodo y detalle con el modelo lineal incremental.

Si en la ecuación de la corriente del diodo (3.60), se considera el caso de conducción directa, el término unitario puede despreciarse con respecto al término exponencial, y la corriente se puede escribir como $i_D = I_s e^{V_D/U_T}$. En este caso, (3.83) puede escribirse como:

$$\boxed{g_d = \frac{I_D}{U_T}} \quad (3.84)$$

En otras palabras, en un entorno del punto de funcionamiento (I_D, V_D) , el diodo puede reemplazarse por una conductancia de valor (3.84), o por una resistencia de valor:

$$\boxed{r_d = \frac{U_T}{I_D}} \quad (3.85)$$

En la región de conducción inversa, el término exponencial ya no es despreciable frente al término -1 , y la conductancia (3.83), teniendo en cuenta que $V_D < -100mV$, resulta $g_d = 0$.

Ejemplo 13 En la región de conducción directa, la resistencia incremental de un diodo puede variar órdenes de magnitud, de acuerdo a la corriente de DC que circule. Para una corriente pequeña, $I_D = 100nA$, la resistencia incremental es:

$$r_d = \frac{25mV}{100nA} = 250K\Omega \quad (3.86)$$

Mientras que para una corriente $I_D = 10mA$, la misma se reduce a:

$$r_d = \frac{25mV}{10mA} = 2,5\Omega \quad (3.87)$$

3.5. Modelo de AC

El modelo de AC del diodo se obtiene añadiendo la capacidad del dispositivo. Esta varía dependiendo del modo de funcionamiento, presentando diferentes comportamientos en directa y en inversa. A continuación se verán ambos casos.

3.5.1. Capacidad en inversa

La capacidad en inversa es producto de la variación de la carga neta acumulada en la zona de vaciamiento. La carga de vaciamiento en el lado P del material es:

$$Q_p(v_D) = -qN_a x_p(v_D) = \quad (3.88)$$

$$= -qN_a x_p(0) \sqrt{1 - \frac{v_D}{\phi_B}} \quad (3.89)$$

La carga de vaciamiento, en el lado N, es:

$$Q_n(v_D) = qN_d x_n(v_D) \quad (3.90)$$

$$= qN_d x_n(0) \sqrt{1 - \frac{v_D}{\phi_B}} \quad (3.91)$$

Y por la neutralidad de carga, ambas satisfacen:

$$Q_p(v_D) = -Q_n(v_D) \quad (3.92)$$

La capacidad incremental puede hallarse entonces como:

$$C_i = \frac{dQ}{dv_D} = \frac{dQ_p}{dv_D} \quad (3.93)$$

o también como:

$$C_i = \frac{-dQ}{-dv_D} = -\frac{dQ_n}{dv_D} \quad (3.94)$$

Esta capacidad se denomina de vaciamiento, y puede calcularse como:

$$C_{vac} = \frac{d}{dv_D} \left(-qN_a x_p(0) \sqrt{1 - \frac{v_D}{\phi_B}} \right) \Big|_{V_D} \quad (3.95)$$

La evaluación de la derivada resulta en:

$$C_{vac} = \frac{qN_a x_p(0)}{2\phi_B \sqrt{1 - \frac{V_D}{\phi_B}}} \quad (3.96)$$

que puede escribirse como

$$C_{vac} = C_{j0} \frac{1}{\sqrt{1 - \frac{V_D}{\phi_B}}} \quad (3.97)$$

donde:

$$C_{j0} = \frac{qN_a x_p(0)}{2\phi_B} \quad (3.98)$$

es el valor de capacidad sin tensión aplicada. Nótese que C_{j0} puede escribirse como:

$$\begin{aligned} C_{j0} &= \frac{qN_a}{2\phi_B} \sqrt{\frac{2\epsilon_{Si} N_d \phi_B}{qN_a(N_a + N_d)}} = \sqrt{\frac{q\epsilon_{Si} N_a N_d}{2\phi_B(N_a + N_d)}} \\ &= \frac{\epsilon_{Si}}{x_B(0)} \quad (3.99) \end{aligned}$$

donde $x_B(0)$ es la longitud de la zona de vaciamiento, sin tensión aplicada, introducida en (3.72). En función de este resultado, (3.97) también puede escribirse como:

$$C_{vac} = \frac{\epsilon_{Si}}{x_B(0) \sqrt{1 - \frac{V_D}{\phi_B}}} = \frac{\epsilon_{Si}}{x_B(V_D)} \quad (3.100)$$

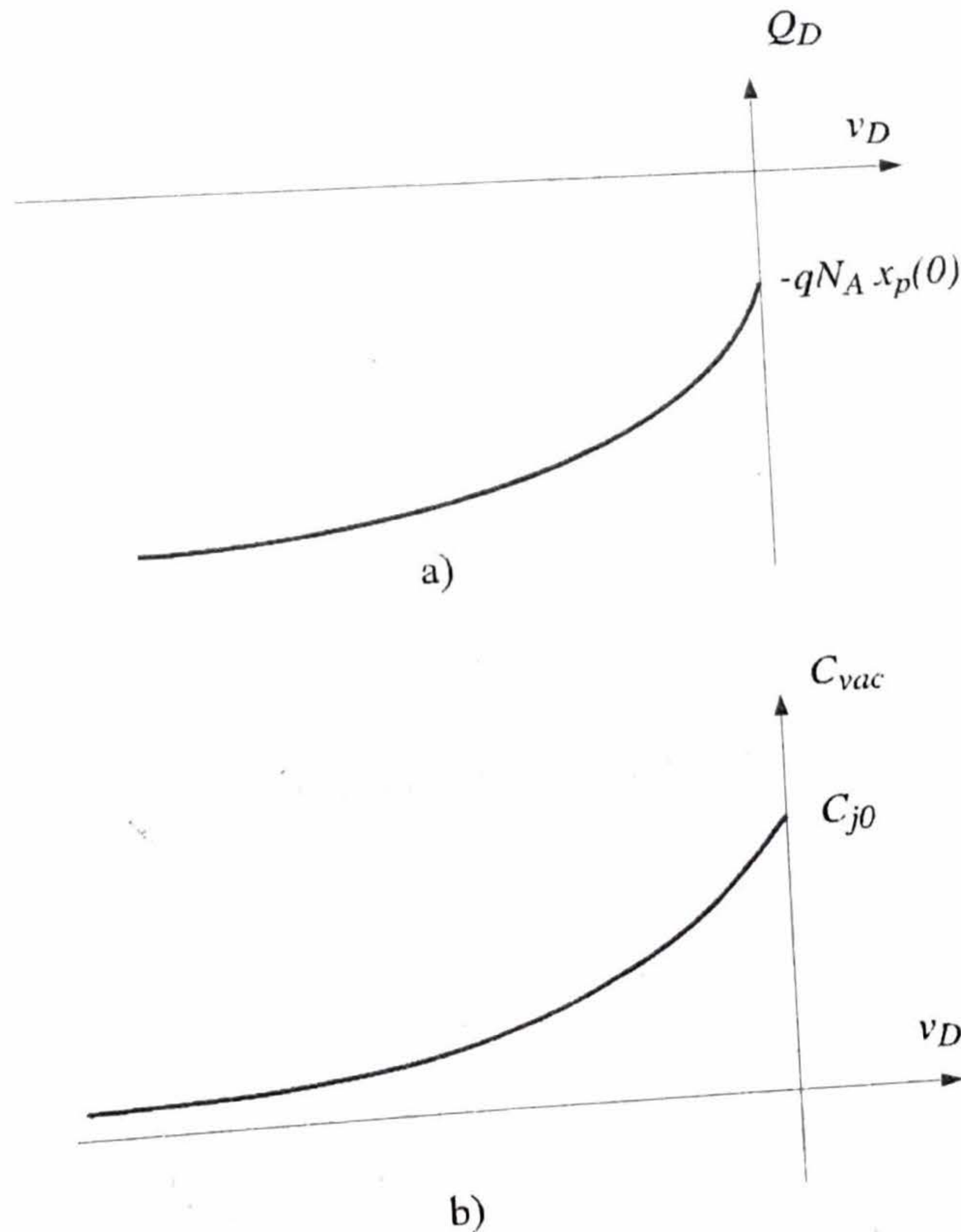


Figura 3.15: a) Carga del diodo versus la tensión sobre el diodo; b) capacidad del diodo en inversa.

Conforme se aplica una tensión inversa $V_D < 0$, la longitud de la zona de vaciamiento se incrementa, y la capacidad se reduce. Esta capacidad puede interpretarse de la siguiente manera, con la ayuda de

la Fig. 3.16. Para una cierta tensión inversa, V_D' , se produce una densidad de cargas como la que se ilustra en la Fig. 3.16-a; a ambos lados de la juntura existen cargas netas $Q_p' = -qN_a x_p'$ y $Q_n' = qN_d x_n'$. Si se varía la tensión inversa a un valor V_D'' , como se ilustra en la Fig. 3.16-b, las cargas netas estarán dadas por: $Q_p'' = -qN_a x_p''$ y $Q_n'' = qN_d x_n''$. Si se toma la diferencia entre ambos casos, se puede ver que ante una variación de tensión:

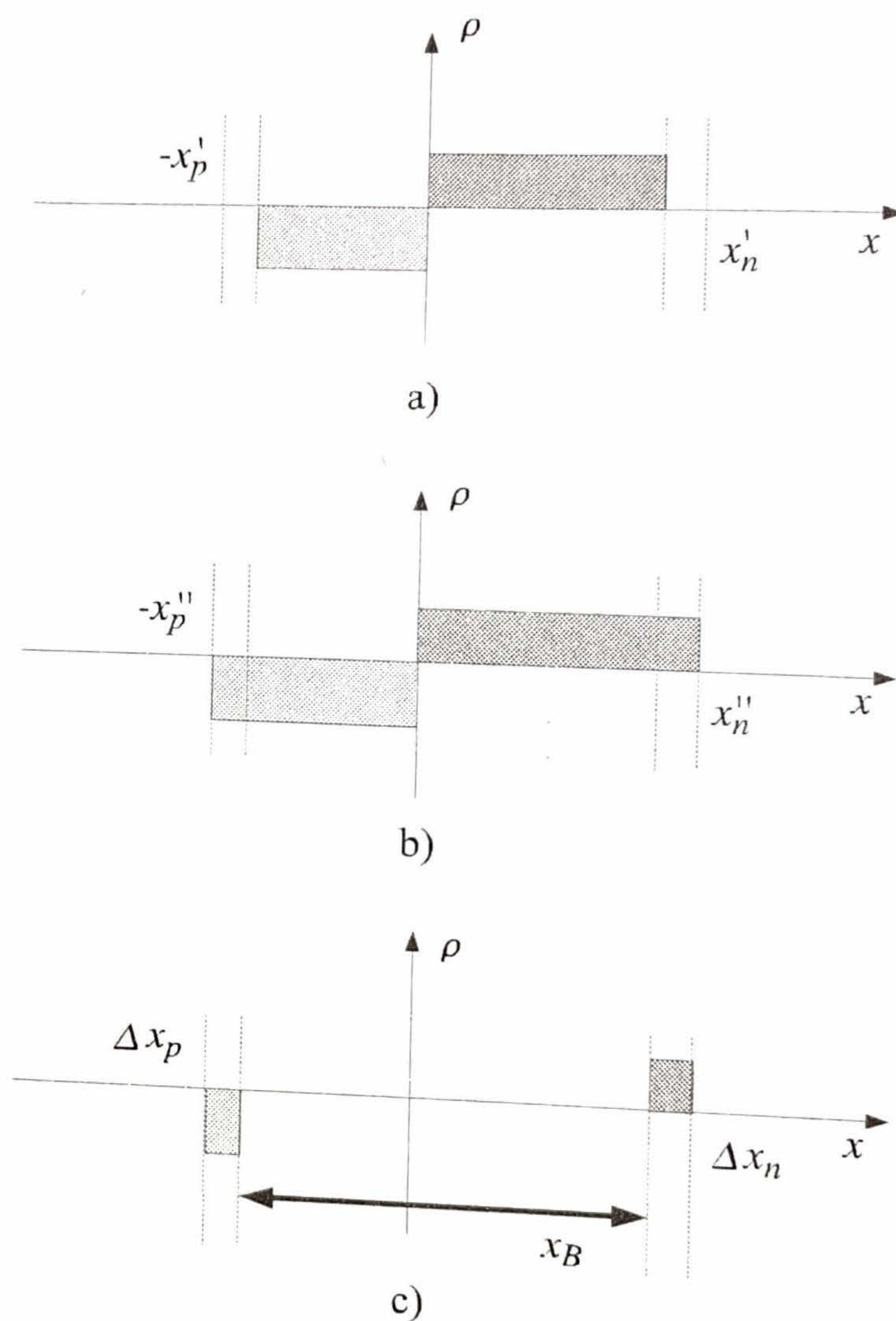


Figura 3.16: a) Densidad de carga para una tensión V_D' ; b) densidad de carga para una tensión V_D'' ; c) diferencia de densidades de carga.

$$\Delta V_D = V_D'' - V_D' \tag{3.101}$$

se produce una variación de carga

$$\Delta Q_p = Q_p'' - Q_p' = -qN_a(x_p'' - x_p') \tag{3.102}$$

del lado P, y una variación de carga

$$\Delta Q_n = Q_n'' - Q_n' = qN_d(x_n'' - x_n') \tag{3.103}$$

del lado N, tal como se muestra en la Fig. 3.16-c. Estas variaciones de carga se producen a una distancia $x_B(V_D)$, donde el dieléctrico intermedio es Silicio, cuya constante dieléctrica es ϵ_{Si} , lo cual

equivale a una capacidad incremental de valor

$$C_{vac} \approx \frac{\Delta Q_p}{\Delta V_D} = -\frac{\Delta Q_n}{\Delta V_D} = \frac{\epsilon_{Si}}{x_B(V_D)} \quad (3.104)$$

Ejemplo 14 Calcule la capacidad de vaciamiento para la juntura del Ejemplo 12, considerando un área de $10\mu m \times 10\mu m$, y una tensión aplicada $V_D = -5V$.

Los valores hallados para la zona de vaciamiento indican $x_p(0) = 0,885\mu m$, con lo cual:

$$C_{j0} = \frac{1,6 \times 10^{-19} \times 10^{15} \times 88,5 \times 10^{-6}}{2 \times 660 \times 10^{-3}} = 10,73 nF/cm^2 \quad (3.105)$$

Para una tensión inversa de $V_D = -5V$,

$$C_{j0} = 10,73 \frac{nF}{cm^2} \times \frac{1}{\sqrt{1 - \left(\frac{-5}{0,66}\right)}} = 10,73 \frac{nF}{cm^2} \times 0,3415 = 3,66 \frac{nF}{cm^2} \quad (3.106)$$

Considerando el área de la juntura, la capacidad total resulta:

$$C = 3,66 \frac{nF}{cm^2} \times (10 \times 10^{-4} cm)^2 = 3,66 fF \quad (3.107)$$

En un dispositivo real, (3.97) toma la forma más general:

$$C_{vac} = C_{j0} \frac{1}{\left(1 - \frac{V_D}{\phi_B}\right)^m} \quad (3.108)$$

donde $m = 1/2$ corresponde al caso de una juntura uniformemente dopada (como el caso descrito), y $m = 1/3$ al de una juntura con un gradiente lineal de dopado. En cualquier otro caso, el coeficiente m debe ser obtenido mediante mediciones.

3.5.2. Capacidad en directa

Para pequeños valores positivos de la tensión del diodo, el efecto de la capacidad de vaciamiento (3.97) sigue siendo notable. Para tensiones mayores a $\phi_B/2$, esta capacidad suele tomarse como constante o despreciarse, debido a la aparición de otra capacidad –llamada de difusión– que es de mayor valor y aumenta en forma más pronunciada, con la tensión directa. Es por ello que, para el modo de conducción directa, la capacidad de vaciamiento se supondrá constante con el siguiente valor:

$$C_{vac} = C_{j0} \frac{1}{\sqrt{1 - \frac{\phi_B}{2\phi_B}}} = \sqrt{2} C_{j0} \quad (3.109)$$

Como ya ha sido explicado, la aplicación de un potencial en directa causa la circulación de una corriente directa de difusión. Tal como se ha visto en la sección anterior, en esta situación se forman perfiles de carga en la zona P y en la zona N, que son los causantes de la corriente de difusión. Esta carga acumulada varía, de acuerdo al potencial aplicado, y por ende, esta variación incremental de carga puede ser interpretada como una capacidad incremental. De hecho, si se quiere reducir la circulación de corriente por el diodo a cero, primero debe eliminarse esta carga, tal como si se tratase de un capacitor. Para hallar esta capacidad, primero debe hallarse la expresión de la carga total.

La carga en lado P. se halla como:

$$\begin{aligned}
 Q_p &= -qA(W_p - x_p) \frac{(n_P(-x_p) + n_P(-W_p))}{2} \\
 &\approx -qAW_p \frac{n_P(1 + e^{v_D/U_T})}{2}
 \end{aligned} \tag{3.110}$$

donde se ha despreciado x_p con respecto a W_p .

La carga en el lado N se halla como:

$$\begin{aligned}
 Q_n &= qA(W_n - x_n) \frac{(p_N(x_n) + p_N(W_n))}{2} \\
 &\approx qAW_n \frac{p_N(1 + e^{v_D/U_T})}{2}
 \end{aligned} \tag{3.111}$$

donde se ha despreciado x_n con respecto a W_n .

La capacidad total es:

$$\begin{aligned}
 C_d &= \frac{-dQ_p}{dv_D} + \frac{dQ_n}{dv_D} \\
 &= qA(W_p n_P + W_n p_N) \frac{e^{v_D/U_T}}{2U_T}
 \end{aligned} \tag{3.112}$$

Suponiendo que, en directa, $e^{v_D/U_T} \approx e^{v_D/U_T} - 1$, (3.112) se puede escribir como:

$$C_d = \frac{qA(W_p n_P + W_n p_N)}{2I_s U_T} I_D \tag{3.113}$$

El factor

$$\tau_T = \frac{qA(W_p n_P + W_n p_N)}{2I_s} \tag{3.114}$$

es el tiempo de transición del diodo, y coincide con el tiempo que les toma a los portadores atravesar las zonas de difusión. Utilizando esta notación, resulta:

$$\boxed{C_d = \frac{\tau_T}{U_T} I_D} \tag{3.115}$$

Nótese de (3.113) ó (3.115), que el valor de esta capacidad es mayor cuanto mayor sea la corriente continua que circula por el diodo.

En el caso particular de una juntura asimétrica, donde un lado está mas dopado que el otro, por ejemplo si $N_D \gg N_A$,

$$C_d = \frac{qAW_p n_P}{2U_T} e^{(v_D/U_T)} \tag{3.116}$$

y

$$I_d = \frac{qAD_n n_P}{W_p} e^{(v_D/U_T)} \tag{3.117}$$

Combinando estas dos últimas ecuaciones, la capacidad se puede escribir como:

$$C_d = \frac{qAn_P D_n}{W_p U_T} e^{(v_D/U_T)} \times \frac{W_p^2}{2D_n} = \frac{I_D}{U_T} \tau_{Tn} = g_m \tau_{Tn} \tag{3.118}$$

donde

$$\tau_{T_n} = \frac{W_p}{\left(\frac{2D_n}{W_p}\right)} \quad (3.119)$$

tiene, claramente, unidades de tiempo (distancia/velocidad), y es el tiempo que le toma a los electrones difundir por la región P, es decir, atravesar la distancia W_p . En el caso opuesto ($N_A \gg N_D$), se puede definir

$$\tau_{T_p} = \frac{W_n}{\left(\frac{2D_p}{W_n}\right)} \quad (3.120)$$

que es el tiempo que le toma a los huecos difundir por la región N, es decir, atravesar la distancia W_n .

Ejemplo 15 La capacidad en directa para el diodo del Ejemplo 14, considerando $W_p = W_n = 5\mu m$, y una corriente $I_D = 1mA$, se calcula de la siguiente manera.

En primer lugar, la capacidad de vaciamiento es:

$$C_{vac} = \sqrt{2} \times 10,73 \frac{nF}{cm^2} \times (10 \times 10^{-4} cm)^2 = 15,15 fF \quad (3.121)$$

Dado que N_A se puede despreciar con respecto a N_D , se puede utilizar (3.119) para calcular el tiempo de tránsito:

$$\tau_{T_N} = \frac{5 \times 10^{-4} cm}{\left(\frac{2 \times 10,93 cm^2/s}{5 \times 10^{-4} cm}\right)} = \frac{5 \times 10^{-4} cm}{1,24 \times 10^5 cm/s} = 4ns \quad (3.122)$$

La capacidad de difusión es, entonces:

$$C_d = \frac{1 \times 10^{-3}}{25 \times 10^{-3}} \times 4 \times 10^{-9} = 160pf \quad (3.123)$$

La capacidad total en directa es la suma de la capacidad de vaciamiento y la de difusión, pero como se puede apreciar de los valores resultantes, la capacidad de difusión domina apreciablemente la capacidad directa.

3.6. Mecanismos de ruptura inversa

Cuando se aumenta lo suficiente la tensión inversa en una juntura, se produce la ruptura o derrumbe ("breakdown") de la misma, lo cual origina una circulación elevada de corriente. Los mecanismos de ruptura son tres [8]: inestabilidad térmica, avalancha y efecto túnel. Estos mecanismos son reversibles, es decir que, si se mantiene la corriente controlada (de manera de no provocar daños), luego de retirar la tensión, el dispositivo continúa operando normalmente. La inestabilidad térmica es importante en Germanio, donde $E_g(Ge)$ es pequeño. En este caso, la disipación de potencia por el aumento de la corriente produce un incremento de la temperatura, que a su vez produce un incremento mayor de la corriente. Este efecto es regenerativo y puede provocar la destrucción del dispositivo, a menos que se limite la corriente con una resistencia externa. Los dos mecanismos restantes, efecto Túnel y Avalancha, son típicos en Silicio, y se discuten a continuación.

3.6.1. Efecto Túnel

El efecto túnel sucede cuando existe un alto campo eléctrico desarrollado sobre la juntura, mayor a $10^6 V/cm$. Campos eléctricos de esta magnitud solo son posibles cuando ambas regiones de la juntura están fuertemente dopadas. Cuando se alcanzan campos eléctricos de esta magnitud, empieza a circular una corriente significativa por efecto túnel entre bandas. Este efecto es dominante para tensiones menores a $4E_g/q = 4,48V$. Este efecto presenta un coeficiente de temperatura negativo. Es decir, cuando la temperatura aumenta, E_g disminuye y la corriente de ruptura se alcanza para menores valores de tensión. En el caso de avalancha, el coeficiente de temperatura es positivo.

3.6.2. Avalancha

El mecanismo de avalancha determina, en la práctica, la tensión máxima de operación de un diodo, la tensión máxima de colector –en el caso del transistor bipolar–, y la tensión máxima de drain –en el caso del transistor MOS.

En conducción inversa, los portadores minoritarios de ambas regiones de la juntura atraviesan la zona de vaciamiento, debido al campo eléctrico aplicado. Conforme atraviesan la zona de vaciamiento, sufren colisiones con la red. Aquí se debe notar que, siendo el camino libre promedio de $10nm$, a lo largo de una zona de vaciamiento típica de algunos micrómetros, se producen cientos a miles de colisiones en la trayectoria de un portador. Para campos eléctricos elevados, los portadores pueden adquirir suficiente energía como para ionizar átomos en la zona de vaciamiento, esto es, para desprender electrones de la banda de valencia. De esta manera, se producen portadores adicionales, que a su vez pueden también ionizar otros átomos, desprendiendo nuevos portadores. Este efecto presenta una característica suave en las cercanías de la tensión de ruptura V_{BR} , y predomina para tensiones mayores a $6E_g/q = 6,72V$. El aumento de la corriente, producto de este efecto, puede modelarse mediante la introducción de un factor de multiplicación M :

$$M \triangleq \frac{i}{i_o} \quad (3.124)$$

donde i_o es la corriente sin multiplicación de portadores. El factor M se modela como:

$$M = \frac{1}{1 - \left(\frac{v_D}{V_{BR}}\right)^m} \quad (3.125)$$

donde m es un valor entre 3 y 6, que depende del material.

La tensión de ruptura para una juntura asimétrica (una región mucho más dopada que la otra), se puede computar numéricamente, planteando la expresión analítica de las corrientes en la juntura, y hallando la tensión para la cual $M \rightarrow \infty$; su solución es [8]:

$$V_{BR} = \frac{x_B \times E_{max}}{2} = \epsilon_{Si} E_{max}^2 \frac{1}{2qN_B} \quad (3.126)$$

donde N_B es el dopaje de la región con menor concentración, y x_B la longitud de la zona de vaciamiento correspondiente.

El campo eléctrico máximo, para Silicio, está dado por:

$$E_{max} = \frac{4 \times 10^5}{1 - \frac{1}{3} \log_{10} \left(\frac{N_B}{10^{16}}\right)} [V/cm] \quad (3.127)$$

En general, para un material arbitrario, se utiliza la siguiente aproximación:

$$V_{BR} = 60 \left(\frac{E_g}{1,1} \right)^{3/2} \left(\frac{N_B}{10^{16}} \right)^{-3/4} \quad (3.128)$$

La Fig. 3.17 muestra las tensiones de ruptura para Si, Ge y GaAs, indicando el límite a partir del cual sucede el efecto túnel por los altos dopados.

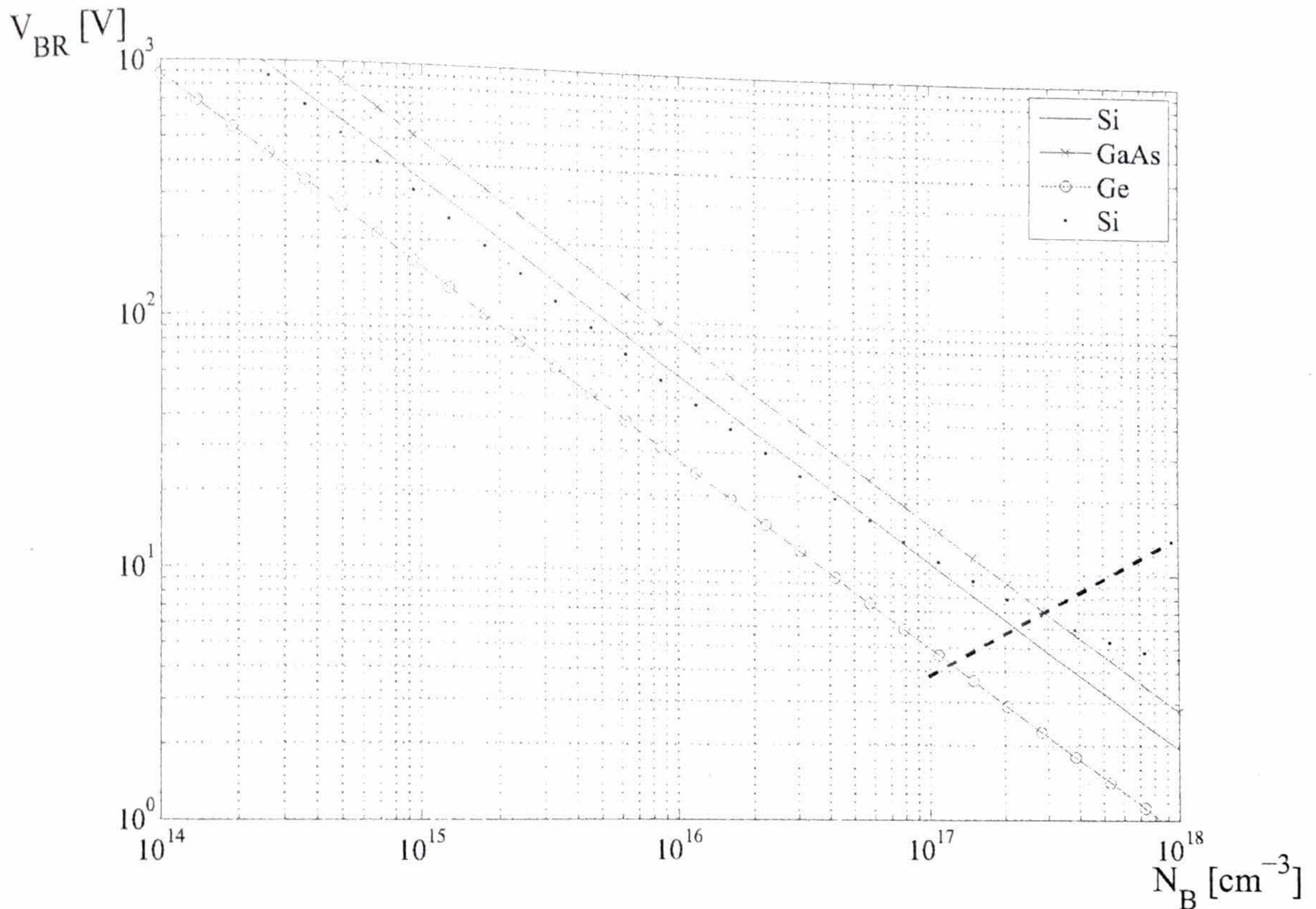


Figura 3.17: Tensiones de ruptura inversa en función del dopado, para distintos materiales. En trazo continuo se muestra la expresión aproximada (3.128) para Si; la sucesión de puntos corresponde a la expresión (3.127); la línea con círculos muestra la expresión aproximada (3.128) para Ge y la línea con cruces muestra la expresión aproximada (3.128) para GaAs. La línea punteada indica el final de la zona de avalancha y el comienzo de la zona de efecto Túnel.

Capítulo 4

Capacitor MOS

4.1. Descripción Cualitativa

Un capacitor Metal-Oxido-Semiconductor, o capacitor MOS, está compuesto por una compuerta o "gate", un dieléctrico, y un sustrato o "bulk" tal como se ilustra en la Fig. 4.1. En la actualidad, el gate se fabrica utilizando Silicio policristalino (esto es, Silicio cristalino por dominios, o regiones) también conocido como *polisilicio*. Para disminuir la resistencia del gate, es habitual doparlo fuertemente. El dieléctrico es una lámina extremadamente delgada de dióxido de Silicio (SiO₂), que se logra oxidando el Silicio mediante la exposición del mismo a una atmósfera con oxígeno, a una cierta temperatura (por ello, este óxido es también llamado óxido térmico). La excelente calidad de la interfaz entre el Silicio y el óxido térmico ha sido uno de los factores principales para el desarrollo de la tecnología MOS. El sustrato es simplemente la oblea o "wafer" de Silicio, que se halla por lo general levemente dopada.

Para el análisis cualitativo, se plantea el caso de un capacitor MOS con un gate de polisilicio fuertemente dopado, tipo N, y un sustrato ligeramente dopado, de tipo P. Si el gate y el sustrato se vinculan eléctricamente mediante una conexión metálica, se puede observar, de acuerdo a lo desarrollado en el Cap. 2, que el polisilicio tendrá un potencial positivo, $\phi_G = \phi_n^+ > 0$, con respecto a Silicio intrínseco, mientras que el sustrato tipo P, tendrá un potencial negativo, $\phi_B = \phi_p < 0$. Esto revela la presencia de cargas positivas en el gate y cargas negativas en el sustrato. Dado que el gate es Silicio dopado tipo N, las cargas positivas son producto de una zona de vaciamiento compuesta por iones positivos. Debido al alto dopado, la concentración de cargas se puede considerar como laminar, es decir, concentrada prácticamente por completo en la cara inferior del gate (en la interfaz con el SiO₂). La carga en el sustrato, que tiene un dopado tipo P, es negativa, lo cual revela la presencia de una zona de vaciamiento, compuesta por iones negativos – aunque en este caso, la longitud de la misma es finita. En este punto inicial del análisis, se ve que existe una tensión inicial entre el gate y el sustrato:

$$\phi_{GB} = \phi_n^+ - \phi_p > 0 \quad (4.1)$$

En esta situación, se dice que el capacitor MOS está en *vaciamiento*, debido a la presencia de una zona de vaciamiento en el sustrato. En este modo de funcionamiento, el capacitor es no lineal, debido a la dependencia de la región de vaciamiento con el potencial aplicado.

Si se aplica una tensión, de igual magnitud y signo opuesto a la existente, dada por (4.1), las zonas de vaciamiento desaparecen, y la densidad de carga a lo largo del dispositivo es nula. Esta condición se denomina de *Banda Plana* o "Flatband", y la tensión que la produce es la tensión de Banda Plana, que se define como:

$$V_{FB} \triangleq -(\phi_n^+ - \phi_p) \quad (4.2)$$

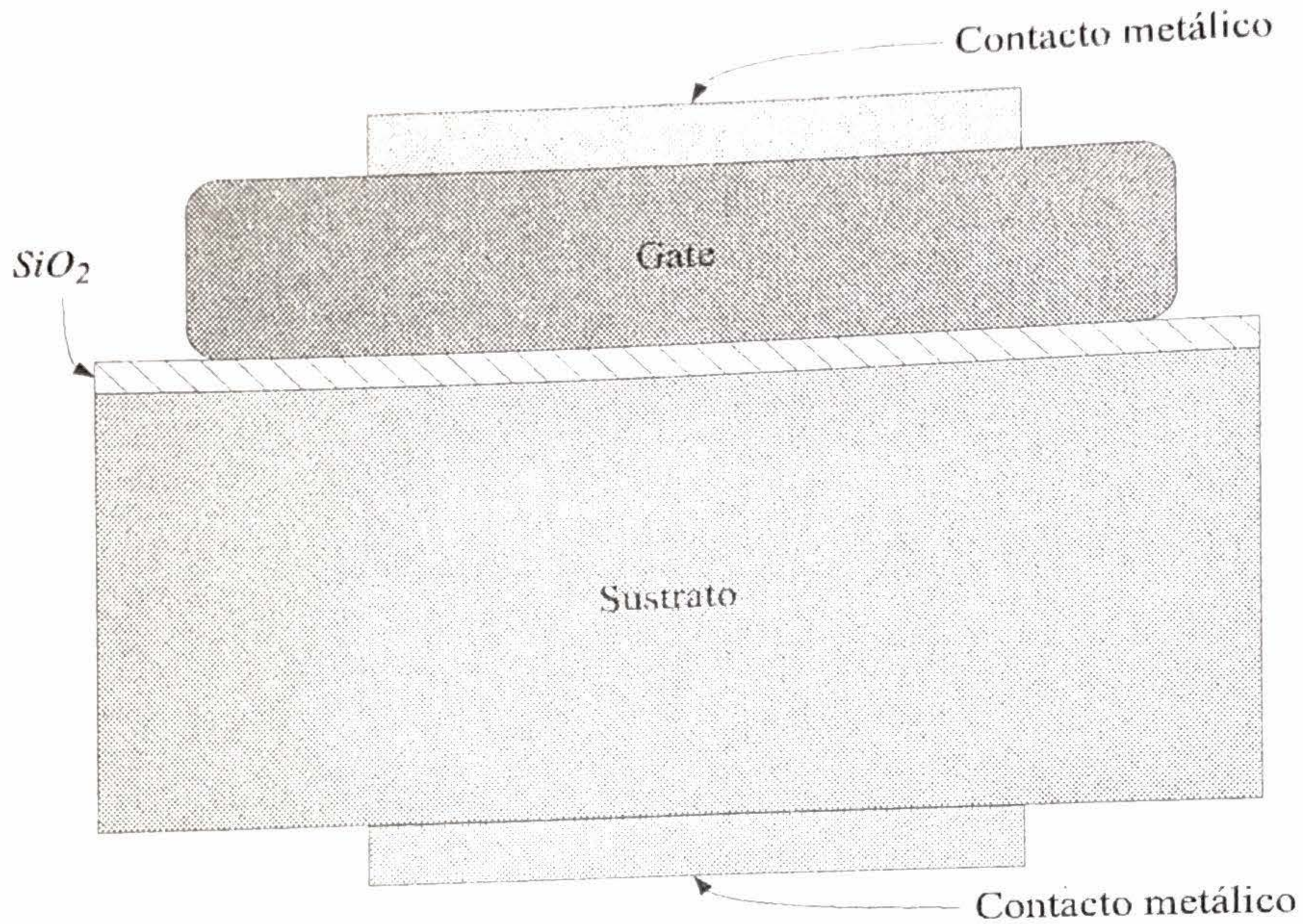


Figura 4.1: Esquema de un capacitor MOS.

Para tensiones menores a la tensión de Banda Plana, se acumulan electrones en el gate y huecos en el sustrato. Ambos portadores son provistos naturalmente por el Polisilicio del gate y por el Silicio dopado P del sustrato. Este modo de funcionamiento se denomina *acumulación*, y en él, el dispositivo se comporta como un capacitor lineal ideal de placas paralelas (conformadas por el gate y el sustrato).

Finalmente, si se aumenta el potencial del gate a partir de la tensión de Banda Plana, la zona de vaciamiento crece hasta que alcanza un valor máximo. La zona de vaciamiento máxima se alcanza para un valor de potencial de gate denominado *tensión de umbral* o "Threshold", V_T . Para potenciales mayores, se forma una lámina de portadores negativos debajo del dieléctrico. Este modo de funcionamiento se denomina *inversión*, dado que aparecen portadores de carga contraria a la esperada (en un material dopado tipo P se esperan portadores positivos). En este modo, el dispositivo se vuelve a comportar como un capacitor lineal ideal.

4.2. Electroestática del capacitor MOS

Para el análisis de la electroestática del capacitor MOS, se considerará una estructura como la de la Fig. 4.2, incluyendo los contactos metálicos. La ecuación de mallas indica que:

$$v_{GB} = \phi_{m,n+} + v_{ox} + v_{vac} + \phi_{p,m} \quad (4.3)$$

donde v_{GB} es la tensión aplicada, v_{ox} la tensión a lo largo del óxido, y v_{vac} la tensión a lo largo de la zona de vaciamiento del sustrato.

Dado que el dispositivo es un capacitor, en estado estacionario, no habrá circulación de corriente continua, y el dispositivo se hallará en equilibrio térmico aún en presencia de potenciales externos aplicados.

4.2.1. Potencial de Banda Plana

El primer caso a analizar es el de Banda Plana, caracterizado por la ausencia de cargas, tanto en el gate, como en el sustrato, es decir:

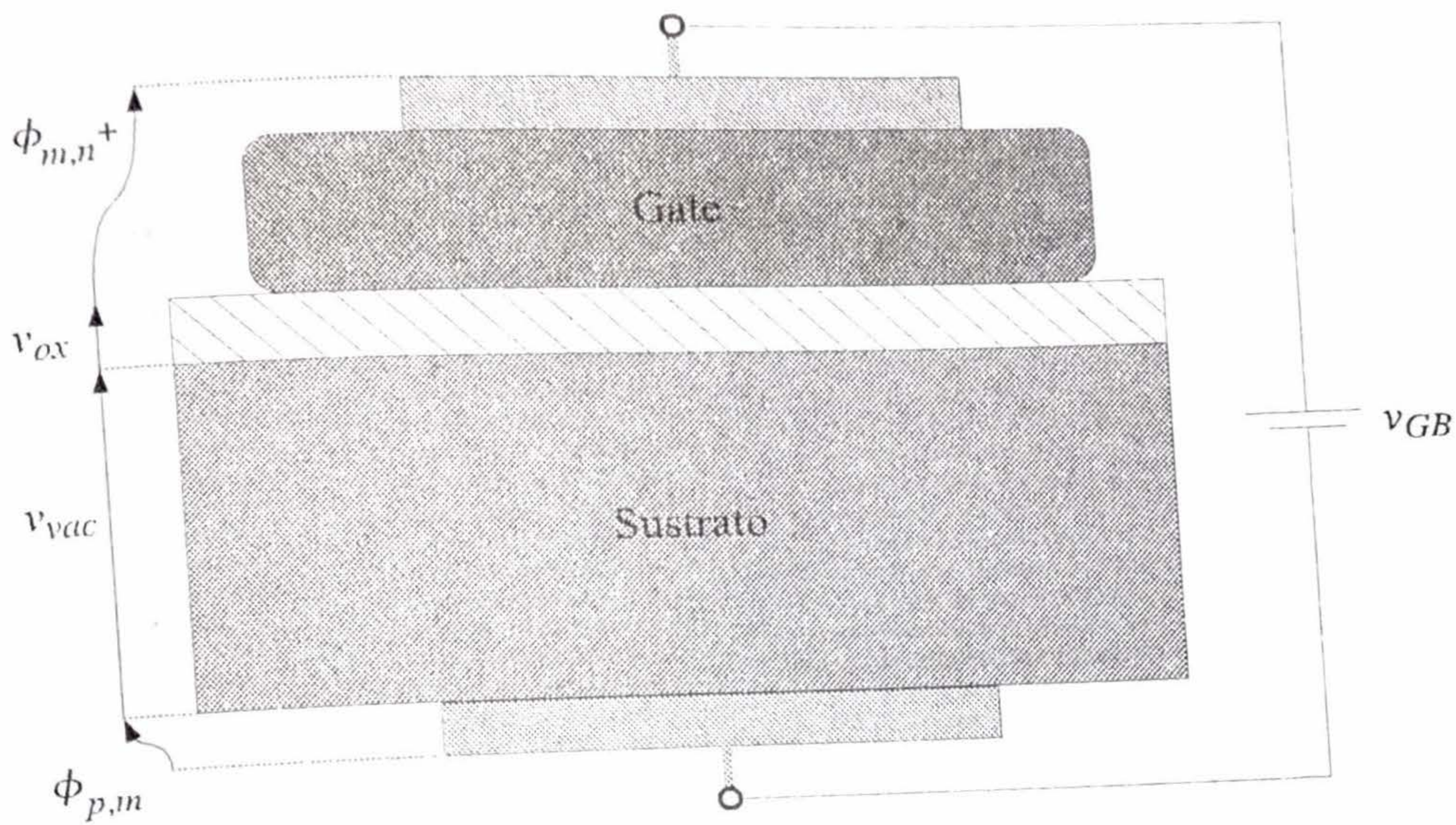


Figura 4.2: Esquema de un capacitor MOS con tensión aplicada.

$$Q_G(V_{FB}) = Q_B(V_{FB}) = 0 \quad (4.4)$$

Debido a (4.4), el campo eléctrico es nulo, y se puede calcular la tensión v_{GB} necesaria, a partir de (4.3), considerando:

$$v_{ox} = 0; \quad v_{vac} = 0 \quad (4.5)$$

En este caso, (4.3) se reduce a:

$$v_{GB} = \phi_{m,n^+} + \phi_{p,m} = -(\phi_{n^+} - \phi_p) = V_{FB} \quad (4.6)$$

de donde se deduce (4.2). Es sencillo ver que (4.3) también puede escribirse de la siguiente manera:

$$v_{GB} = v_{ox} + v_{vac} + V_{FB} \quad (4.7)$$

La Fig. 4.3 ilustra la densidad de carga, el campo eléctrico y el potencial en esta situación.

4.2.2. Acumulación

Si se disminuye la tensión a partir de la tensión de Banda Plana, es decir, para:

$$v_{GB} < V_{FB} \quad (4.8)$$

el capacitor MOS se encuentra en el modo de acumulación. La concentración de carga consiste en dos impulsos, de igual magnitud y signo contrario, a ambos lados del óxido, como se muestra en la Fig. 4.4-a. El campo eléctrico en el óxido se puede hallar a partir de:

$$\epsilon_{ox} \int_{-t_{ox}}^x dE = \int_{-t_{ox}}^x \rho d\xi \quad (4.9)$$

donde $-t_{ox} \leq x < 0$, $\epsilon_{ox} = 3,9\epsilon_0$ y $\epsilon_0 = 8,854 \times 10^{-14}$ F/cm. Si se reconoce el término derecho de (4.9), considerando $x < 0$, como la carga en el gate, (4.9) puede escribirse como:

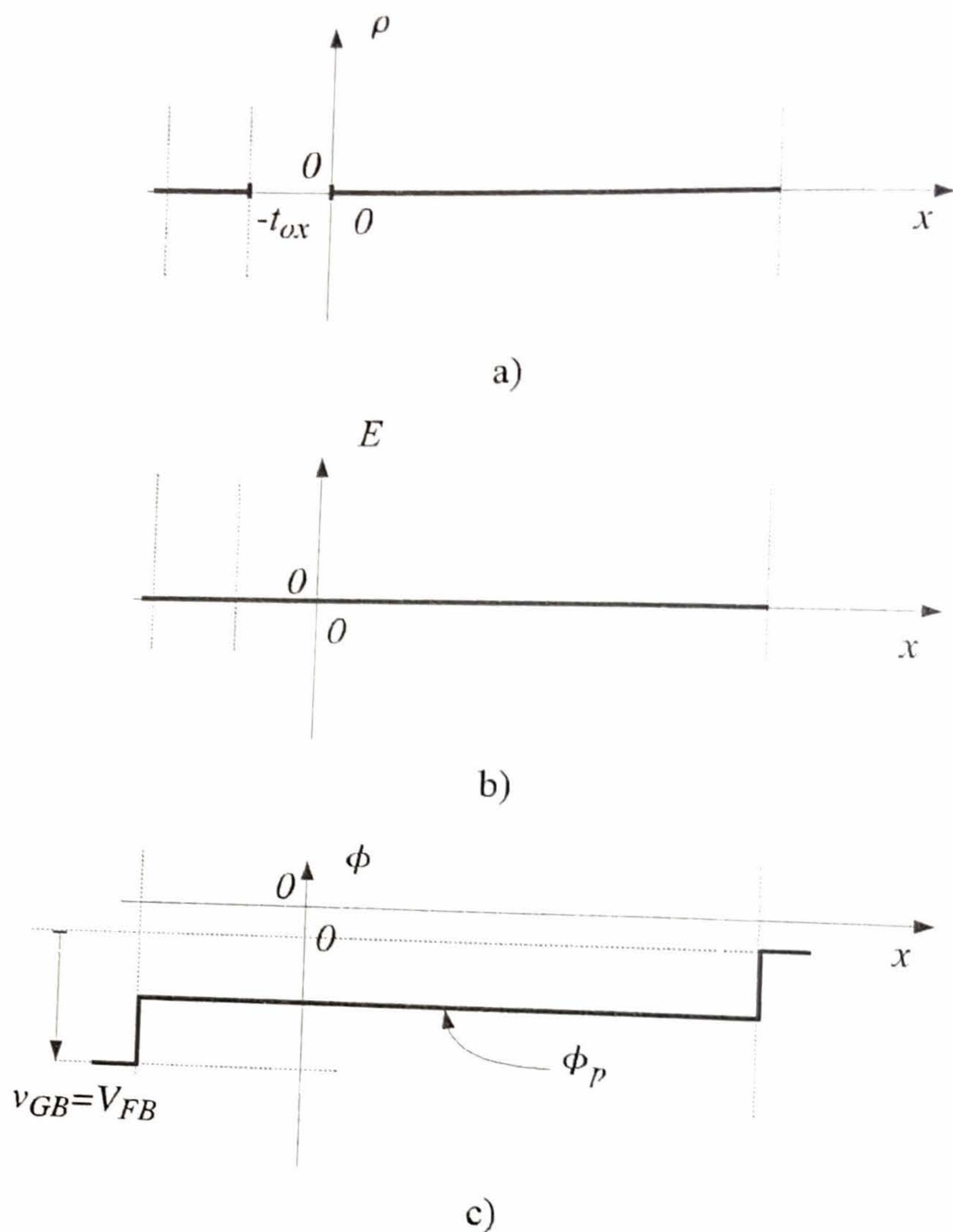


Figura 4.3: Electroestática de un capacitor MOS en Banda Plana: a) densidad de carga; b) campo eléctrico; c) potencial.

$$\varepsilon_{ox} E(x) = Q_G \quad (4.10)$$

La Ec. (4.10) indica que el campo eléctrico es constante a lo largo del óxido, con lo cual se está en presencia de un capacitor lineal (Fig. 4.4-b). Para valores mayores de x , es decir, $x \geq 0$, la carga neta en el intervalo $(-t_{ox}, x)$ es nula; por lo tanto, el término derecho de (4.9) se anula y el campo eléctrico también. El potencial en un punto x , dentro del óxido, puede hallarse a partir de:

$$\phi(x) = \phi(-t_{ox}) + \int_{-t_{ox}}^x -E(\xi) d\xi = \phi(-t_{ox}) + \int_{-t_{ox}}^x -\frac{Q_G}{\varepsilon_{ox}} d\xi \quad (4.11)$$

De esta última ecuación resulta:

$$\phi(x) = \phi(-t_{ox}) - \frac{Q_G}{\varepsilon_{ox}}(x + t_{ox}) \quad (4.12)$$

para $-t_{ox} \leq x \leq 0$. El potencial a lo largo del óxido v_{ox} , tal como se indica en la Fig. 4.2, está definido como:

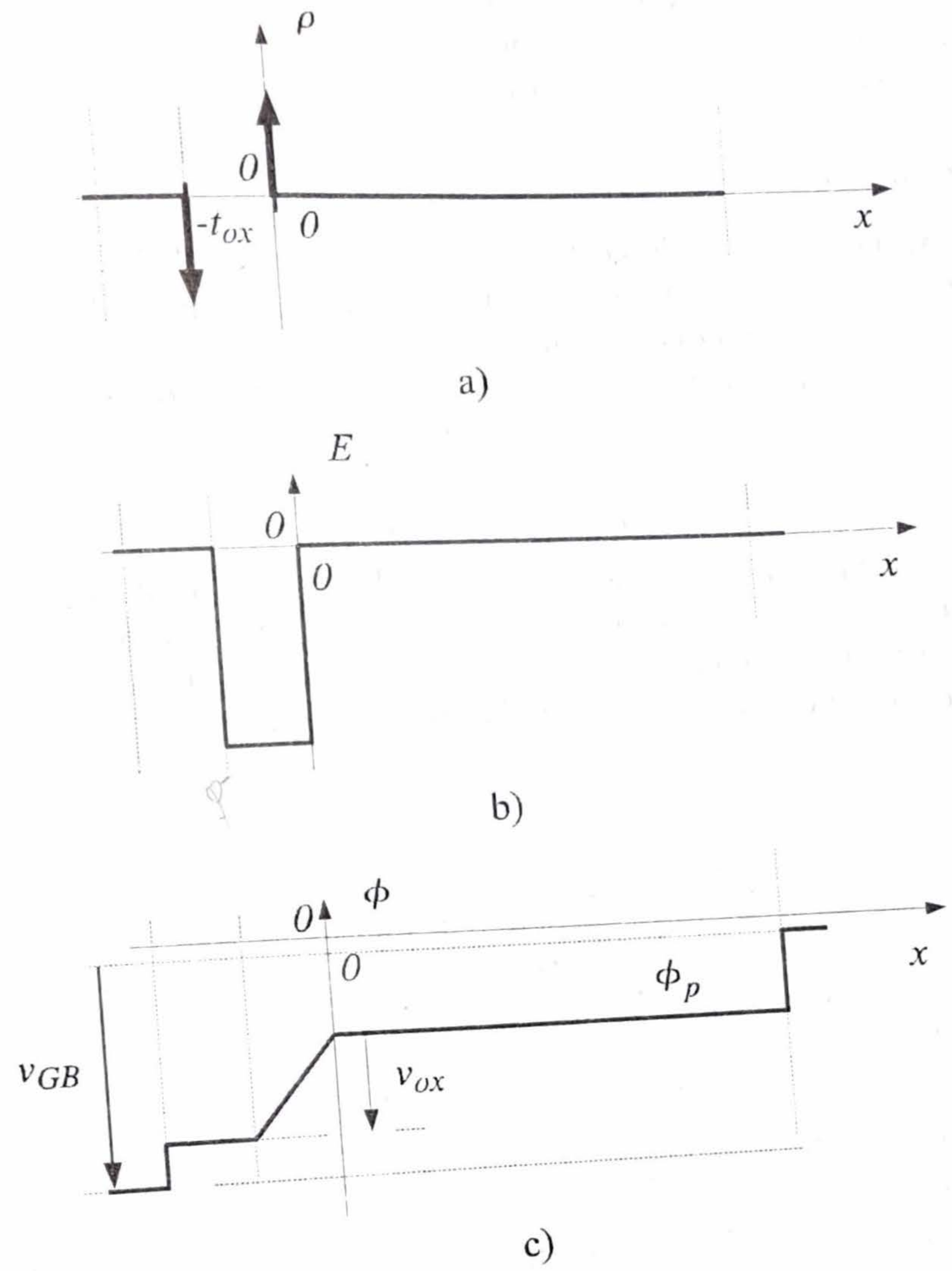


Figura 4.4: Electroestática de un capacitor MOS en Acumulación: a) densidad de carga; b) campo eléctrico; c) potencial.

$$v_{ox} \triangleq \phi(-t_{ox}) - \phi(0) \tag{4.13}$$

y resulta igual a:

$$v_{ox} = \frac{Q_G}{\epsilon_{ox}} t_{ox} = \frac{Q_G}{C_{ox}} \tag{4.14}$$

donde

$$C_{ox} \triangleq \frac{\epsilon_{ox}}{t_{ox}} \tag{4.15}$$

es la capacidad del óxido por unidad de área.

Dado que $v_{vac} = 0$, por la ausencia de carga de vaciamiento en el sustrato, entonces, de (4.7) y (4.12), resulta:

$$Q_G = C_{ox}(v_{GB} - V_{FB}) \tag{4.16}$$

La Ec. (4.16) pone de manifiesto la naturaleza lineal del capacitor formado por un par de placas paralelas, a una distancia t_{ox} , y separadas por un dieléctrico de constante dieléctrica ϵ_{ox} . El potencial a lo largo de la estructura se muestra en la Fig. 4.4-c.

4.2.3. Vaciamiento

El modo de vaciamiento sucede cuando $v_{GB} > V_{FB}$. En este caso, la tensión positiva del gate indica la existencia de una carga positiva, constituida por una lámina de iones positivos, que repele los huecos debajo del óxido, produciendo una zona de vaciamiento, tal como se ilustra en la Fig. 4.5. Para el análisis, se supondrá en el gate la existencia de una carga Q_G . En el sustrato, la concentración de cargas está dada por:

$$\rho_B = q(-N_a - n(x) + p(x)) \quad (4.17)$$

En la zona de vaciamiento, $0 \leq x \leq x_{vac}$, es razonable suponer que no hay huecos, es decir $p(x) = 0$, y los portadores negativos, al ser minoritarios, pueden despreciarse¹, es decir $n(x) = 0$. De hecho, se supondrá que, mientras la tensión aplicada sea menor que la tensión de umbral, V_T , no habrá electrones en la zona de vaciamiento del sustrato. Luego, la carga en el sustrato (4.17) está dada por:

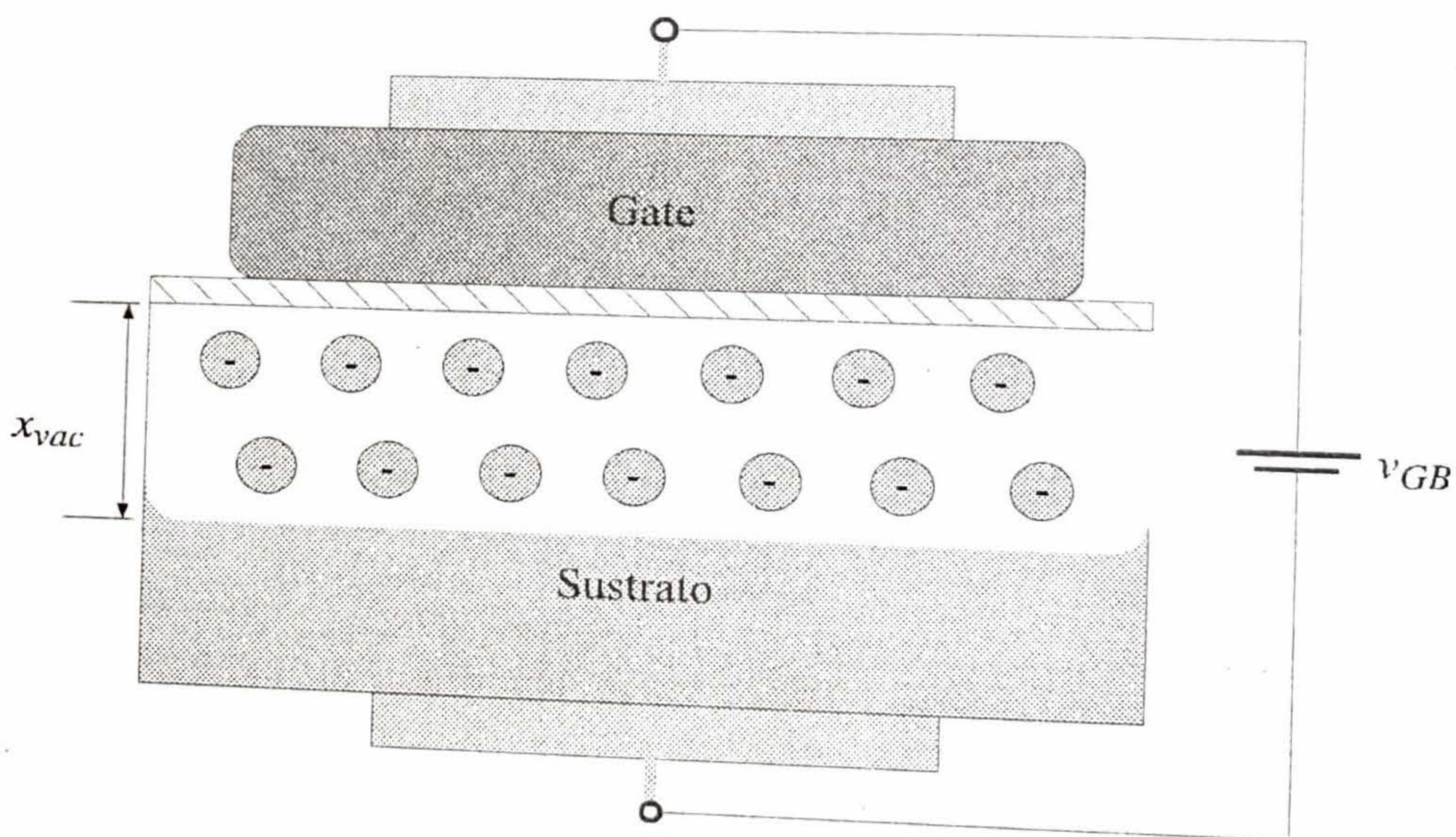


Figura 4.5: Esquema de un capacitor MOS funcionando en el modo de vaciamiento.

$$\rho_B = -qN_a \quad (4.18)$$

para $0 \leq x \leq x_{vac}$, y la conservación de cargas requiere:

$$Q_G = -Q_B = qN_a x_{vac} \quad (4.19)$$

La densidad de carga a lo largo del dispositivo se muestra en la Fig. 4.6-a. El campo eléctrico, en el interior del óxido, se puede hallar a partir de la integración de la carga, como en el caso anterior:

$$E(x) = \frac{Q_G}{\epsilon_{ox}} \quad (4.20)$$

¹En vaciamiento, hay una cantidad despreciable de electrones en el sustrato, que se vuelve significativa en un entorno de la tensión de umbral, y para valores mayores a la tensión de umbral domina el comportamiento.

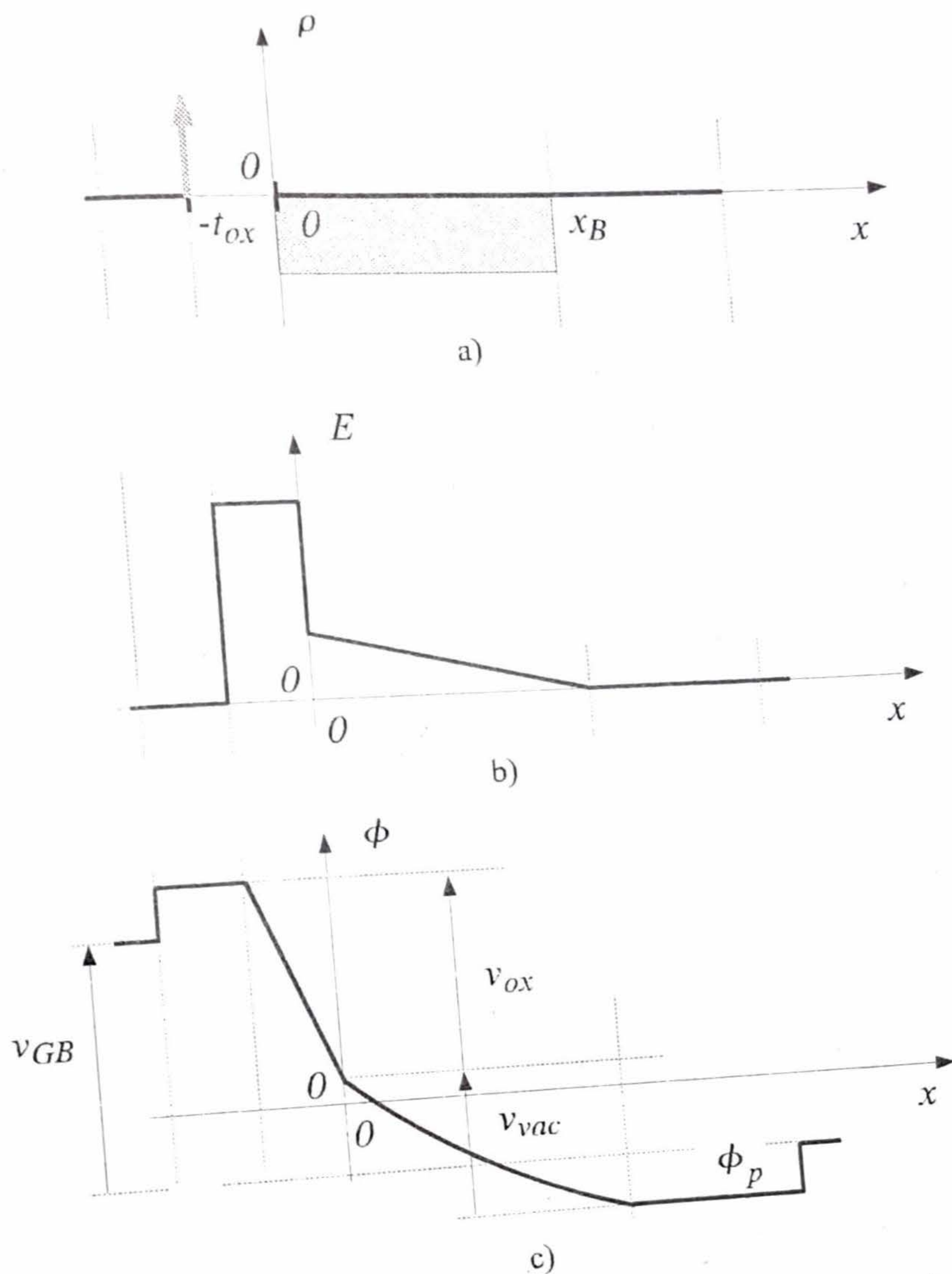


Figura 4.6: Electroestática de un capacitor MOS en vaciamiento: a) densidad de carga; b) campo eléctrico; c) potencial.

para $-t_{ox} \leq x \leq 0$. El potencial en un punto x , dentro del óxido, es:

$$\phi(x) = \phi(-t_{ox}) - \frac{Q_G}{\epsilon_{ox}}(x + t_{ox}) \quad (4.21)$$

para $-t_{ox} \leq x \leq 0$. El potencial total, desarrollado a lo largo del óxido, es:

$$v_{ox} = \phi(-t_{ox}) - \phi(0) = \frac{1}{C_{ox}} Q_G \quad (4.22)$$

En la zona de vaciamiento, el análisis prosigue de la siguiente manera. En la interfaz, el campo eléctrico debe verificar:

$$\epsilon_{ox} E_{ox} = \epsilon_{Si} E_{Si} \quad (4.23)$$

donde $\epsilon_{ox} = 3,9\epsilon_0$ y $\epsilon_{Si} = 11,8\epsilon_0$. Nótese que, en virtud de (4.23), el campo eléctrico en el Silicio es aproximadamente tres veces menor que en el óxido:

$$E_{Si} = \frac{\epsilon_{ox}}{\epsilon_{Si}} E_{ox} \approx \frac{1}{3} E_{ox} = \frac{Q_G}{\epsilon_{Si}} \quad (4.26)$$

A partir de la interfaz, el campo eléctrico satisface:

$$\begin{aligned} \epsilon_{Si} \int_0^x dE &= \int_0^x \rho d\xi \\ \epsilon_{Si} (E(x) - E(0^+)) &= -qN_a x \end{aligned} \quad (4.27)$$

donde $0 \leq x \leq x_{vac}$. La expresión del campo eléctrico es, entonces:

$$E(x) = \frac{Q_G}{\epsilon_{Si}} - \frac{qN_a}{\epsilon_{Si}} x \quad (4.28)$$

Al llegar al final de la zona de vaciamiento, para $x = x_{vac}$, la carga total encerrada es nula, y el campo eléctrico debe hacerse nulo, por lo tanto, se verifica:

$$Q_G = qN_a x_{vac} \quad (4.29)$$

La Fig. 4.6-b ilustra el campo eléctrico.

El potencial en un punto x , dentro de la zona de vaciamiento, es:

$$\begin{aligned} \phi(x) &= \phi(0) - \int_0^x \left(\frac{Q_G}{\epsilon_{Si}} - \frac{qN_a}{\epsilon_{Si}} \xi \right) d\xi \\ &= \phi(0) - \frac{Q_G}{\epsilon_{Si}} x + \frac{qN_a}{2\epsilon_{Si}} x^2 \end{aligned} \quad (4.30)$$

para $0 \leq x \leq x_{vac}$. El potencial total desarrollado a lo largo de la zona de vaciamiento, v_{vac} , tal como se indica en la Fig. 4.2, se define como:

$$\begin{aligned} v_{vac} &\triangleq \phi(0) - \phi(x_{vac}) \\ &= \frac{Q_G}{\epsilon_{Si}} x_{vac} - \frac{qN_a}{2\epsilon_{Si}} x_{vac}^2 \end{aligned} \quad (4.31)$$

Reemplazando Q_G en función de x_{vac} , de (4.29), el potencial v_{vac} puede escribirse también como:

$$v_{vac} = \frac{qN_a}{\epsilon_{Si}} x_{vac}^2 - \frac{qN_a}{2\epsilon_{Si}} x_{vac}^2 = \frac{qN_a}{2\epsilon_{Si}} x_{vac}^2 \quad (4.32)$$

La Fig. 4.6-c muestra el potencial desarrollado a lo largo del dispositivo.

El valor de x_{vac} , puede hallarse a partir de la ecuación de mallas (4.7) y de las ecuaciones de los potenciales individuales, (4.22) y (4.30), que se repiten por comodidad:

$$\begin{cases} v_{GB} = v_{ox} + v_{vac} + V_{FB} \\ v_{ox} = \frac{qN_a x_{vac}}{C_{ox}}; \quad v_{vac} = \frac{qN_a}{2\epsilon_{Si}} x_{vac}^2 \end{cases} \quad (4.33)$$

A partir de estas ecuaciones, puede plantearse la ecuación cuadrática para despejar x_{vac} :

$$\frac{qN_a}{2\epsilon_{Si}} x_{vac}^2 + \frac{qN_a}{C_{ox}} x_{vac} + (V_{FB} - v_{GB}) = 0 \quad (4.34)$$

La solución general a esta ecuación es:

$$x_{vac} = \frac{-\frac{qN_a}{C_{ox}} \pm \sqrt{\left(\frac{qN_a}{C_{ox}}\right)^2 - 2\frac{qN_a}{\epsilon_{Si}}(V_{FB} - v_{GB})}}{\frac{qN_a}{\epsilon_{Si}}} \quad (4.33)$$

que puede reducirse a:

$$\begin{aligned} x_{vac} &= -\frac{\epsilon_{Si}}{C_{ox}} \pm \sqrt{\left(\frac{\epsilon_{Si}}{C_{ox}}\right)^2 + 2\frac{\epsilon_{Si}(v_{GB} - V_{FB})}{qN_a}} \\ &= \frac{\epsilon_{Si}}{C_{ox}} \left(-1 \mp \sqrt{1 + 2\frac{C_{ox}^2(v_{GB} - V_{FB})}{qN_a\epsilon_{Si}}} \right) \end{aligned} \quad (4.34)$$

La solución para x_{vac} es la raíz positiva de (4.34), es decir:

$$x_{vac} = \frac{\epsilon_{Si}}{C_{ox}} \left(-1 + \sqrt{1 + 2\frac{C_{ox}^2(v_{GB} - V_{FB})}{qN_a\epsilon_{Si}}} \right) \quad (4.35)$$

Es directo verificar que, para la condición de Banda Plana, es decir $v_{GB} = V_{FB}$, esta ecuación produce una carga nula. Reemplazando esta última ecuación en (4.19) se puede obtener la carga en el gate y en el sustrato:

$$Q_B = -Q_G = -qN_a \frac{\epsilon_{Si}}{C_{ox}} \left(-1 + \sqrt{1 + 2\frac{C_{ox}^2(v_{GB} - V_{FB})}{qN_a\epsilon_{Si}}} \right) \quad (4.36)$$

De ser necesario, reemplazando (4.36) en (4.30), se obtiene la tensión del sustrato, y reemplazando (4.36) en (4.22), se obtiene la tensión del óxido, para un valor dado de v_{GB} .

4.2.4. Inversión

Conforme se aumenta v_{GB} , la zona de vaciamiento, de longitud x_{vac} , continúa aumentando, aunque cada vez más lentamente, debido a la dependencia de la raíz cuadrada de v_{GB} , como se puede ver de (4.35). Cuando la tensión v_{GB} alcanza la tensión de umbral, es decir $v_{GB} = V_T$, la zona de vaciamiento alcanza un valor máximo x_{vacM} , y la zona de vaciamiento no varía más². Si se aumenta la tensión v_{GB} por encima de V_T , se introduce más carga positiva en el gate, que debe ser balanceada por la carga debajo del óxido. Dado que la zona de vaciamiento no varía más, la única posibilidad para balancear la carga es que aparezcan portadores negativos por debajo del óxido. Estos portadores son inducidos por la tensión positiva del gate, y son provistos por la fuente de alimentación. En esta situación, se dice que el sustrato tipo P se ha invertido, dado que ahora posee portadores negativos. Estos portadores negativos forman un canal, como se ilustra en la Fig. 4.7.

El análisis que se hará en esta sección es simplificado y se basa en las siguientes suposiciones:

1. La carga en el sustrato, en el modo de vaciamiento, es solo debido a iones positivos

²En realidad, continúa aumentando, pero en magnitudes extremadamente pequeñas. Para observar esta situación, es necesario utilizar la expresión completa de la carga (4.17) tal como se hace en [8].

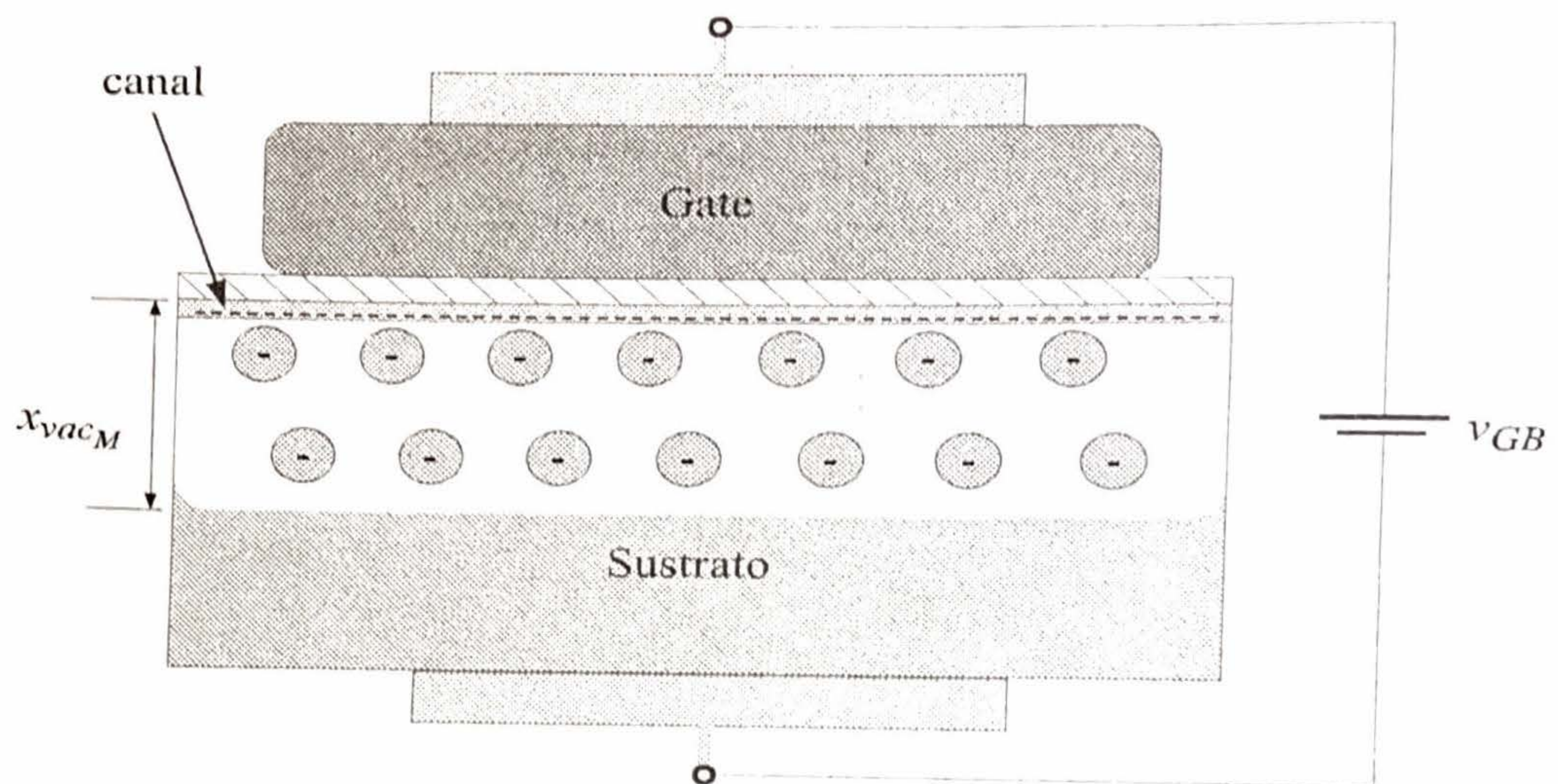


Figura 4.7: Esquema de un capacitor MOS funcionando en el modo de inversión.

2. El sustrato se invierte para $v_{GB} = V_T$, y solo aparece carga de inversión para $v_{GB} > V_T$

Resulta conveniente comenzar el análisis a partir del punto de inversión. Esta situación se produce, por definición, cuando el potencial en el canal, también llamado potencial superficial $\phi_s = \phi(0)$, se vuelve igual a $-\phi_p$, es decir:

$$\phi_s = -\phi_p \quad (4.37)$$

La observación de los portadores en el canal haría pensar que el sustrato es tipo N, con un dopado N_D de la misma magnitud que N_A , es decir, con la misma concentración de impurezas, pero de distinto tipo. Es por ello que esta condición se denomina inversión. En este caso, el potencial desarrollado sobre la zona de vaciamiento es:

$$v_{vac} = \phi(0) - \phi(x_{vac_M}) = -2\phi_p \quad (4.38)$$

Reemplazando (4.38) en la expresión de la zona de vaciamiento (4.30), se puede despejar el valor de x_{vac_M} :

$$x_{vac_M} = \sqrt{\frac{2(-2\phi_p)\epsilon_{Si}}{qN_a}} \quad (4.39)$$

El lector puede verificar que (4.39) y (4.35), son consistentes³, es decir:

$$\begin{aligned} x_{vac_M} &= \sqrt{\frac{2(-2\phi_p)\epsilon_{Si}}{qN_a}} \\ &= \frac{\epsilon_{Si}}{C_{ox}} \left(-1 + \sqrt{1 + 2\frac{C_{ox}^2(V_T - V_{FB})}{qN_a\epsilon_{Si}}} \right) \end{aligned} \quad (4.40)$$

³Ayuda: se sugiere reemplazar la expresión de V_T , (4.43), en (4.35), y completar cuadrados, para expresar el término bajo la raíz cuadrada como $\left(\frac{\epsilon_{Si}}{C_{ox}} + \sqrt{\frac{2\epsilon_{Si}(-2\phi_p)}{qN_a}} \right)^2$

A partir del valor de la zona de vaciamiento, se puede hallar el potencial desarrollado sobre el óxido:

$$v_{ox} = \frac{qN_a}{C_{ox}} \sqrt{\frac{2(-2\phi_p)\epsilon_{Si}}{qN_a}} = \frac{1}{C_{ox}} \sqrt{2qN_a\epsilon_{Si}(-2\phi_p)} \quad (4.41)$$

Reemplazando este valor en la ecuación de mallas, y teniendo en cuenta que $v_{GB} = V_T$, es decir:

$$V_T = v_{ox} + v_{vac} + V_{FB} \quad (4.42)$$

se obtiene la expresión de la tensión de umbral:

$$V_T = \frac{1}{C_{ox}} \sqrt{2qN_a\epsilon_{Si}(-2\phi_p)} - 2\phi_p + V_{FB} \quad (4.43)$$

Para valores de tensión mayores que la tensión de umbral, es decir $v_{GB} > V_T$, surge una carga pelicular o canal, constituida por electrones, que se notará Q_i . Por el balance de cargas, la carga en el gate deberá balancear las cargas en el sustrato, tanto la de inversión como la de vaciamiento, cumpliendo:

$$Q_G(v_{GB}) = -Q_i(v_{GB}) - Q_{vac}(v_{GB}) \quad (4.44)$$

La carga en la zona de vaciamiento se obtiene como:

$$Q_{vac}(v_{GB}) = Q_{vac_M} = -qN_ax_{vac_M} = -\sqrt{2qN_a\epsilon_{Si}(-2\phi_p)} \quad (4.45)$$

Como la carga de vaciamiento, Q_{vac} , permanece constante para tensiones mayores a la de umbral, entonces $Q_{vac}(v_{GB}) = Q_{vac}(V_T)$, para $v_{GB} \geq V_T$. Por otro lado, en el punto de inversión, $Q_{vac}(V_T) = -Q_G(V_T)$. Teniendo en cuenta estas consideraciones, la carga de inversión se puede escribir, a partir de (4.44), de la siguiente manera:

$$\begin{aligned} Q_i(v_{GB}) &= -Q_G(v_{GB}) - Q_{vac}(v_{GB}) = -Q_G(v_{GB}) - Q_{vac}(V_T) \\ &= -Q_G(v_{GB}) + Q_G(V_T) \end{aligned} \quad (4.46)$$

Si se considera que:

$$Q_G(v_{GB}) = C_{ox}v_{ox} = C_{ox}(v_{GB} - v_{vac} - V_{FB}) \quad (4.47)$$

$$Q_G(V_T) = C_{ox}v_{ox} = C_{ox}(V_T - v_{vac} - V_{FB}) \quad (4.48)$$

y se reemplazan (4.47)-(4.48), en (4.46), la carga de inversión puede escribirse como:

$$Q_i(v_{GB}) = -C_{ox}(v_{GB} - V_T) \quad (4.49)$$

De acuerdo a (4.44), la carga total vista desde el gate es igual a:

$$Q_G(v_{GB}) = C_{ox}(v_{GB} - V_T) + \sqrt{2qN_a\epsilon_{Si}(-2\phi_p)} \quad (4.50)$$

La Fig. 4.8 muestra la distribución de cargas, el campo eléctrico, y el potencial para esta situación.

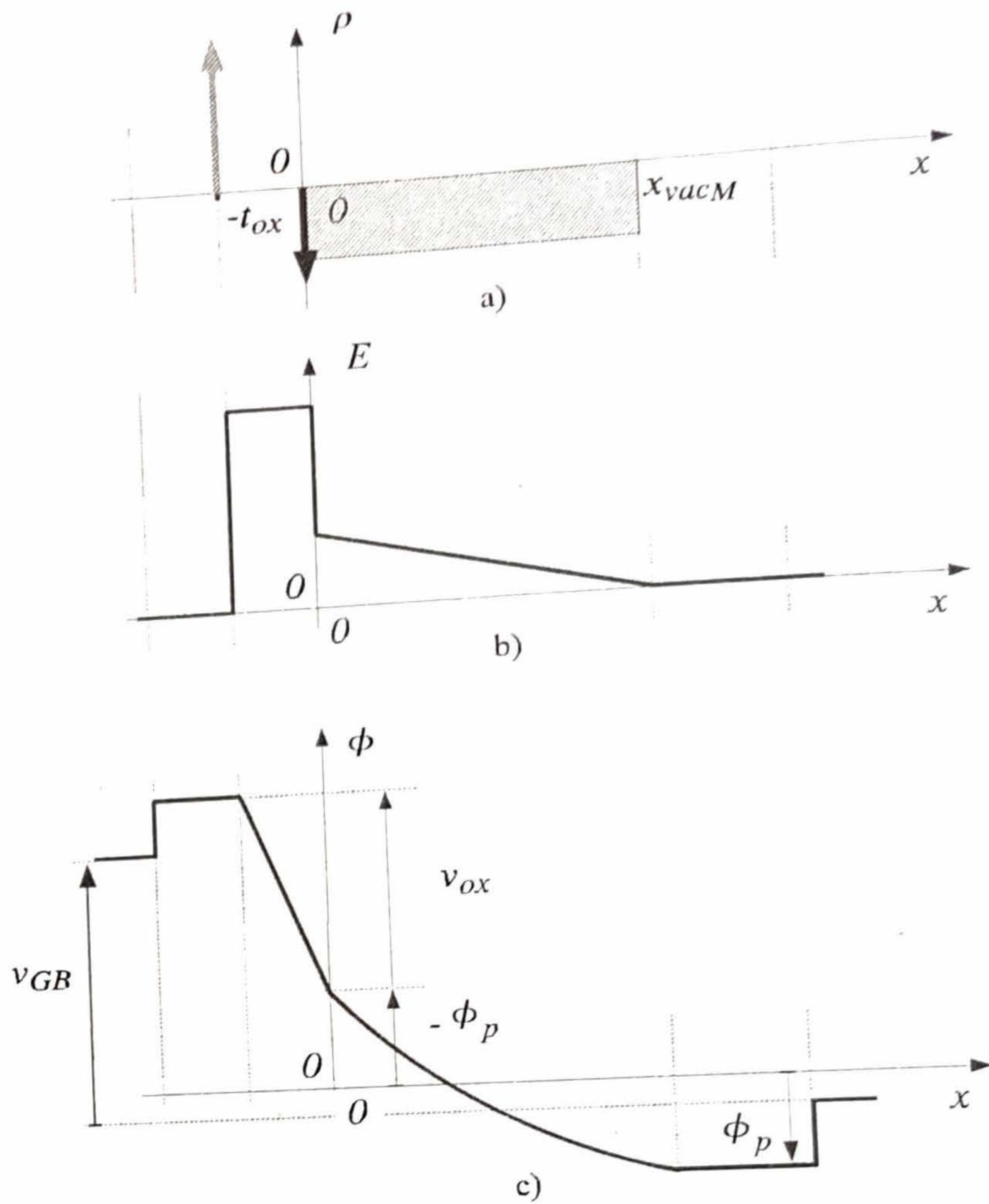


Figura 4.8: Electroestática de un capacitor MOS en inversión: a) densidad de carga; b) campo eléctrico; c) potencial.

Ejemplo 16 Un capacitor MOS se realiza con un dopado de gate $N_d = 10^{19}$, un dopado de sustrato $N_a = 10^{17}$, y un espesor de óxido $t_{ox} = 14\text{nm}$. Calcule V_{FB} , V_T , y la longitud máxima de la zona de vaciamiento.

La tensión de Flatband es: $V_{FB} = -(540\text{mV} - (-420\text{mV})) = -960\text{mV}$. La longitud máxima de la zona de vaciamiento es:

$$x_{vacM} = \sqrt{\frac{2 \times 0,96 \times 11,8 \times 8,854 \times 10^{-14}}{1,6 \times 10^{-19} \times 10^{17}}} = 0,1\mu\text{m} \quad (4.51)$$

La capacidad del óxido por unidad de área es: $C_{ox} = 0,245\mu\text{F}/\text{cm}^2$. La caída de tensión en el óxido es:

$$v_{ox} = \frac{\sqrt{2 \times 1,6 \times 10^{-19} \times 10^{17} \times 11,8 \times 8,854 \times 10^{-14} \times 0,96}}{2,45 \times 10^{-7}} = 0,684\text{V} \quad (4.52)$$

La caída de tensión en el sustrato es $v_{vacM} = 0,84\text{V}$, con lo cual, la tensión de umbral es:

$$V_T = 0,684\text{V} + 0,84\text{V} - 0,96\text{V} = 0,564\text{V} \quad (4.53)$$

4.3. Modelo de AC

Habiendo obtenido la expresión de la carga, tanto en el óxido como en el sustrato, ambas en función de v_{GB} , ahora es posible hallar un modelo que describa la capacidad incremental del dispositivo. La expresión de la carga es no lineal con respecto a v_{GB} y puede resumirse, para cada una de las zonas de operación, de la siguiente forma:

$$Q_G = \begin{cases} C_{ox}(v_{GB} - V_{FB}) & \text{en acumulación} \\ \frac{qN_a\epsilon_{Si}}{C_{ox}} \left(-1 + \sqrt{1 + 2 \frac{C_{ox}^2 (V_{GB} - V_{FB})}{qN_a\epsilon_{Si}}} \right) & \text{en vaciamiento} \\ C_{ox}(v_{GB} - V_T) + \sqrt{2qN_a\epsilon_{Si}(-2\phi_p)} & \text{en inversión} \end{cases} \quad (4.54)$$

Un diagrama de la carga, en función de la tensión v_{GB} , se muestra en la Fig. 4.9. Para cada región de operación, dependiendo del valor de la tensión (de DC) V_{GB} , la capacidad incremental se puede hallar como:

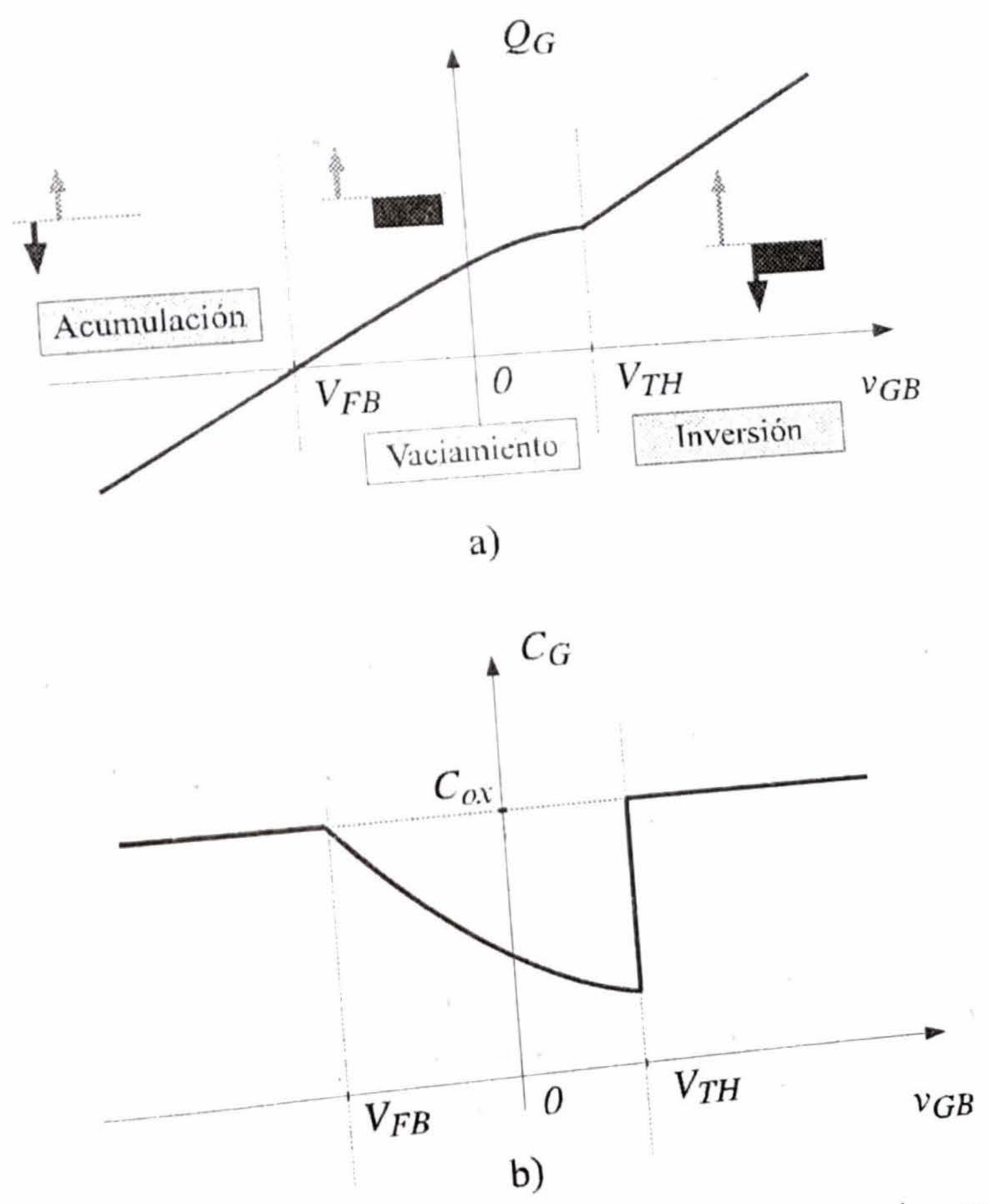


Figura 4.9: Capacitor MOS con gate dopado N sobre un sustrato tipo P: a) carga acumulada en el gate vs. la tensión v_{GB} ; b) capacidad incremental en función de v_{GB} .

$$C = \left. \frac{dQ(v_{GB})}{dv_{GB}} \right|_{V_{GB}} \quad (4.55)$$

A continuación se describe cada caso.

Acumulación Si el dispositivo está funcionando en la región de acumulación, $v_{GB} < V_{BF}$, la capacidad está dada por:

$$\boxed{C = C_{ox}} \quad (4.56)$$

Es decir, en este modo de funcionamiento, desde el gate se aprecia solamente la capacidad producta del óxido.

Vaciamiento En la región de vaciamiento, la derivada de Q_G , con respecto a v_{GB} , resulta:

$$C = C_{ox} \frac{1}{\sqrt{1 + 2 \frac{C_{ox}^2 (V_{GB} - V_{FB})}{qN_a \epsilon_{Si}}}} \quad (4.57)$$

Esta ecuación puede escribirse de la siguiente manera:

$$C = \frac{\epsilon_{Si}}{\frac{\epsilon_{Si}}{C_{ox}} \left(-1 + \sqrt{1 + 2 \frac{C_{ox}^2 (V_{GB} - V_{FB})}{qN_a \epsilon_{Si}}} \right) + \frac{\epsilon_{Si}}{C_{ox}}} \quad (4.58)$$

que, luego de reconocer el término $x_{vac}(v_{GB})$, puede re-escribirse como:

$$C = \frac{\epsilon_{Si}}{x_{vac}(v_{GB}) + \frac{\epsilon_{Si}}{C_{ox}}} = \frac{\left(\frac{\epsilon_{Si}}{x_{vac}(v_{GB})} \right) C_{ox}}{C_{ox} + \left(\frac{\epsilon_{Si}}{x_{vac}(v_{GB})} \right)} \quad (4.59)$$

Si se nota de (3.100), que:

$$\frac{\epsilon_{Si}}{x_{vac}(v_{GB})} = C_{vac} \quad (4.60)$$

es la capacidad de la zona de vaciamiento, la capacidad total, en esta región de funcionamiento, está dada por:

$$\boxed{C = \frac{C_{vac} C_{ox}}{C_{vac} + C_{ox}} < C_{ox}} \quad (4.61)$$

y es producto de dos capacidades en serie: la del óxido y la de la zona de vaciamiento. En consecuencia, en esta región, la capacidad se reduce con respecto al valor de capacidad, C_{ox} , en acumulación. Si se tiene en cuenta que $Q_{vac} = -qN_a x_{vac}$, y se utiliza la expresión para la zona de vaciamiento (4.30), la carga de vaciamiento se puede expresar como:

$$Q_{vac} = -qN_a \sqrt{\frac{2\epsilon_{Si} v_{vac}}{qN_a}} = \sqrt{2\epsilon_{Si} qN_a} \sqrt{v_{vac}} \quad (4.62)$$

Luego, la capacidad de vaciamiento también puede expresarse como:

$$C_{vac} = - \frac{dQ_{vac}}{dv_{vac}} \Big|_{V_{vac}} = \sqrt{2\epsilon_{Si} qN_a} \frac{1}{2\sqrt{v_{vac}}} \quad (4.63)$$

Inversión En la región de inversión, $v_{GB} \geq V_T$, la capacidad vuelve al valor de acumulación:

$$C = C_{ox} \quad (4.64)$$

En este modo de operación, la carga del sustrato no varía más con el potencial aplicado, por lo cual la capacidad resultante es sólo la del óxido.
 La Fig. 4.10 ilustra resultados experimentales⁴ de un capacitor MOS, en un proceso con $t_{ox} = 70 \text{ \AA}$.

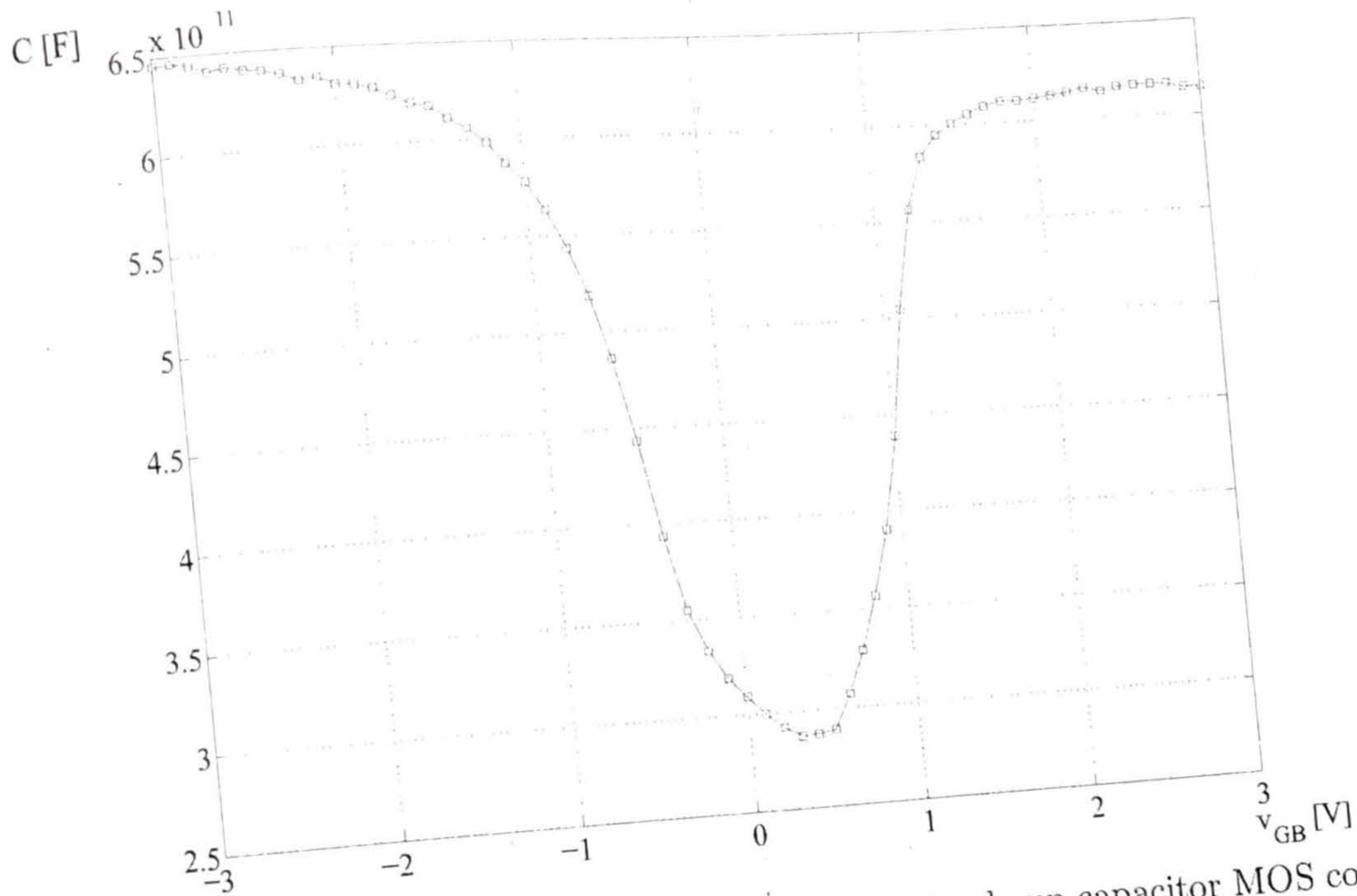


Figura 4.10: Capacidad incremental, medida experimentalmente, de un capacitor MOS con un espesor $t_{ox} = 70 \text{ \AA}$.

Ejemplo 17 Calcule la capacidad, en cada modo de operación, de un capacitor con los parámetros del Ejemplo 16, y un área de $10 \mu\text{m} \times 10 \mu\text{m}$.
 Tanto en inversión, como en acumulación, la capacidad es:

$$C = C_{ox} \times A = 0,245 \times (0,01)^2 = 24,5 \text{ pF} \quad (4.65)$$

En vaciamiento, el valor mínimo de la capacidad de la zona de vaciamiento es:

$$C = C_{vac} \times A = \frac{11,8 \times 8,854 \times 10^{-14}}{1 \times 10^5} \times (0,01)^2 = 9,97 \text{ pF} \quad (4.66)$$

con lo cual, en esta región de operación, la capacidad total se reduce de $C = A \times C_{ox}$ hasta el valor:

$$C = \frac{24,5 \times 10^{-12} \times 9,97 \times 10^{-12}}{24,5 \times 10^{-12} + 9,97 \times 10^{-12}} = 7,1 \text{ pF} \quad (4.67)$$

4.4. Otras configuraciones

A continuación, se describen, por completitud, las diferentes configuraciones posibles teniendo en cuenta el dopado del sustrato y del gate.

⁴Datos cortesía del Dr. Felix Palumbo, CNEA

4.4.1. Capacitor MOS sobre un sustrato P y gate P

En el caso de un capacitor MOS sobre un sustrato ligeramente dopado P, con el gate fuertemente dopado P, la tensión de Banda Plana será:

$$V_{FB} = -\phi_{GB} = -(\phi_{p+} - \phi_p) = \phi_p - \phi_{p+} \quad (4.68)$$

Por ejemplo, si los dopados del gate y el sustrato, son, respectivamente, $N_G = 10^{19}$ y $N_B = 10^{13}$, entonces, $\phi_{p+} = -540mV$, $\phi_p = -180mV$, y $V_{FB} = 360mV$. Dado que, salvo casos particulares, el dopado del gate es mayor al del sustrato, en este caso, la tensión de Banda Plana siempre será positiva.

La ecuación para la tensión de umbral mantiene el mismo valor que (4.43):

$$V_T = \frac{1}{C_{ox}} \sqrt{2qN_a \epsilon_{Si} (-2\phi_p)} - 2\phi_p + V_{FB} \quad (4.69)$$

Nótese que como $-2\phi_p$ y el valor de la raíz son positivos, resulta $V_{FB} < V_T$, y las zonas de funcionamiento tienen la misma secuencia, es decir: a) si $v_{GB} \leq V_{FB}$, el dispositivo está en acumulación; b) si $V_{FB} \leq v_{GB} \leq V_T$, el dispositivo está en vaciamiento; c) y si $V_T \leq v_{GB}$, el dispositivo está en inversión. En resumen, comparando este caso con el anterior, solo se experimenta una traslación de las tensiones de umbral y de Banda Plana hacia la derecha (valores más positivos). La situación es similar a la mostrada en la Fig. 4.9, con un corrimiento de V_{FB} y V_T .

4.4.2. Capacitor MOS sobre un sustrato N y gate N

En el caso de un capacitor MOS sobre un sustrato ligeramente dopado N, con el gate fuertemente dopado N, la tensión de Banda Plana será:

$$V_{FB} = -\phi_{GB} = -(\phi_{n+} - \phi_n) = \phi_n - \phi_{n+} \quad (4.70)$$

Por ejemplo, si los dopados del gate y el sustrato, son, respectivamente, $N_G = 10^{19}$ y $N_B = 10^{13}$, entonces, $\phi_{n+} = 540mV$, $\phi_n = 180mV$, y $V_{FB} = -360mV$. En este caso la tensión de Banda Plana siempre será negativa.

Nótese que, en este caso, al ser el sustrato tipo N, si se hace más positiva la tensión V_{GB} , a partir de la tensión de Banda Plana, se inducen cargas positivas en el gate, y negativas en el sustrato N. Debido a esto, para tensiones $v_{GB} \geq V_{FB}$ el dispositivo se halla en acumulación, de manera contraria a lo que sucede en el caso de un sustrato tipo P.

La expresión de la carga en el gate, para este modo de funcionamiento, permanece igual:

$$Q_G = C_{ox}(v_{GB} - V_{FB}) \quad (4.71)$$

con $v_{GB} \geq V_{FB}$.

La zona de vaciamiento se produce cuando $v_{GB} \leq V_{FB}$. Para ello, debe reducirse la tensión del gate, razón por la cual, la carga en el gate será negativa, formándose una zona de vaciamiento en el sustrato N, compuesta por iones positivos. El signo de las cargas será opuesto con respecto al caso del sustrato P:

$$Q_G = -Q_B = -qN_d x_{vac} \quad (4.72)$$

El potencial desarrollado sobre el óxido será negativo:

$$v_{ox} = \frac{1}{C_{ox}} Q_G = -\frac{qN_d x_{vac}}{C_{ox}} \quad (4.73)$$

al igual que el potencial desarrollado sobre el sustrato:

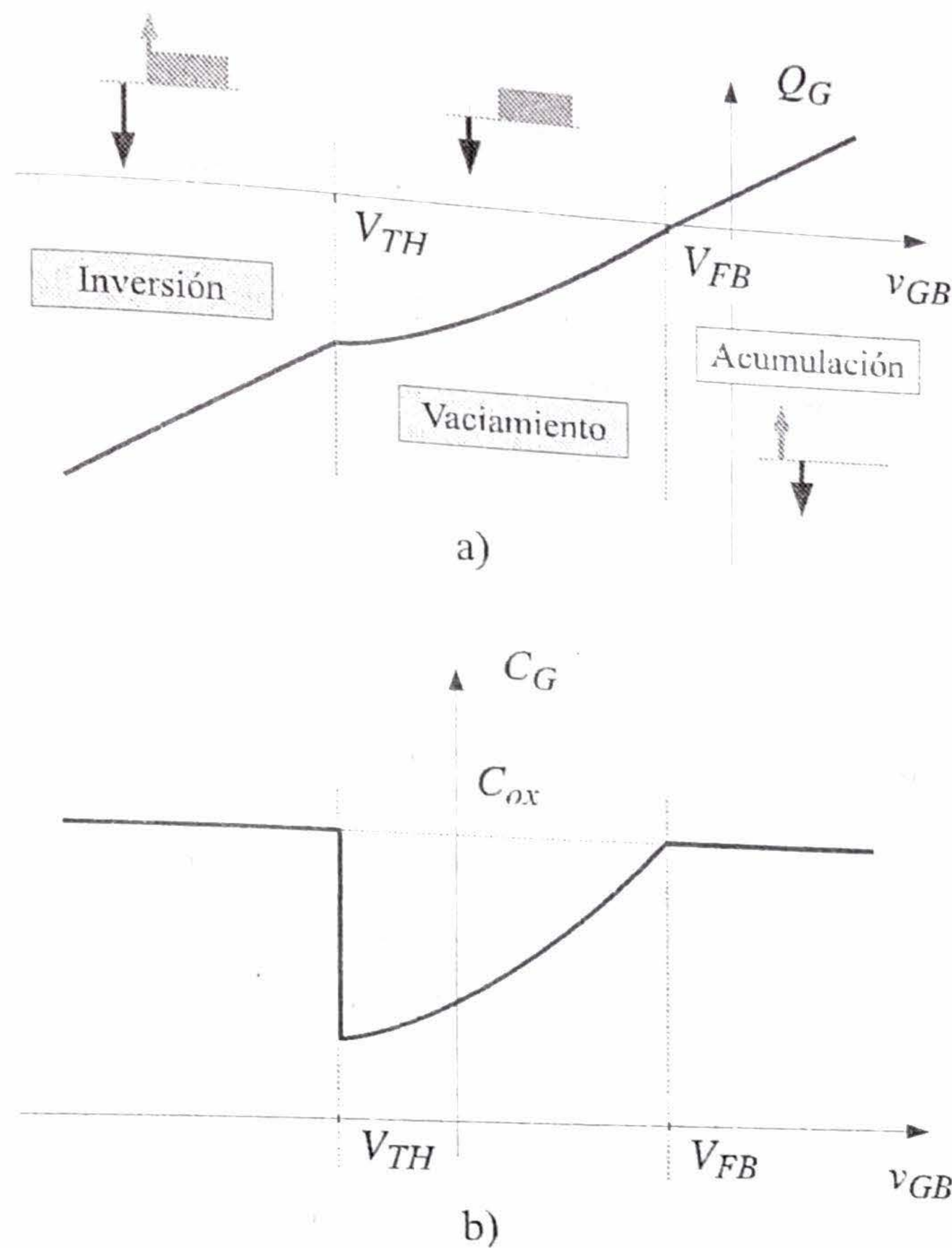


Figura 4.11: Capacitor MOS con gate dopado N sobre un sustrato tipo N: a) carga acumulada en el gate vs. la tensión v_{GB} ; b) capacidad incremental en función de v_{GB} .

$$v_{vac} = -\frac{qN_d}{2\epsilon_{Si}} x_{vac}^2 \quad (4.74)$$

El valor de la zona de vaciamiento satisface:

$$x_{vac} = \frac{\epsilon_{Si}}{C_{ox}} \left(-1 + \sqrt{1 - 2 \frac{C_{ox}^2 (v_{GB} - V_{FB})}{qN_d \epsilon_{Si}}} \right) \quad (4.75)$$

Cuando sucede la inversión, el potencial superficial es:

$$\phi_s = -\phi_n \quad (4.76)$$

y el potencial desarrollado sobre el sustrato es:

$$v_{vac} = -2\phi_n \quad (4.77)$$

El potencial desarrollado sobre el óxido es:

$$v_{ox} = -\frac{1}{C_{ox}} \sqrt{2qN_d \epsilon_{Si} (2\phi_n)} \quad (4.78)$$

y el valor de tensión de umbral:

$$\begin{aligned}
 V_T &= v_{ox} + v_{vac} + V_{FB} \\
 &= -\frac{1}{C_{ox}} \sqrt{2qN_d \epsilon_{Si} (-2\phi_n)} - 2\phi_n + V_{FB}
 \end{aligned} \tag{4.79}$$

De (4.79), se ve claramente que la tensión de umbral es menor que la tensión de Banda Plana, por lo cual, la región de vaciamiento se desarrolla en el intervalo:

$$V_T \leq v_{GB} \leq V_{FB} \tag{4.80}$$

La zona de inversión, a su vez, sucede cuando:

$$v_{GB} \leq V_T \tag{4.81}$$

donde la carga de inversión es positiva, y toma el valor:

$$Q_i(v_{GB}) = -C_{ox}(v_{GB} - V_T) \tag{4.82}$$

La Fig. 4.11 muestra la carga en el gate y la capacidad incremental para este caso. Nótese que al cambiar el tipo de dopado del sustrato, la secuencia de regiones se invierte, es decir: a) si $v_{GB} \geq V_{FB}$, el dispositivo está en acumulación; b) si $V_T \leq v_{GB} \leq V_{FB}$, el dispositivo está en vaciamiento; c) y si $V_T \geq v_{GB}$, el dispositivo está en inversión.

4.4.3. Capacitor MOS sobre un sustrato N y gate P

El último caso es el de un capacitor MOS sobre un sustrato ligeramente dopado N, con el gate fuertemente dopado P. En este caso, la tensión de Banda Plana será:

$$V_{FB} = -\phi_{GB} = -(\phi_{p+} - \phi_n) = \phi_n - \phi_{p+} \tag{4.83}$$

Por ejemplo, si los dopados del gate y el sustrato, son, respectivamente, $N_G = 10^{19}$ y $N_B = 10^{13}$, entonces, $\phi_{p+} = -540mV$, $\phi_n = 180mV$, y $V_{FB} = 720mV$. En este caso, la tensión de Banda Plana siempre será positiva.

La ecuación para la tensión de umbral mantiene el mismo valor que (4.79):

$$V_T = -\frac{1}{C_{ox}} \sqrt{2qN_d \epsilon_{Si} (2\phi_n)} - 2\phi_n + V_{FB} \tag{4.84}$$

De los análisis previos resulta sencillo ver que, en este caso, la secuencia de las regiones de funcionamiento es la correspondiente a un sustrato tipo N (es decir, inversión, vaciamiento y acumulación, para tensiones en orden creciente, dado que $V_T < V_{FB}$) y solo se experimenta una traslación de la tensión de umbral y de Banda Plana hacia la derecha (valores más positivos). Es decir, la situación de la Fig. 4.11 sigue siendo válida, con un corrimiento en los valores de V_T y V_{FB} . La Tabla 4.1 resume los valores de tensión de umbral, tensión de Banda Plana, y las secuencias de regiones para los cuatro tipos posibles de capacitores MOS.

Cuadro 4.1: Características principales de los diferentes capacitores MOS

Sustrato	Gate	Secuencia	V_T	V_{FB}
P	N	$V_{FB} < V_T$	$C_{ox}^{-1} \sqrt{2qN_a \epsilon_{Si}} (-2\phi_p) - 2\phi_p + V_{FB}$	$\phi_p - \phi_{n^+}$
P	P	$V_{FB} < V_T$	$C_{ox}^{-1} \sqrt{2qN_a \epsilon_{Si}} (-2\phi_p) - 2\phi_p + V_{FB}$	$\phi_p - \phi_{p^+}$
N	N	$V_T < V_{FB}$	$-C_{ox}^{-1} \sqrt{2qN_d \epsilon_{Si}} (2\phi_n) - 2\phi_n + V_{FB}$	$\phi_n - \phi_{n^+}$
N	P	$V_T < V_{FB}$	$-C_{ox}^{-1} \sqrt{2qN_d \epsilon_{Si}} (2\phi_n) - 2\phi_n + V_{FB}$	$\phi_n - \phi_{p^+}$

Capítulo 5

Transistor MOS

5.1. Descripción cualitativa

La descripción del funcionamiento del transistor MOS puede realizarse naturalmente a partir del capacitor MOS, introducido en el capítulo anterior. Constructivamente, para transformar un capacitor MOS en un transistor MOS, es necesario proveer dos puntos de acceso al canal formado debajo del óxido, uno a cada lado del dispositivo. Si bien la estructura original de un transistor MOS data de los años treinta, cuando J. E. Lilienfeld publica la patente [11], no es hasta los años sesenta que estos dispositivos se comienzan a fabricar de manera rutinaria.

Del análisis ya realizado, se sabe que cuando el capacitor MOS funciona en el régimen de inversión, debajo del óxido se generan portadores de signo contrario a los existentes en el sustrato. Esta capa de portadores se denomina canal, y es la base de la conducción del transistor MOS. El objetivo de los dos contactos introducidos, a ambos lados del dispositivo, es hacer contacto con los extremos del canal de conducción, formado en inversión. Luego, en el caso de un sustrato P, los portadores en inversión serán electrones, y los dos contactos deberán ser zonas de (fuerte) dopado N, que contengan abundantes cantidades de electrones. En el caso de un sustrato N, los portadores en inversión serán huecos, y los dos contactos deberán ser zonas de (fuerte) dopado P, que contengan abundantes cantidades de huecos.

Una vez establecido un canal de portadores (mediante una tensión de gate superior a la de umbral), se aplica una tensión entre los dos nuevos contactos, lo cual produce una circulación de los portadores por arrastre, como se ilustra en la Fig. 5.1, debido al campo eléctrico horizontal E_y . Esta corriente se puede modular, tanto por la tensión aplicada al gate, que produce una variación del campo eléctrico vertical E_x , como por la diferencia de potencial aplicada a lo largo del canal. Según el tipo de portadores del canal de inversión, el transistor MOS recibe el nombre de transistor de canal N, o NMOS (canal de electrones); o transistor de canal P, o PMOS (canal de huecos). Los contactos a ambos lados del canal reciben el nombre de "source" (S) y "drain" (D). El source se define como el contacto de partida de los portadores, y el drain como el contacto de llegada, cuando los mismos son arrastrados por medio del campo eléctrico horizontal. Los símbolos comúnmente utilizados para estos dispositivos se muestran en la Fig. 5.2.

5.2. El transistor NMOS

5.2.1. Principio básico de funcionamiento

Para este análisis, se considerará la estructura cuyo corte se muestra en la Fig. 5.3. Se supone que el transistor tiene dimensiones W (ancho) y L (longitud), como se puede apreciar en la vista superior del transistor, en la Fig. 5.4. El potencial de los terminales de source y drain siempre deberá ser positivo,

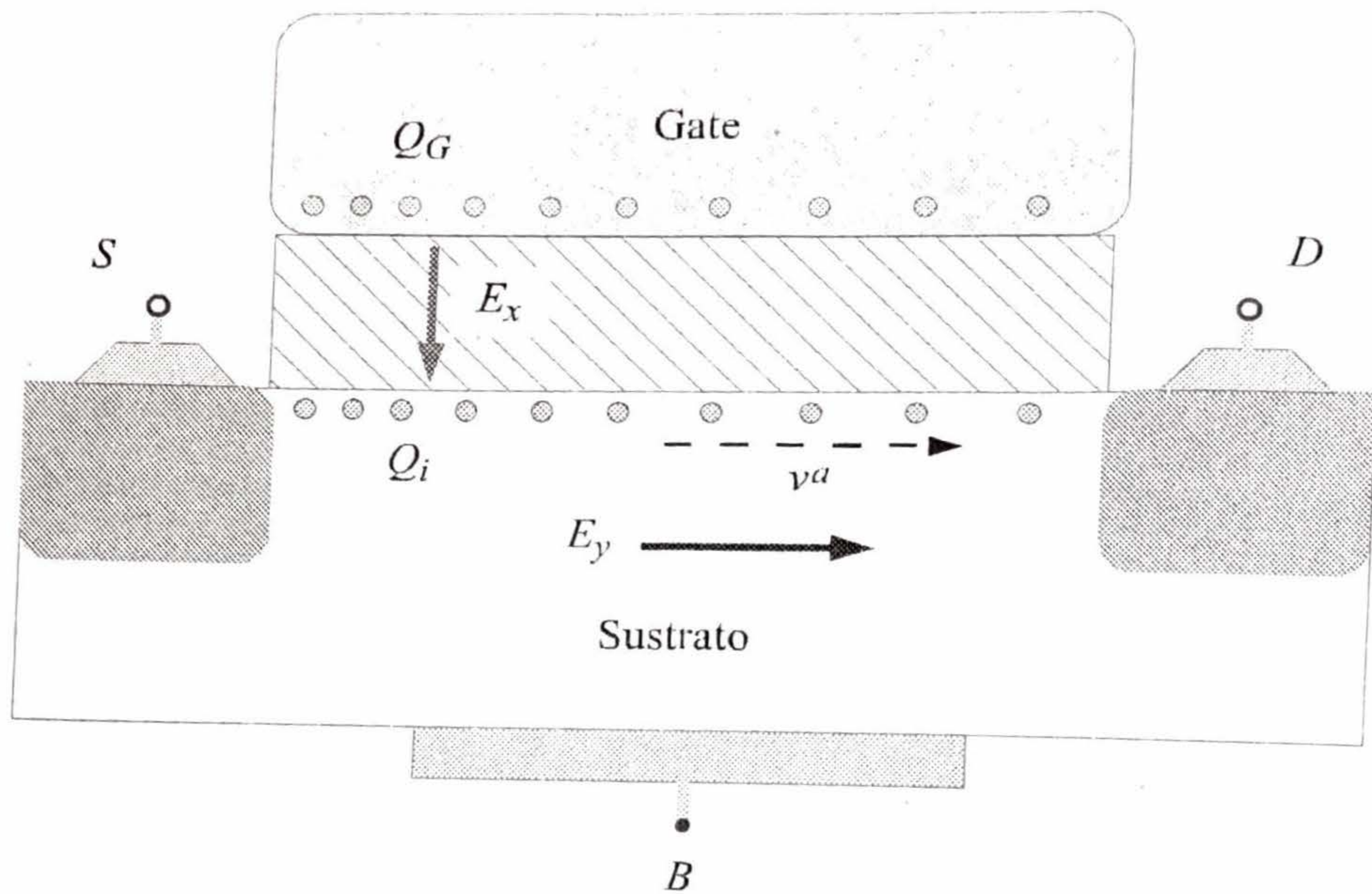


Figura 5.1: Corte longitudinal de un transistor MOS ilustrando la carga de inversión, provocada por el campo eléctrico vertical E_x , y la velocidad de los portadores de source (S) a drain (D), provocada por el campo eléctrico horizontal E_y .

de manera que las junturas (diodos) que se forman, entre ellos y el sustrato, estén en inversa en todo momento, es decir:

$$v_{DB} \geq 0; \quad v_{SB} \geq 0 \quad (5.1)$$

Bajo estas condiciones de funcionamiento, la corriente de sustrato será despreciable, pudiendo suponerse nula, es decir:

$$i_B \approx 0 \quad (5.2)$$

Por otro lado, dado que el gate está conectado en serie con el capacitor de gate, de valor C_{ox} , la corriente de entrada de DC, en estado estacionario, será nula:

$$I_G = 0 \quad (5.3)$$

La conducción de corriente es producto del arrastre de la carga de inversión (electrones), producido por el campo eléctrico horizontal. Debido a esto, se comienza el análisis a partir de la ecuación de corriente de arrastre, planteada en el Cap. 2 (2.50):

$$j_n^a = -qn(x, y)v_n^a(y) \quad (5.4)$$

donde la variable y mide la longitud del canal, comenzando en el terminal de "source"; x mide la profundidad del canal, comenzando en la interfaz entre el óxido y el sustrato; y z mide el ancho del canal (perpendicular al plano). Al plantear (5.4), se han considerado dos hechos: la densidad de cargas n puede variar según la longitud y profundidad en el dispositivo, pero permanece constante según z , dado que las variables eléctricas no cambian con respecto a z ; la velocidad v_n^a , solo depende del campo eléctrico vertical, por lo tanto, puede variar a lo largo de la longitud del canal, pero permanece constante según x y z .

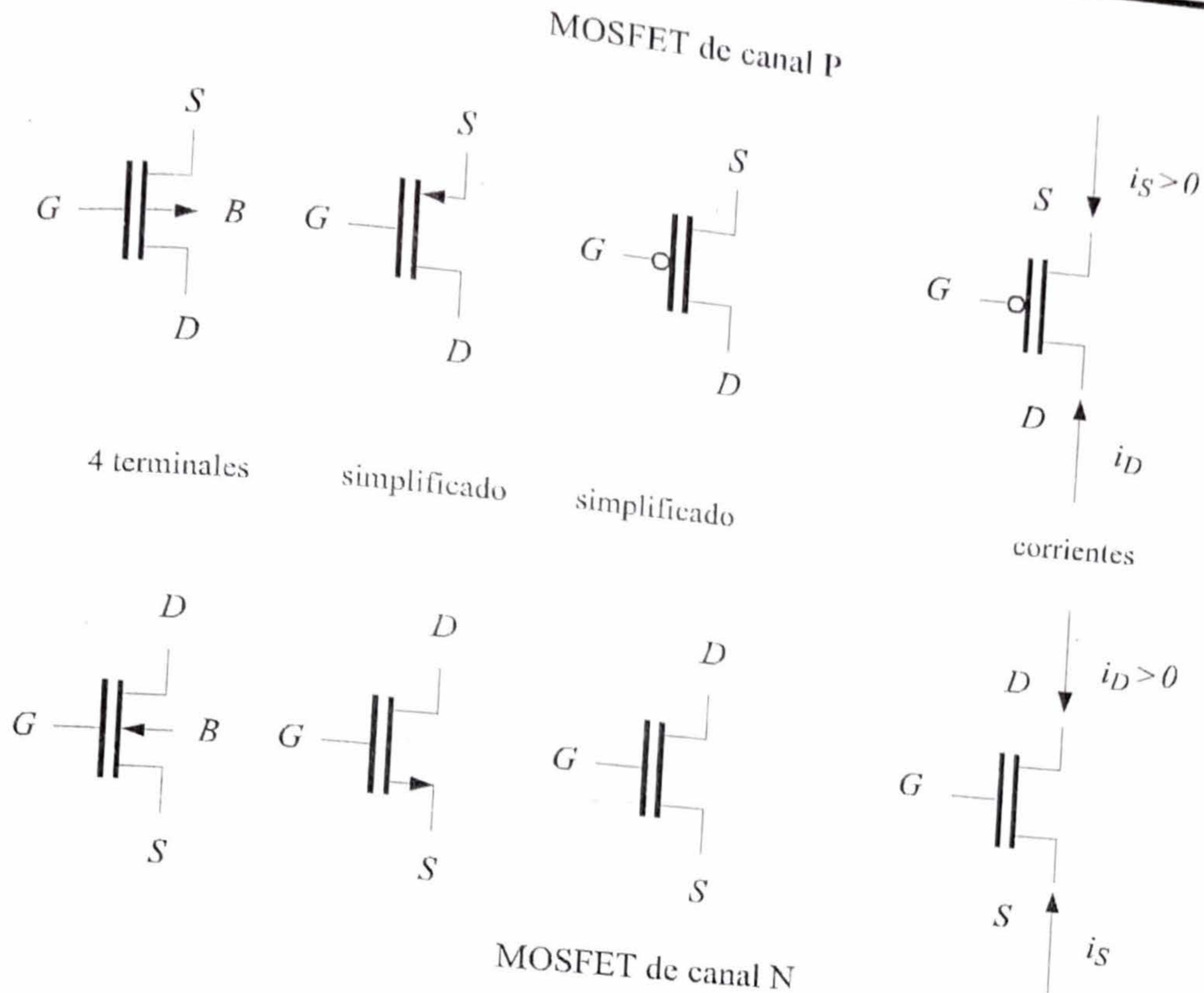


Figura 5.2: Símbolos utilizados comúnmente para transistores MOS, de canal N y P, de cuatro y tres terminales, respectivamente.

Típicamente, la distribución de carga por debajo del canal, $n(x, y)$, se concentra en una película extremadamente delgada, de algunos pocos nanómetros, cuya longitud se notará Δx , como se muestra en el detalle de la Fig. 5.3.

La corriente del dispositivo se obtiene a partir de (5.4), integrando con respecto a x y a z :

$$\begin{aligned}
 i_n^a &= \int_0^W \int_0^{\Delta x} -qn(x, y)v_n^a(y)dx dz \\
 &= Wv_n^a(y) \int_0^{\Delta x} -qn(x, y)dx
 \end{aligned}
 \tag{5.5}$$

Si se nota que el término integral en (5.5) es la carga de inversión en el canal, para un punto dado y , es decir:

$$Q_i(y) = \int_0^{\Delta x} -qn(x, y)dx
 \tag{5.6}$$

la corriente puede escribirse como:

$$\boxed{i_n^a = Wv_n^a(y)Q_i(y)}
 \tag{5.7}$$

La Ec. (5.7) es la expresión general de la corriente en el dispositivo, y pone de manifiesto que la magnitud de la corriente depende del producto de la velocidad de los portadores y la carga de inversión. La corriente circulando por el dispositivo tiene un valor único, que es el mismo para cada punto y a lo largo del canal. Por el contrario, la velocidad de los portadores y la carga de inversión, en general,

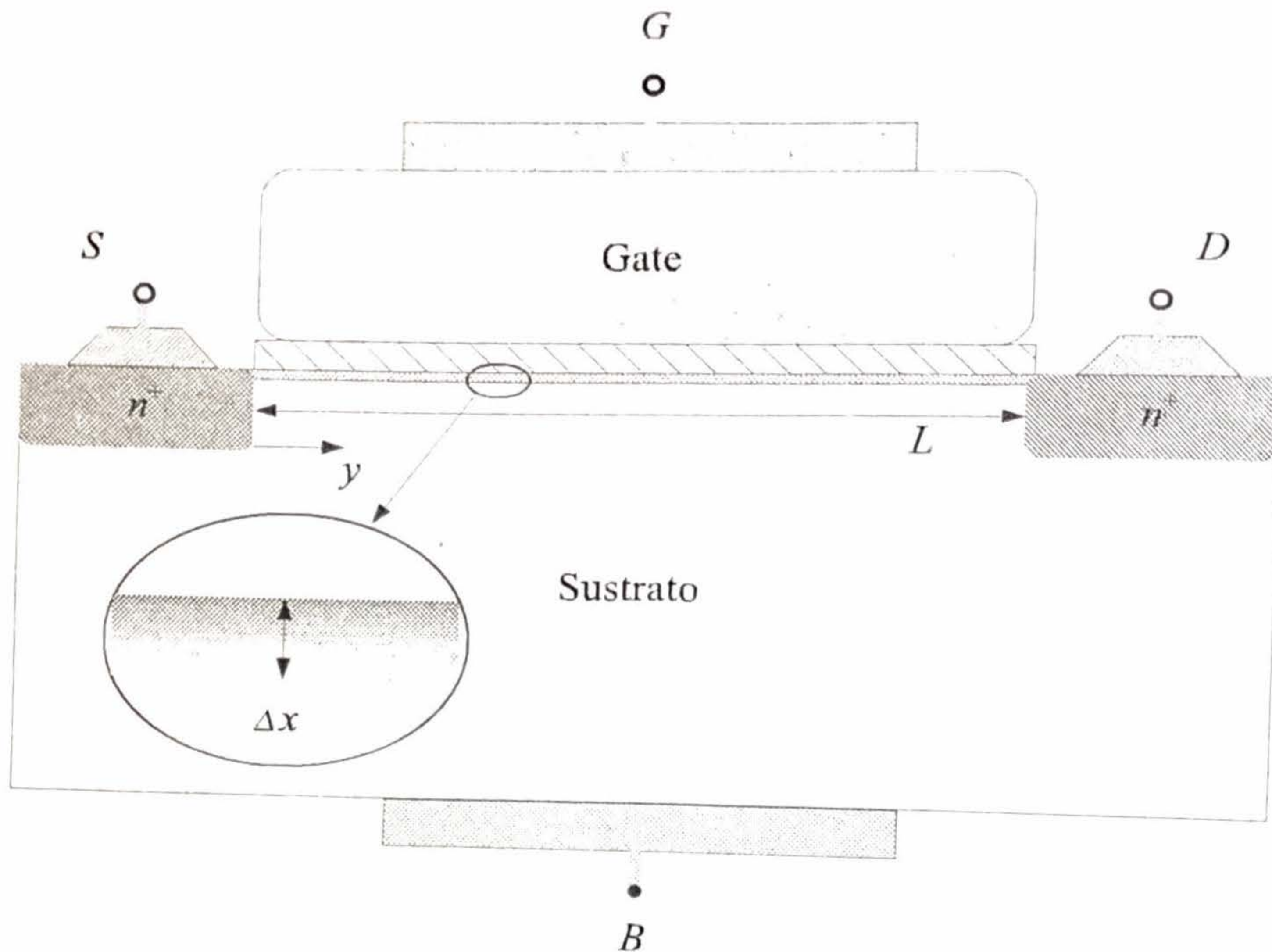


Figura 5.3: Corte longitudinal ($W = cte$) de un transistor NMOS de longitud de canal L , con un detalle del canal de conducción inducido debajo del gate.

experimentarán variaciones a lo largo del dispositivo (en función de y), pero manteniendo su producto constante, de acuerdo a (5.7).

A partir de la relación entre la velocidad de los portadores y el campo eléctrico, si se tiene en cuenta que $E = -d\phi_s/dy$, donde ϕ_s es el potencial de los portadores en el canal, resulta:

$$v_n^a = -\mu_n E = \mu_n \frac{d\phi_s}{dy} \quad (5.8)$$

y la corriente puede escribirse de la siguiente manera:

$$i_n^a = W Q_i(y) \mu_n \frac{d\phi_s}{dy} \quad (5.9)$$

Como la corriente de drain circula en sentido opuesto al eje y , su expresión es:

$$i_D = -i_n^a = -W \mu_n Q_i(y) \frac{d\phi_s}{dy} \quad (5.10)$$

La Ec. (5.10) relaciona la corriente en el transistor con la carga de inversión y la variación del potencial en el canal. En este punto, resulta necesario relacionar la carga de inversión Q_i y la variación de la tensión de los portadores en el canal, con las tensiones de los terminales. Esto puede hacerse tomando como terminal de referencia el sustrato o el source. A continuación, se explican las dos alternativas.

5.2.2. Modelo referido al Sustrato

El modelo referido al sustrato es atractivo dado que conserva la simetría del dispositivo entre los terminales de source y drain con respecto al sustrato, y surge como una continuación natural del

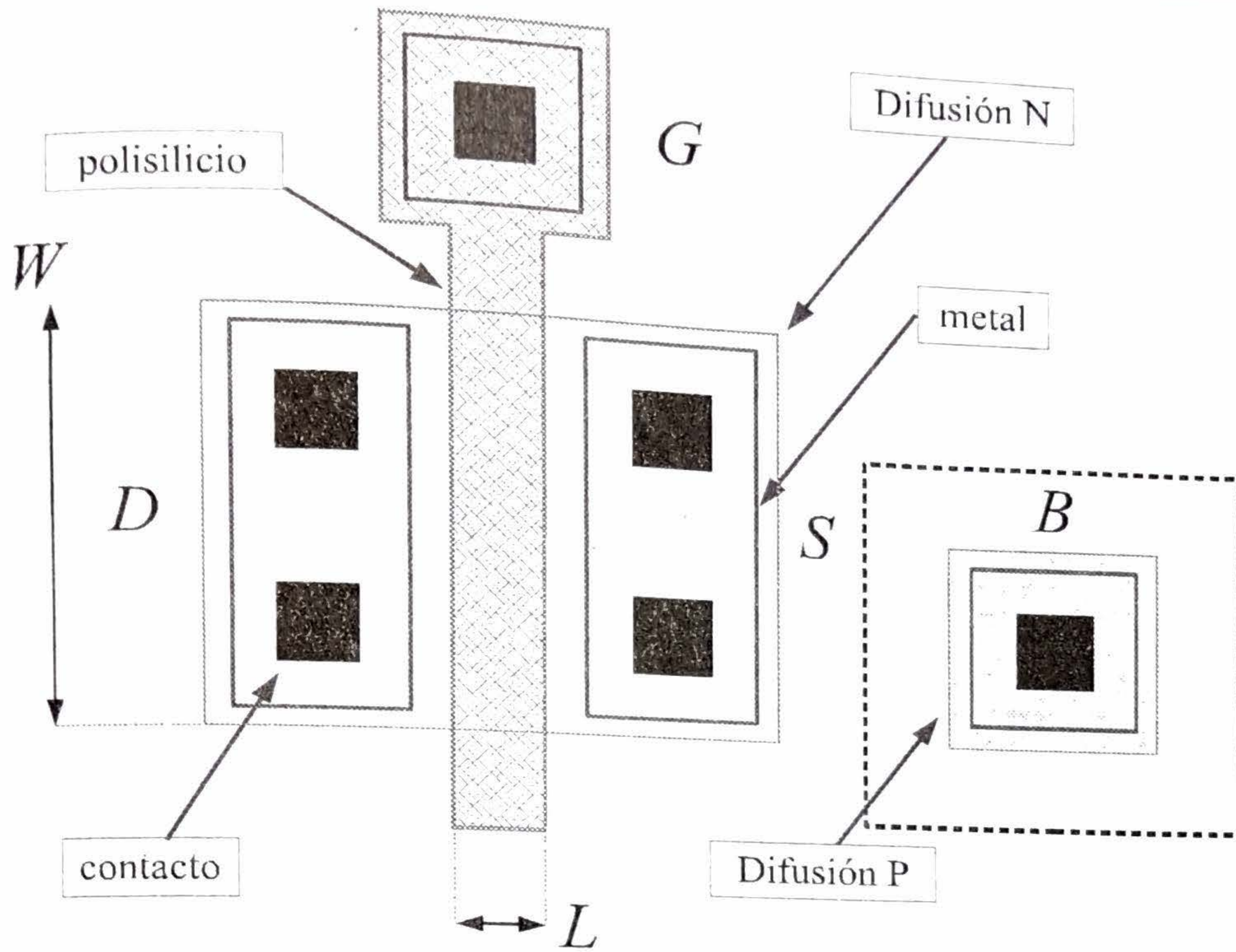


Figura 5.4: Esquema de un transistor NMOS de longitud L y ancho W .

capacitor MOS. Su uso es de preferencia en circuitos analógicos, con especial énfasis en el régimen de bajas corrientes [12].

Para el cálculo de la corriente, durante esta sección, se tomará el terminal de sustrato como punto de referencia de las tensiones aplicadas, tal como se muestra en la Fig. 5.5.

En el capítulo anterior, se estableció que, una vez en inversión, el potencial superficial ϕ_s de un capacitor MOS se mantiene fijo en el valor $\phi_s = -\phi_p$. En el caso del transistor MOS, aparece un nuevo ingrediente dado por los dos contactos. Estos dos contactos permiten acceder al canal y fijar su potencial en los extremos (source y drain); mediante una fuente externa. Debido a esto, el valor de la tensión superficial experimentará una variación a lo largo del canal, desde el valor inicial en el source, hasta el valor final en el drain; esto implica que la tensión superficial será función de la posición en el canal, y , y se notará $\phi_s(y)$. El potencial desarrollado a lo largo de la zona de vaciamiento, para un punto arbitrario del canal, y , se notará $v_{vac}(y)$.

Si en principio se cortocircuitan los terminales de source y drain con el terminal de sustrato, el potencial neto entre los terminales de source/drain y sustrato, será nulo. Suponiendo que la tensión del gate satisface $v_G \geq V_T$, en los extremos del canal, se cumple:

$$v_{vac}(0) = -2\phi_p \quad (5.11)$$

$$v_{vac}(L) = -2\phi_p \quad (5.12)$$

Si ahora se aplican potenciales a los terminales de source y drain, con respecto al sustrato, introduciendo dos fuentes de tensión, v_S y v_D , respectivamente, como se muestra en la Fig. 5.5, la tensión de la zona de vaciamiento en los extremos del canal aumentará en forma directa:

$$v_{vac}(0) = -2\phi_p + v_S \quad (5.13)$$

$$v_{vac}(L) = -2\phi_p + v_D \quad (5.14)$$

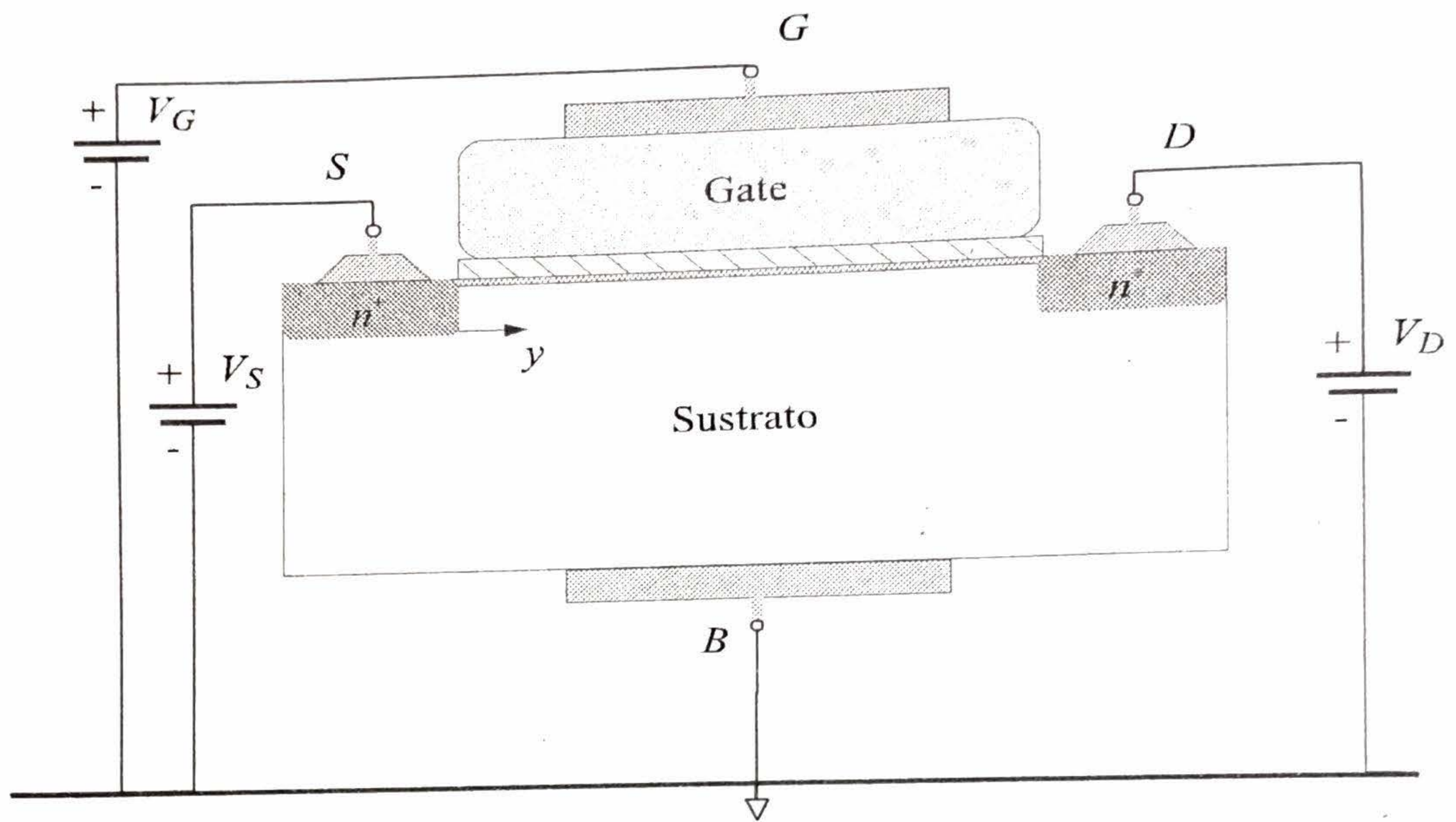


Figura 5.5: Transistor NMOS con tensiones referidas al sustrato.

Si las tensiones en los extremos del canal se modifican, indefectiblemente, la tensión en los puntos intermedios a lo largo del canal, se deberán modificar apropiadamente. En un punto y cualquiera, a lo largo del canal, el potencial desarrollado en la zona de vaciamiento, se puede escribir de la forma:

$$v_{vac}(y) = -2\phi_p + v_c(y) \quad (5.15)$$

tal como se ilustra en la Fig. 5.6, donde la variable $v_c \triangleq \phi_s - (-\phi_p)$ describe el potencial de los portadores en el canal con respecto a su valor en inversión¹, dado por $-\phi_p$. Nótese aquí, que el potencial de canal v_c satisface $v_c(0) = v_S$, y $v_c(L) = v_D$.

La longitud correspondiente de la zona de vaciamiento, a una distancia y del source, es:

$$x_{vac}(y) = \sqrt{\frac{2\epsilon_{Si}(-2\phi_p + v_c(y))}{qN_a}} \quad (5.16)$$

y la carga asociada es:

$$Q_{vac}(y) = -\sqrt{2qN_a\epsilon_{Si}(-2\phi_p + v_c(y))} \quad (5.17)$$

Es interesante notar el efecto de un incremento del potencial del canal, producto del aumento de los potenciales externos: de acuerdo a (5.16) y (5.17), al incrementarse $v_{vac}(y)$, la carga en el sustrato y la longitud de la zona de vaciamiento también se incrementan.

Para que un punto genérico del canal, y , esté justo en el inicio de inversión, la carga del gate debe ser de igual magnitud y signo contrario a la carga del sustrato, es decir:

$$Q_G = -Q_{vac}(y) \quad (5.18)$$

Debido a esto, y en virtud de (5.17), la tensión a lo largo del óxido, en este caso particular, estará dada por:

¹Nótese que ϕ_s representa la tensión del canal con respecto a Silicio intrínscico, y que $\phi_s - \phi_p + \phi_{p,m}$ representa la tensión del canal con respecto al terminal común de sustrato.

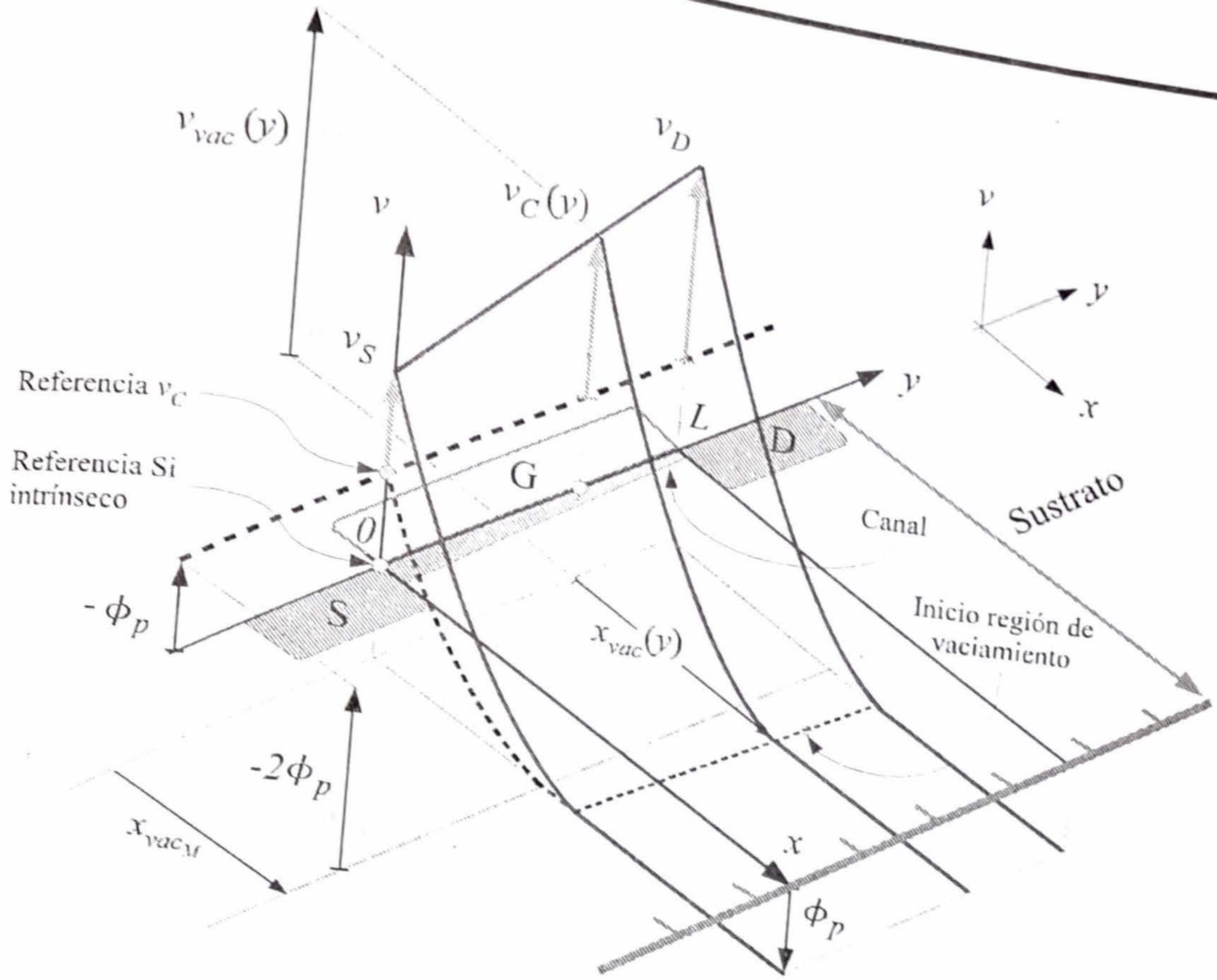


Figura 5.6: Corte transversal de un transistor NMOS, indicando el potencial del canal, desde el source hasta el drain.

$$v_{ox}^*(y) = \frac{Q_G}{C_{ox}} = \frac{1}{C_{ox}} \sqrt{2qN_a\epsilon_{Si}(-2\phi_p + v_c(y))} \quad (5.19)$$

Planteando mallas:

$$\begin{aligned} v_G^* &= v_{ox}^* + v_{vac}(y) + V_{FB} \\ &= \frac{1}{C_{ox}} \sqrt{2qN_a\epsilon_{Si}(-2\phi_p + v_c(y))} + \\ &\quad + (-2\phi_p + v_c(y)) + V_{FB} \end{aligned} \quad (5.20)$$

Si se define:

$$V_T(y) \triangleq -2\phi_p + V_{FB} + \frac{1}{C_{ox}} \sqrt{2qN_a\epsilon_{Si}(-2\phi_p + v_c(y))} \quad (5.21)$$

se puede reescribir (5.20) como:

$$v_G^* = V_T(y) + v_c(y) \quad (5.22)$$

El incremento de la tensión del canal produce un aumento de la carga en el sustrato, que como se puede ver de (5.21) produce un aumento de la tensión de umbral.
A partir del valor de tensión de gate (5.22), se producirá carga de inversión, la cual estará dada por:

$$Q_i(y) = -C_{ox} (v_G - v_G^*) = -C_{ox} (v_G - V_T(y) - v_c(y)) \quad (5.23)$$

La existencia de corriente se dará cuando haya carga de inversión a lo largo de todo el canal, es decir, cuando la tensión de gate satisfaga:

$$v_G > v_G^* = V_T(y) + v_c(y) \quad (5.24)$$

para todo punto y , a lo largo del canal ($0 \leq y \leq L$). En tal caso, para hallar la expresión de la corriente, se debe plantear, a partir de (5.10):

$$i_D dy = \mu_n C_{ox} W (v_G - V_T(y) - v_c(y)) dv_c \quad (5.25)$$

Integrando ambos miembros de la igualdad:

$$\int_0^L i_D dy = \int_{v_S}^{v_D} \mu_n C_{ox} W (v_G - V_T(y) - v_c(y)) dv_c \quad (5.26)$$

resulta la expresión genérica de la corriente:

$$i_D = \mu_n C_{ox} \frac{W}{L} \int_{v_S}^{v_D} (v_G - V_T(y) - v_c(y)) dv_c \quad (5.27)$$

En este punto, es necesario prestar atención a la dependencia funcional de $V_T(y)$ con respecto a $v_c(y)$, dada por (5.21). La solución exacta a la integral (5.27), resulta en la siguiente expresión ([13], [14]):

$$i_D = \mu_n C_{ox} \frac{W}{L} \left\{ (v_G - V_{FB})(v_D - v_S) - \frac{1}{2} [(v_D - 2\phi_p)^2 - (v_S - 2\phi_p)^2] - \frac{2}{3} \gamma [(v_D - 2\phi_p)^{3/2} - (v_S - 2\phi_p)^{3/2}] \right\} \quad (5.28)$$

Esta expresión resulta complicada para análisis y diseño, a raíz de los factores elevados a la potencia $3/2$, que surgen de la integral de la tensión de umbral, por su proporcionalidad con la raíz cuadrada del potencial en la zona de vaciamiento. Por el contrario, una linealización de V_T conduce a resultados sencillos. Si se expande, por medio de la expansión de Taylor, la expresión de V_T en función de la tensión de canal, resulta:

$$V_T(y) = V_{T0} + \left. \frac{dV_T(y)}{dv_c} \right|_{v_c=0} v_c = V_{T0} + \frac{\gamma}{2\sqrt{-2\phi_p}} v_c \quad (5.29)$$

donde V_{T0} se denomina tensión de umbral (extrapolada²) con respecto a la tensión gate-sustrato:

$$V_{T0} \triangleq -2\phi_p + V_{FB} + \frac{1}{C_{ox}} \sqrt{2qN_a\epsilon_{Si}(-2\phi_p)} \quad (5.30)$$

y se define el factor de sustrato o "back-gate", γ , como:

$$\gamma \triangleq \frac{\sqrt{2qN_a\epsilon_{Si}}}{C_{ox}} \quad (5.31)$$

²Se denomina extrapolada porque si se realiza una gráfica de $\sqrt{i_D}$ versus v_G (que resulta una recta para $v_G > V_{T0}$), V_{T0} es el punto donde la extrapolación de esta recta cruza el eje v_G (dado por $\sqrt{i_D} = 0$).

Luego, la carga de inversión está dada por:

$$Q_i(y) = -C_{ox} (v_G - V_{T0} - nv_c) \quad (5.32)$$

donde la variable n se define³ como:

$$n \triangleq 1 + \frac{\gamma}{2\sqrt{-2\phi_p}} \quad (5.36)$$

y el argumento de la integral (5.27), se puede escribir como:

$$v_G - V_T(y) - v_c(y) = v_G - V_{T0} - nv_c \quad (5.37)$$

Utilizando (5.37), la expresión de la corriente se reduce a:

$$i_D = \mu_n C_{ox} \frac{W}{L} \int_{v_S}^{v_D} (v_G - V_{T0} - nv_c) dv_c \quad (5.38)$$

La evaluación de esta integral se puede expresar de la siguiente manera:

$$i_D = \mu_n C_{ox} \frac{W}{L} \left((v_G - V_{T0})v_c \Big|_{v_S}^{v_D} - \frac{1}{2}nv_c^2 \Big|_{v_S}^{v_D} \right) \quad (5.39)$$

de donde resulta la expresión de la corriente en la región de triodo:

$$i_D = \mu_n C_{ox} \frac{W}{L} \left((v_G - V_{T0})(v_D - v_S) - \frac{1}{2}n(v_D^2 - v_S^2) \right) \quad (5.40)$$

La corriente puede interpretarse gráficamente, a partir de (5.38), como el área entre la línea horizontal v_G y la recta $V_{T0} + v_c(y)$, dentro de los límites impuestos por v_S y v_D , tal como se muestra en la Fig. 5.7.

De (5.38), y de la Fig. 5.7, se puede ver que cuando el potencial en el drain alcanza un valor $V_{D(sat)}$, para el cual $v_G = V_{T0} + nv_c$, la corriente no varía más. Este valor de potencial, $V_p = V_{D(sat)}$, se denomina de estrangulamiento o "pinch-off", y se define como:

$$V_p \triangleq \frac{v_G - V_{T0}}{n} \quad (5.41)$$

Para este potencial, la carga de inversión en el drain ($y = L$) se iguala a cero:

³El término n puede escribirse como

$$n = 1 + \frac{\gamma}{2\sqrt{-2\phi_p}} \quad (5.33)$$

$$= 1 + \frac{\gamma C_{ox}}{2\sqrt{-2\phi_p}} \frac{1}{C_{ox}} = 1 + \frac{C_{vac}}{C_{ox}} = \frac{C_{ox} + C_{vac}}{C_{ox}} \quad (5.34)$$

donde se debe reconocer la capacidad incremental del sustrato:

$$C_{vac} = - \frac{dQ_{vac}}{dv_{vac}} \Big|_{v_c=0} = \frac{\gamma C_{ox}}{2\sqrt{-2\phi_p}} \quad (5.35)$$

dado que de (4.63) se deduce que $Q_{vac} = -\gamma C_{ox} \sqrt{v_{vac}}$. El factor n puede interpretarse como el divisor capacitivo, dado por la capacidad del óxido y la capacidad del sustrato en serie, en el extremo del source. Se puede ver entonces que elevar el potencial de superficie ϕ_s , de $-2\phi_p$ a $-2\phi_p + v_c$, al incrementar la tensión en los contactos laterales (drain y/o source), es equivalente a disminuir en nv_c el potencial del gate. Esto se ve claramente a partir de la ecuación de la carga de inversión (5.32).

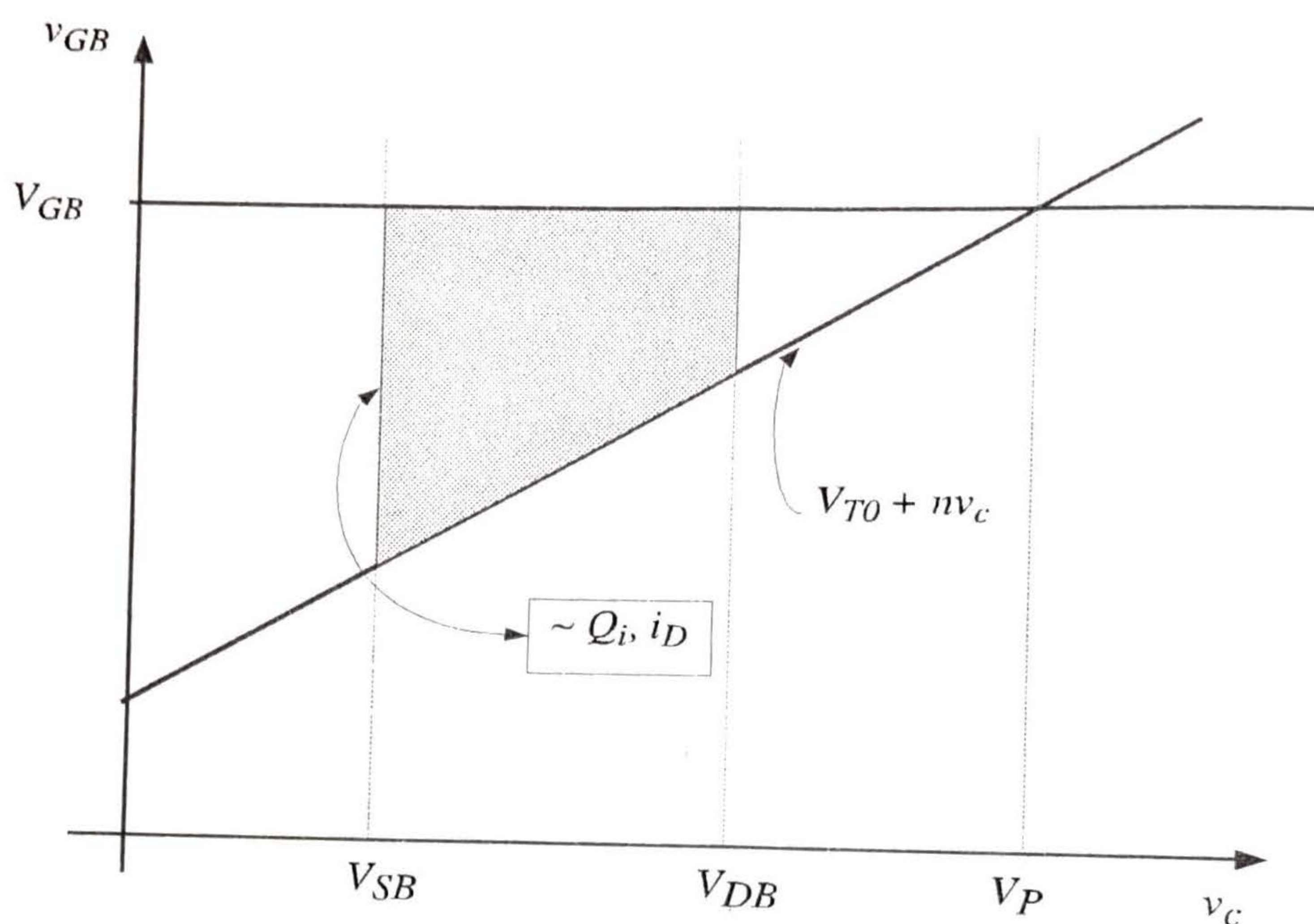


Figura 5.7: Gráfica de v_G y $V_{T0} + nv_c$. El área resaltada es proporcional a la carga de inversión, y a la corriente.

$$Q_i = -C_{ox}(v_G - V_{T0} - nV_p) = 0 \quad (5.42)$$

Es por ello, que aumentos posteriores de la tensión de drain, $v_D > V_{D(sat)}$, no producen ninguna variación de la corriente. La región de funcionamiento, en la cual $v_D \leq V_{D(sat)}$, se denomina zona lineal, o de triodo, y se caracteriza por la dependencia de la corriente, tanto de la tensión de drain v_D , como de la tensión de gate v_G . Para valores mayores a $V_{D(sat)}$, el dispositivo está en la zona de funcionamiento denominada de saturación, la cual se caracteriza por la independencia de la corriente con respecto a la tensión de drain. Reemplazando v_D por su valor de saturación (5.41), en (5.40), se obtiene:

$$\begin{aligned} i_D &= \mu_n C_{ox} \frac{W}{L} \left\{ (v_G - V_{T0}) \left(\frac{v_G - V_{T0}}{n} - v_S \right) - \right. \\ &\quad \left. - \frac{n}{2} \frac{(v_G - V_{T0})^2}{n^2} - \frac{n}{2} v_S^2 \right\} \\ &= \mu_n C_{ox} \frac{W}{L} \left(\frac{(v_G - V_{T0})^2}{2n} - (v_G - V_{T0})v_S + \frac{n}{2} v_S^2 \right) \end{aligned} \quad (5.43)$$

Operando algebraicamente y completando binomios, resulta:

$$i_D = \frac{\mu_n C_{ox} W}{2n L} (v_G - V_{T0} - nv_S)^2 \quad (5.44)$$

En resumen, las expresiones de la corriente del transistor MOS, en triodo (5.40) y saturación (5.44), son:

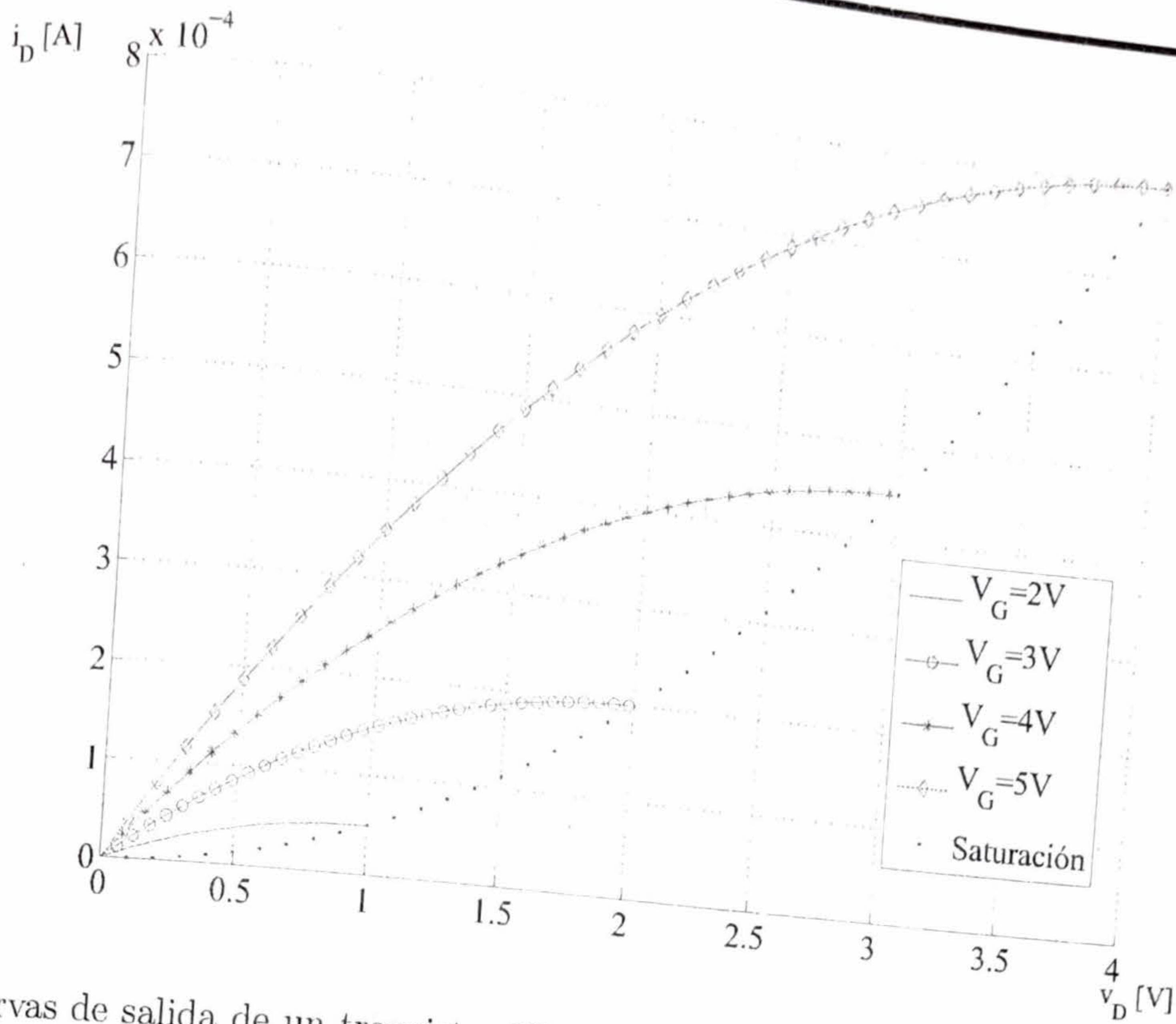


Figura 5.8: Curvas de salida de un transistor NMOS en la zona de triodo ($K = 100\mu A/V^2$, $V_{T0} = 1$, $n = 1$) considerando $V_S = 0V$. La línea punteada muestra el límite de la región de triodo, dado por (5.50).

$$i_D = \begin{cases} K_n \left((v_G - V_{T0})v_{DS} - \frac{1}{2}n(v_D^2 - v_S^2) \right) & \text{si } v_D < V_{D(sat)} \\ \frac{K_n}{2n} (v_G - V_{T0} - nv_S)^2 & \text{si } v_D \geq V_{D(sat)} \end{cases} \quad (5.45)$$

donde se han introducido los parámetros

$$K_n = \frac{W}{L} K'_n \quad (5.46)$$

$$K'_n = \mu_n C_{ox} \quad (5.47)$$

y la tensión de saturación está dada por:

$$V_{D(sat)} = \frac{v_G - V_{T0}}{n} \quad (5.48)$$

Un caso de particular interés resulta cuando la tensión de source es igual a cero ($v_S = 0$). En este caso, las expresiones de triodo y saturación se reducen a:

$$i_D = \begin{cases} K_n \left((v_G - V_{T0})v_D - \frac{nv_D^2}{2} \right) & \text{si } v_D < V_{D(sat)} \\ \frac{K_n}{2n} (v_G - V_{T0})^2 & \text{si } v_D \geq V_{D(sat)} \end{cases} \quad (5.49)$$

Las curvas de corriente de salida para distintos valores de v_G se ilustran en la Fig. 5.8. En la región de triodo, se puede ver que las curvas de salida comienzan en el origen y aumentan en forma cuadrática con v_D , para una $v_G (> V_{T0})$ constante, alcanzando su valor máximo cuando $v_D = V_{D(sat)} = (v_G - V_{T0})/n$.

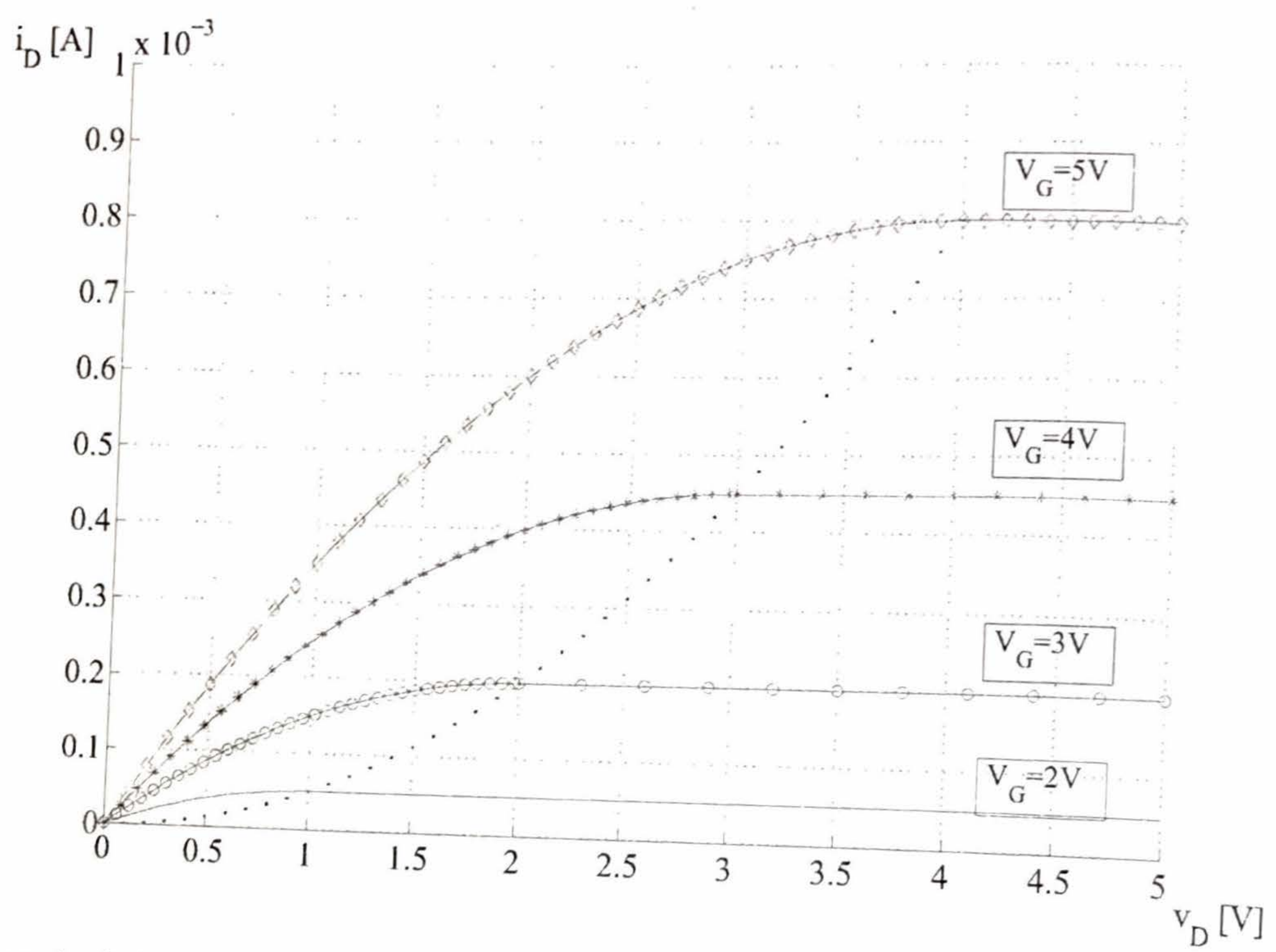


Figura 5.9: Característica de salida de un transistor NMOS ($K_n = 100\mu A/V^2$, $V_{T0} = 1$, $n = 1$), considerando $V_S = 0V$.

Si se incrementa el valor de v_G , se observa que la curva de corriente es mayor, al igual que la tensión de drain de saturación, que ahora ocurre más a la derecha. A partir de (5.49), es sencillo ver que el conjunto de puntos de la curva de salida para los cuales se alcanza la saturación satisface:

$$i_D = \frac{K}{2n} V_{D(sat)}^2 \tag{5.50}$$

Esta parábola determina el límite entre las zonas de triodo y saturación.

En la zona de saturación, las curvas de salida son constantes con respecto a la tensión v_D , como se puede ver de la Fig. 5.9. La Fig. 5.10 muestra curvas experimentales de salida para un transistor de $W = 5\mu m \times L = 5\mu m$, en una tecnología de $0,5\mu m$.

A medida que se aumenta v_G , se obtienen curvas paralelas cuyo valor de corriente es:

$$i_D = \frac{K}{2n} (v_G - V_{T0})^2 \tag{5.51}$$

por lo cual, se ve un aumento cuadrático de la corriente con respecto a la tensión de gate. Dado que para esta región de funcionamiento, la corriente solo depende de la tensión de entrada v_G , se puede producir una curva de transferencia en los ejes i_D vs. v_G . La misma se ilustra en la Fig. 5.11.

Para valores de $V_S > 0$ se pueden obtener similares conjuntos de curvas de salida y transferencia, salvo que la corriente i_D será menor por la influencia de v_S , tal como se desprende de (5.45).

Ejemplo 18 Un transistor NMOS ($W = 20\mu m$ y $L = 5\mu m$) tiene los siguientes parámetros⁴: $t_{ox} = 14,2nm$, $V_{T0} = 0,64V$, $\mu_0 = 472cm^2/Vs$, $N_a = 1,7 \times 10^{17}/cm^3$. Halle C_{ox} , K'_n y γ .
De acuerdo a estos datos,

⁴Datos adaptados en base al proceso C5N de $0,5\mu m$ de MOSIS, corrida T82Q

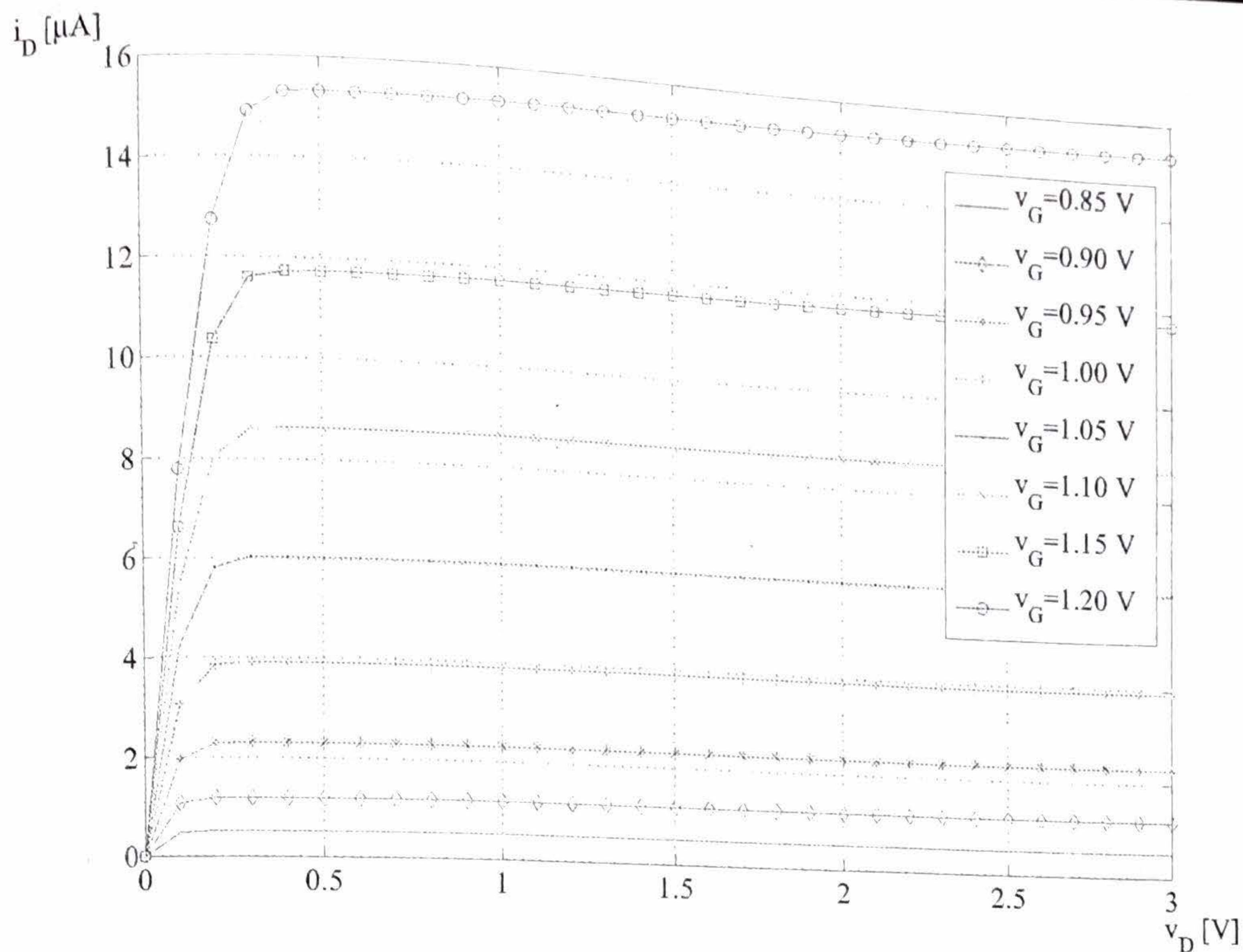


Figura 5.10: Curvas experimentales de salida de un transistor NMOS de $W = 5\mu m$ y $L = 5\mu m$ en un proceso de $0,5\mu m$.

$$C_{ox} = \frac{3,9 \times 8,854 \times 10^{-14}}{1,42 \times 10^{-14}} = 0,243\mu F/cm^2 \quad (5.52)$$

$$K'_n = 472 \times 2,43 \times 10^{-7} = 114,78\mu A/V^2 \quad (5.53)$$

$$\gamma = \frac{\sqrt{2 \times 1,6 \times 10^{-19} \times 1,7 \times 10^{17} \times 11,8 \times 8,854 \times 10^{-14}}}{2,43 \times 10^{-7}} = 0,98\sqrt{V} \quad (5.54)$$

y

$$n = 1 + \frac{0,98}{2 \times \sqrt{-2 \times -0,4334}} = 1,52 \quad (5.55)$$

Ejemplo 19 El transistor del ejemplo anterior funciona con $V_G = 3V$ y $V_S = 0,5V$. Calcule la tensión de saturación de drain, la máxima corriente del dispositivo, y la corriente del dispositivo si $V_D = V_{D(sat)}/2$.

La tensión de saturación es $V_{D(sat)} = (3 - 0,64)/1,52 = 1,55V$, y $K_n = 20/4 \times 114,78\mu/V^2 = 573,9\mu A/V^2$. La corriente máxima en saturación se calcula como:

$$I_D = \frac{573,9}{2 \times 1,52} (3 - 0,64 - 1,52 \times 0,5)^2 = 483,28\mu A \quad (5.56)$$

Si $V_D = V_{D(sat)}/2 = 0,775V$, entonces el dispositivo se encuentra en triodo. En este punto de operación la corriente está dada por:

$$I_D = 573,9 \left((3 - 0,64) \times (0,775 - 0,5) - \frac{1,52}{2} (0,775^2 - 0,5^2) \right) = 219,53\mu A \quad (5.57)$$

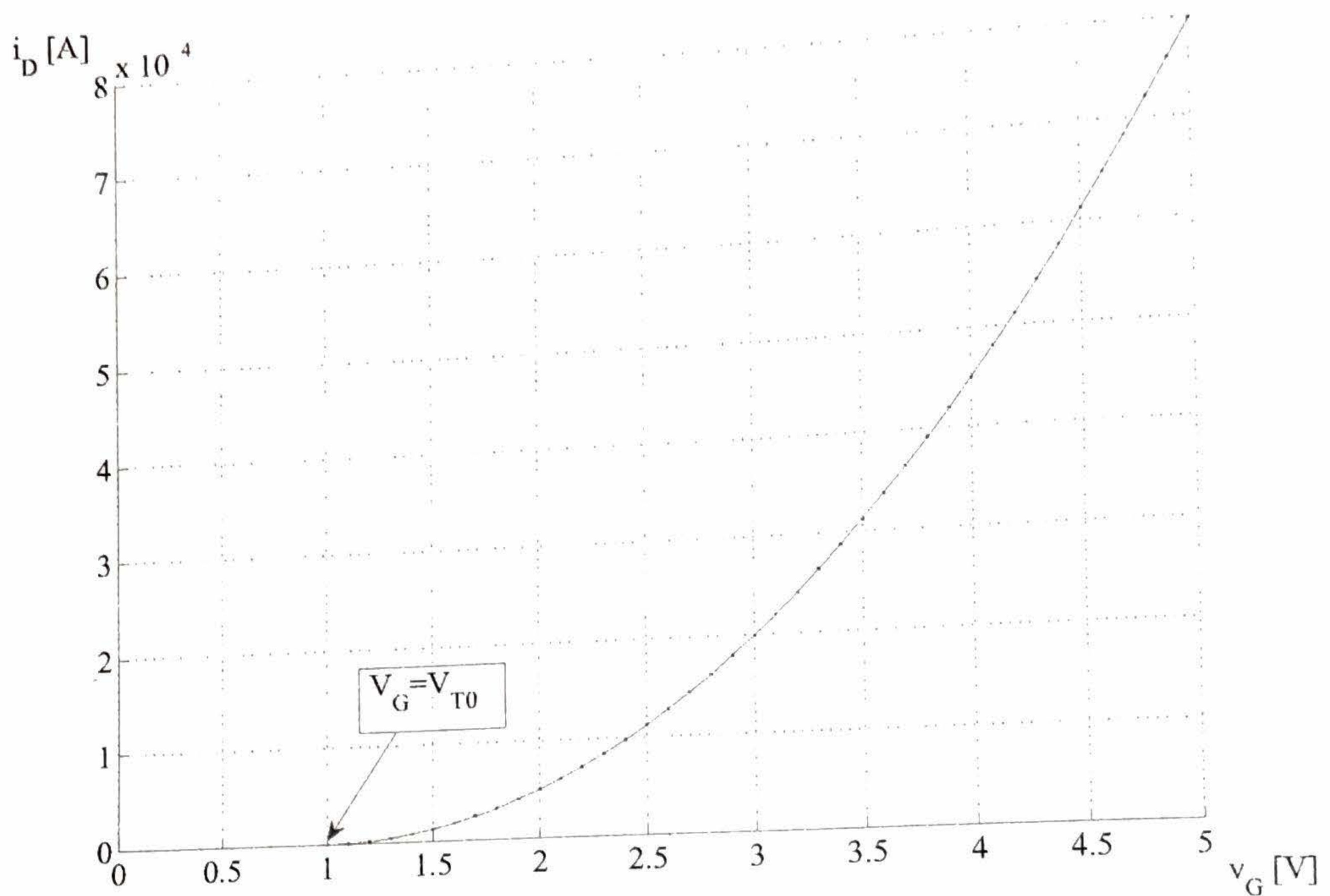


Figura 5.11: Característica de transferencia de un transistor NMOS ($K_n = 100\mu A/V^2$, $V_{T0} = 1$, $n = 1$), con $V_S = 0V$.

5.2.3. Modelo referido al Source

Los modelos referidos al source han sido históricamente utilizados para la descripción de transistores bipolares (donde la base reemplaza al gate y el emisor al source), transistores de efecto de campo de juntura (JFET), y aún transistores valvulares (donde la grilla reemplaza al gate, y el cátodo al source). El diagrama de tensiones para un transistor MOS referido al terminal de source se ilustra en la Fig. 5.12.

Tanto el modelo descrito en la sección anterior, en el cual las tensiones se toman con respecto al sustrato, como este modelo, pueden utilizarse para resolver las corrientes y tensiones de un transistor, en un circuito dado, y ambos producirán resultados adecuados. En el modelo referido al source, se utilizan las tensiones v_{GS} , v_{DS} y v_{BS} . Este modelo tiene la ventaja que relaciona en forma directa la corriente de drain con sus dos causas principales: la carga de inversión, producto de la tensión v_{GS} , y el campo eléctrico lateral, dado por v_{DS} [12]. Otra diferencia con el modelo referido al sustrato es que la dependencia de la corriente con respecto a v_{SB} se manifiesta indirectamente a través de la variación de V_T . Por otro lado, en este modelo, los dos terminales de control del transistor, el principal dado por el gate, y el secundario dado por el sustrato (también llamado “back-gate”) aparecen de manera natural.

En este caso, el potencial a lo largo de la zona de vaciamiento está dado por:

$$v_{vac}(y) = -2\phi_p + \tilde{v}_c(y) + v_{SB} \quad (5.58)$$

donde ahora el potencial del canal \tilde{v}_c varía en el intervalo $0 \leq \tilde{v}_c(y) \leq v_{DS}$, y satisface $\tilde{v}_c(0) = 0V$, y $\tilde{v}_c(L) = v_{DS}$, de tal manera que (5.58) coincide con (5.15).

Por otro lado, la carga asociada⁵ está dada por:

⁵Esta expresión coincide con (5.6), si se tiene en cuenta la diferencia entre los límites de variación de $\tilde{v}_c(y)$ y $v_c(y)$.

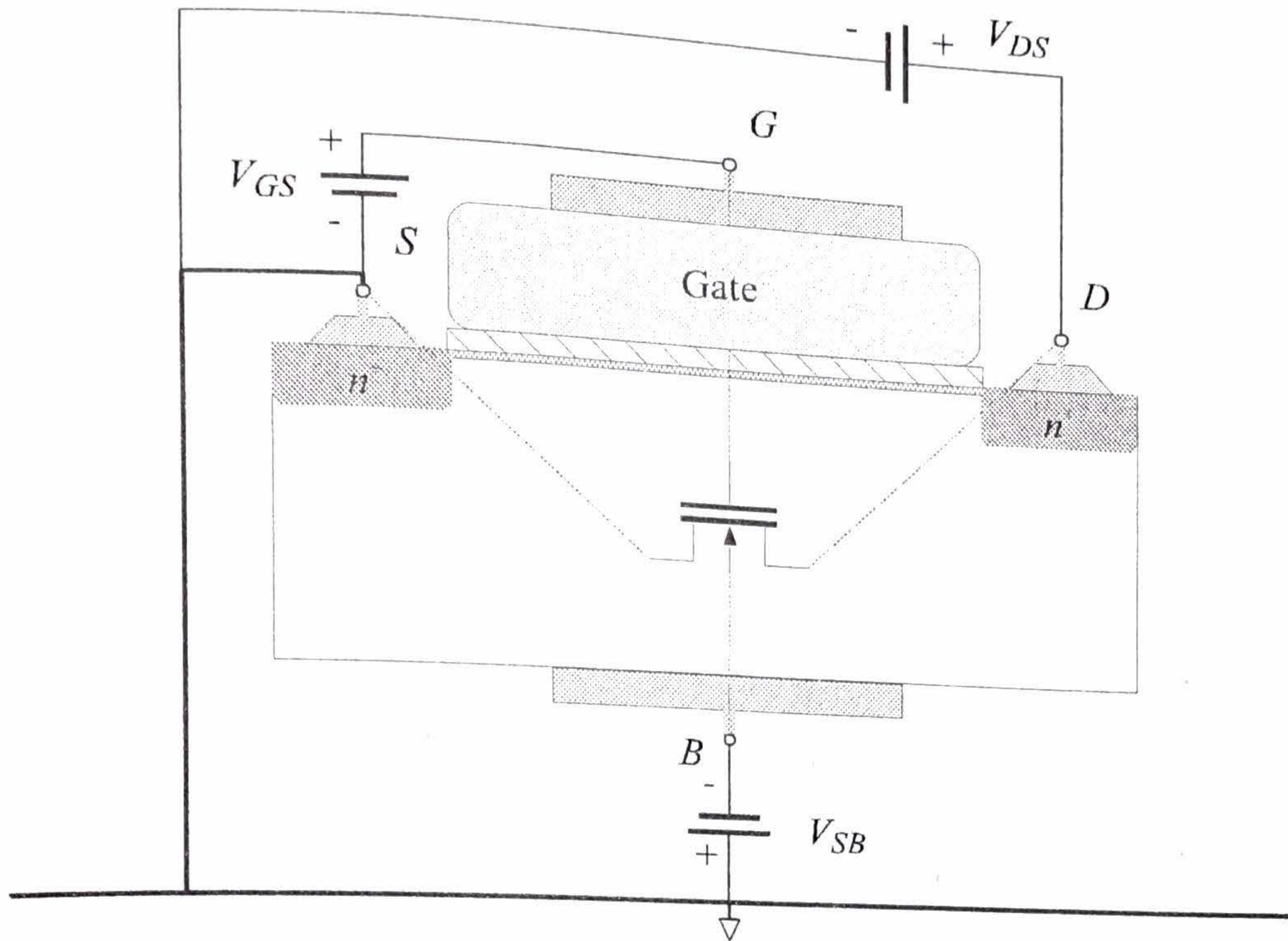


Figura 5.12: Transistor NMOS referido al potencial de source.

$$Q_{vac}(y) = -\sqrt{-2\phi_p + \tilde{v}_c(y) + v_{SB}} \quad (5.59)$$

En el inicio de inversión, la tensión v_{GB} satisface:

$$v_{GB}^* = V_{FB} - 2\phi_p + \tilde{v}_c(y) + v_{SB} + \gamma\sqrt{-2\phi_p + \tilde{v}_c(y) + v_{SB}} \quad (5.60)$$

y considerando que $v_{GB}^* = v_{GS}^* + v_{SB}$, la carga de inversión, dada por (5.23), se puede escribir como:

$$Q_i(y) = -C_{ox} (v_{GS} - v_{GS}^*) = -C_{ox} (v_{GS} - V_T(y) - \tilde{v}_c(y)) \quad (5.61)$$

donde:

$$V_T(y) = V_{FB} - 2\phi_p + \gamma\sqrt{-2\phi_p + \tilde{v}_c(y) - v_{BS}} \quad (5.62)$$

se denomina tensión de umbral referida a la tensión de gate-source.

La tensión $V_T(y)$ puede linealizarse en un entorno de $\tilde{v}_c(y) = 0$, obteniéndose:

$$V_T(y) = V_T + \frac{\gamma}{2\sqrt{-2\phi_p + V_{SB}}} \tilde{v}_c(y) \quad (5.63)$$

donde

$$V_T = V_{FB} - 2\phi_p + \gamma\sqrt{-2\phi_p - V_{BS}} \quad (5.64)$$

La corriente puede obtenerse integrando la carga de inversión, tal como en (5.38), de donde se obtiene:

$$i_D = \mu_n C_{ox} \frac{W}{L} \int_0^{v_{DS}} (v_{GS} - V_T - \alpha \tilde{v}_c) d\tilde{v}_c \quad (5.65)$$

donde

$$\alpha \triangleq 1 + \frac{\gamma}{2\sqrt{-2\phi_p - V_{BS}}} \quad (5.66)$$

La solución de la integral puede expresarse como:

$$i_D = \mu_n C_{ox} \frac{W}{L} \left((v_{GS} - V_T) \tilde{v}_c \Big|_0^{v_{DS}} - \frac{1}{2} \alpha \tilde{v}_c^2 \Big|_0^{v_{DS}} \right) \quad (5.67)$$

cuya solución es:

$$i_D = \mu_n C_{ox} \frac{W}{L} \left((v_{GS} - V_T) v_{DS} - \frac{1}{2} \alpha v_{DS}^2 \right) \quad (5.68)$$

y representa la expresión de la corriente en la zona de triodo.
La saturación se alcanza cuando

$$\frac{di_D}{dv_{DS}} = K_n (v_{GS} - V_T - \alpha v_{DS}) = 0 \quad (5.69)$$

de donde se obtiene el valor de la tensión de saturación:

$$V_{DS(sat)} = \frac{v_{GS} - V_T}{\alpha} \quad (5.70)$$

A partir de este valor de tensión, el transistor abandona la zona de triodo y pasa a funcionar en la región de saturación, con un valor de corriente:

$$i_D = \frac{K_n}{2\alpha} (v_{GS} - V_T)^2 \quad (5.71)$$

En el caso que se considere la corrección por el efecto Early, la corriente es:

$$i_D = \frac{K_n}{2\alpha} (v_{GS} - V_T)^2 \left(1 + \frac{v_{DS} - V_{DS(sat)}}{V_A} \right) \quad (5.72)$$

Una forma alternativa de escribir la tensión de umbral (5.64), comúnmente utilizada en la literatura, es la siguiente:

$$\begin{aligned} V_T &= -2\phi_p + V_{FB} + \gamma\sqrt{-2\phi_p} - \gamma\sqrt{-2\phi_p} + \gamma\sqrt{-2\phi_p - V_{BS}} \\ &= V_{T0} + \gamma \left(\sqrt{-2\phi_p - V_{BS}} - \sqrt{-2\phi_p} \right) \end{aligned} \quad (5.73)$$

Como se ve claramente de (5.73), el factor γ , relaciona los cambios de v_{BS} con los cambios en la tensión de umbral. Dado que un cambio en la tensión de umbral produce un cambio en la corriente, el terminal de sustrato puede considerarse como un segundo gate, comúnmente denominado "back-gate". La Fig. 5.13 muestra la característica de transferencia medida de un transistor NMOS, para distintos valores de tensión $V_{SB} = -V_{BS}$. Puede notarse de esta figura como la tensión de umbral aumenta conforme aumenta V_{SB} .

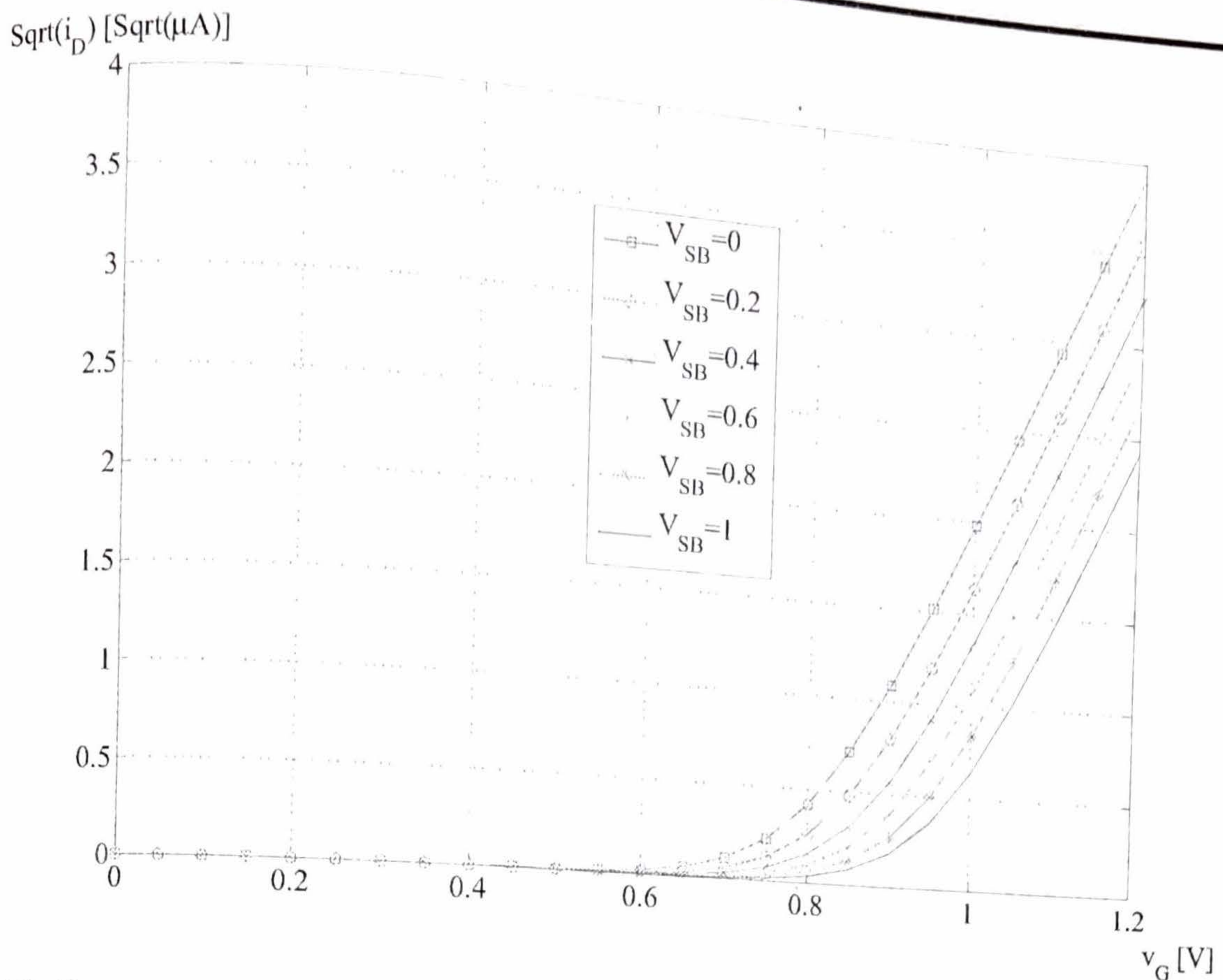


Figura 5.13: Curvas experimentales de transferencia $\sqrt{i_D}$ vs. v_{GS} , para distintos valores de V_{SB} , de un transistor NMOS de $W = 5\mu\text{m}$ y $L = 5\mu\text{m}$, en una tecnología de $0,5\mu\text{m}$.

5.2.4. Desviaciones del comportamiento ideal

Modulación de la longitud del canal

En los desarrollos de la corriente del transistor MOS de las secciones pasadas, se consideró que cuando la carga de inversión se anulaba, $Q_i(y) = 0$, lo hacía en el extremo del drain, $y = L$, y a partir de allí, incrementos posteriores de la tensión de drain no modificaban la longitud del canal $y = L$, ni la carga de inversión, por lo tanto, la corriente de drain permanecía constante.

Las curvas de salida de un transistor real, en la zona de saturación, presentan un fenómeno denominado *modulación de la longitud del canal*, que provocan una desviación del comportamiento ideal. En efecto, cuando se produce la saturación del transistor ($v_D = V_{D(sat)}$), la carga de inversión en el extremo del drain se hace nula. A partir de este punto, si se aumenta la tensión de drain, se forma una región de vaciamiento horizontal en el extremo del drain, que soporta la diferencia de tensión $v_D - V_{D(sat)}$. Es importante notar que la desaparición de la carga de inversión, con la subsecuente aparición de la región de vaciamiento, no impide el paso de los portadores de source a drain, por lo cual la corriente continúa circulando⁶. La zona de vaciamiento formada, provoca un acortamiento efectivo del canal, tal como se ilustra en la Fig. 5.14-b, de la forma:

$$L = L - \Delta L \quad (5.74)$$

Una expansión de Taylor de $1/L$ da como resultado:

⁶A modo de ejemplo, el lector puede recordar que la juntura semiconductor tiene una zona de vaciamiento por la cual circula corriente, tanto en conducción directa (corriente directa), como en conducción inversa (corriente de fuga).

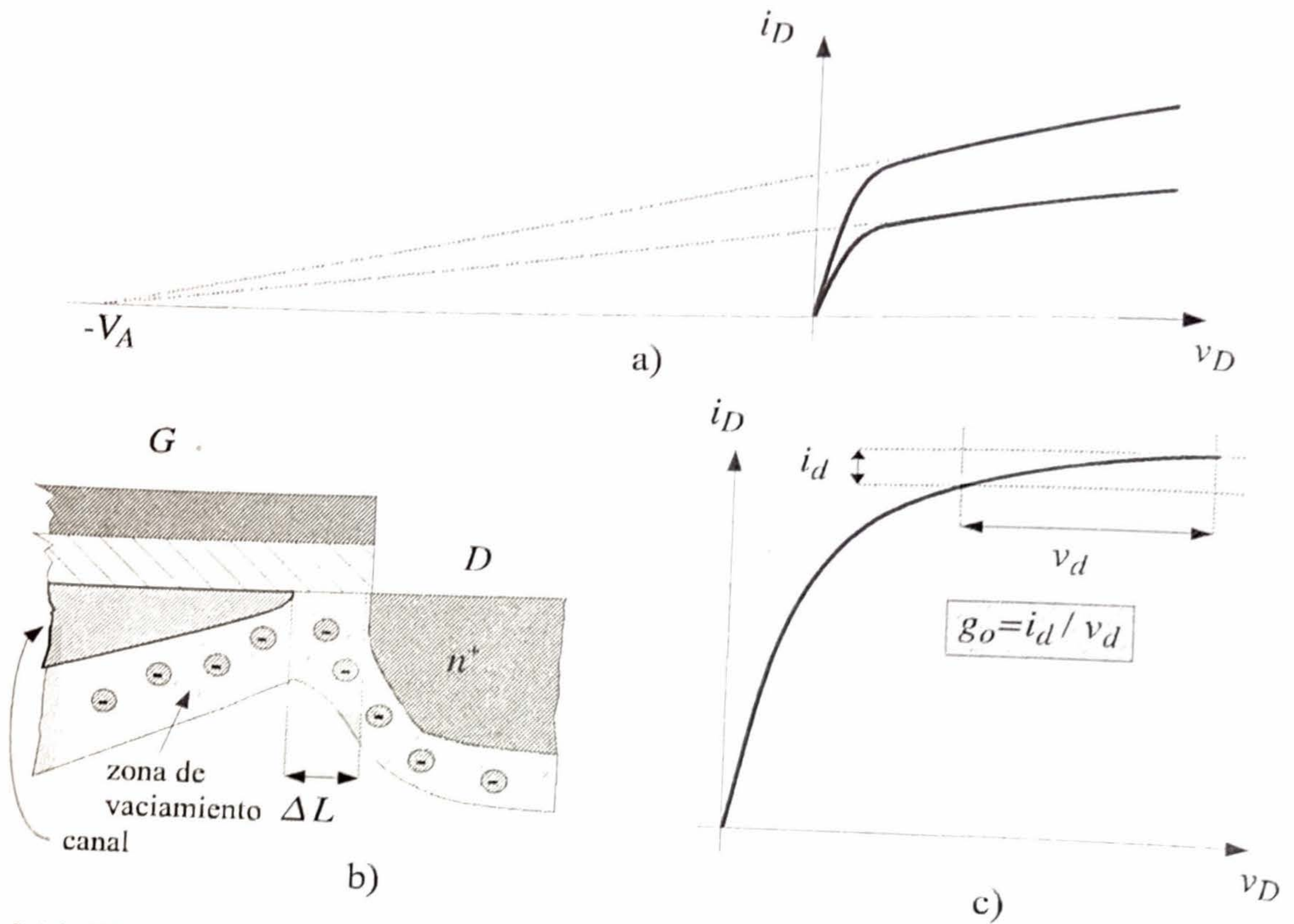


Figura 5.14: Efecto de modulación de canal en un transistor NMOS: a) curvas de salida, resaltando la intersección imaginaria de las mismas en $v_D = -V_A$; b) acortamiento del canal en el extremo del drain; c) incrementos de tensión y corriente para calcular la conductancia de salida.

$$\frac{1}{L} = \frac{1}{L} - \frac{1}{L^2} \Delta L = \frac{1}{L} \left(1 - \frac{\Delta L}{L} \right) \quad (5.75)$$

Por otro lado, la magnitud del acortamiento del canal es proporcional a la diferencia entre la tensión de drain y la de saturación, es decir $\Delta L = -\xi (v_D - V_{D(sat)})$, con lo cual⁷

$$1 - \frac{\Delta L}{L} = 1 + \frac{\xi (v_D - V_{D(sat)})}{L} = 1 + \frac{v_D - V_{D(sat)}}{L/\xi} \quad (5.77)$$

Se define la tensión de Early [16] de la siguiente manera⁸:

$$V_A \triangleq \frac{L}{\xi} \quad (5.78)$$

Luego, la corriente en saturación se puede escribir como:

⁷El valor de ξ puede aproximarse de la siguiente manera [15]:

$$\xi^{-1} = \phi_A \sqrt{\frac{N_c}{10^{16}}} \quad (5.76)$$

donde ϕ_A depende del proceso, y N_c es el dopado efectivo en el canal, que suele ser algo superior a N_a debido a la introducción intencional de impurezas. Un valor de compromiso es $\phi_A = 16V$, y $N_c \approx N_a$.

⁸El efecto de Early debe su nombre al ingeniero James M. Early (1922-2004). Es posible también hallar en la literatura el factor recíproco de la tensión de Early: $\lambda \triangleq 1/V_A$

$$i_D = \frac{K}{2n}(v_G - V_{T0})^2 \left(1 - \frac{\Delta L}{L}\right) = \frac{K}{2n}(v_G - V_{T0})^2 \left(1 + \frac{v_D - V_{D(sat)}}{V_A}\right) \quad (5.79)$$

El efecto de la modulación de la longitud del canal, tal como se ve de (5.79), es el aumento de la corriente con la tensión v_D , que se refleja en el aumento de la pendiente de las curvas de salida, tal como se muestra en la Fig. 5.14-c. La tensión de Early se interpreta gráficamente como el punto de cruce de la prolongación de las curvas de corriente de salida con el eje $i_D = 0$, tal como se muestra en la Fig. 5.14-a.

Ejemplo 20 Recalcule el valor de la corriente de saturación para el Ejemplo 22, teniendo en cuenta $V_D = 5V$. La tensión de saturación es $V_{D(sat)} = 1,55V$. La tensión de Early puede aproximarse utilizando:

$$\xi^{-1} = 16 \times \sqrt{\frac{1,7 \times 10^{17}}{10^{16}}} = 65,96 \mu m^{-1} \quad (5.80)$$

con lo cual

$$V_A = 65,96 \times 5 = 329,84V \quad (5.81)$$

La corriente máxima en saturación se calcula como:

$$I_D = 483,28 \times \left(1 + \frac{5 - 1,55}{329,84}\right) = 488,33 \mu A \quad (5.82)$$

En este caso, el acortamiento de canal es:

$$\Delta L = (5 - 1,55)/65,96 = 0,052 \mu m \quad (5.83)$$

Efectos de la temperatura

En un transistor MOS, la temperatura afecta la movilidad de los portadores y la tensión de umbral. La movilidad efectiva se reduce con la temperatura, de acuerdo a la siguiente ecuación:

$$\mu(T) = \mu(T_{ref}) \left(\frac{T}{T_{ref}}\right)^{-k_1} \quad (5.84)$$

donde T es la temperatura absoluta, T_{ref} es la temperatura ambiente absoluta, y k_1 es una constante, cuyo valor satisface $1,5 \leq k_1 \leq 2$.

Por otro lado, la tensión de umbral reduce su valor con la temperatura, de acuerdo a:

$$V_T(T) = V_T(T_{ref}) - k_2(T - T_{ref}) \quad (5.85)$$

donde la constante k_2 satisface $0,5mV/K \leq k_2 \leq 4mV/K$. El aumento de temperatura tiende a producir un aumento de la corriente de drain, debido a la disminución de la tensión de umbral, y una disminución de la corriente, debido a la disminución de la movilidad. Estos dos efectos son contrapuestos. A pequeñas corrientes, la disminución de la tensión de umbral prevalece y la corriente aumenta con la temperatura. A corrientes mayores, el efecto de la disminución de la movilidad prevalece y la corriente disminuye con la temperatura. Existe un punto intermedio, donde ambos efectos se cancelan, produciendo variación nula de la corriente.

Corriente sub-umbral

El modelo desarrollado predice que, para $v_{GS} = V_T$, la corriente de drain i_D se hace nula. En realidad, para valores menores a la tensión de umbral, existe conducción de corriente por difusión de portadores [17], y la corriente es de la forma:

$$i_D = I_0 \exp\left(\frac{v_{GS}}{nU_T}\right) \quad (5.86)$$

En este régimen, caracterizado por corrientes de operación extremadamente bajas, la característica del transistor MOS es similar a la de un transistor bipolar, y presenta una tensión de saturación de salida independiente de la corriente, $V_{DS(sat)} \approx 4U_T$. Estas características hacen este régimen especialmente atractivo para diseños de bajo consumo de potencia y bajas tensiones de alimentación.

La corriente I_0 , en (5.86), es la corriente cuando $v_{GS} = 0V$, y su valor es del orden de los picoamperes. A esta corriente, no prevista en el modelo desarrollado, debe añadirse la corriente de pérdida de las junturas (formadas por las regiones de drain y source con el sustrato), cuyo valor, como ya se vio, depende fuertemente de la temperatura.

5.3. El transistor PMOS

5.3.1. Principio básico de funcionamiento

Para el análisis se considerará el transistor PMOS cuyo corte se muestra en la Fig. 5.15, con dimensiones W (ancho) y L (longitud) como se indica en la Fig. 5.16. En este caso, también se parte de una pieza de Silicio que tiene inicialmente un dopado leve tipo P, y a los efectos de formar el sustrato adecuado se define, por compensación, una zona tipo N denominada "well"⁹. El well surge a partir de la necesidad de fabricar ambos tipos de transistores sobre una misma pieza de Silicio. No se hará ninguna suposición sobre el dopado del gate (se dejará la tensión de Flatband expresada en forma genérica).

En el caso del transistor PMOS, las tensiones v_S y v_D son negativas, de manera que la juntura que se forma entre las regiones de source y drain (ánodo) y el sustrato (cátodo) están en inversa. Para que los portadores (huecos) vayan de source a drain, será necesario cumplir $v_D < v_S$.

La ecuación de corriente de arrastre de huecos es:

$$j_p^a = qp(x, y)v_p^a(y) \quad (5.87)$$

La corriente se obtiene a partir de (5.87), integrando con respecto a x y a z :

$$\begin{aligned} i_p^a &= \int_0^W \int_0^{\Delta x} qp(x, y)v_p^a(y) dx dz \\ &= Wv_p^a(y) \int_0^{\Delta x} qp(x, y) dx \end{aligned} \quad (5.88)$$

El término integral en (5.88) es la carga de inversión en el canal, para un punto y dado, es decir:

$$Q_i(y) = \int_0^{\Delta x} qn(x, y) dx \quad (5.89)$$

⁹En el caso en que la oblea esté dopada inicialmente tipo N, los transistores PMOS se pueden fabricar directamente sobre el sustrato, mientras que los NMOS requieren de la realización de un well tipo P.

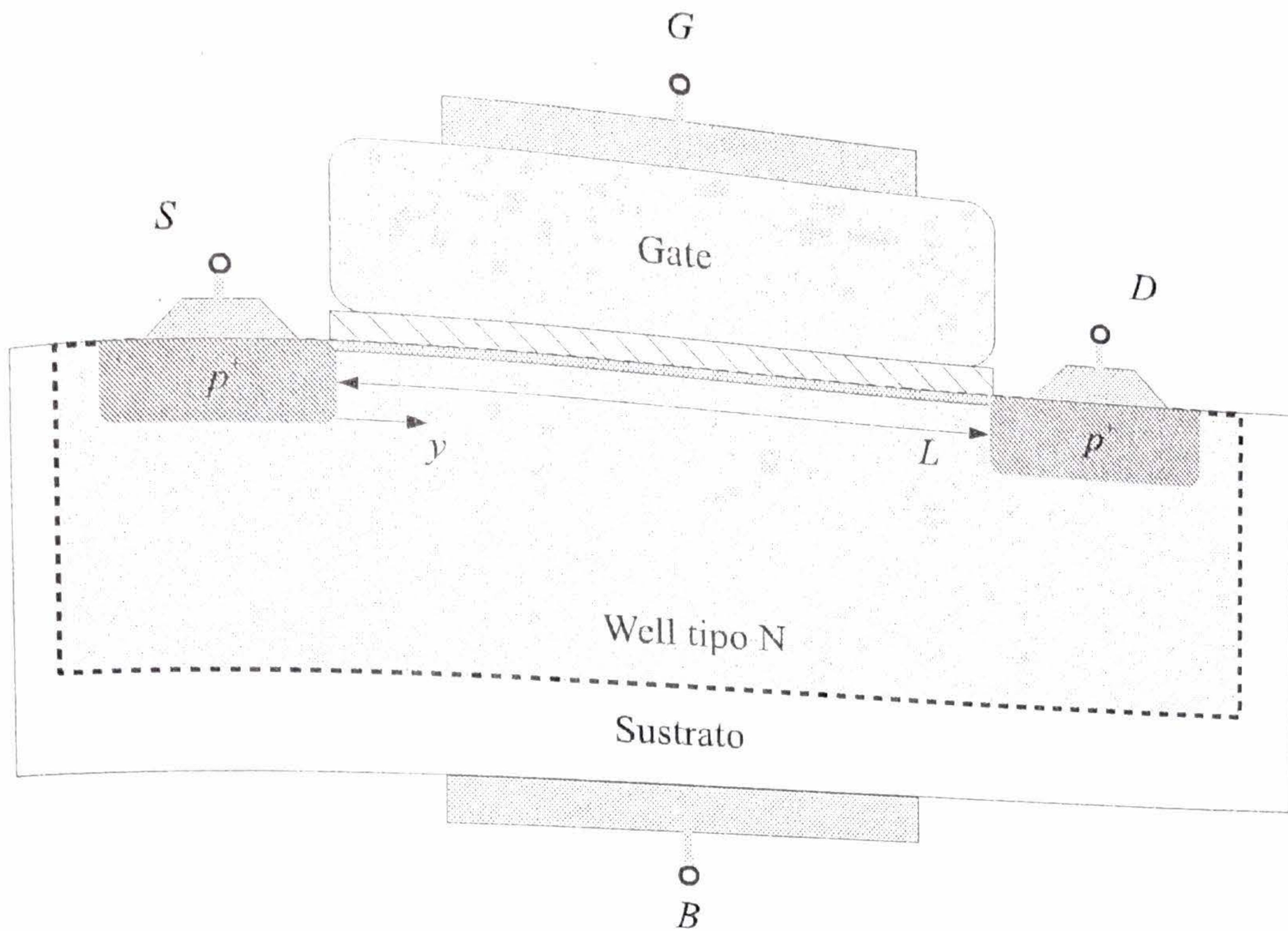


Figura 5.15: Corte longitudinal de un transistor PMOS de longitud L .

En este caso, esta carga es positiva y corresponde a huecos. La corriente puede expresarse de manera general como:

$$i_p^a = W v_p^a(y) Q_i(y) \quad (5.90)$$

Teniendo en cuenta ahora la relación entre la velocidad de los portadores y el campo eléctrico, y que $E = -d\phi_s/dy$, donde ϕ_s es el potencial de los portadores en el canal, resulta:

$$v_p^a = \mu_p E = -\mu_p \frac{d\phi_s}{dy} \quad (5.91)$$

y la corriente puede escribirse como:

$$i_p^a = -W Q_i(y) \mu_p \frac{d\phi_s}{dy} \quad (5.92)$$

Teniendo en cuenta que la corriente de drain circula en sentido opuesto al eje y , su expresión es:

$$i_D = -i_p^a = W Q_i(y) \mu_p \frac{d\phi_s}{dy} \quad (5.93)$$

Al igual que en el caso anterior, se procede, a continuación, a hallar una relación entre la carga de inversión Q_i y la variación del potencial del canal, en función de la tensión de los terminales. Se describirán también los dos modelos posibles, resultantes de utilizar el sustrato, o el source, como terminal de referencia.

5.3.2. Modelo referido al Sustrato

Para el cálculo de la corriente, durante esta sección, se tomará el sustrato como punto de referencia de las tensiones aplicadas, tal como se muestra en la Fig. 5.17.

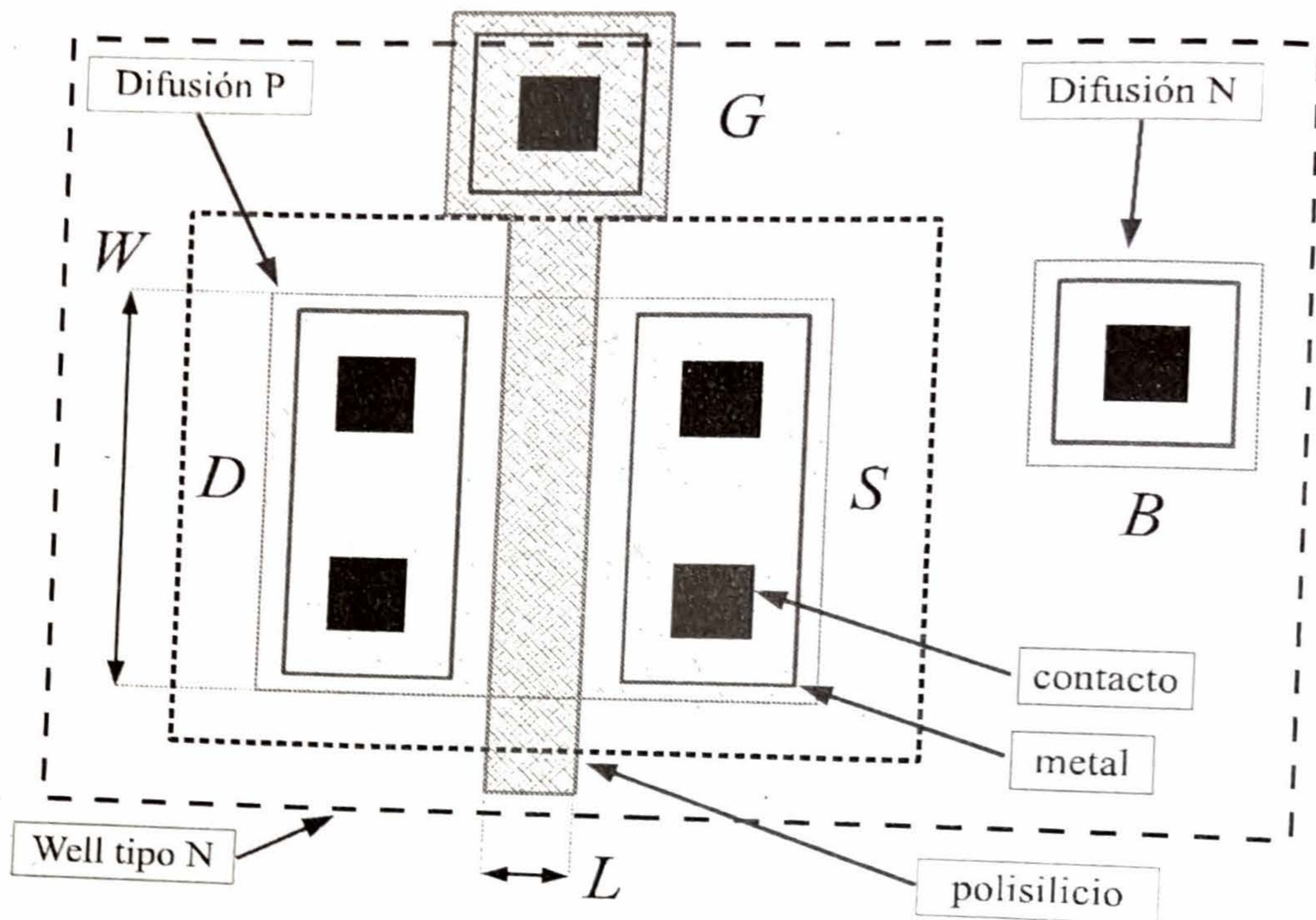


Figura 5.16: Esquema de un transistor PMOS de longitud L y ancho W .

En el caso de un transistor PMOS sin potencial aplicado a los terminales de source/drain, y suponiendo que el mismo se encuentra en inversión, es decir, que la tensión $v_G \leq V_T$ ¹⁰, se cumple que:

$$v_{vac}(0) = -2\phi_n \quad (5.94)$$

$$v_{vac}(L) = -2\phi_n \quad (5.95)$$

Al aplicar potenciales a los terminales de source y drain, v_S y v_D , respectivamente, el potencial resultante en la zona de vaciamiento es¹¹:

$$v_{vac}(y) = -2\phi_n + v_c(y) \quad (5.96)$$

donde $v_c = \phi_s - (-\phi_n)$ describe el potencial de los portadores en el canal con respecto a su valor en inversión, dado por $-\phi_n$, y satisface:

$$v_{vac}(0) = -2\phi_n + v_S \quad (5.97)$$

$$v_{vac}(L) = -2\phi_n + v_D \quad (5.98)$$

La longitud de la zona de vaciamiento está dada por:

$$x_{vac}(y) = \sqrt{\frac{2\epsilon_{Si}(2\phi_n - v_c(y))}{qN_d}} \quad (5.99)$$

¹⁰La tensión de umbral para un transistor PMOS es negativa, y la inversión se da cuando la tensión de gate es aún más negativa.

¹¹Nótese que v_c es negativo al igual que v_S y v_D .

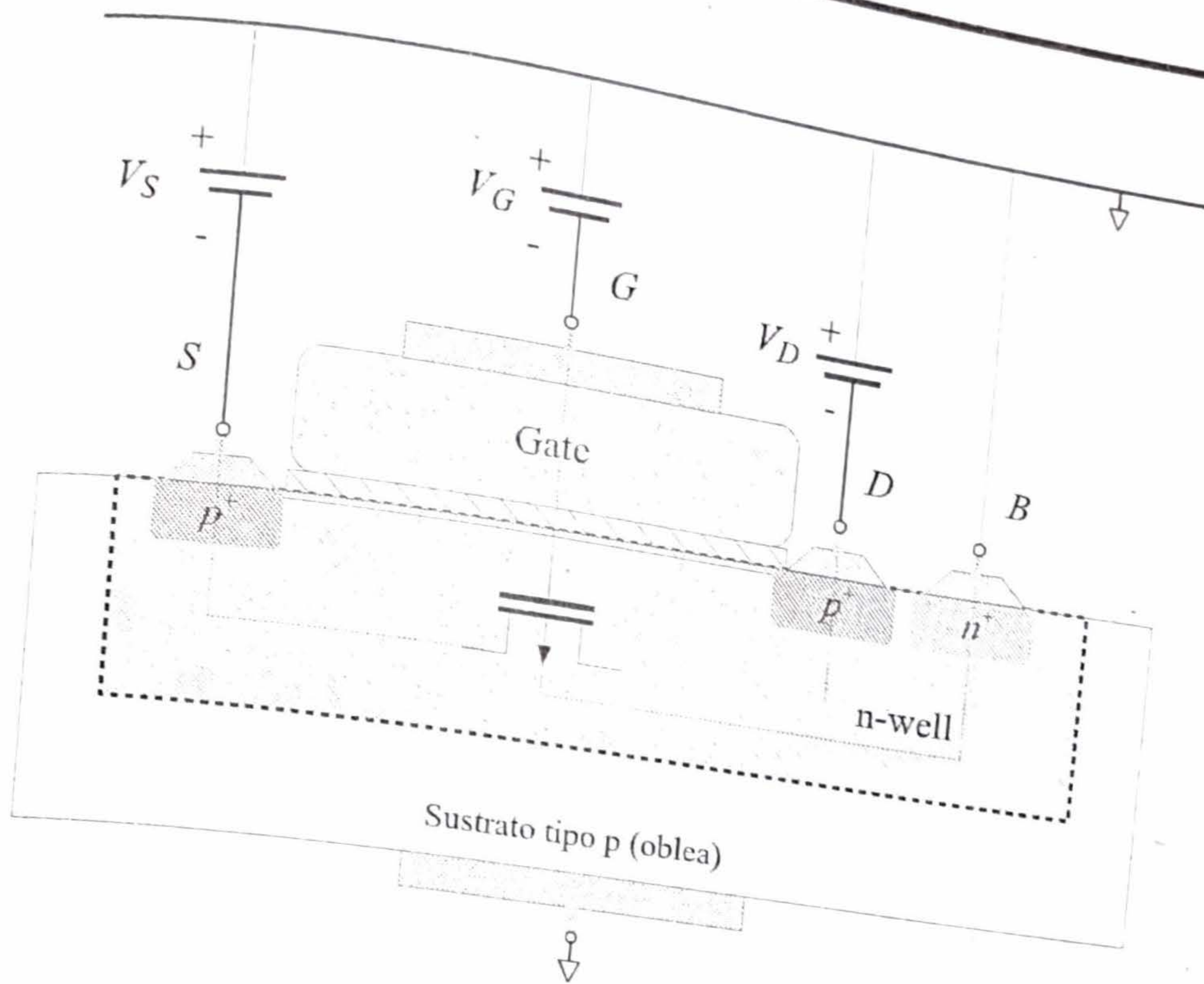


Figura 5.17: Transistor PMOS con tensiones referidas al sustrato.

donde N_d es la concentración de impurezas en el sustrato del transistor (N-well). La carga de vaciamiento es:

$$Q_{vac}(y) = \sqrt{2qN_d\epsilon_{Si}(2\phi_n - v_c(y))} \quad (5.100)$$

Para que un punto genérico del canal, y , esté justo en el inicio de inversión, la tensión a lo largo del óxido (negativa) debe ser:

$$v_{ox}^*(y) = -\frac{1}{C_{ox}} \sqrt{2qN_d\epsilon_{Si}(2\phi_n - v_c(y))} \quad (5.101)$$

Planteando mallas:

$$\begin{aligned} v_G^* &= v_{ox}^* + v_{vac}(y) + V_{FB} \\ &= -\frac{1}{C_{ox}} \sqrt{2qN_d\epsilon_{Si}(2\phi_n - v_c(y))} + (-2\phi_n + v_c(y)) + V_{FB} \end{aligned} \quad (5.102)$$

En este caso, la tensión de umbral se define de la siguiente manera:

$$V_T(y) \triangleq -2\phi_n + V_{FB} - \frac{1}{C_{ox}} \sqrt{2qN_d\epsilon_{Si}(2\phi_n - v_c(y))} \quad (5.103)$$

Luego, la Ec. (5.102) puede escribirse como:

$$v_G^* = V_T(y) + v_c(y) \quad (5.104)$$

y la carga de inversión, que en este caso es positiva, resulta de la forma:

$$Q_i(y) = -C_{ox} (v_G - V_T(y) - v_c(y)) \quad (5.105)$$

La expresión genérica de la corriente de drain es:

$$i_D = -\mu_p C_{ox} \frac{W}{L} \int_{V_S}^{V_D} (v_G - V_T(y) - v_c(y)) dv_c \quad (5.106)$$

Tal como en el caso del transistor NMOS, la corriente puede interpretarse como el área entre la recta v_G y la curva $V_T(y) - v_c(y)$, dentro de los límites impuestos por V_S y V_D . La única diferencia es que todos los valores, en este caso, son negativos.

Linealizando la expresión de la tensión de umbral se llega a:

$$V_T(y) = V_{T0} - \frac{\gamma}{\sqrt{2\phi_n}} v_c \quad (5.107)$$

donde

$$V_{T0} \triangleq -2\phi_n + V_{FB} - \frac{1}{C_{ox}} \sqrt{2qN_d\epsilon_{Si}2\phi_n} \quad (5.108)$$

y el factor de back-gate es:

$$\gamma \triangleq \frac{\sqrt{2qN_d\epsilon_{Si}}}{C_{ox}} \quad (5.109)$$

Luego, el argumento de la integral (5.106), se puede escribir como:

$$v_G - V_T(y) - v_c(y) = v_G - V_{T0} - nv_c \quad (5.110)$$

Utilizando (5.110), la corriente resulta:

$$i_D = -\mu_p C_{ox} \frac{W}{L} \int_{v_S}^{v_D} (v_G - V_{T0} - nv_c) dv_c \quad (5.111)$$

Para hallar la solución de la integral, se debe plantear:

$$i_D = -\mu_p C_{ox} \frac{W}{L} \left((v_G - V_{T0})v_c \Big|_{v_S}^{v_D} - \frac{1}{2}nv_c^2 \Big|_{v_S}^{v_D} \right) \quad (5.112)$$

cuyo desarrollo resulta en:

$$i_D = -\mu_p C_{ox} \frac{W}{L} \left((v_G - V_{T0})(v_D - v_S) - \frac{1}{2}n(v_D^2 - v_S^2) \right) \quad (5.113)$$

que es la expresión de la corriente en la región de triodo.

El potencial de estrangulamiento o "pinch-off" es:

$$V_p \triangleq \frac{v_G - V_{T0}}{n} \quad (5.114)$$

Para este potencial, la carga de inversión en el drain se iguala a cero:

$$Q_i = -C_{ox}(v_G - V_{T0} - nV_p) = 0 \quad (5.115)$$

La región de funcionamiento en la cual $v_D \geq V_{D(sat)}$ ¹², se denomina zona lineal, o de triodo, y se caracteriza por la dependencia de la corriente, tanto de la tensión de drain v_D , como de la tensión de gate v_G . Para valores menores a $V_{D(sat)}$, el dispositivo está en la zona de funcionamiento denominada de saturación, la cual se caracteriza por la independencia de la corriente con respecto a la tensión de drain:

$$i_D = -\frac{\mu_p C_{ox} W}{2n L} (v_G - V_T - n v_S)^2 \quad (5.116)$$

En el caso que la tensión de source sea igual a cero, las expresiones de triodo y saturación se reducen a:

$$i_D = \begin{cases} K_p \left((v_G - V_T) v_D - \frac{n v_D^2}{2} \right) & \text{si } v_D > V_{D(sat)} \\ \frac{K_p}{2n} (v_G - V_T)^2 & \text{si } v_D \leq V_{D(sat)} \end{cases} \quad (5.117)$$

donde

$$V_{D(sat)} = \frac{v_G - V_T}{n} \quad (5.118)$$

y

$$K'_p = -\mu_p C_{ox} \quad (5.119)$$

$$K_p = \frac{W}{L} K'_p \quad (5.120)$$

En la región de triodo, que se ilustra en la Fig. 5.18, las curvas de salida comienzan en el origen, y disminuyen en forma cuadrática con v_D , para una $v_G < V_T$ constante, alcanzando su valor mínimo cuando $v_D = V_{D(sat)} = (v_G - V_T)/n$. Si se disminuye el valor de v_G , se observa que la curva de corriente es menor, al igual que el punto de saturación, que ahora ocurre más a la izquierda. A partir de (5.117), es sencillo ver que el conjunto de puntos de la curva de salida, para los cuales se alcanza la saturación, satisface:

$$i_D = \frac{K_p}{2n} V_{D(sat)}^2 \quad (5.121)$$

Esta parábola determina el límite entre las zonas de triodo y saturación.

En la zona de saturación, las curvas de salida son constantes con respecto a la tensión v_D , como se puede ver de la Fig. 5.19. A medida que se aumenta v_G se obtienen curvas paralelas cuyo valor de corriente es:

$$i_D = \frac{K_p}{2n} (v_G - V_T)^2 \quad (5.122)$$

La curva de transferencia se ilustra en la Fig. 5.20.

5.3.3. Modelo referido al Source

En esta sección, se describe en forma sintética el modelo del transistor PMOS referido al terminal de source, cuya configuración se ilustra en la Fig. 5.21.

El potencial a lo largo de la zona de vaciamiento está dado por:

$$v_{vac}(y) = -2\phi_n + \tilde{v}_c(y) - v_{BS} \quad (5.123)$$

¹²Tanto v_D como $V_{D(sat)}$ son tensiones negativas.

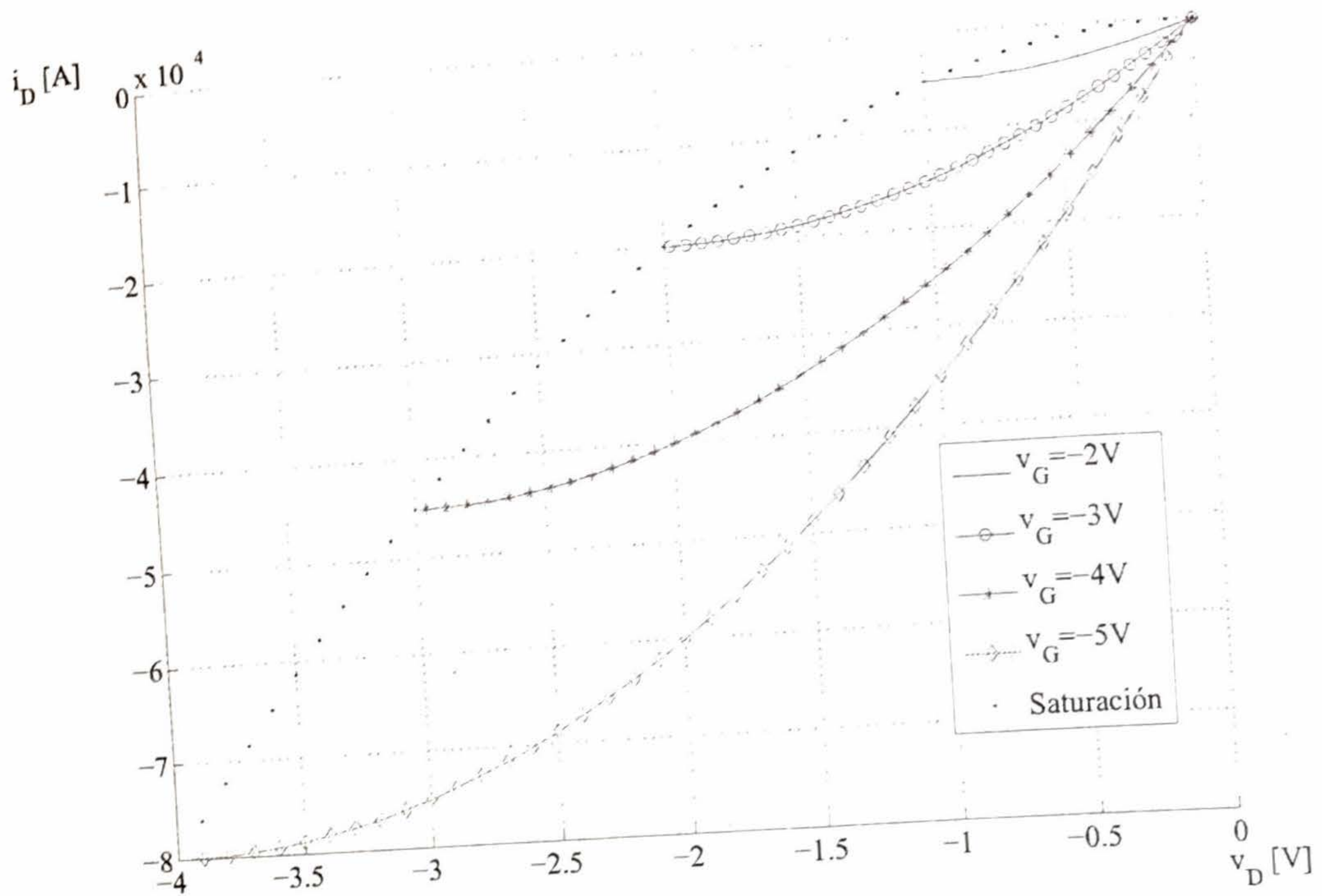


Figura 5.18: Curvas de salida de un transistor PMOS en la zona de triodo ($K = -100\mu\text{A}/\text{V}^2$, $V_T = -1\text{V}$, $n = 1$), considerando $v_S = 0\text{V}$. La línea punteada muestra el límite de la región de triodo, dado por (5.121).

donde $v_{BS} > 0$, y la carga asociada¹³ es:

$$Q_{vac}(y) = \sqrt{2qN_d\epsilon_{Si}(2\phi_n - \tilde{v}_c(y) + v_{BS})} \quad (5.124)$$

donde ahora, el potencial del canal satisface $\tilde{v}_c(0) = 0\text{V}$, y $\tilde{v}_c(L) = v_{DS}$, de tal manera que (5.123) coincide con (5.96).

En el inicio de inversión, la tensión v_{GB} satisface:

$$v_{GB} = V_{FB} - 2\phi_n + \tilde{v}_c(y) + v_{SB} - \gamma\sqrt{2\phi_n - \tilde{v}_c(y) + v_{BS}} \quad (5.125)$$

y considerando que $v_{GB} = v_{GS} + v_{SB}$, la carga de inversión, dada por (5.105), se puede escribir como:

$$Q_i(y) = -C_{ox}(v_{GS} - V_T(y) - \tilde{v}_c(y)) \quad (5.126)$$

donde:

$$V_T(y) = V_{FB} - 2\phi_n - \gamma\sqrt{2\phi_n - \tilde{v}_c(y) + v_{BS}} \quad (5.127)$$

se denomina tensión de umbral referida a la tensión de gate-source.

La tensión $V_T(y)$ puede linealizarse en un entorno de $\tilde{v}_c(y) = 0$, obteniéndose:

$$V_T(y) = V_T - \frac{\gamma}{2\sqrt{2\phi_n + V_{BS}}}\tilde{v}_c(y) \quad (5.128)$$

¹³Esta expresión coincide con (5.100), si se tiene en cuenta la diferencia entre los límites de variación de $\tilde{v}_c(y)$ y $v_c(y)$.

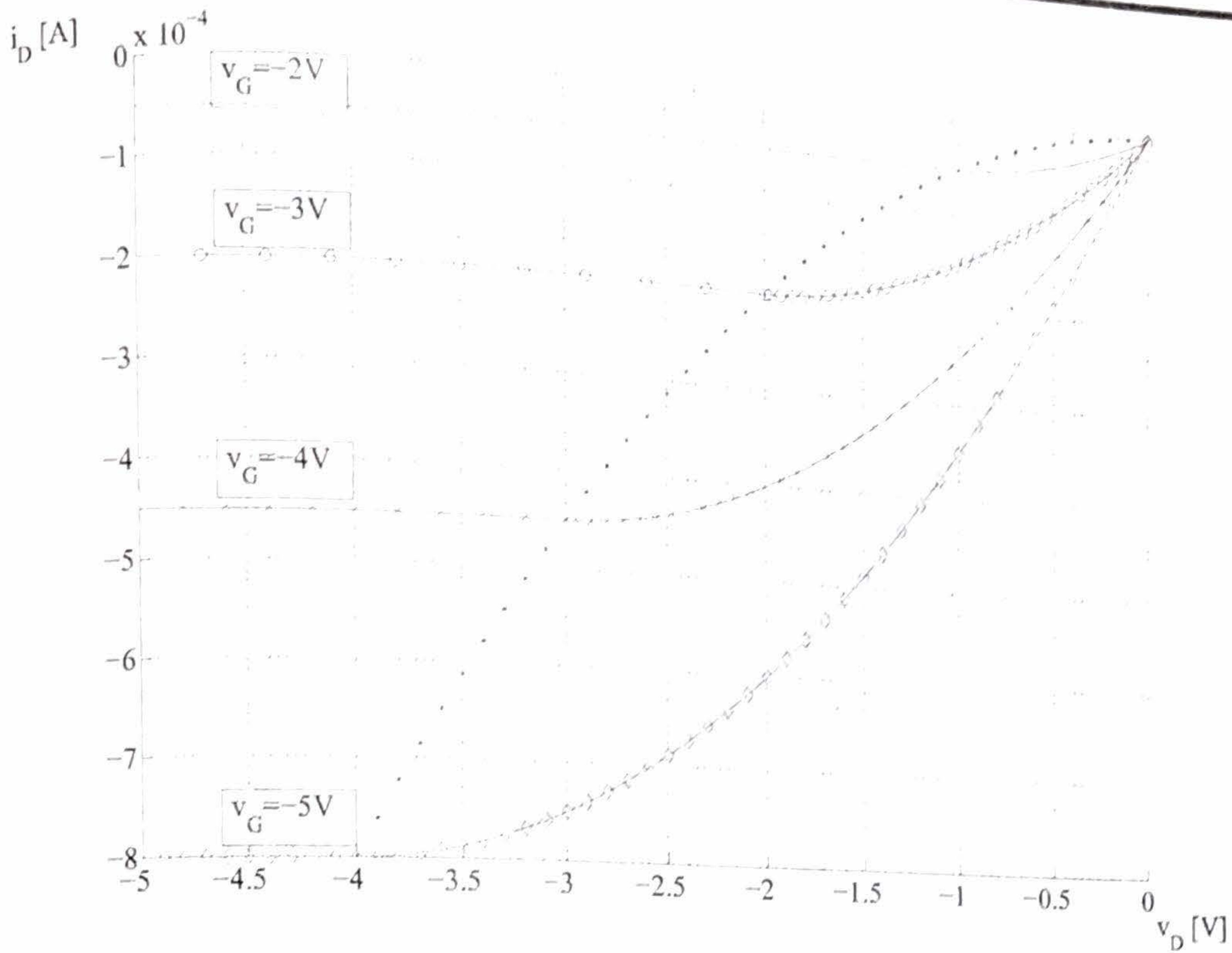


Figura 5.19: Característica de salida de un transistor PMOS ($K = -100\mu A/V^2$, $V_T = -1V$, $n = 1$), considerando $v_S = 0V$.

donde

$$V_T = V_{FB} - 2\phi_n - \gamma\sqrt{2\phi_n + V_{BS}} \quad (5.129)$$

La corriente puede obtenerse integrando la carga de inversión, tal como en (5.111), de donde resulta:

$$i_D = -\mu_p C_{ox} \frac{W}{L} \int_0^{v_{DS}} (v_{GS} - V_T - \alpha \tilde{v}_c) d\tilde{v}_c \quad (5.130)$$

donde

$$\alpha \triangleq 1 + \frac{\gamma}{2\sqrt{2\phi_n + V_{BS}}} \quad (5.131)$$

Para hallar la solución de la integral, se debe plantear:

$$i_D = -\mu_p C_{ox} \frac{W}{L} \left((v_{GS} - V_T) \tilde{v}_c \Big|_0^{v_{DS}} - \frac{1}{2} \alpha \tilde{v}_c^2 \Big|_0^{v_{DS}} \right) \quad (5.132)$$

cuyo desarrollo resulta en:

$$i_D = -\mu_p C_{ox} \frac{W}{L} \left((v_{GS} - V_T) v_{DS} - \frac{1}{2} \alpha v_{DS}^2 \right) \quad (5.133)$$

que es la expresión de la corriente en la zona de triodo.

La saturación se alcanza cuando

$$\frac{di_D}{dv_{DS}} = K_p (v_{GS} - V_T - \alpha v_{DS}) = 0 \quad (5.134)$$

de donde se obtiene el valor de la tensión de saturación:

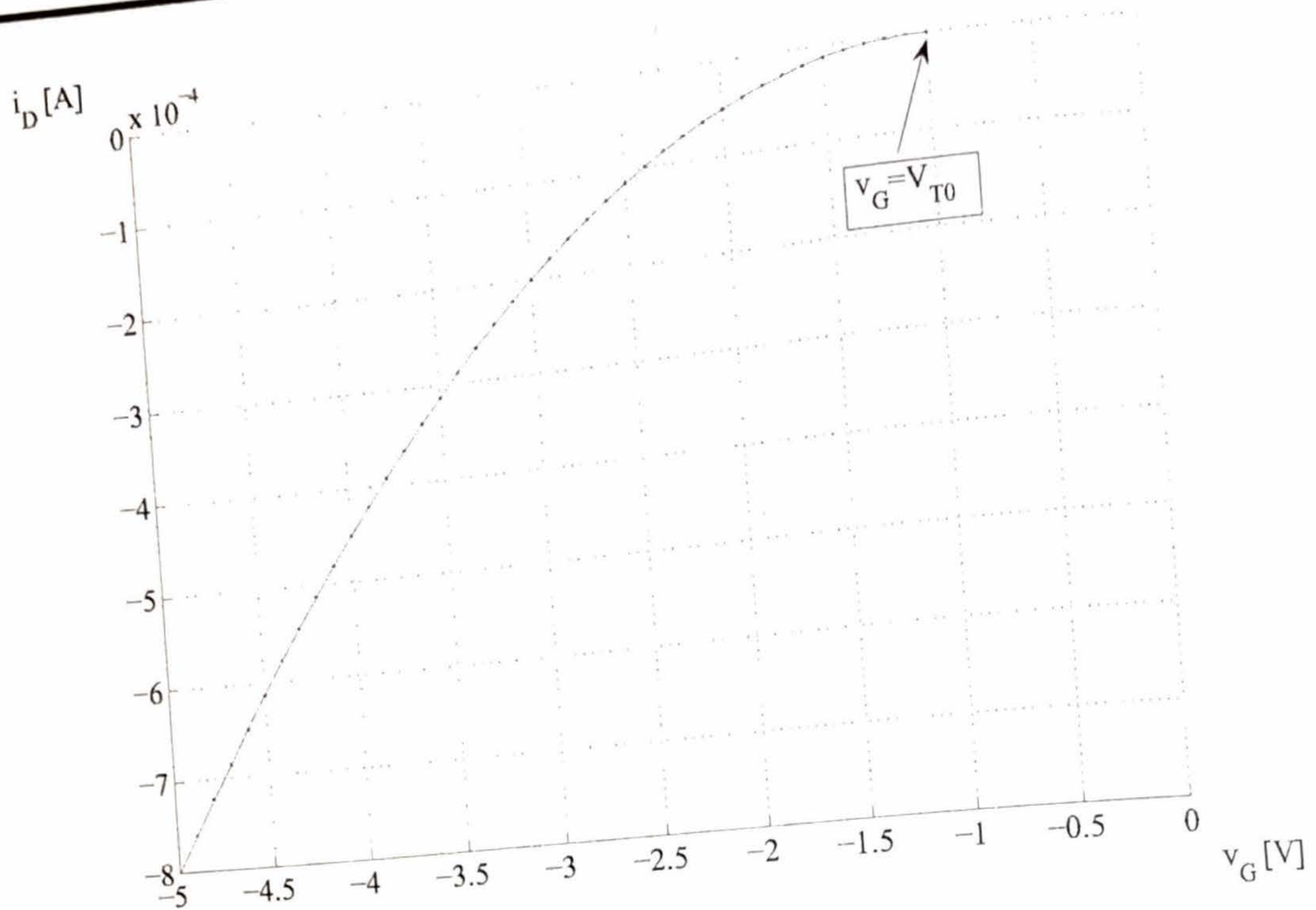


Figura 5.20: Curva de transferencia de un transistor PMOS ($K = -100\mu\text{A}/\text{V}^2$, $V_T = -1\text{V}$, $n = 1$), considerando $v_S = 0\text{V}$.

$$V_{DS(sat)} = \frac{v_{GS} - V_T}{\alpha} \quad (5.135)$$

A partir de este valor de tensión, el transistor abandona la zona de triodo y pasa a funcionar en la región de saturación, con un valor de corriente:

$$i_D = \frac{K_p}{2\alpha} (v_{GS} - V_T)^2 \quad (5.136)$$

En el caso que se considere la corrección por el efecto Early, la corriente es:

$$i_D = \frac{K_p}{2\alpha} (v_{GS} - V_T)^2 \left(1 + \frac{v_{DS} - V_{DS(sat)}}{V_A} \right) \quad (5.137)$$

Una forma alternativa de escribir la tensión de umbral (5.129), comúnmente utilizada en la literatura, es la siguiente:

$$\begin{aligned} V_T &= -2\phi_n + V_{FB} - \gamma\sqrt{2\phi_n} - \gamma\sqrt{2\phi_n + V_{BS}} \\ &= V_{T0} + \gamma \left(\sqrt{2\phi_n + V_{BS}} - \sqrt{2\phi_n} \right) \end{aligned} \quad (5.138)$$

La Fig. 5.22 muestra la característica de transferencia medida de un transistor PMOS, para distintos valores de tensión V_{BS} . En esta figura puede notarse claramente como la tensión de umbral disminuye (aumenta en magnitud) conforme aumenta V_{BS} .

5.3.4. Desviaciones del comportamiento ideal

Un transistor PMOS presenta también modulación de la longitud del canal, lo cual afecta la conducción de la corriente de drenaje, resultando de la forma:

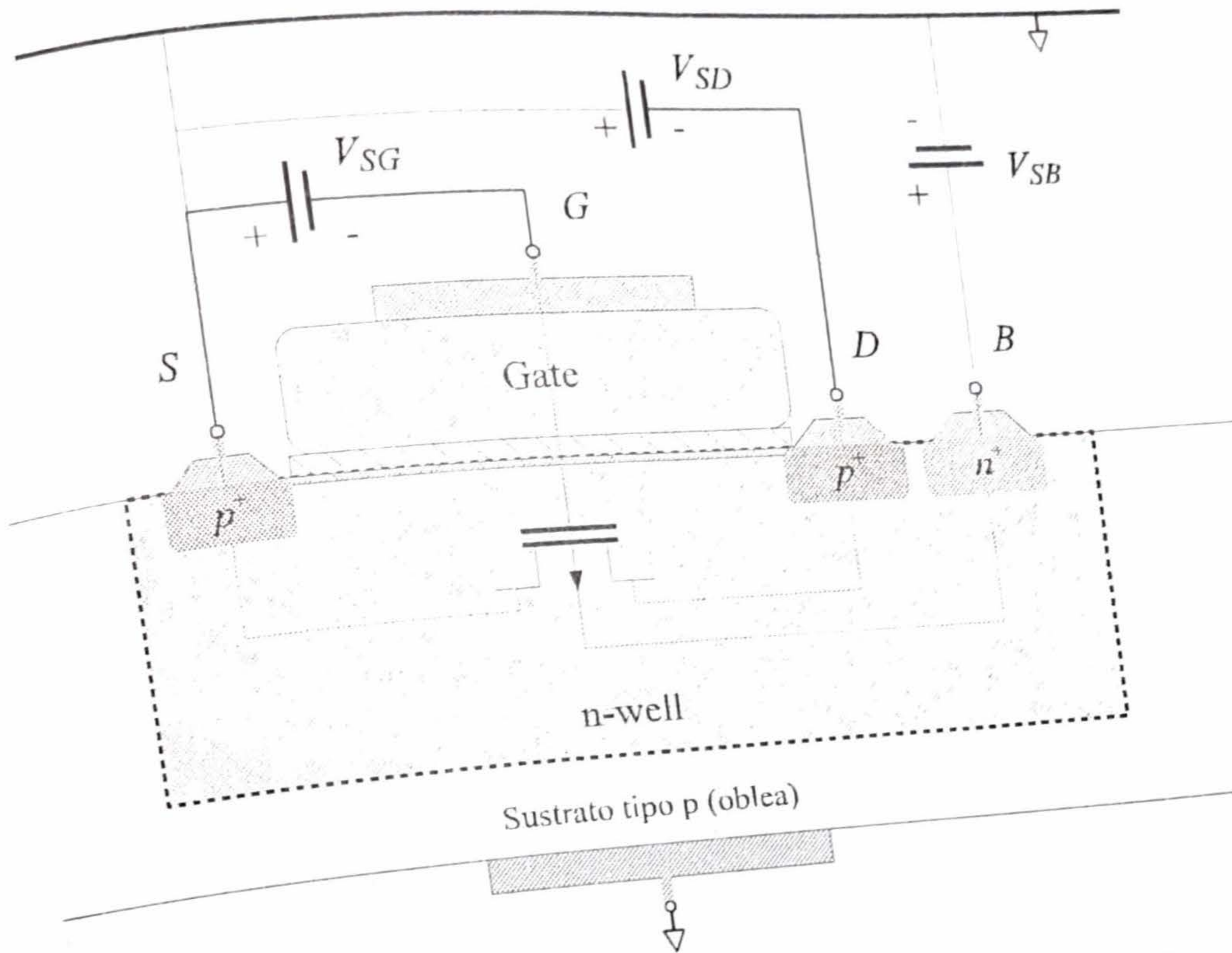


Figura 5.21: Transistor PMOS referido al potencial de source.

$$i_D = \frac{K_p}{2n} (v_G - V_T)^2 \left(1 + \frac{v_D - V_{D(sat)}}{V_A} \right) \quad (5.139)$$

donde ahora la tensión de Early es negativa:

$$V_A = \frac{L}{\xi} \quad (5.140)$$

La Fig. 5.23 muestra las curvas experimentales obtenidas de la medición de un transistor PMOS, de $W = 5\mu\text{m}$ y $L = 5\mu\text{m}$, en un proceso de $0,5\mu\text{m}$. En este caso, es posible apreciar con mayor claridad el efecto de Early.

5.4. Modelo lineal incremental

Para el desarrollo del MLI no se harán diferencias en cuanto al tipo de dispositivo (NMOS o PMOS), dado que al estar regidos por el mismo tipo de ecuación, tienen el mismo modelo lineal incremental.

5.4.1. MLI referido al sustrato

El modelo a utilizar para la obtención del MLI se repite, por comodidad, a continuación. Para la zona de triodo se utilizará:

$$i_D = K \left((v_G - V_{T0})(v_D - v_S) - \frac{1}{2}n(v_D^2 - v_S^2) \right) \quad (5.141)$$

mientras que en saturación, se utilizará:

$$i_D = \frac{K}{2n} (v_G - V_{T0} - nv_S)^2 \quad (5.142)$$

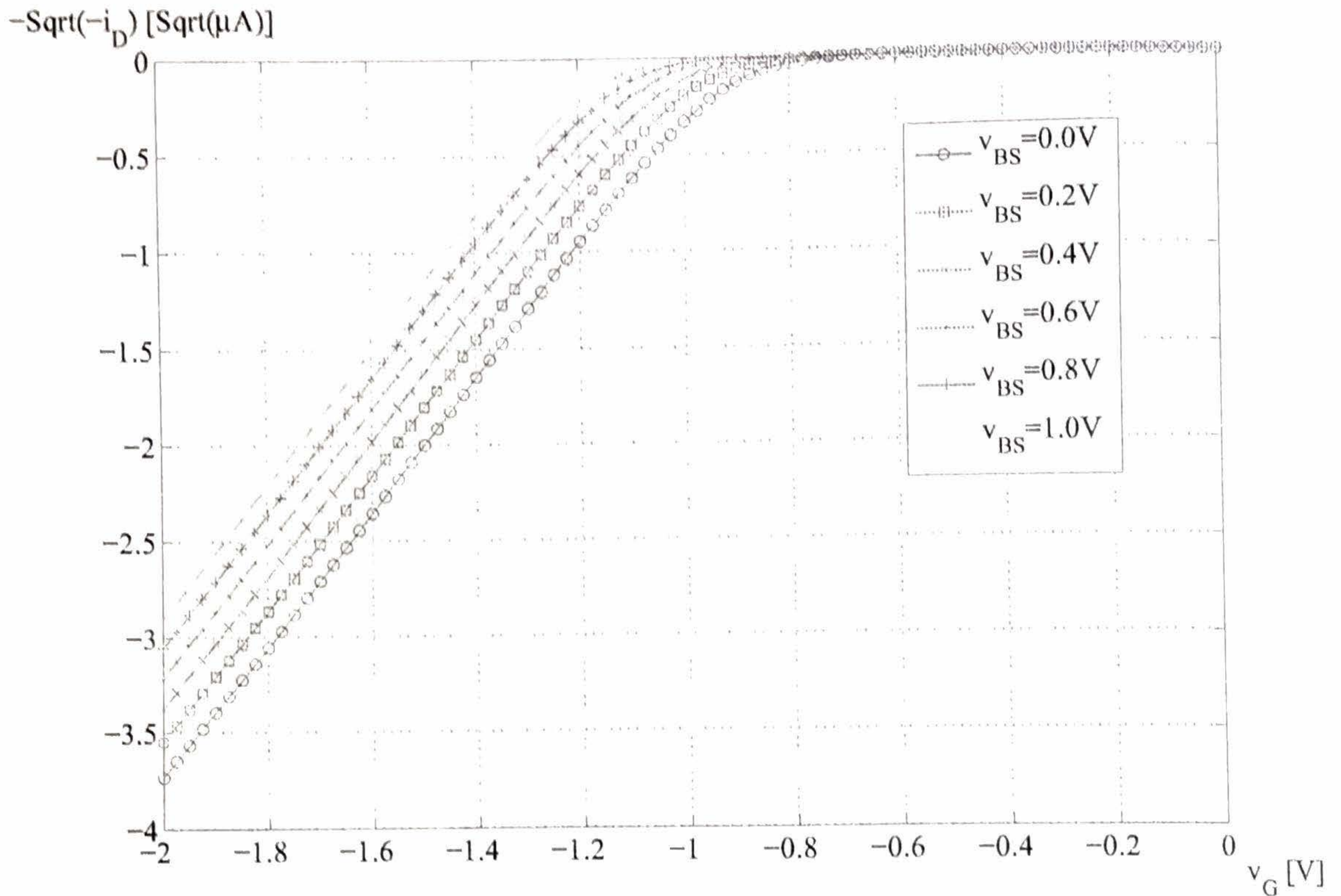


Figura 5.22: Curvas experimentales de transferencia $-\sqrt{-i_D}$ vs. v_{GS} , para distintos valores de v_{BS} , en un transistor PMOS de $W = 5\mu\text{m}$ y $L = 5\mu\text{m}$, en una tecnología de $0,5\mu\text{m}$.

salvo para la obtención de la impedancia de salida, en cuyo caso, resulta *imprescindible* tomar en cuenta el efecto de la modulación de canal:

$$i_D = \frac{K}{2n}(v_G - V_{T0} - nv_S)^2 \left(1 + \frac{v_D - V_{D(sat)}}{V_A} \right) \quad (5.143)$$

Si se analiza el dispositivo desde el terminal de gate, dado que la corriente de gate satisface $I_G = 0$, por la presencia de la capacidad de óxido C_{ox} , el modelo es un circuito abierto, por lo que resulta:

$$g_i = \left. \frac{di_G}{dv_G} \right|_Q = 0 \quad (5.144)$$

donde g_i es la conductancia de entrada.

Zona de triodo

Desde los terminales de salida, drain y source, el modelo varía dependiendo de la zona de funcionamiento donde opera el dispositivo. En la región de triodo, para pequeños valores de las tensiones de source y drain, los términos cuadráticos de (5.141) son despreciables frente al término lineal, y se obtiene:

$$i_D \doteq K(v_G - V_{T0})(v_D - v_S) = K(v_G - V_{T0})v_{DS} \quad (5.145)$$

Esto implica la existencia de una conductancia de salida:

$$g_o \triangleq \frac{i_d}{v_{ds}} = \left. \frac{di_D}{dv_{DS}} \right|_Q = K(v_G - V_{T0}) \quad (5.146)$$

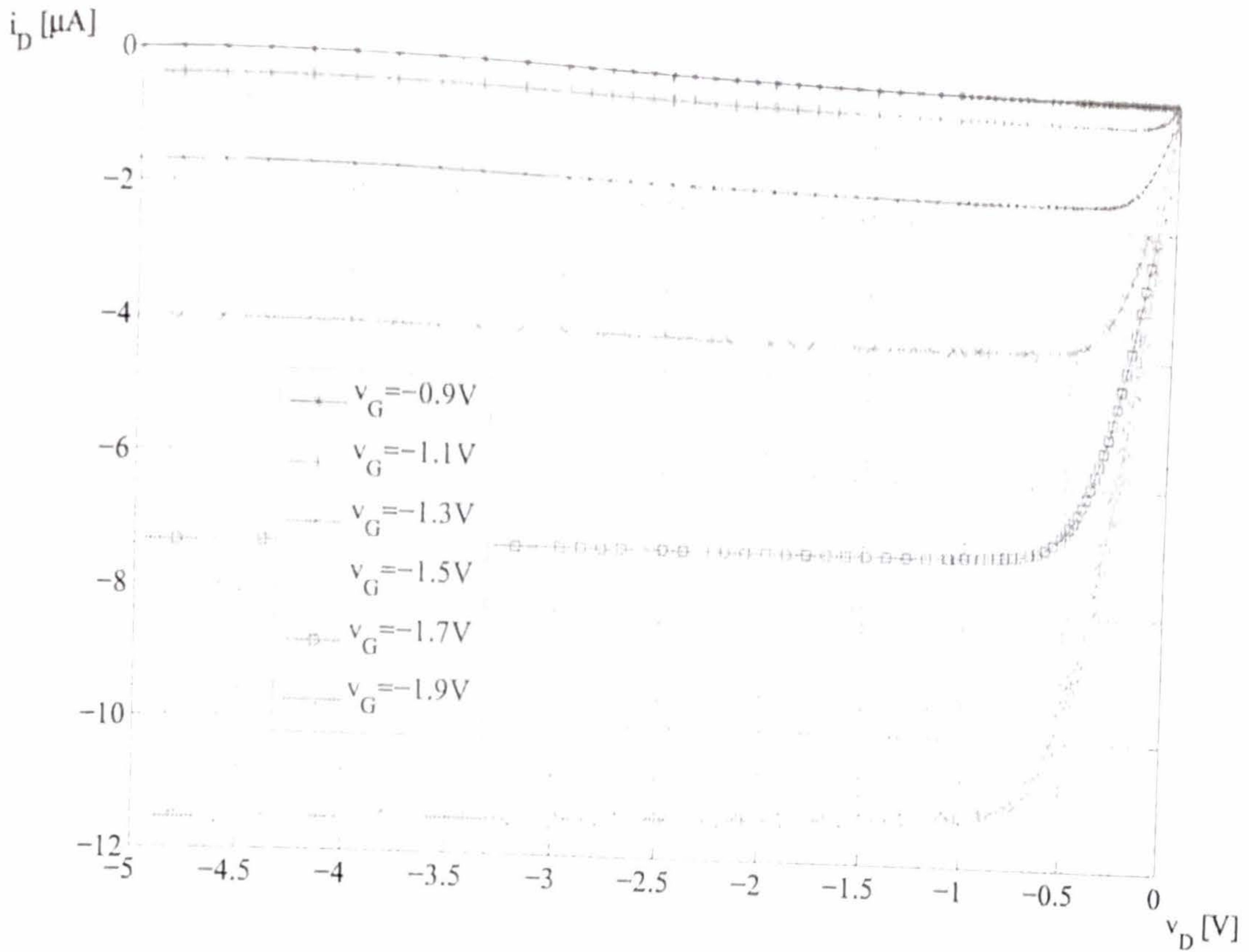


Figura 5.23: Curvas experimentales de salida de un transistor PMOS, de $W = 5\mu\text{m}$ y $L = 5\mu\text{m}$, en un proceso de $0,5\mu\text{m}$.

la cual depende de la tensión de gate, es decir, es una conductancia controlada por tensión, cuyo valor mínimo es cero para $v_G = V_{T0}$, y aumenta proporcionalmente a $v_G - V_{T0}$.

Para valores mayores de la tensión de drain:

$$g_o = \frac{i_d}{v_d} = \left. \frac{di_D}{dv_D} \right|_Q = K(V_G - V_{T0} - nV_D) \quad (5.147)$$

Para valores pequeños de V_D , la conductancia de salida coincide con la hallada en (5.146), y a medida que la tensión de drain aumenta, se va reduciendo, hasta alcanzar un valor nulo en el límite con la región de saturación (dado que allí se cumple: $V_G - V_{T0} = nV_{D(sat)}$). En ese punto, la conductancia nula (o resistencia infinita), indica que la corriente de salida i_d no presenta cambios ante variaciones de la tensión de salida v_d . De hecho, si se sale de la región de triodo, y se ingresa en saturación ($v_D > V_{D(sat)}$), el modelo (5.142) predice que la conductancia de salida sigue siendo nula:

$$g_o = \frac{i_d}{v_d} = \left. \frac{di_D}{dv_D} \right|_Q = 0 \quad (5.148)$$

ya que la corriente i_D no depende de v_D .

Si se analiza el efecto de la variación de v_G sobre la corriente i_D , se obtiene la transconductancia con respecto al gate:

$$g_{mG} \triangleq \frac{i_d}{v_g} = \left. \frac{di_D}{dv_G} \right|_Q = K(V_D - V_S) \quad (5.149)$$

Dado que la corriente i_d es producida por la tensión v_g , que es la tensión de un terminal (el terminal de gate) por el cual *no circula* i_d , este término representa una fuente de corriente controlada por tensión (FCCT). La constante de proporcionalidad, g_{mG} , tiene unidades de conductancia, y dado que relaciona una variable de salida con otra de entrada, se denomina "transconductancia". Nótese que el valor de esta transconductancia depende del punto de funcionamiento, con v_D variando en el rango $0 < v_D < V_{D(sat)}$, y con $v_S \leq v_D$; por lo tanto, la transconductancia tiene un valor mínimo, igual a cero, para $V_{DS} = 0$, y luego crece monótonamente hasta alcanzar su valor máximo en saturación: $g_{mG} = (K/n)(V_{D(sat)} - V_S)$. Nótese que $g_{mG} \geq 0$, tanto para un transistor NMOS, como para un transistor PMOS.

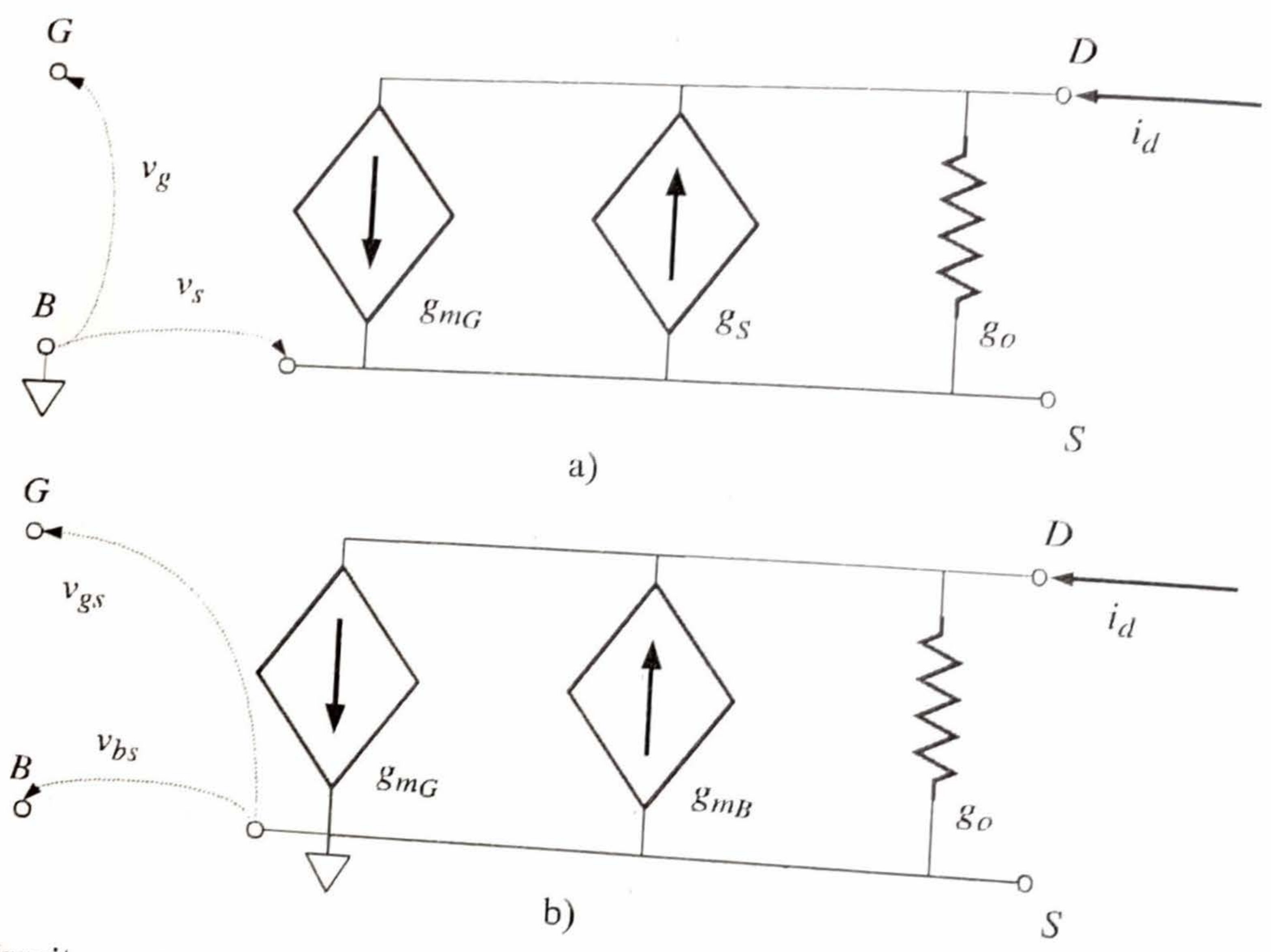


Figura 5.24: Circuito equivalente del MLI del transistor MOS: a) modelo referido al terminal de sustrato; b) modelo referido al terminal de source.

Zona de saturación

En la zona de saturación, la transconductancia está dada por:

$$g_{mG} = \frac{i_d}{v_g} = \left. \frac{di_D}{dv_G} \right|_Q = \frac{K}{n} (\bar{V}_G - V_{T0} - nV_S) \tag{5.150}$$

Si se considera que, en esta zona de funcionamiento, se cumple:

$$V_G - V_{T0} - nV_S = \sqrt{\frac{2nI_D}{K}} \tag{5.151}$$

entonces, (5.150) puede escribirse también como:

$$g_{mG} = \sqrt{\frac{2K}{n}} \sqrt{I_D} \tag{5.152}$$

De (5.150) y (5.152), se ve que la transconductancia, en esta zona, es proporcional a la tensión de gate utilizada y, por lo tanto, es proporcional a la raíz cuadrada de la corriente continua. Debido a esto último, es importante notar que para duplicar el valor de la transconductancia, es necesario cuadruplicar el valor de corriente continua.

Para el cálculo de la transconductancia, la utilización de la corrección de Early (5.143) no causa modificaciones apreciables, dado que $(v_D - V_{D(sat)})/V_A$ es pequeño con respecto al factor 1. Algunos valores típicos, como $V_A = 50V$, $v_D = 5V$, y $V_{D(sat)} = 1V$ permiten ver que, en efecto, $(v_D - V_{D(sat)})/V_A = 0,08$ resulta menor que 1 y tiene un impacto pequeño sobre el valor de la corriente.

Sin embargo, es imprescindible utilizar la corrección de Early (5.143) para el cálculo de la conductancia de salida en saturación, dado que el modelo ideal (5.148) predice $g_o = 0$. Considerando la expresión (5.143), la conductancia de salida es:

$$g_o = \frac{i_d}{v_d} = \left. \frac{di_D}{dv_D} \right|_Q = \frac{1}{V_A} \frac{K}{2n} (v_G - V_{T0} - nV_S)^2 \quad (5.153)$$

que puede escribirse también como:

$$g_o = \frac{I_D}{V_A} \quad (5.154)$$

En caso que se varíe la tensión de source, la variación correspondiente de corriente, i_d , será:

$$g_S = \frac{i_d}{v_s} = \left. \frac{di_D}{dv_S} \right|_Q = -K(V_G - V_{T0} - nV_S) \quad (5.155)$$

$$= -ng_{mG} \quad (5.156)$$

El signo negativo se debe a que un aumento de la tensión de source produce una disminución de la corriente de drain. Las variaciones de corriente de drain ante variaciones de la tensión de source son mayores, en un factor n , que las variaciones de corriente provocadas por la tensión de gate. Esto se debe a que solo una fracción de la variación de tensión de gate, dada por el divisor capacitivo entre la capacidad de óxido y la de sustrato, produce una variación de la tensión del canal, mientras que las variaciones de tensiones de source se aplican directamente sobre el canal.

El diagrama circuital correspondiente al MLI se muestra en la Fig. 5.24-a.

Ejemplo 21 Un transistor NMOS, con los parámetros del Ejemplo 18, está funcionando con $V_S = 0V$, $V_D = 4V$, y $V_G = 2V$. Qué valores tienen la transconductancia y la resistencia de salida? Si se aplica una señal, $v_{in} = 10mV$, al terminal de gate, cuál es la corriente de salida de señal resultante?

Para los valores de DC mencionados, el dispositivo se halla en saturación ($V_D = 4V > V_{D(sat)} = 1,55V$), y el valor de la corriente de DC es:

$$I_D = \frac{573,9}{2 \times 1,52} (2 - 0,64)^2 = 347\mu A \quad (5.157)$$

La transconductancia de gate es:

$$g_{mG} = \sqrt{\frac{2 \times 573,9}{1,52}} \times \sqrt{349} = 513,36\mu A/V \quad (5.158)$$

y la conductancia de salida es:

$$g_o = \frac{347}{329,84} = 1,05\mu A/V \quad (5.159)$$

Esta conductancia corresponde a una resistencia de salida $r_o = 952K$. Para una señal de gate $v_g = 10mV$, la corriente de salida será $i_d = 513,36 \times 10^{-6} \times 10 \times 10^{-3} = 5,13\mu A$. En el caso que esta corriente de señal circule por la resistencia de salida, se produciría una tensión de señal $v_o = 5,13\mu A / 1,05\mu A/V = 4,88V$.

5.4.2. MLI referido al source

El modelo a utilizar para este caso se resume a continuación. En la zona de triodo, la expresión de la corriente de drain es:

$$i_D = K \left((v_{GS} - V_T)v_{DS} - \frac{1}{2}\alpha v_{DS}^2 \right) \quad (5.160)$$

mientras que, en la zona de saturación, es:

$$i_D = \frac{K}{2\alpha} (v_{GS} - V_T)^2 \quad (5.161)$$

con

$$V_T = V_{FB} - 2\phi_p + \gamma \sqrt{-2\phi_p - V_{BS}} \quad (5.162)$$

El análisis de la transconductancia de gate, y de la conductancia de salida, es similar al realizado para el modelo referido al sustrato, y se repite de manera abreviada.

La mayor diferencia entre el modelo referido al source y el referido al sustrato radica en que, en el modelo referido al source, la variación de la tensión $v_{BS} = v_B - v_S = v_B$ modifica la corriente i_D a través de la variación de la tensión de umbral, y no directamente, como en el modelo referido al sustrato, donde la variación es directa con la tensión $v_{SB} = v_S - v_B = v_S$.

Al igual que en el caso anterior, la presencia de la capacidad de óxido implica que $I_G = 0$.

Zona de triodo

Para pequeños valores de la tensión v_{DS} resulta:

$$i_d = K(v_{GS} - V_T)v_{DS} \quad (5.163)$$

En este caso, el MLI coincide con (5.146), reemplazando V_G por V_{GS} :

$$g_o = \frac{i_d}{v_{ds}} = \left. \frac{di_D}{dv_{DS}} \right|_Q = K(V_{GS} - V_T) \quad (5.164)$$

Para valores mayores, el término cuadrático no es despreciable, y el MLI coincide con (5.147), reemplazando V_G por V_{GS} , y V_D por V_{DS} :

$$g_o = K(V_{GS} - V_T - \alpha V_{DS}) \quad (5.165)$$

Por último, la contribución de v_{GS} a la corriente i_D coincide también con (5.149):

$$g_{mG} = \frac{i_d}{v_{gs}} = \left. \frac{di_D}{dv_{GS}} \right|_Q = KV_{DS} \quad (5.166)$$

Zona de saturación

En la zona de saturación, la transconductancia está dada por:

$$g_{mG} = \frac{K}{\alpha} (V_{GS} - V_T) \quad (5.167)$$

Si se considera que, en esta zona de funcionamiento, se cumple:

$$V_{GS} - V_T = \sqrt{\frac{2\alpha I_D}{K}} \quad (5.168)$$

entonces, (5.167) puede escribirse como:

$$g_{mG} = \sqrt{\frac{2K}{\alpha}} \sqrt{I_D} \quad (5.169)$$

La expresión de la conductancia de salida, tomando en cuenta la corrección por el término de Early, es:

$$g_o = \frac{i_d}{v_{ds}} = \left. \frac{di_D}{dv_{DS}} \right|_Q = \frac{1}{V_A} \frac{K}{2\alpha} (V_{GS} - V_T)^2 \quad (5.170)$$

que puede escribirse como:

$$g_o = \frac{I_D}{V_A} \quad (5.171)$$

Por último, es necesario analizar el efecto de variaciones de la tensión v_{BS} (efecto de back-gate), dado que la corriente de drain es sensible a las variaciones de la tensión de sustrato, v_B , con respecto al terminal de source. Esta variación de la corriente también se modela con una FCCT, la cual es controlada por la tensión de sustrato. Se define, para este caso, una transconductancia de sustrato:

$$g_{mB} \triangleq \frac{i_d}{v_{bs}} = \left. \frac{di_D}{dv_{BS}} \right|_Q \quad (5.172)$$

El valor de este parámetro se puede hallar, mediante la regla de la cadena, de la siguiente manera:

$$\begin{aligned} g_{mB} &= \left. \frac{di_D}{dV_T} \times \frac{dV_T}{dv_{BS}} \right|_Q \\ &= -g_{mG}(\alpha - 1) \end{aligned} \quad (5.173)$$

El signo negativo se debe a que un aumento de la tensión de sustrato v_{BS} produce una disminución de la corriente. En general, la transconductancia de sustrato g_{mB} resulta menor que la transconductancia de gate g_{mG} , lo cual puede verse de (5.173), teniendo en cuenta que los valores típicos de α son cercanos a $\alpha = 1,5$. Nótese que

$$\alpha - 1 = \frac{\gamma}{2\sqrt{-2\phi_p - V_{BS}}} \frac{C_{ox}}{C_{ox}} = C_{vac}/C_{ox} \quad (5.174)$$

de acuerdo a (4.63) y (5.35), por lo tanto, la situación $\alpha \approx 1,5$, corresponde a valores de $C_{ox} > C_{vac}$.

La Fig. 5.24-b muestra el circuito del MLI, donde el terminal de source se utiliza como referencia para los potenciales aplicados.

5.5. Modelo de AC

Para arribar a un modelo de AC, es menester derivar las capacidades propias del dispositivo, para agregarlas a los modelos ya desarrollados. En el caso del transistor MOS, existen tres tipos de capacidades. La más compleja es la capacidad del gate que, dependiendo del modo de funcionamiento, puede verse de distintas maneras: entre gate y source, entre gate y drain, o entre gate y sustrato. Además, están las capacidades de difusión, producto de las junturas existentes entre source/drain y sustrato. Estas capacidades corresponden a junturas en inversa, las cuales se han descrito en el tratamiento de la juntura. Por último, es necesario considerar las llamadas capacidades de solapamiento: una de ellas, es producto del solapamiento entre el gate y las zonas de drain/source, dado que al momento de la difusión de impurezas, se produce una difusión lateral desde source y drain hacia la zona por debajo del óxido; la otra, es producto del solapamiento entre gate y sustrato, como consecuencia de la extensión del gate hacia los lados del transistor (en este caso el dieléctrico intermedio es óxido grueso).

5.5.1. Capacidad de gate

El análisis de la capacidad de gate se basa en el capacitor MOS descrito en el Cap. 4. En el caso del transistor, aparecen además, los terminales de source y drain a ambos lados. De acuerdo al modo de operación (corte, triodo o saturación), la capacidad de gate tomará distintos valores y podrá verse distribuida de distintas maneras. Desde el gate, interesa evaluar la capacidad con respecto a source (C_{gs}), drain (C_{gd}) y sustrato (C_{gb}).

Corte

En la región de corte, $|v_G| < |V_T|$, el dispositivo estará, en general, en la zona de vaciamiento (pudiendo también estar en la zona de acumulación¹⁴). En este caso, la capacidad de gate está dada por (4.61):

$$C_{gb} = C_v(v_G) \parallel C_{ox} = \frac{C_v(v_G)C_{ox}}{C_v(v_G) + C_{ox}} < C_{ox} \quad (5.175)$$

Esta capacidad se da entre el gate y el sustrato, por lo cual no hay componentes de capacidad entre gate y source/drain, es decir:

$$C_{gs} = 0 \quad C_{gd} = 0 \quad (5.176)$$

Triodo

En la zona de triodo, cuando ya se ha formado el canal de inversión (para $|v_G| \geq |V_T|$), la capacidad del óxido aparece entre el terminal de gate y el canal. Si se asume que la distribución de carga en el canal es homogénea, el canal se puede considerar como un vínculo metálico entre drain y source. Por lo tanto, la capacidad del óxido se halla repartida, en partes iguales, entre los terminales de source y drain, es decir:

$$C_{gs} = C_{gd} = \frac{1}{2}C_{ox} \quad (5.177)$$

¹⁴Esto depende del valor de la tensión V_{FB} que se define en el proceso de fabricación de acuerdo a los dopados del gate y el sustrato.

Dado que una vez en inversión, la zona de vaciamiento no varía más con respecto a la tensión aplicada en el gate, la capacidad de la zona de vaciamiento es nula¹⁵. En este modo de operación, el canal puede interpretarse como una lámina metálica que aísla al terminal de gate del sustrato. Por lo dicho, resulta

$$C_{gb} = 0 \quad (5.178)$$

Saturación

En la zona de saturación, el canal se hace nulo sobre el extremo del drain. En esta situación, el perfil de carga en el canal no es homogéneo, y es necesario hallar la carga en el mismo, para luego evaluar su derivada con respecto a v_G .

La expresión de la carga de inversión (5.32), para una cierta distancia y , desde el source, es:

$$Q_i(y) = -C_{ox} (v_G - V_T - nv_c(y)) \quad (5.179)$$

La expresión de la carga total es:

$$Q_i = - \int_0^L C_{ox} (v_G - V_T - nv_c(y)) dy \quad (5.180)$$

Si se halla la expresión de dy en función de dv_c , utilizando (5.25), resulta:

$$dy = \frac{\mu_n C_{ox} W}{i_D} (v_G - V_T(y) - v_c(y)) dv_c \quad (5.181)$$

Reemplazando esta expresión en (5.180), se obtiene:

$$\begin{aligned} Q_i &= - \frac{\mu_n C_{ox}^2 W^2}{i_D} \int_{V_S}^{V_{D(sat)}} (v_G - V_T - nv_c(y))^2 dv_c \\ &= - \frac{\mu_n C_{ox}^2 W^2}{i_D} \left\{ \frac{(nv_c - V_G + V_T)^3}{3} \right\} \Big|_{V_S}^{V_{D(sat)}} \\ &= - \frac{\mu_n C_{ox}^2 W^2 (V_G - V_T - nV_S)^3}{3} \end{aligned} \quad (5.182)$$

Si se reemplaza la expresión de la corriente i_D en función de $V_G - V_T - nV_S$, dada por (5.44), en (5.182), y se simplifica, se obtiene la expresión final de la carga de inversión:

$$Q_i = - \frac{2C_{ox} W L}{3} (V_G - V_T - nV_S) \quad (5.183)$$

La carga del gate, de (4.44), se halla como:

$$\begin{aligned} Q_G(v_G) &= -Q_i(v_G) - Q_{vac}(v_G) \\ &= \frac{2C_{ox} W L}{3} (V_G - V_T - nV_S) - Q_{vac} \end{aligned} \quad (5.184)$$

donde Q_{vac} no varía con v_G por estar en la zona de inversión. La capacidad total desde el terminal del gate se halla de la siguiente manera:

¹⁵Esto es cierto si se mantiene v_B constante. Si v_B se varía, la capacidad de la región de vaciamiento se manifiesta en ambos terminales, source y drain.

$$C_{GS} = \left. \frac{dQ_G}{dv_G} \right|_Q = \frac{2C_{ox}WL}{3} \quad (5.185)$$

mientras que la capacidad por unidad de área es:

$$C_{gs} = \frac{1}{WL} = \left. \frac{dQ_G}{dv_G} \right|_Q = \frac{2C_{ox}}{3} \quad (5.186)$$

Nótese también que Q_G no depende de la tensión de drain, por lo tanto $C_{gd} = 0$. La Tabla 5.1 resume los valores hallados.

Cuadro 5.1: Capacidad del gate por unidad de área de acuerdo a la zona de funcionamiento

	C_{gb}	C_{gs}	C_{gd}
Corte	$C_{vac} C_{ox}$	0	0
Triodo	0	$C_{ox}/2$	$C_{ox}/2$
Saturación	0	$2C_{ox}/3$	0

5.5.2. Capacidad de juntas

Tal como se vió en la Sección 3.5.1, la capacidad en inversa de una junta está dada por (3.108), que se repite por conveniencia:

$$C_{vac} = C_{j0} \frac{1}{\left(1 - \frac{V_D}{\phi_B}\right)^m} \quad (5.187)$$

Tanto el drain como el source forman una junta en inversa con el sustrato, por lo tanto, presentan una capacidad de la forma (5.187). Sin pérdida de generalidad, se asumirá que tanto el drain como el source tienen idénticas dimensiones, dadas por: ancho $W = W_d = W_s$, longitud $L_d = L_s$, y profundidad x_b . La junta entre la difusión y el sustrato se forma en cada una de las caras comunes, tal como se ilustra en la Fig. 5.26. La cara del fondo presenta una capacidad C_{jb} , y las otras cuatro caras presentan una capacidad perimetral denominada C_{jsw} , de tal manera que la capacidad total es:

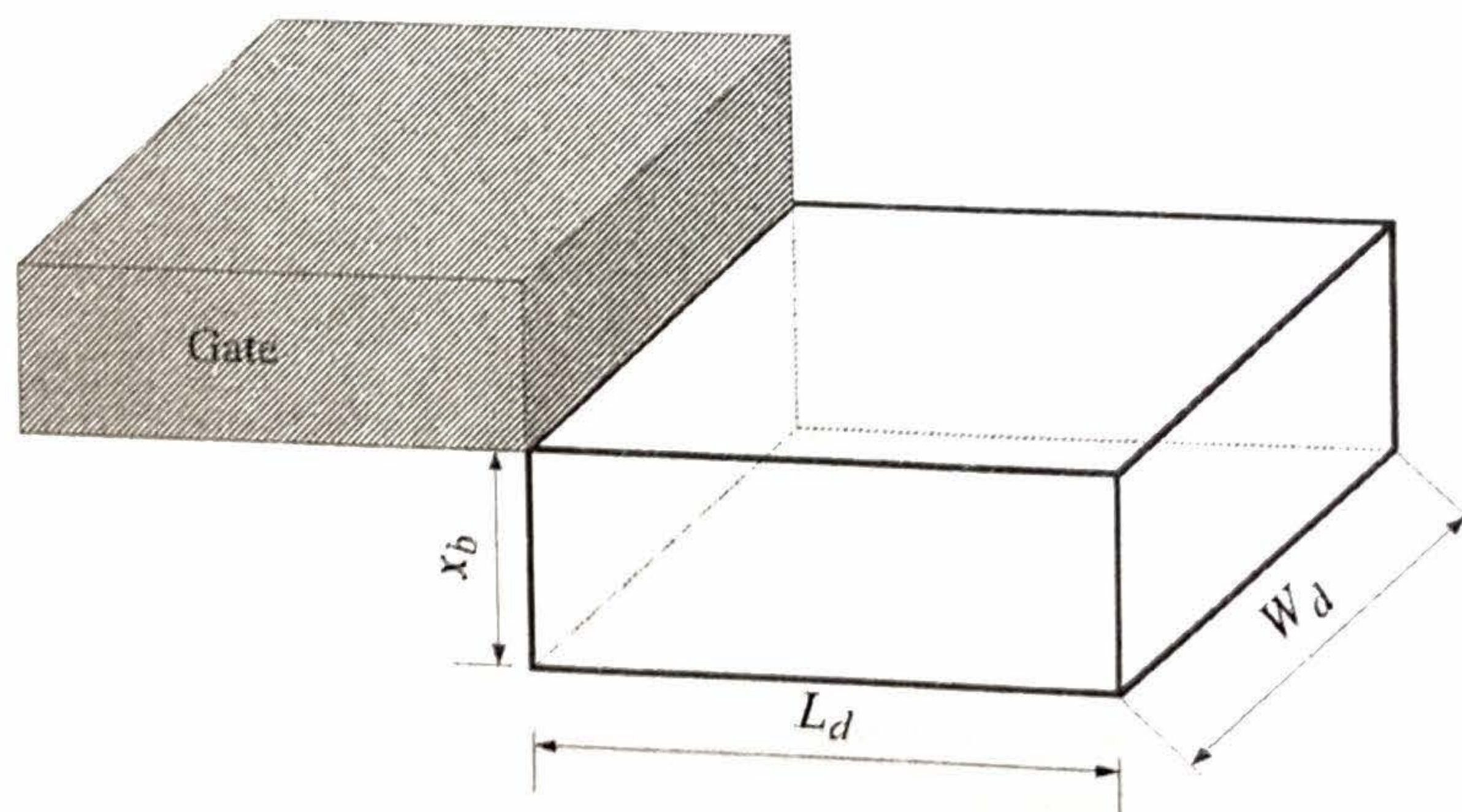


Figura 5.25: Geometría de las capacidades de junta del source/drain.

$$C_{JT} = C_{JB} + C_{JSW} \quad (5.188)$$

La capacidad de la cara inferior se obtiene multiplicando el área de la difusión por la expresión de la capacidad de la juntura en inversa (5.187):

$$\begin{aligned} C_{JB} &= A_{jb} C_{jb0} \frac{1}{\left(1 - \frac{V_D}{\phi_{B_b}}\right)^{m_b}} \\ &= W_d L_d C_{jb0} \frac{1}{\left(1 - \frac{V_D}{\phi_{B_b}}\right)^{m_b}} \end{aligned} \quad (5.189)$$

Nótese que se han utilizado subíndices particulares para los diferentes parámetros de esta juntura: C_{jb0} , el potencial ϕ_{B_b} , y el exponente m_b .

Para el cálculo de la capacidad perimetral, es necesario tener en cuenta que tres de las caras de la difusión (de área A_{psw}), se enfrentan al sustrato y la cuarta (de área A_{pswg}), se enfrenta al sustrato debajo del gate. Debido a esto, la capacidad perimetral es:

$$\begin{aligned} C_{JSW} &= A_{psw} C_{j0} \frac{1}{\left(1 - \frac{V_D}{\phi_{B_{sw}}}\right)^{m_{sw}}} + A_{pswg} C_{jg0} \frac{1}{\left(1 - \frac{V_D}{\phi_{B_{swg}}}\right)^{m_{swg}}} \\ &= (W_d + 2L_d) \frac{x_b C_{j0}}{\left(1 - \frac{V_D}{\phi_{B_{sw}}}\right)^{m_{sw}}} + W_d \frac{x_b C_{jg0}}{\left(1 - \frac{V_D}{\phi_{B_{swg}}}\right)^{m_{swg}}} \\ &= (W_d + 2L_d) \frac{C_{jsw}}{\left(1 - \frac{V_D}{\phi_{B_{sw}}}\right)^{m_{sw}}} + W_d \frac{C_{jswg}}{\left(1 - \frac{V_D}{\phi_{B_{swg}}}\right)^{m_{swg}}} \end{aligned} \quad (5.190)$$

donde $C_{jswg} = C_{jg0} x_b$ es la capacidad perimetral por unidad de longitud del lado del gate, y $\phi_{B_{swg}}$, m_{swg} , son los valores correspondientes del potencial de juntura y su exponente; $C_{jsw} = C_{j0} x_b$ es la capacidad perimetral por unidad de longitud para los otros tres lados, y $\phi_{B_{sw}}$, m_{sw} , los valores del potencial de juntura y su exponente. En general, estos coeficientes son distintos a los correspondientes a la cara inferior, dados por ϕ_{B_b} y m_b .

Ejemplo 22 Un transistor NMOS ($W = 5\mu m$, $L = 5\mu m$), en el proceso del Ejemplo 18, tiene los siguientes parámetros: $C_{jb0} = 4,26 \times 10^{-4} F/m^2$, $C_{jsw} = 3,24 \times 10^{-10} F/m$, $C_{jswg} = 4,26 \times 10^{-10} F/m$. Las áreas de drain y source son iguales, y tienen una longitud $L_d = L_s = 1,8\mu m$. Calcule los valores máximos de las capacidades de juntura.

Los valores máximos de capacidad resultan cuando la tensión aplicada es nula. Las capacidades respectivas son:

$$\begin{aligned} C_{JB} &= 5 \times 10^{-6} \times 1,8 \times 10^{-6} \times 4,26 \times 10^{-4} = 3,83 fF \\ C_{JSW} &= (5 + 2 \times 1,8) \times 10^{-6} \times 3,24 \times 10^{-10} + \\ &\quad + 5 \times 10^{-6} \times 4,26 \times 10^{-10} = 2,78 fF + 0,82 fF = 3,60 fF \end{aligned} \quad (5.191)$$

5.5.3. Capacidad de solapamiento

Como ya se ha mencionado, si bien el gate actúa como una máscara para la difusión, o implantación de impurezas, los procesos de calentamiento ("annealing") que sufren los circuitos integrados durante el proceso de fabricación, hacen que la difusión continúe una vez implantadas las impurezas. Esto produce una difusión lateral, que provoca que tanto las impurezas del drain como las del source penetren una distancia x_d debajo del óxido, acortando el canal a $L - 2x_d$. Por otro lado, el gate sobresale también una distancia x_g por sobre el óxido grueso, provocando otra capacidad adicional. Ambos casos se muestran en la Fig. 5.26.

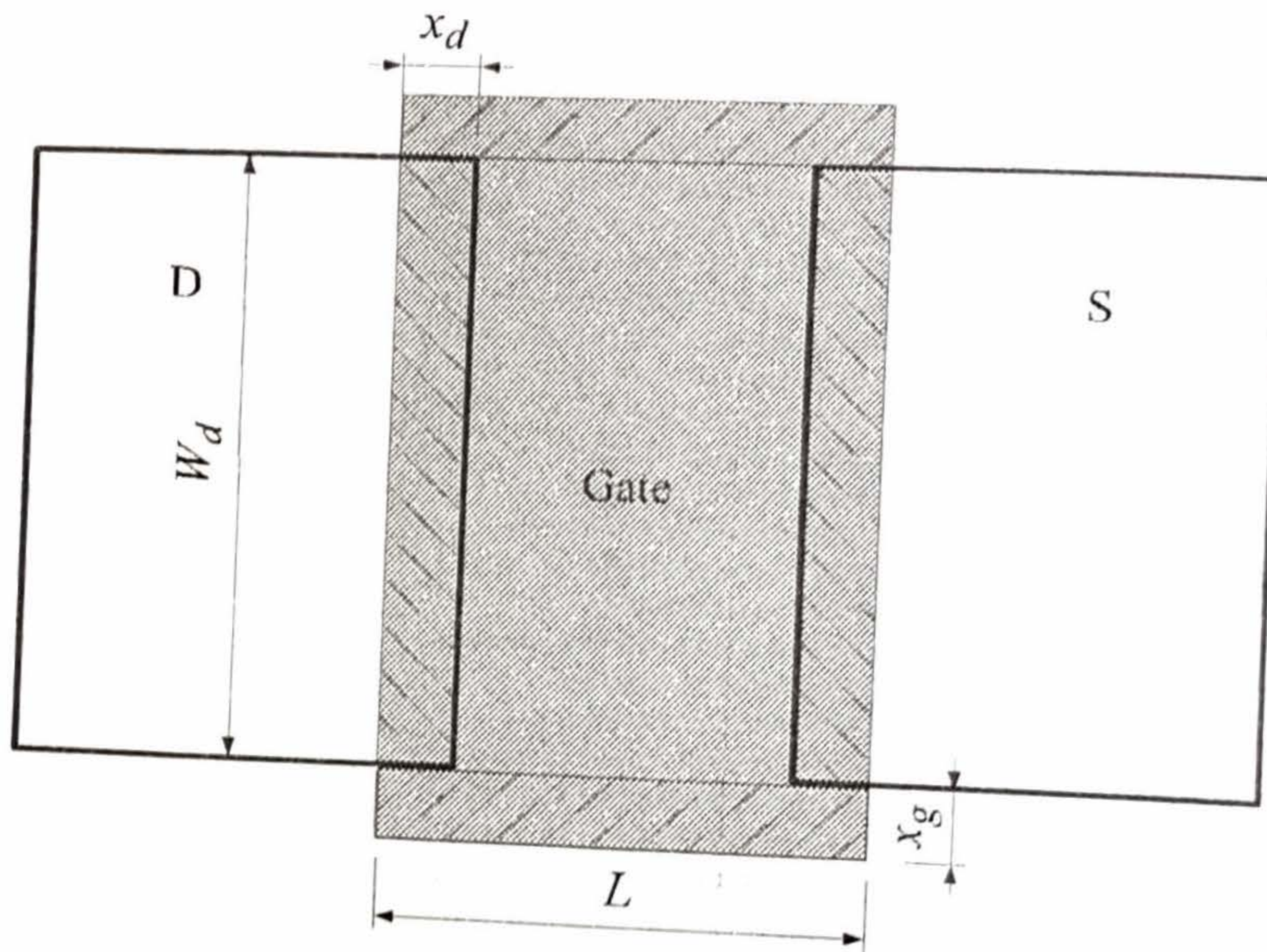


Figura 5.26: Geometría de las capacidades de solapamiento del gate con respecto a source/drain y al óxido grueso.

La difusión lateral produce un par de capacitores lineales, de valor:

$$C_{GS} = W_s x_d C_{ox} \quad C_{GD} = W_d x_d C_{ox} \quad (5.192)$$

Es habitual que se especifiquen los valores por unidad de longitud, C_{gso} y C_{gdo} , de manera que:

$$C_{GS_o} = W_s C_{gso} \quad C_{GD_o} = W_d C_{gso} \quad (5.193)$$

El solapamiento del gate sobre sustrato produce una capacidad lineal a ambos costados del transistor, de valor:

$$C_{GB_o} = L x_g \frac{\epsilon_{ox}}{t_{FOX}} \quad (5.194)$$

donde t_{FOX} es el espesor del óxido grueso, que típicamente es cercano a los 500nm, produciendo un valor de capacidad relativamente pequeño. En este caso, también es habitual que se especifique el valor por unidad de longitud C_{gbo} , de manera que:

$$C_{GB_o} = L C_{gbo} \quad (5.195)$$

Ejemplo 23 Calcule las capacidades de solapamiento para el transistor NMOS del ejemplo anterior, sabiendo que: $C_{gdo} = 2,26 \times 10^{-10} \text{ F/m}$, $C_{gso} = 2,26 \times 10^{-10} \text{ F/m}$, y $C_{gbo} = 1,00 \times 10^{-9} \text{ F/m}$.

5.5. Modelo de AC

Los valores de las capacidades son:

$$\begin{aligned}
 C_{GD_o} &= 2,26 \times 10^{-10} \times 1,8 \times 10^{-6} = 1,13 \text{ fF} \\
 C_{GS_o} &= 2,26 \times 10^{-10} \times 1,8 \times 10^{-6} = 1,13 \text{ fF} \\
 C_{GB_o} &= 1,00 \times 10^{-9} \times 5 \times 10^{-6} = 5,00 \text{ fF}
 \end{aligned}
 \tag{5.196}$$

Por último, la Fig. 5.27 muestra todas las capacidades enumeradas. El agregado de estas capacidades, a los circuitos del MLI de la Fig. 5.24, produce el MLI de AC para el transistor MOS.

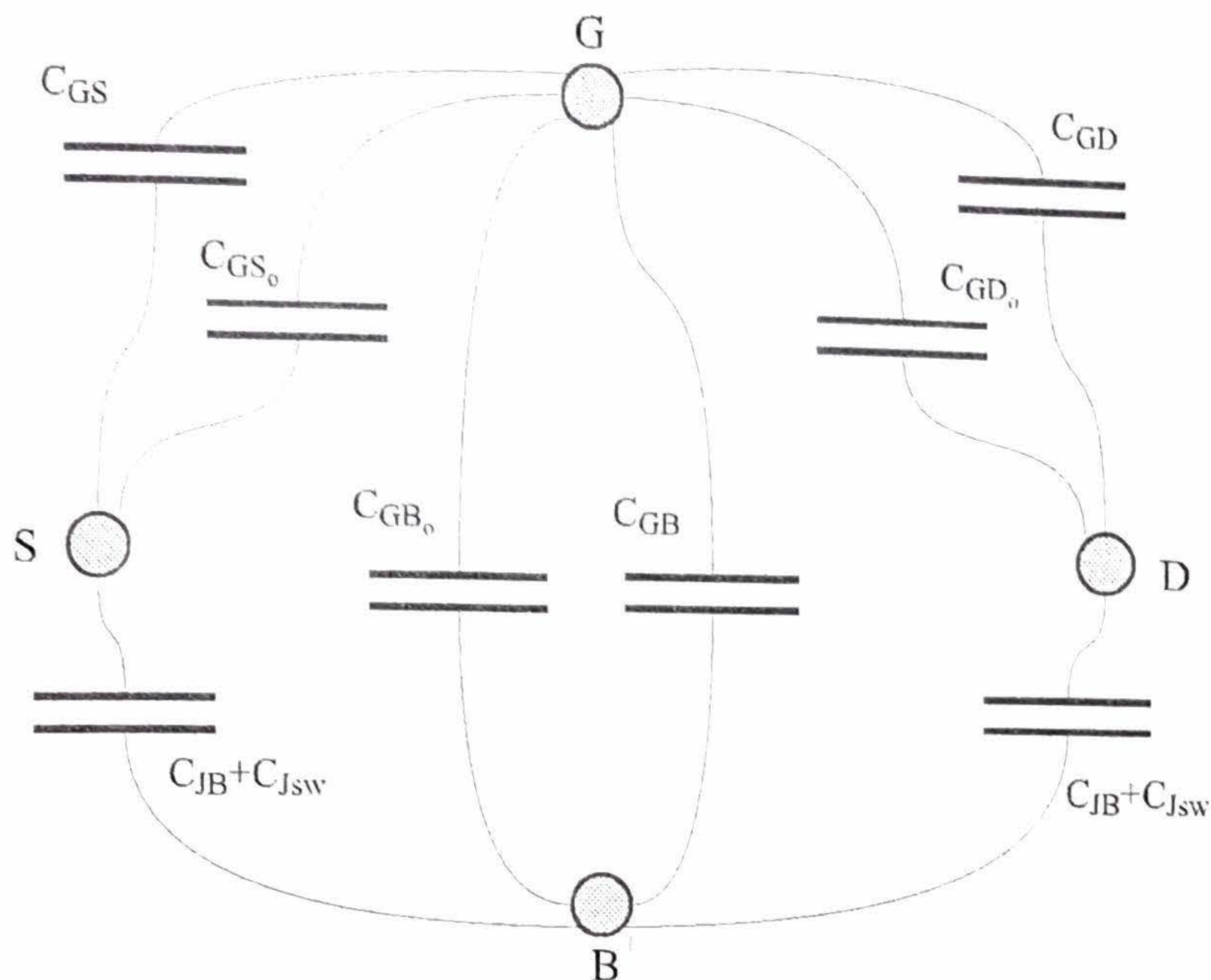


Figura 5.27: Circuito esquemático con las capacidades de un transistor MOS.

5.5.4. Límite de validez del modelo de AC

Conforme la frecuencia se incrementa, la impedancia de la capacidad de gate hace que la corriente de entrada i_g se vuelva apreciable. Se define la frecuencia de transición f_T del transistor como la frecuencia para la cual la magnitud de la corriente de gate se vuelve igual a la corriente de drain. Para el cálculo de la misma, se utiliza el MLI de AC de la Fig. 5.28-a, donde se coloca una fuente de tensión sinusoidal de entrada, de valor $v = A \sin(\omega t)$, tal como se muestra en la Fig. 5.28-b. Si se cortocircuitan las fuentes de tensión de DC, y se considera que el dispositivo está en saturación (por lo cual $C_{DS} = 0$), y además, se desprecia la capacidad de solapamiento C_{GD_o} , resulta el circuito lineal incremental de AC de la Fig. 5.28-c.

En este circuito, la corriente de entrada está dada por:

$$i_g = C \frac{dv_{in}}{dt} = A\omega(C_{GS} + C_{GB}) \cos(\omega t)
 \tag{5.197}$$

y la corriente de drain es:

$$i_d = g_m v_{gs} = g_m A \sin(\omega t)
 \tag{5.198}$$

El cociente del módulo entre ambas corrientes define la ganancia de corriente del circuito A_i :

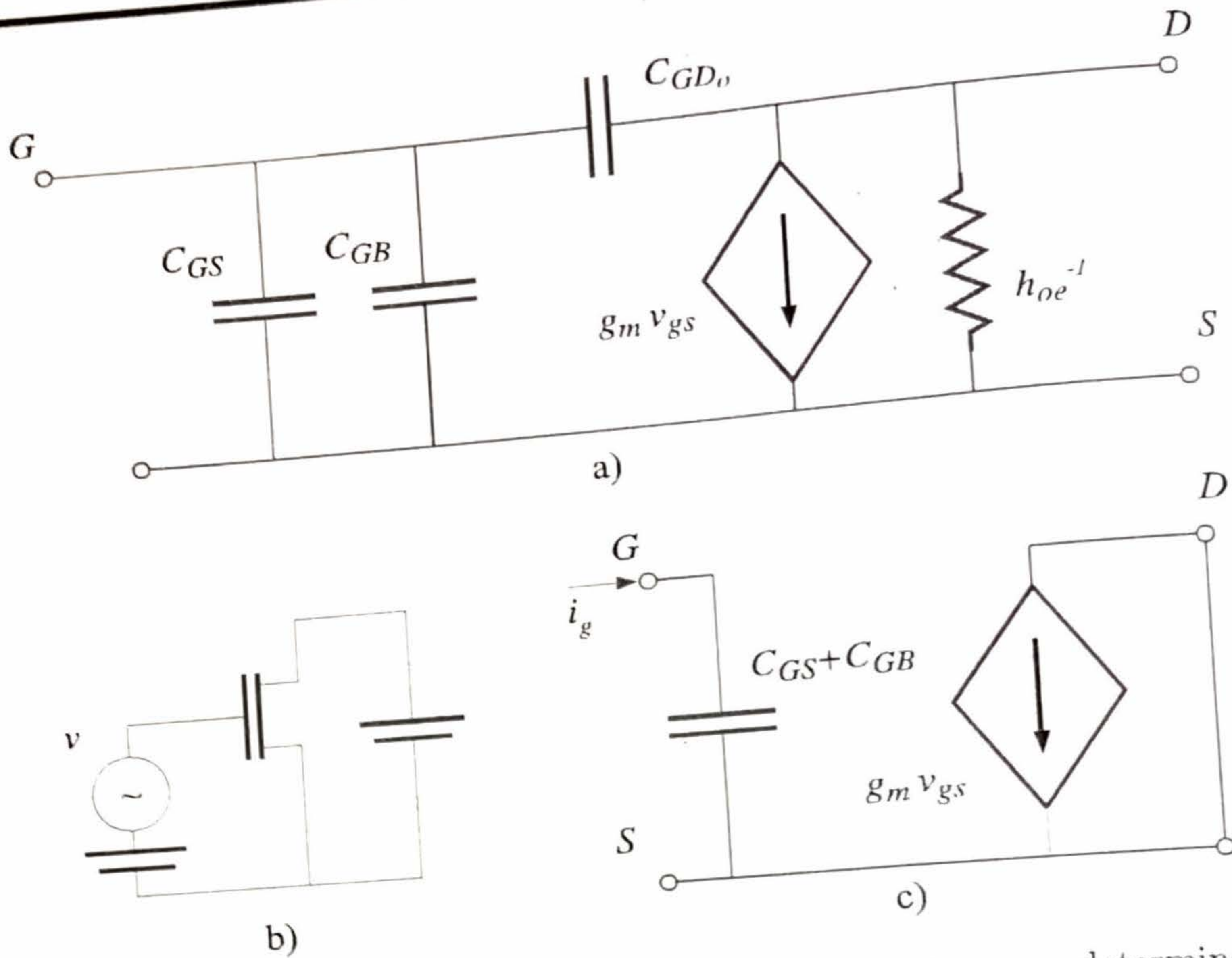


Figura 5.28: a) Modelo lineal incremental con capacidades; b) circuito para determinar la frecuencia de transición; c) circuito de pequeña señal correspondiente al circuito b.

$$A_i \triangleq \frac{\|i_d\|}{\|i_g\|} = \frac{g_m}{2\pi f(C_{GS} + C_{GB})} \quad (5.199)$$

La frecuencia de transición se define como la frecuencia para la cual la ganancia de corriente se iguala a 1, es decir:

$$A_i \Big|_{f=f_T} = 1 \quad (5.200)$$

y de acuerdo a (5.199) es igual a:

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_{GB})} \quad (5.201)$$

Si se desprecia la capacidad \$C_{GB}\$ frente a \$C_{GS}\$, y se considera el valor de \$C_{GS}\$ en saturación, la frecuencia de transición resulta:

$$\begin{aligned} f_T &\approx \frac{g_m}{2\pi C_{GS}} = \frac{(\mu C_{ox}/\alpha)(W/L)(V_{GS} - V_T)}{2\pi(2/3)WLC_{ox}} \\ &= \frac{3\mu(V_{GS} - V_T)}{2\alpha \cdot 2\pi L^2} \end{aligned} \quad (5.202)$$

De aquí se puede ver que para extender la operación del transistor en frecuencia, de acuerdo a (5.202), es necesario utilizar tensiones de gate grandes con respecto a \$V_T\$ (lo cual implica mayores corrientes) y transistores de la menor longitud posible.

La frecuencia \$f_T\$ da una idea aproximada de la frecuencia a partir de la cual el transistor pierde sus características de amplificación. Para frecuencias cercanas a \$f_T\$ la corriente de gate es significativa

5.5. Modelo de AC

y es necesario incluir efectos resistivos. Asimismo, la carga en el canal empieza a exhibir un desfase con respecto a la tensión aplicada al gate, por lo cual deben utilizarse otros modelos.

Por último, es relevante señalar que el modelo lineal incremental de AC es preciso, en general, hasta una frecuencia $f = f_T/5$ -salvo que se requiera precisión sobre la fase, en cuyo caso es recomendable utilizarlo solamente hasta $f = f_T/10$. Es por esta razón que el valor (5.201) debe tomarse como una aproximación de f_T , dado que el modelo utilizado para obtener (5.201), en rigor de verdad, no es válido para $f = f_T$. Sin embargo, a los fines de predecir el rango de funcionamiento, el valor de f_T dado por (5.201) es ampliamente utilizado.

Ejemplo 24 La f_T para el transistor NMOS del Ejemplo 22, suponiendo que $V_{GS} = 5V$, está dada por:

$$f_T = \frac{3 \times 472 \times (5 - 0,64)}{2 \times 1,5 \times 2 \times \pi \times (5 \times 10^{-4})^2} = 1,31GHz \quad (5.203)$$

Capítulo 6

Transistores Bipolares

6.1. Descripción Cualitativa

Constructivamente, un transistor bipolar se logra disponiendo un material con un tipo de dopado entre dos materiales del dopado opuesto. La región intermedia de semiconductor dopado se denomina base, y las regiones de dopado opuesto, dispuestas a ambos lados de la base, se denominan emisor y colector. En el caso en que la base sea de tipo N, el colector y el emisor son de tipo P, y el transistor se llama PNP (debido a la secuencia de materiales del emisor, la base y el colector). En el caso que la base es de tipo P, el colector y el emisor son de tipo N, y el transistor resultante se denomina NPN. En ambos dispositivos, el principio de funcionamiento es el mismo. La juntura resultante entre la base y el emisor se polariza en conducción directa mediante una fuente de tensión externa, lo cual provoca una inyección de los portadores mayoritarios del emisor hacia la base. Por otro lado, la juntura base colector se polariza en inversa mediante una tensión entre sus terminales. La base se diseña lo suficientemente delgada como para que los portadores inyectados desde el emisor, que en la base son minoritarios, no se recombinen y alcancen la zona de vaciamiento de la juntura en inversa, en la cual son arrastrados por el campo eléctrico hacia el colector.

En el caso de un transistor PNP se inyectan huecos del emisor a la base, los cuales continúan casi en su mayoría hacia el colector. También se inyectan electrones de la base al emisor, que son recogidos por la fuente de alimentación. El transistor bipolar PNP se construye en forma especial para que la corriente de electrones de base a emisor sea muy pequeña. Como resultado, en este dispositivo se produce una corriente significativa de huecos que van de emisor a colector, y una corriente muy pequeña de electrones que entran por la base y salen por el emisor. Como se verá en la próxima sección, la corriente que circula por el colector es, bajo ciertas condiciones de operación, proporcional y mucho mayor que la corriente que circula por la base, por lo cual el dispositivo se comporta como una fuente de corriente (colocada entre emisor y colector) controlada por una pequeña corriente de base¹.

En el caso de un transistor NPN, por el contrario, se inyectan electrones del emisor a la base, los cuales continúan casi en su mayoría hacia el colector. También se inyectan huecos de la base al emisor, que son recogidos por la fuente de alimentación. El transistor NPN se construye especialmente para que la corriente de huecos de la base al emisor sea muy pequeña. En este dispositivo, se logra una conducción significativa de electrones que van del emisor al colector, y una corriente muy pequeña de huecos que entran por la base y salen por el emisor. Como en el caso del transistor PNP, la corriente que circula por el colector es, bajo ciertas condiciones de operación, proporcional y mucho mayor que la corriente que circula por la base, por lo cual el dispositivo se comporta como una fuente de corriente (colocada entre emisor y colector) controlada por una pequeña corriente de base.

¹También puede considerarse que la variable de control es la tensión entre base y emisor.

En ambos dispositivos, el efecto final es el mismo, y consiste en la posibilidad de controlar el flujo de corriente principal entre emisor y colector, mediante una pequeña corriente de base.

La Fig. 6.1 muestra los símbolos comúnmente utilizados para representar a los transistores bipolares y la convención de corrientes.

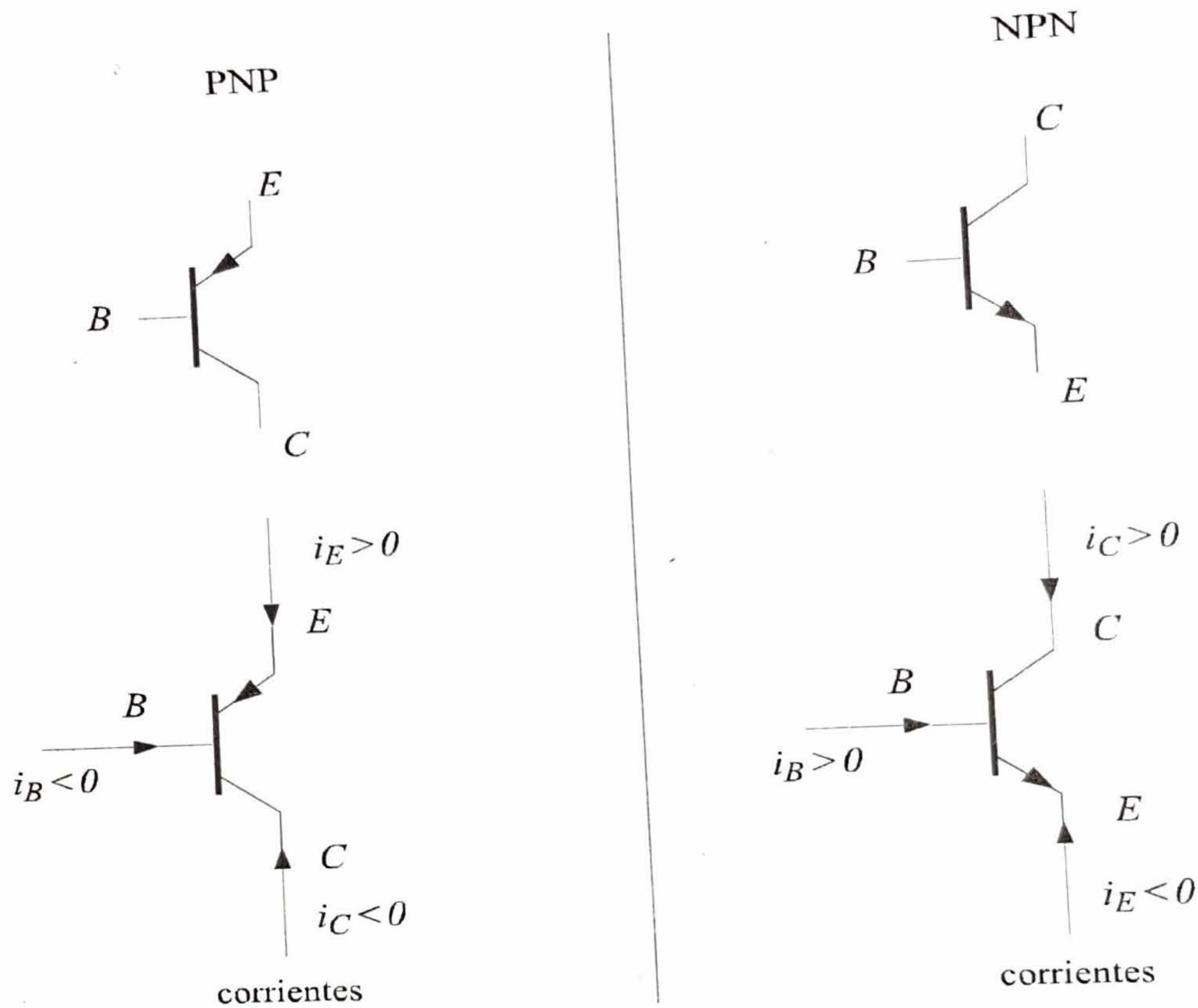


Figura 6.1: Símbolos de los transistores bipolares y convenciones de corriente. Se muestra el signo de las corrientes para el modo de operación estándar.

6.2. Modelo de DC del transistor PNP

Constructivamente, un transistor bipolar PNP se logra haciendo una implantación profunda de tipo P, fuertemente dopada, sobre un sustrato de Silicio para obtener un colector de buena conductividad. A continuación, mediante un proceso epitaxial se hace crecer Silicio dopado P, y sobre éste se realiza el implante de base tipo N, y el implante de emisor tipo P fuertemente dopado. La situación se ilustra en la Fig. 6.2. Estos pasos adicionales, no requeridos para la realización de transistores MOS, hacen que la integración de transistores bipolares sea más compleja y por ende, más cara.

6.2.1. Región de conducción activa directa

Para este análisis, se utilizará la estructura que se muestra en la Fig. 6.3, donde la longitud del emisor es W_E , la de la base W_B y la del colector W_C . En el caso de conducción *activa directa* ("forward active"), el potencial del emisor es el mayor de los tres. El potencial de la base es menor que el del emisor, de tal manera que la juntura EB esté polarizada en conducción directa. A su vez, el

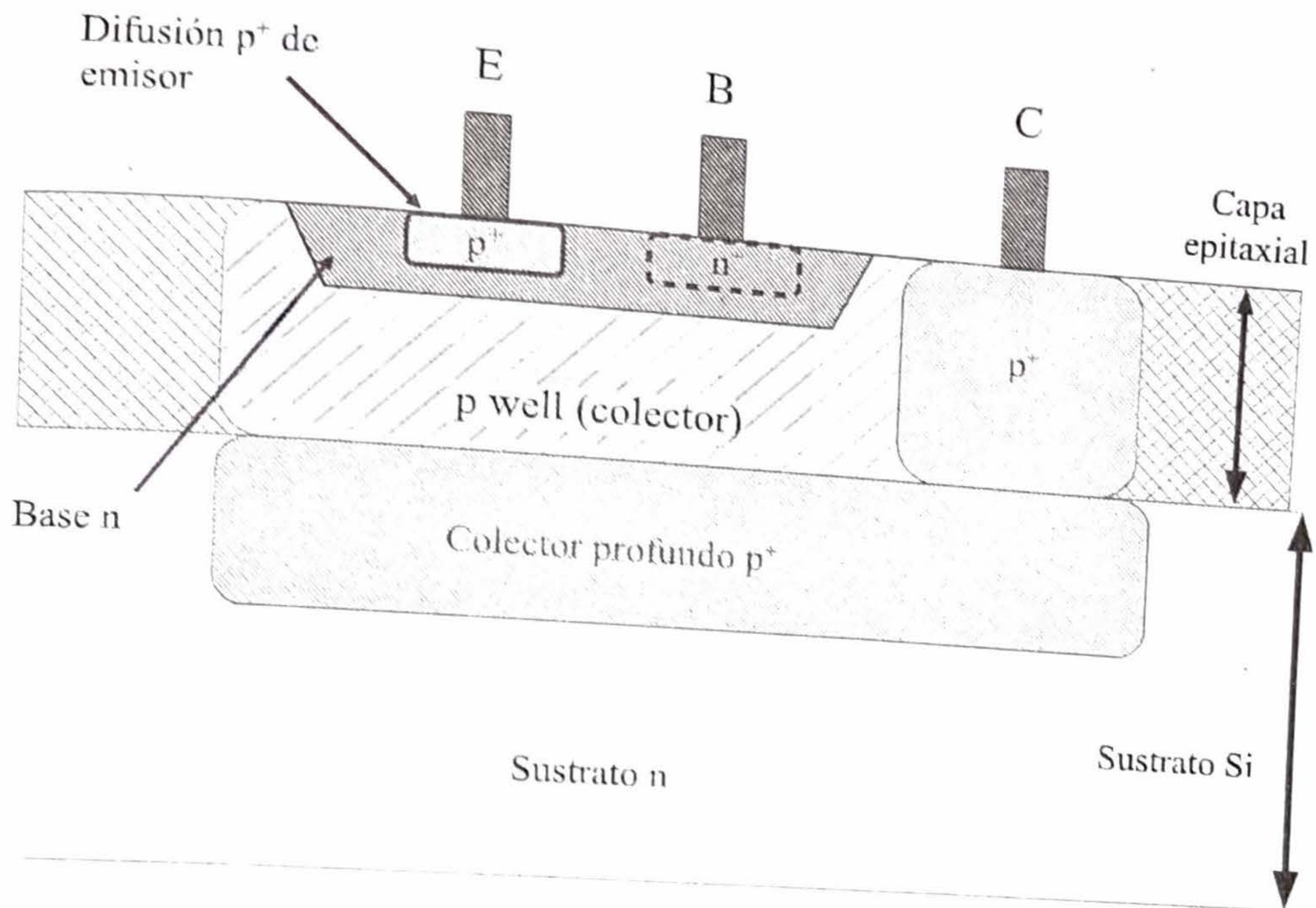


Figura 6.2: Esquema de un transistor PNP integrado. La implantación de colector profundo requiere un paso especial de fabricación.

potencial del colector es menor que el de la base, de tal manera que la juntura CB esté en conducción inversa. Se verifica entonces

$$v_{EB} > 0 \quad v_{CB} < 0 \tag{6.1}$$

De igual manera que con el análisis de la juntura, se utilizará la aproximación de juntura corta, es decir, se supondrá que W_E , W_B , y W_C son de una longitud lo suficientemente pequeña como para que los portadores se trasladen sin sufrir recombinación. Asimismo, para simplificar la notación, se despreciarán las longitudes de las zonas de vaciamiento. Del análisis de la juntura se puede ver que, debido a la polarización en directa de la juntura EB, los niveles de portadores a ambos lados de esta zona de vaciamiento se incrementan en forma exponencial con la tensión v_{EB} . Por otro lado, la polarización en inversa de la juntura CB hace que los niveles de portadores a ambos lados de la zona de vaciamiento se reduzcan a cero. Esta situación se ilustra en la Fig. 6.4.

La generación de corriente se origina en la juntura EB, dado que la misma está en conducción directa, mientras que la juntura CB, que está en conducción inversa se limita a transportar los portadores emitidos por la primera².

El primer paso para calcular las corrientes del transistor es establecer el valor de las corrientes J_1 y J_2 , de la Fig. 6.3. La corriente de difusión de huecos J_1 , proveniente del emisor, se puede escribir en función de la concentración de portadores en la base como:

$$J_1 = -qD_p \frac{p(W_B^-) - p(0^+)}{W_B} \tag{6.2}$$

donde:

²En realidad, también se genera una pequeña corriente inversa en la juntura CB, producto de la diferencia de concentraciones en ambos extremos de la región del colector, $(n_{PC} - 0)$, como se puede ver en la Fig. 6.4. Sin embargo, esta componente es muy pequeña y será despreciada para el presente análisis. En algunos casos especiales es necesario tener en cuenta esta corriente, por lo cual, se la desarrolla en detalle en la Sección 6.2.5

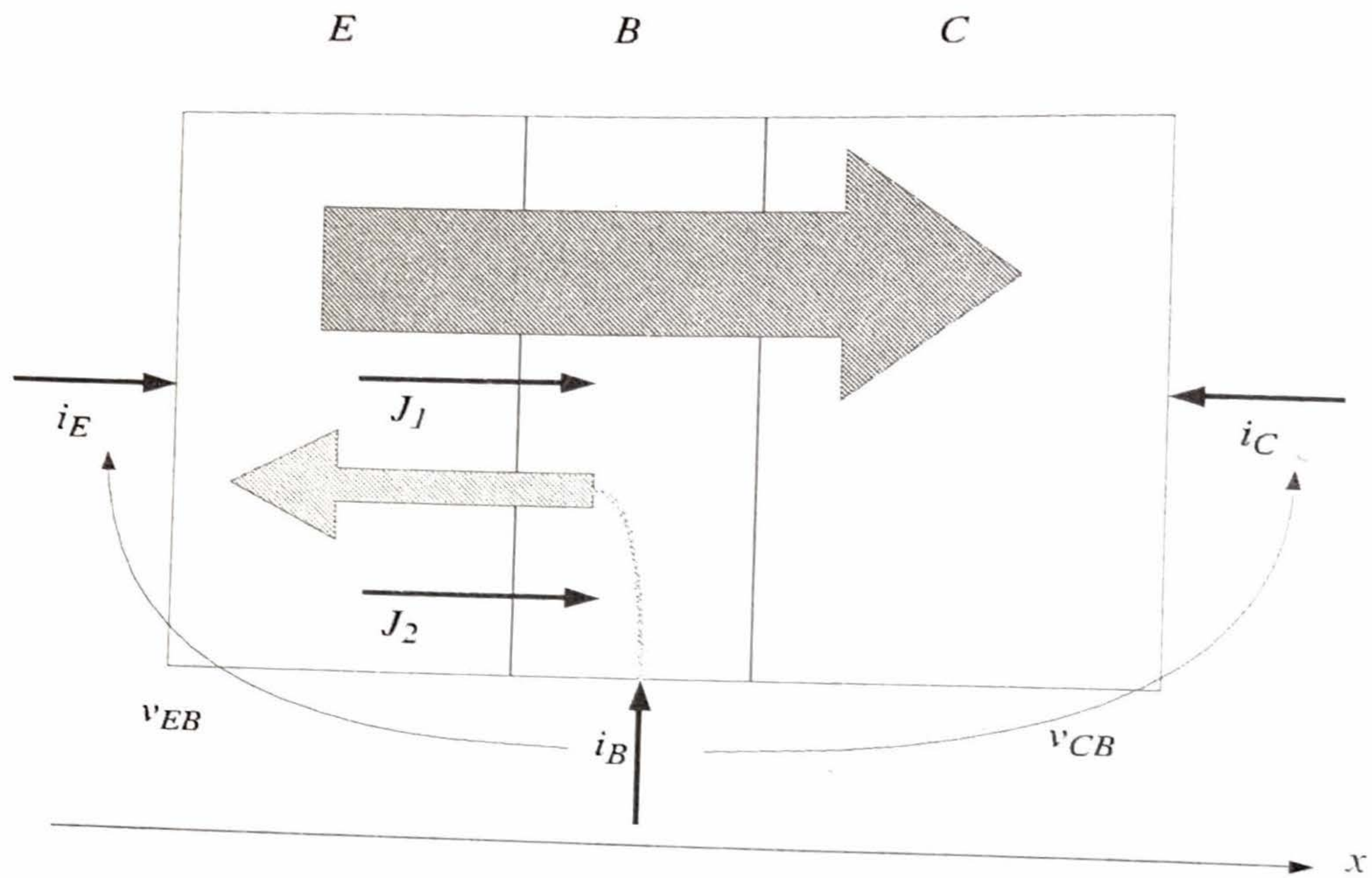


Figura 6.3: Esquema de un transistor bipolar PNP y las corrientes en el modo de conducción activo directo.

$$p(W_B^-) = 0 \quad (6.3)$$

y considerando que, $\phi_{J_{EB}} = \phi_{B_{EB}} - v_{EB}$, y $p_{NB} = N_{aE} e^{-\phi_{B_{EB}}/U_T}$, resulta:

$$p(0^+) = N_{aE} e^{-\phi_{J_{EB}}/U_T} = p_{NB} e^{v_{EB}/U_T} \quad (6.4)$$

Utilizando estas dos ecuaciones en (6.2), resulta:

$$J_1 = qD_p \frac{p_{NB} e^{v_{EB}/U_T}}{W_B} \quad (6.5)$$

La corriente de difusión de electrones J_2 , que ocurre en el material P, se puede escribir como:

$$J_2 = qD_n \frac{n(0^-) - n(-W_E)}{W_E} \quad (6.6)$$

donde:

$$n(0^-) = N_{dB} e^{-\phi_{J_{EB}}/U_T} = n_{PE} e^{v_{EB}/U_T} \quad (6.7)$$

y

$$n(-W_E) = n_{PE} \quad (6.8)$$

La corriente J_2 resulta de la siguiente forma:

$$J_2 = qD_n \frac{n_{PE} e^{v_{EB}/U_T} - n_{PE}}{W_E} = qD_n \frac{n_{PE} (e^{v_{EB}/U_T} - 1)}{W_E} \quad (6.9)$$

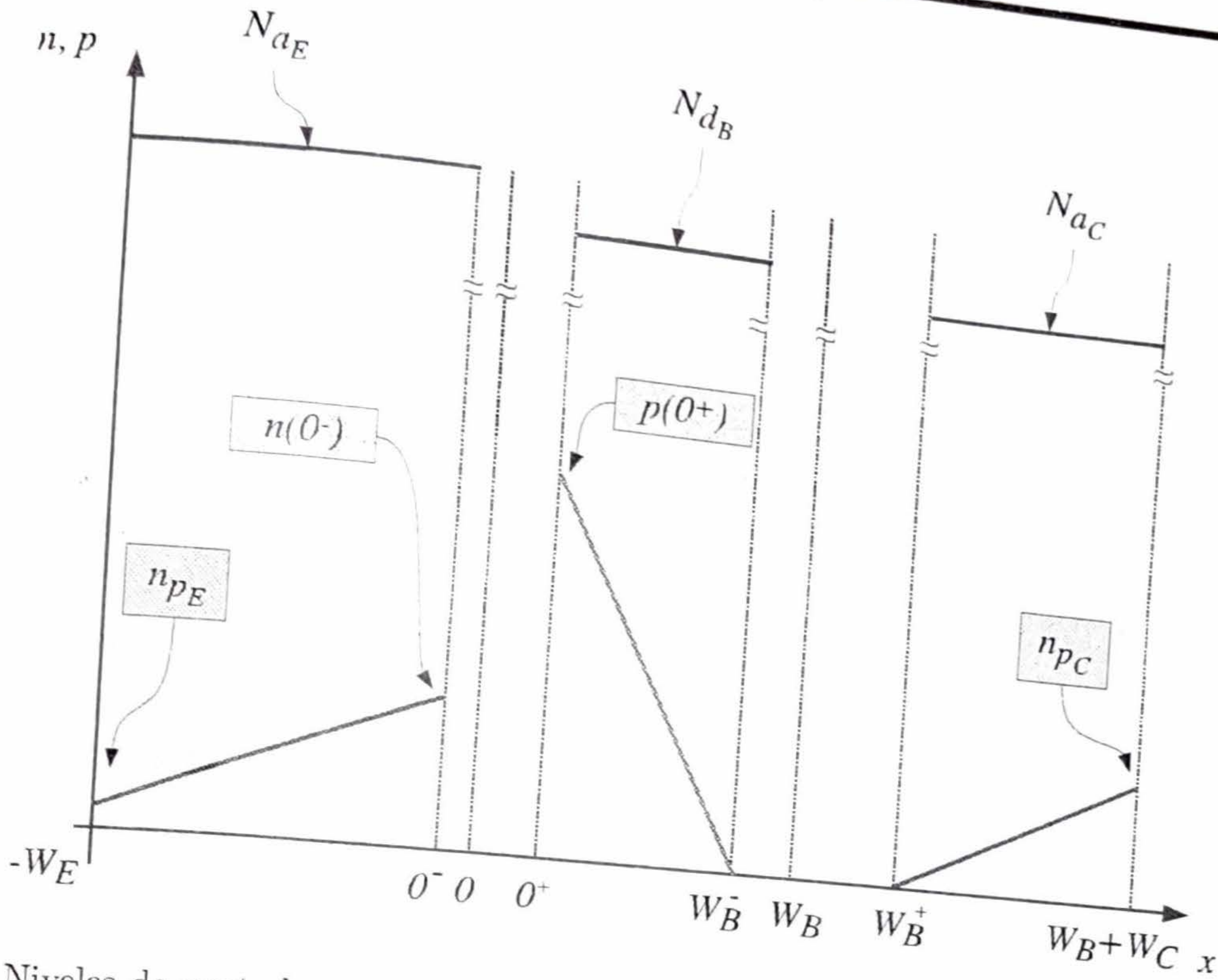


Figura 6.4: Niveles de portadores en un transistor PNP en el modo de conducción activo directo.

En base a estas dos corrientes, es posible resolver las corrientes de los terminales del transistor. De hecho, planteando nodos resulta:

$$\begin{cases} i_E = A(J_1 + J_2) \\ i_B = -AJ_2 \\ i_C = -AJ_1 \end{cases} \quad (6.10)$$

A partir de (6.10) se pueden calcular las expresiones de las corrientes individuales. La corriente de emisor es la suma de las contribuciones de los huecos, que van de emisor a colector, y de los electrones, que van de base a emisor, razón por la cual es la mayor de las tres en magnitud:

$$i_E = AqD_p \frac{p_{NB} e^{v_{EB}/U_T}}{W_B} + AqD_n \frac{n_{PE} (e^{v_{EB}/U_T} - 1)}{W_E} \quad (6.11)$$

Considerando que en conducción directa el término -1 es muy pequeño con respecto al término exponencial, (6.11) puede escribirse como:

$$i_E = A \left(\frac{qD_p p_{NB}}{W_B} + \frac{qD_n n_{PE}}{W_E} \right) e^{v_{EB}/U_T} \quad (6.12)$$

Nótese que con la convención elegida, en la cual la corriente de emisor entra al dispositivo, en conducción directa, es decir con $v_{EB} > 0$, resulta $i_E > 0$.

La corriente de base está compuesta solamente por los electrones que circulan hacia el emisor. Estos electrones circulan de base a emisor y retornan a la base gracias a la fuente de alimentación. La expresión de esta corriente es:

$$i_B = -A \frac{qD_n n_{PE}}{W_E} (e^{v_{EB}/U_T} - 1) \quad (6.13)$$

Por la convención elegida, la corriente de base resulta negativa, $i_B < 0$.

Por último, la corriente de colector está compuesta solo por los huecos que salen del emisor, atraviesan la base, y llegan al colector. La diferencia entre la corriente de colector y la de emisor es la corriente de electrones que circula por la base; en otras palabras, la corriente de colector es ligeramente menor en módulo que la corriente de emisor:

$$i_C = -A \frac{qD_p p_{NB}}{W_B} e^{v_{EB}/U_T} \quad (6.14)$$

Esta corriente también resulta negativa con la convención elegida.

El cociente entre la corriente de colector y la corriente de emisor, en activa directa, es un parámetro importante del dispositivo, y se define de la siguiente manera:

$$\begin{aligned} \alpha_F &\triangleq -\frac{i_C}{i_E} \\ &= \frac{\left(\frac{qD_p p_{NB}}{W_B}\right)}{\left(\frac{qD_p p_{NB}}{W_B}\right) + \left(\frac{qD_n n_{PE}}{W_E}\right)} \\ &= \frac{1}{1 + \frac{D_n n_{PE} W_B}{D_p p_{NB} W_E}} \end{aligned} \quad (6.15)$$

Dado que los niveles de portadores minoritarios satisfacen:

$$p_{NB} = \frac{n_i^2}{N_{dB}} \quad n_{PE} = \frac{n_i^2}{N_{aE}} \quad (6.16)$$

α_F puede escribirse como:

$$\alpha_F = \frac{1}{1 + \frac{D_n N_{dB} W_B}{D_p N_{aE} W_E}} \quad (6.17)$$

El término α_F es siempre menor a 1, como se puede deducir de (6.17) para valores típicos de los parámetros. Por ejemplo, si se considera $N_{aE} = 10^{18} \text{ cm}^{-3}$, $N_{dB} = 10^{16} \text{ cm}^{-3}$, y $W_B \approx W_E$, se obtiene $D_p \approx 1,8D_n$, y el valor de α_F es:

$$\alpha_F = \frac{1}{1 + \frac{1}{1,8 \times 100 \times 1}} = 0,9945 \quad (6.18)$$

Otro parámetro importante es la relación entre la corriente de colector y la corriente de base:

$$\begin{aligned} \beta_F &\triangleq \frac{i_C}{i_B} \\ &= \frac{i_C}{-i_C - i_E} = \frac{\alpha_F}{1 - \alpha_F} \\ &= \frac{D_p p_{NB} W_E}{D_n n_{PE} W_B} \end{aligned} \quad (6.19)$$

6.2. Modelo de DC del transistor PNP

Reemplazando en función de los niveles de dopado:

$$\beta_F = \frac{D_p N_{aE} W_E}{D_n N_{dB} W_B} \quad (6.20)$$

Para los mismos valores típicos, el valor de β_F resulta:

$$\beta_F = 1,8 \times 100 \times 1 = 180 \quad (6.21)$$

Si se interpreta la corriente de colector como corriente de salida, y la corriente de base como corriente de entrada, el parámetro β_F define la ganancia de corriente. Por esta razón, interesa que el mismo sea lo más grande posible. En función de (6.20), se puede ver que para lograr un β_F grande se deben hacer dos cosas:

1. Dopar el emisor en mayor medida que la base:

$$N_{aE} \gg N_{dB} \quad (6.22)$$

De esta manera, el flujo de huecos del emisor a la base será mucho mayor que el flujo de electrones de la base al emisor, y la corriente de base será pequeña con respecto a i_E e i_C .

2. Hacer la base angosta:

$$W_B \ll W_E \quad (6.23)$$

Esto hace que el perfil de portadores en la base sea más pronunciado, provocando una mayor corriente de difusión. Por otro lado, al hacer la base corta se evita que los huecos provenientes del emisor se recombinen con los electrones de la base.

El término D_p/D_n es menor a uno, producto de la menor difusividad y/o movilidad de los huecos con respecto a los electrones. Dado que en el transistor PNP la corriente principal es de huecos, β_F será menor que en los transistores NPN.

La Ec. (6.14) representa la curva de transferencia, que relaciona la corriente de salida versus la tensión de entrada, como se ilustra en la Fig. 6.5 (para $v_{CE} = 380\text{mV}$). En la zona de funcionamiento activa directa, la corriente de colector depende de la corriente de base, o de la tensión v_{EB} , y no depende de la tensión v_{CB} . Entonces, en esta región de funcionamiento el transistor puede considerarse como una fuente de corriente controlada por la corriente de base i_B , o controlada por la tensión v_{EB} .

Si en (6.14), se reemplaza la corriente de colector por la de base, se obtiene la curva de entrada del transistor bipolar:

$$i_B = -\frac{1}{\beta_F} \frac{AqD_p p_{NB}}{W_B} \left(e^{v_{EB}/U_T} - 1 \right) \quad (6.24)$$

que es igual a la curva de transferencia, salvo por el factor de escala β_F .

Comúnmente, se toma como tensión de salida del transistor a la tensión v_{CE} (en lugar de la tensión v_{CB}). En la región activa directa, las curvas de salida ideales del dispositivo, i_C versus v_{CE} , son rectas horizontales, parametrizadas en función de la tensión v_{EB} , o en función de la corriente de base i_B . La Fig. 6.6 muestra las curvas experimentales de salida de un transistor PNP, donde la región $v_{CE} < -0,6V$ corresponde a la región activa directa⁴. Para valores mayores (hacia la derecha) que $v_{CE} = -0,6V$, o

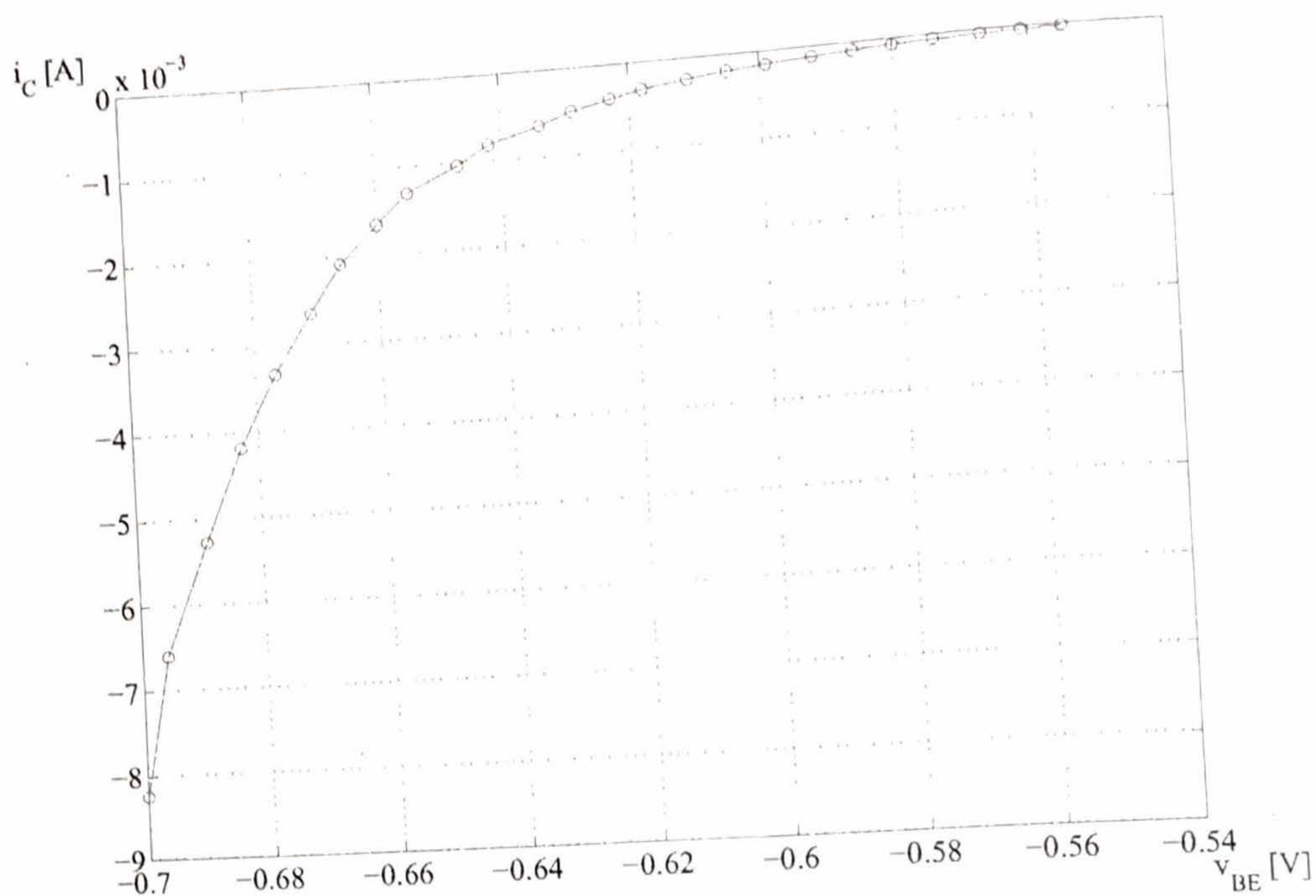


Figura 6.5: Curva experimental de transferencia de un transistor PNP BC558 para $v_{CE} = 380\text{mV}$.

$v_{CB} = 0\text{V}$, la juntura de CB comienza a estar en zona directa, y el transistor pasa a otra región de funcionamiento, denominada de saturación.

En un dispositivo real, al igual que en el caso de los transistores MOS, la corriente experimenta una variación cuando la tensión v_C se varía. De hecho, al hacerse la tensión de colector más negativa con respecto a la base, la zona de vaciamiento de la juntura CB aumenta, reduciendo la longitud efectiva de la base. Esto hace que la pendiente del perfil de portadores sea mayor, lo cual implica un aumento de la corriente. Esta variación de la corriente en función de v_{CB} , se modela de la misma manera que en el transistor MOS, utilizando una tensión de Early V_A ($V_A < 0$):

$$i_C = \beta_F i_B \left(1 + \frac{v_{CB}}{V_A} \right) \tag{6.27}$$

ó

$$i_C = -A \frac{qD_p p_{NB}}{W_B} \left(e^{v_{EB}/U_T} - 1 \right) \left(1 + \frac{v_{CB}}{V_A} \right) \tag{6.28}$$

³Dado que la corriente de salida en activa directa es función de la tensión v_{CE} por el efecto de Early, la curva de transferencia debe hallarse idealmente para una tensión $v_{CE} = V_{CE(sat)}$. En este punto, la corriente no es alterada por el efecto de Early y la corriente solo es función de v_{EB} .

⁴En algunos casos, los fabricantes pueden especificar las curvas de corriente i_C versus v_{CB} . En estos casos, si se plantea la ecuación de mallas en el dispositivo, resulta:

$$v_{CB} + v_{BE} + v_{EC} = 0 \tag{6.25}$$

con lo cual, la tensión v_{CE} puede expresarse como:

$$v_{CE} = -v_{EB} + v_{CB} \tag{6.26}$$

Nótese que los puntos de corriente, para los cuales $v_{CB} = 0$, se trasladan a $v_{CE} = -v_{EB}$. En activa directa, la juntura EB está en directa, y su valor de tensión es aproximadamente $v_{EB} = 0,6\text{V}$. Debido a esto, $v_{CE} \approx v_{CB} - 0,6\text{V}$, y la curva de salida en función de v_{CB} es una versión desplazada hacia la derecha de la curva versus v_{CE} .

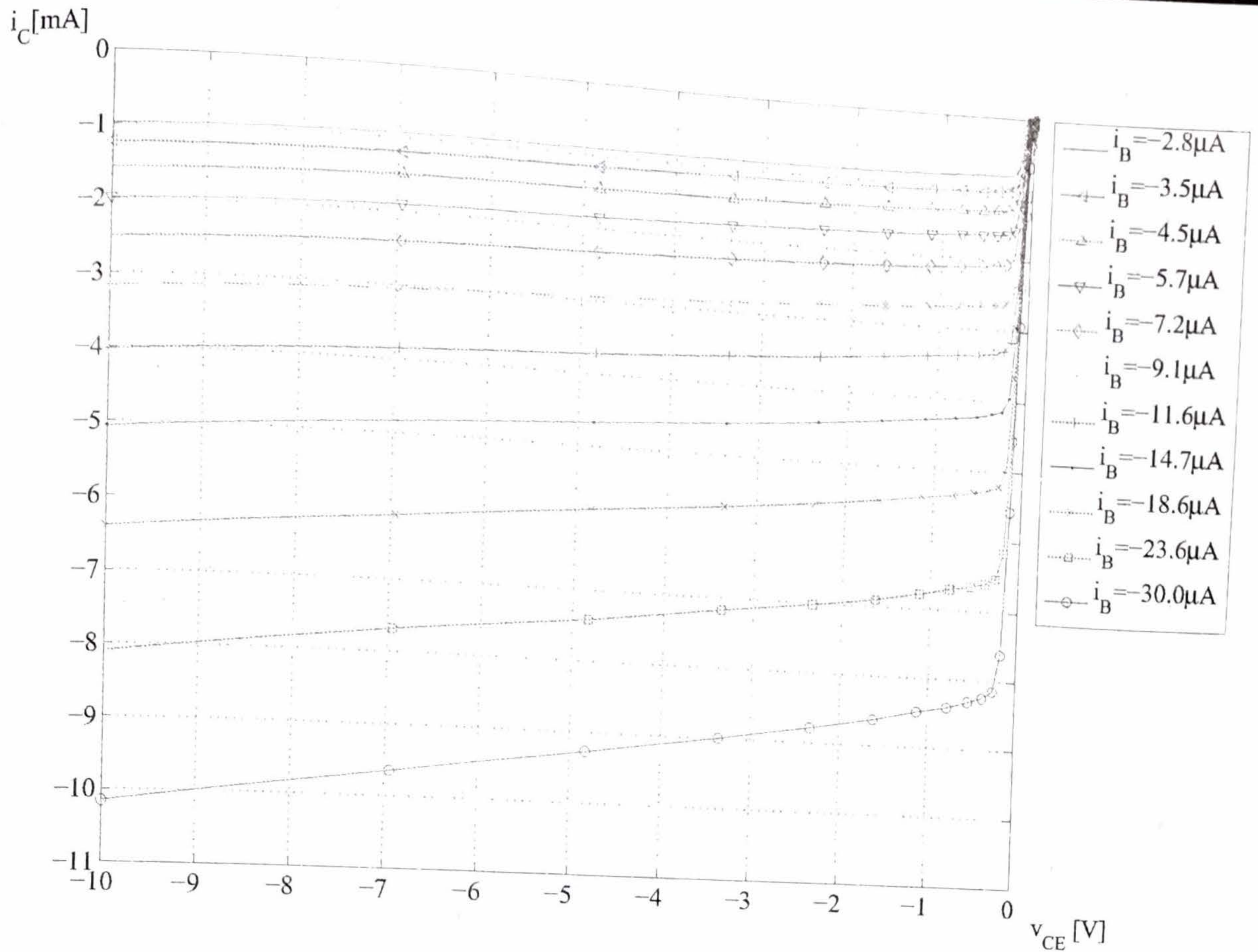


Figura 6.6: Curvas experimentales de salida de un transistor PNP BC558 para varios valores de corriente de base.

Similares expresiones son válidas, en caso de utilizar v_{CE} :

$$i_C = \beta i_B \left(1 + \frac{v_{CE} - V_{CE(sat)}}{V_A} \right) \quad (6.29)$$

ó

$$i_C = -A \frac{q D_{pPNB}}{W_B} \left(e^{v_{EB}/U_T} - 1 \right) \left(1 + \frac{v_{CE} - V_{CE(sat)}}{V_A} \right) \quad (6.30)$$

donde $V_{CE(sat)}$ es el valor de v_{CE} a partir del cual el dispositivo está en activa directa.

El efecto de la tensión v_{CB} sobre la zona de vaciamiento de la juntura CB se puede ver con el auxilio de las Ecs. (3.27) y (3.28). La zona de vaciamiento de la juntura BC, en la región de la base, está dada por:

$$x_{nB} = \sqrt{\frac{2\epsilon S_i(\phi_B - v_{CB})}{q}} \sqrt{\frac{N_{aC}}{N_{dB}(N_{aC} + N_{dB})}} \quad (6.31)$$

mientras que en la región del colector:

$$x_{pC} = \sqrt{\frac{2\epsilon S_i(\phi_B - v_{CB})}{q}} \sqrt{\frac{N_{dB}}{N_{aC}(N_{aC} + N_{dB})}} \quad (6.32)$$

La relación entre ambas es:

$$\frac{x_{nB}}{x_{pC}} = \sqrt{\frac{N_{aC}^2}{N_{dB}^2}} = \frac{N_{aC}}{N_{dB}} \quad (6.33)$$

La variación de la corriente, en el caso del transistor bipolar, se produce por la reducción de la base, que cambia el perfil de portadores y aumenta la corriente. De (6.33), se puede ver claramente que para minimizar la reducción de la base, se debe satisfacer:

$$N_{aC} \ll N_{dB} \quad (6.34)$$

En consecuencia, de (6.22) y (6.34), se deduce que para los transistores bipolares la relación de dopado es descendente a medida que se recorre el dispositivo desde el emisor hacia el colector:

$$\boxed{N_{aE} \gg N_{dB} \gg N_{aC}} \quad (6.35)$$

Esto hace que –en la práctica– el dispositivo no tenga características simétricas. De todas maneras, los terminales de colector y emisor pueden intercambiarse y el dispositivo sigue funcionando, como se muestra en la sección siguiente, aunque con peor desempeño.

6.2.2. Región de conducción activa inversa

La región *activa inversa* se caracteriza por tener la juntura EB en conducción inversa, y la juntura CB en conducción directa, es decir:

$$v_{EB} < 0 \quad v_{CB} > 0 \quad (6.36)$$

Si el dopado de las regiones de colector y emisor fuese el mismo, este modo de funcionamiento sería indistinguible del modo activo directo, y los terminales de colector y emisor serían intercambiables.

Para el desarrollo de este caso, se utilizará la estructura de la Fig. 6.7, donde se ve una conducción de huecos, del colector hacia la base, que son arrastrados por el campo eléctrico de la zona de vaciamiento EB hacia el emisor. Por otro lado, hay también una conducción de electrones desde la base hacia el colector. El perfil resultante de portadores se ilustra en la Fig. 6.8.

La corriente de huecos, que se dirigen del colector al emisor, se calcula como:

$$J_1 = -qD_p \frac{p(W_B^-) - p(0^+)}{W_B} \quad (6.37)$$

donde ahora:

$$p(W_B^-) = N_{aC} e^{-\phi_{J_{CB}}/U_T} = p_{NB} e^{v_{CB}/U_T} \quad (6.38)$$

y

$$p(0^+) = 0 \quad (6.39)$$

Considerando que $\phi_{J_{CB}} = \phi_{B_{CB}} - v_{CB}$, y $p_{NB} = N_{aC} e^{-\phi_{B_{CB}}/U_T}$, (6.37) se puede escribir de la siguiente forma:

$$J_1 = -qD_p \frac{p_{NB} e^{v_{CB}/U_T}}{W_B} \quad (6.40)$$

La corriente de difusión de electrones, que se dirigen de la base al colector, se puede escribir como:

$$J_2 = qD_n \frac{n(W_B + W_C) - n(W_B^+)}{W_C} \quad (6.41)$$

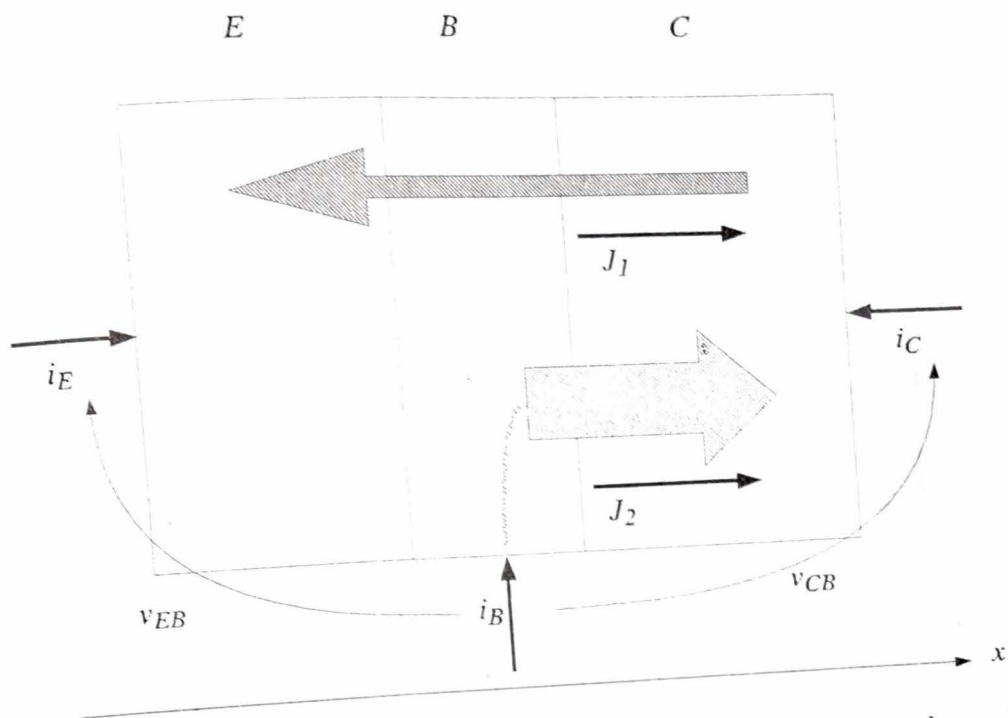


Figura 6.7: Esquema de un transistor bipolar PNP y las corrientes en el modo de conducción activo inverso.

donde:

$$n(W_B^+) = N_{dB} e^{-\phi_{J_{CB}}/U_T} = n_{PC} e^{v_{CB}/U_T} \quad (6.42)$$

y

$$n(W_B + W_C) = n_{PC} \quad (6.43)$$

La corriente J_2 resulta de la siguiente forma:

$$J_2 = -qD_n \frac{n_{PC} (e^{v_{CB}/U_T} - 1)}{W_C} \quad (6.44)$$

En función de estas dos corrientes, se pueden calcular las corrientes de los terminales. La corriente de colector es la suma de las contribuciones de los huecos, que van de colector a emisor, y de los electrones, que van de base a colector, razón por la cual es la mayor de las tres en magnitud:

$$i_C = AqD_p \frac{p_{NB} e^{v_{CB}/U_T}}{W_B} + AqD_n \frac{n_{PC} (e^{v_{CB}/U_T} - 1)}{W_C} \quad (6.45)$$

y considerando que, en conducción directa, el término -1 es muy pequeño con respecto al término exponencial, (6.45) puede escribirse como:

$$i_C = A \left(\frac{qD_p p_{NB}}{W_B} + \frac{qD_n n_{PC}}{W_C} \right) (e^{v_{CB}/U_T} - 1) \quad (6.46)$$

Nótese que con la convención elegida, en la cual la corriente de colector entra al dispositivo, en conducción inversa, es decir con $v_{CB} > 0$, resulta $i_C > 0$.

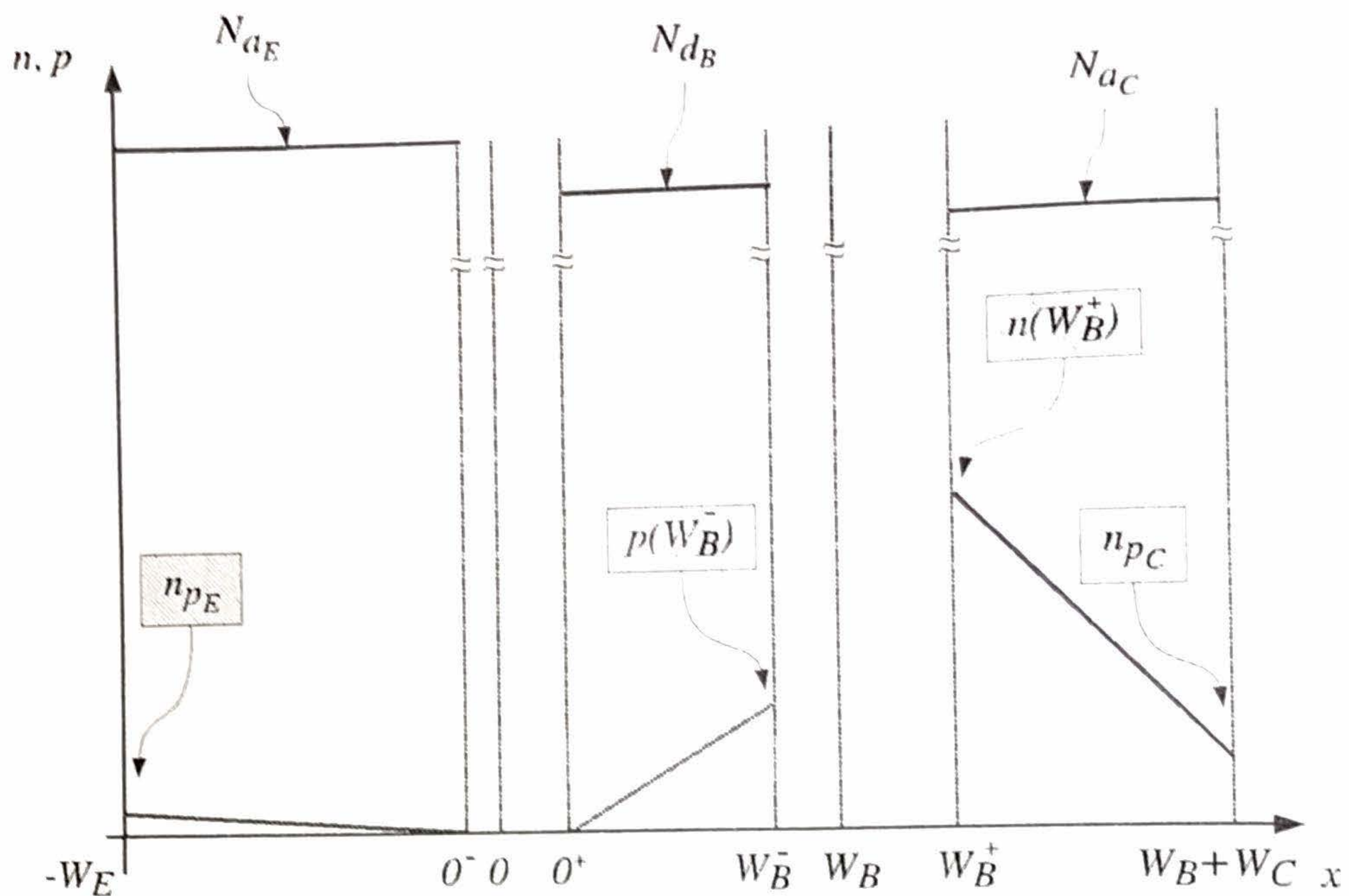


Figura 6.8: Niveles de portadores en un transistor PNP en el modo de conducción activo inverso.

La corriente de base está compuesta solamente por los electrones que circulan hacia el colector. Estos electrones circulan de base a colector y son provistos por la fuente de alimentación. La expresión de esta corriente es:

$$i_B = -A \frac{qD_n n_{PC}}{W_C} \left(e^{v_{CB}/U_T} - 1 \right) \quad (6.47)$$

La corriente de base, en este caso, sigue siendo negativa, $i_B < 0$.

Por último, la corriente de emisor está compuesta solo por los huecos que salen del colector, atraviesan la base, y llegan al emisor, y es ligeramente menor en magnitud que la corriente de colector:

$$i_E = -A \frac{qD_p p_{NB}}{W_B} \left(e^{v_{CB}/U_T} - 1 \right) \quad (6.48)$$

Esta corriente también resulta negativa con la convención elegida.

En este caso, el cociente entre corriente de emisor y corriente de colector, se define de la siguiente manera:

$$\begin{aligned} \alpha_R &\triangleq -\frac{i_E}{i_C} \\ &= \frac{\left(\frac{qD_p p_{NB}}{W_B} \right)}{\left(\frac{qD_p p_{NB}}{W_B} \right) + \left(\frac{qD_n n_{PC}}{W_C} \right)} \\ &= \frac{1}{1 + \frac{D_n n_{PC} W_B}{D_p p_{NB} W_C}} \end{aligned} \quad (6.49)$$

y reemplazando en función de los niveles de dopado:

$$\alpha_R = \frac{1}{1 + \frac{D_n N_{dB} W_B}{D_p N_{aC} W_C}} \quad (6.50)$$

El término α_R es también menor a 1; con valores típicos ($N_{aC} = 10^{15} \text{cm}^{-3}$, $N_{dB} = 10^{16} \text{cm}^{-3}$, y $W_B \approx W_C$), $D_n \approx 3D_p$ y :

$$\alpha_R = \frac{1}{1 + 3 \times 10} = 0,032 \quad (6.51)$$

El cociente entre la corriente de emisor y la corriente de base, en activa reversa, se define de la siguiente manera:

$$\begin{aligned} \beta_R &\triangleq \frac{i_E}{i_B} \\ &= \frac{i_E}{-i_E - i_C} = \frac{\alpha_R}{1 - \alpha_R} \\ &= \frac{D_p p_{NB} W_C}{D_n n_{PC} W_B} \end{aligned} \quad (6.52)$$

En función de los niveles de dopado, el parámetro β_R , se puede expresar como:

$$\beta_R = \frac{D_p N_{aC} W_C}{D_n N_{dB} W_B} \quad (6.53)$$

Para los valores típicos indicados, resulta:

$$\beta_R = \frac{1}{3 \times 10} = 0,033 \quad (6.54)$$

La Ec. (6.53) pone en evidencia una diferencia *sustancial* entre los modos directo e inverso. Dado que $N_{aC} \ll N_{dB}$, resulta $\beta_R \ll \beta_F$. Es sencillo ver de (6.52) que si $\alpha_R \ll 1$, entonces $\beta_R \approx \alpha_R$; por lo tanto, β_R resulta un valor muy pequeño. Esto resulta en una pequeña corriente de emisor, y una corriente de colector aproximadamente igual a la corriente de base.

Las curvas de salida, en este caso, expresan la variación de i_E versus v_{EC} , y se ven afectadas en mayor medida por el efecto de Early ($V_A < 0$ resulta menor en valor absoluto con respecto al modo activo directo) dado que el dopado del emisor es mayor que el de la base:

$$i_E = \beta_R i_B \left(1 + \frac{v_{EB}}{V_A} \right) \quad (6.55)$$

$$\text{ó} \quad i_E = -A \frac{q D_p p_{NB}}{W_B} \left(e^{v_{CB}/U_T} - 1 \right) \left(1 + \frac{v_{EB}}{V_A} \right) \quad (6.56)$$

Similares expresiones son válidas, en caso de utilizar v_{EC}

$$i_E = \beta_R i_B \left(1 + \frac{v_{EC} - V_{EC(sat)}}{V_A} \right) \quad (6.57)$$

$$\text{ó} \quad i_E = -A \frac{q D_p p_{NB}}{W_B} \left(e^{v_{CB}/U_T} - 1 \right) \left(1 + \frac{v_{EC} - V_{EC(sat)}}{V_A} \right) \quad (6.58)$$

donde $V_{EC(sat)}$ es el valor de v_{EC} a partir del cual el dispositivo está en activa directa.

6.2.3. Región de saturación y el modelo de Ebers-Moll

En el caso en que la juntura EB está en directa y la juntura CB queda ligeramente en conducción directa, el dispositivo opera en la zona de funcionamiento denominada *saturación*. Esta región está caracterizada por:

$$v_{EB} > 0 \quad v_{CB} > 0 \quad (6.59)$$

En la práctica, la conducción directa de la juntura CB comienza a apreciarse cuando $v_{CE} \approx -0,2V$ y $v_{CB} \approx 0,5V$.

Para el análisis de esta región, se utilizará la estructura de la Fig. 6.9. Aquí se puede apreciar la presencia de dos corrientes adicionales J'_1 y J'_2 , con respecto al modo de conducción directa activa. Estas dos corrientes son producto de la conducción directa de la juntura CB. Es importante notar que ambas junturas, aún estando en conducción directa, tienen una zona de vaciamiento con un campo eléctrico interno. En el caso de la juntura CB, este campo eléctrico arrastra los huecos inyectados (J_1) por la juntura EB hacia el colector. En el caso de la juntura EB, el campo eléctrico arrastra los huecos inyectados (J'_1) por la juntura CB hacia el emisor.

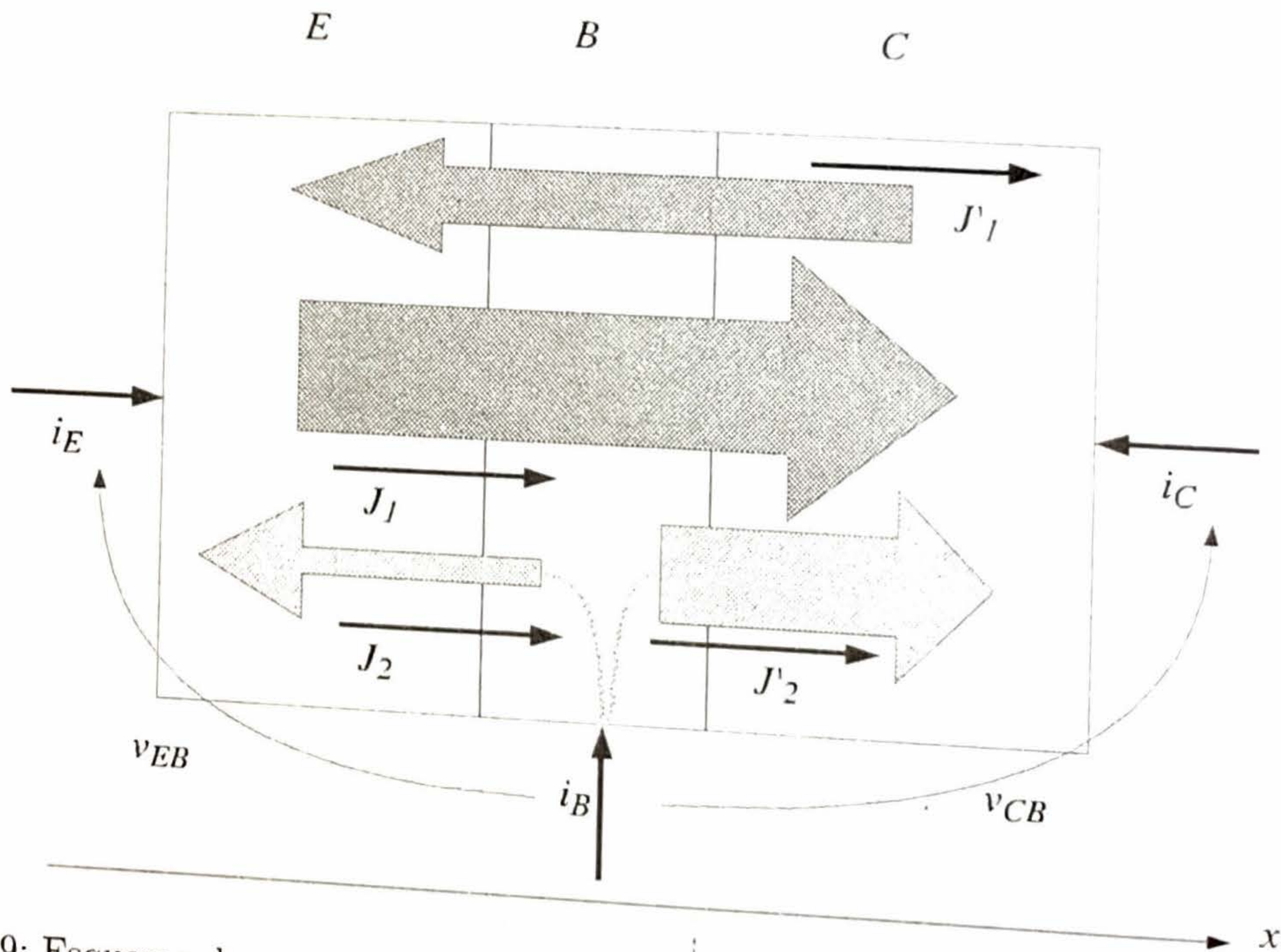


Figura 6.9: Esquema de un transistor bipolar PNP y las corrientes en el modo de saturación.

En este modo de funcionamiento, las corrientes en el dispositivo pueden hallarse haciendo una superposición de los casos de conducción directa activa e inversa activa, es decir, considerando las corrientes individuales J_1 , J_2 , J'_1 y J'_2 , de (6.5), (6.9), (6.40), y (6.44), respectivamente.

Planteando las ecuaciones de nodos en el emisor y el colector:

$$i_E = A_E(J_1 + J_2 + J'_1) \quad (6.60)$$

$$i_C = -A_C(J_1 + J'_1 + J'_2) \quad (6.61)$$

Reemplazando las expresiones (6.5), (6.9), (6.40) y (6.44), las corrientes de emisor y colector resultan:

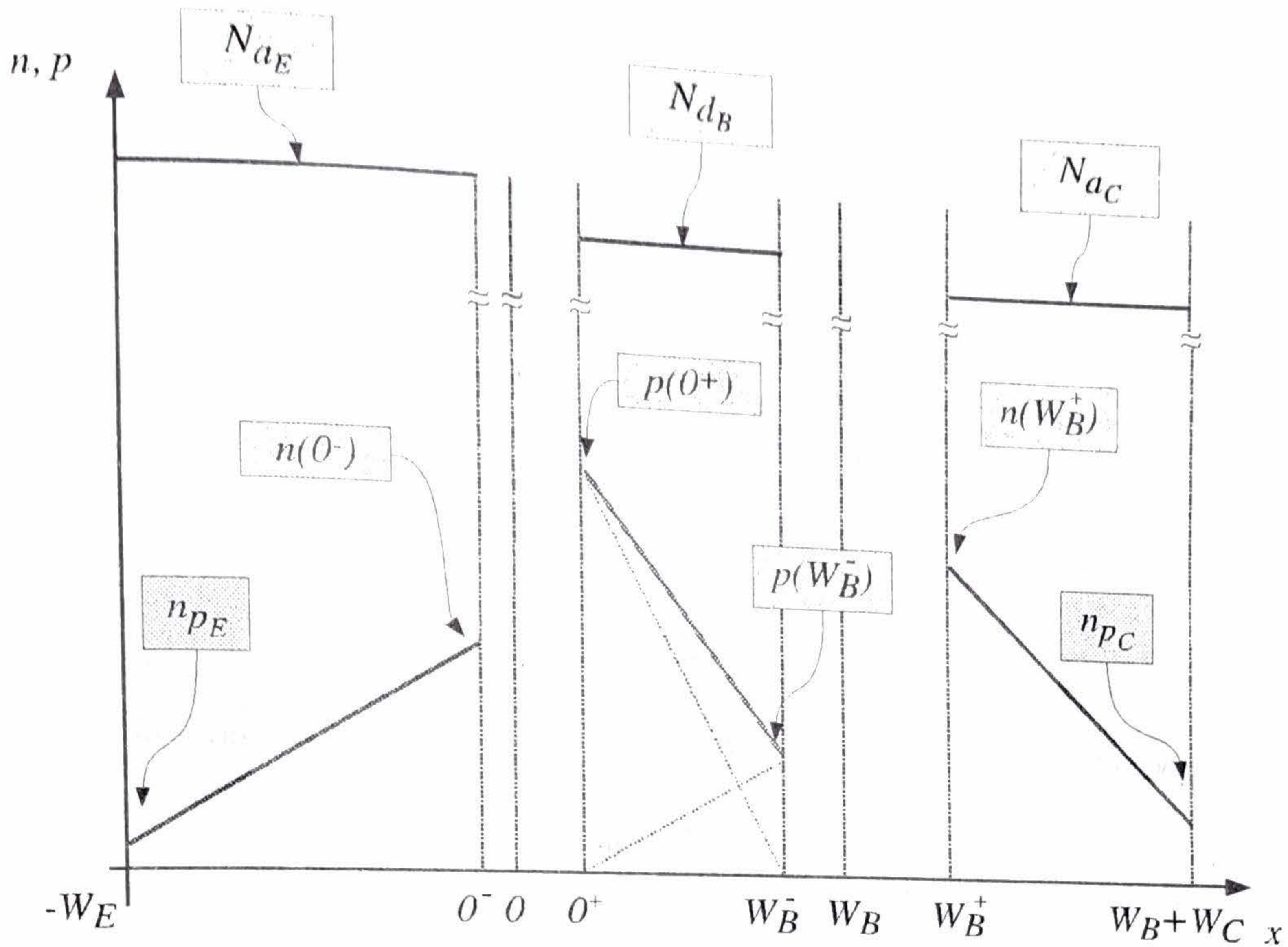


Figura 6.10: Niveles de portadores en un transistor PNP en el modo de saturación.

$$\begin{aligned}
 i_E &= qA_E \left(\frac{D_p p_{NB} e^{v_{EB}/U_T}}{W_B} + \frac{D_n n_{PE} (e^{v_{EB}/U_T} - 1)}{W_E} - \frac{D_p p_{NB} e^{v_{CB}/U_T}}{W_B} \right) \\
 &= qA_E \left(\frac{D_p p_{NB} (e^{v_{EB}/U_T} - 1)}{W_B} + \frac{D_n n_{PE} (e^{v_{EB}/U_T} - 1)}{W_E} - \frac{D_p p_{NB} (e^{v_{CB}/U_T} - 1)}{W_B} \right) \\
 &= qA_E \left(\left(\frac{D_p p_{NB}}{W_B} + \frac{D_n n_{PE}}{W_E} \right) (e^{v_{EB}/U_T} - 1) - \frac{D_p p_{NB}}{W_B} (e^{v_{CB}/U_T} - 1) \right) \quad (6.62)
 \end{aligned}$$

$$\begin{aligned}
 i_C &= qA_C \left(-\frac{D_p p_{NB} e^{v_{EB}/U_T}}{W_B} + \frac{D_p p_{NB} e^{v_{CB}/U_T}}{W_B} + \frac{D_n n_{PC} (e^{v_{CB}/U_T} - 1)}{W_C} \right) \\
 &= qA_C \left(-\frac{D_p p_{NB} (e^{v_{EB}/U_T} - 1)}{W_B} + \frac{D_p p_{NB} (e^{v_{CB}/U_T} - 1)}{W_B} + \frac{D_n n_{PC} (e^{v_{CB}/U_T} - 1)}{W_C} \right) \\
 &= qA_C \left(\left(\frac{D_p p_{NB}}{W_B} + \frac{D_n n_{PC}}{W_C} \right) (e^{v_{CB}/U_T} - 1) - \frac{D_p p_{NB}}{W_B} (e^{v_{EB}/U_T} - 1) \right) \quad (6.63)
 \end{aligned}$$

donde, en ambos casos, se ha sumado y restado el factor $D_p p_{NB}/W_B$ para completar los términos exponenciales con el factor -1 .

Si se considera $A_E = A_C = A$, luego de definir:

$$I_{ES} \triangleq qA \left(\frac{D_p p_{NB}}{W_B} + \frac{D_n n_{PE}}{W_E} \right) \quad (6.64)$$

$$I_{CS} \triangleq qA \left(\frac{D_p p_{NB}}{W_B} + \frac{D_n n_{PC}}{W_C} \right) \quad (6.65)$$

y sacando factor común, las corrientes de emisor y colector pueden re-escribirse como:

$$i_E = I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) - \alpha_R I_{CS} \left(e^{v_{CB}/U_T} - 1 \right) \quad (6.66)$$

$$i_C = I_{CS} \left(e^{v_{CB}/U_T} - 1 \right) - \alpha_F I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \quad (6.67)$$

Las Ecs. (6.66)-(6.67) constituyen el "Modelo de Ebers-Moll" [18], que presenta la notable característica de ser válido, no sólo en la región de saturación, sino también en las regiones activa directa y activa inversa.

Nótese de (6.62) y (6.63), que

$$\alpha_R I_{CS} = \alpha_F I_{ES} = \frac{D_p P_{NB}}{W_B} \quad (6.68)$$

Esta relación se denomina de *reciprocidad* y determina que, en realidad, solo tres de los cuatro parámetros del modelo de Ebers-Moll son independientes.

Si se define una corriente directa i_F , que circula por una juntura entre emisor y base, y una corriente inversa i_R , que circula por una juntura entre colector y emisor, de la siguiente manera:

$$i_F \triangleq I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \quad (6.69)$$

$$i_R \triangleq I_{CS} \left(e^{v_{CB}/U_T} - 1 \right) \quad (6.70)$$

el modelo de Ebers-Moll puede expresarse como:

$$\begin{aligned} i_E &= i_F(v_{EB}) - \alpha_R i_R(v_{CB}) \\ i_C &= i_R(v_{CB}) - \alpha_F i_F(v_{EB}) \end{aligned} \quad (6.71)$$

El circuito correspondiente al modelo de Ebers-Moll se muestra en la Fig. 6.11. El mismo está compuesto por dos junturas con corrientes i_F e i_R y por dos FCCC de valores α_F y α_R .

La corriente de base puede obtenerse a partir de las expresiones de las corrientes de emisor y colector, como:

$$i_B = -(1 - \alpha_F) i_F(v_{EB}) - (1 - \alpha_R) i_R(v_{CB}) \quad (6.72)$$

Para analizar la saturación es conveniente situarse en el punto de funcionamiento en el cual la tensión de colector (negativa) se ha elevado lo suficiente como para que $v_{CB} = 0$. Este punto marca el comienzo de la zona de saturación. En este caso, se verifica:

$$\begin{aligned} i_E &= i_F = I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \\ i_C &= -\alpha_F i_F = -\alpha_F I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \\ i_B &= -(1 - \alpha_F) i_F \end{aligned} \quad (6.73)$$

donde es importante notar que la corriente de base es una pequeña fracción de la corriente de emisor y el valor de i_R es cero. Para el análisis, se puede despreciar la corriente de la fuente controlada $\alpha_R i_R$ (aún cuando sea $i_r \neq 0$) debido al pequeño valor de α_R . A partir de este punto, un incremento

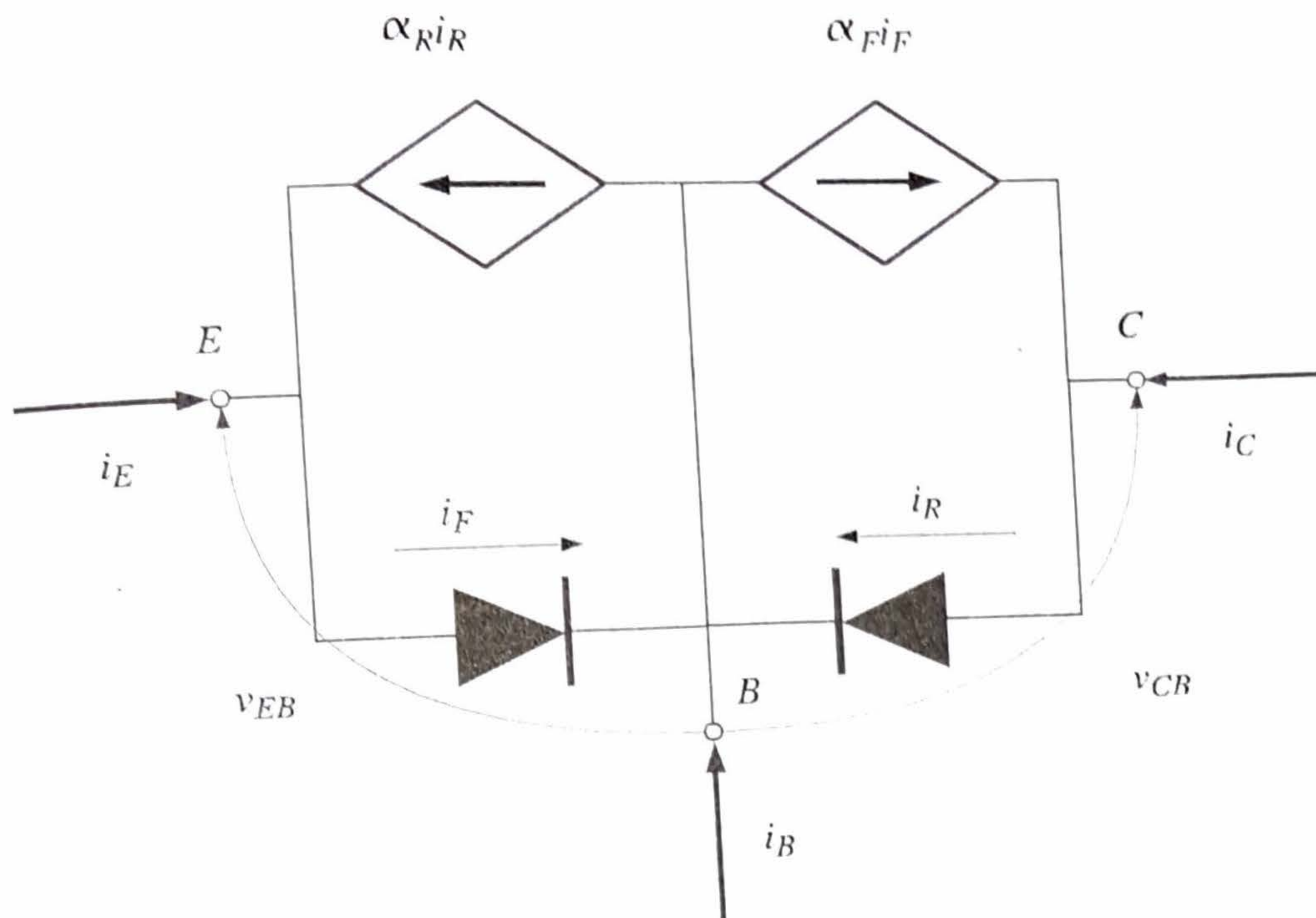


Figura 6.11: Circuito correspondiente al modelo de Ebers-Moll para un transistor PNP.

en v_C (o v_{CB}) produce un incremento de la componente de corriente Δi_C a través del diodo CB, con lo cual $\Delta i_C = \Delta i_R$. Si se supone que este incremento es pequeño en comparación a la magnitud de i_F , entonces i_E permanece prácticamente constante y la corriente de base i_B aumenta en módulo Δi_R debido al término $(1 - \alpha_R)i_R$ en (6.72). Si se continúa aumentando la tensión v_{CB} , eventualmente la componente de corriente de colector debida a la corriente directa del diodo CB, dada por i_R igualará a la inyección de huecos de la juntura EB dada por $\alpha_F i_F$. En este momento la corriente de colector es nula, así como la ganancia de corriente i_C/i_B , y la corriente de emisor completa su recorrido por la base, es decir:

$$\begin{aligned} i_E &= (1 - \alpha_F \alpha_R) I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \\ i_C &= 0 \\ i_B &= -i_E \end{aligned} \tag{6.74}$$

La zona de saturación comprende los valores de v_{CB} , tal que $0 \leq v_{CB} \leq V_{CB(sat)}$; es decir, desde que la corriente de colector comienza su reducción (situación descrita por la ecuaciones en (6.73)), hasta que se anula (situación descrita por la ecuaciones en (6.74)). En esta región, se verifica que:

$$\sigma \triangleq \frac{i_C}{\beta_F i_B} = \frac{i_R - \alpha_F i_F}{-(1 - \alpha_F) i_F - (1 - \alpha_R) i_R} \times \frac{1}{\beta_F} < 1 \tag{6.75}$$

donde σ se define como el índice de saturación.

Para hallar la tensión v_{CE} , en esta región, se puede proceder como sigue. En saturación, (6.66) y (6.67) pueden utilizarse para resolver las tensiones v_{CB} y v_{EB} . Despreciando los términos -1 , por estar ambas junturas en directa, y luego de algunas manipulaciones algebraicas, resulta:

$$v_{CB} = U_T \ln \frac{i_C + \alpha_F i_E}{(1 - \alpha_F \alpha_R) I_{CS}} \quad (6.76)$$

$$v_{EB} = U_T \ln \frac{i_E + \alpha_R i_C}{(1 - \alpha_F \alpha_R) I_{ES}} \quad (6.77)$$

$$(6.78)$$

Restando ambas ecuaciones para hallar $v_{CE} = v_{CB} - v_{EB}$, resulta:

$$V_{CE(sat)} = U_T \ln \frac{(i_C + \alpha_F i_E) I_{ES}}{(i_E + \alpha_R i_C) I_{CS}} \quad (6.79)$$

Reemplazando $I_{ES}/I_{CS} = \alpha_R/\alpha_F$, $i_E = -i_B - i_C$ y operando algebraicamente:

$$\begin{aligned} V_{CE(sat)} &= U_T \ln \frac{\alpha_R(1 - \alpha_F)(i_C - \beta_F i_B)}{\alpha_F((\alpha_R - 1)i_C - i_B)} \\ &= U_T \ln \frac{1 + (1 - \alpha_R)(i_C/i_B)}{\alpha_R(1 - i_C/(\beta_F i_B))} \\ &= U_T \ln \frac{1 + (1 - \alpha_R)\beta_F \sigma}{\alpha_R(1 - \sigma)} \end{aligned} \quad (6.80)$$

La Ec. (6.80) relaciona $V_{CE(sat)}$ con σ , para un dado α_R y para una dada temperatura.

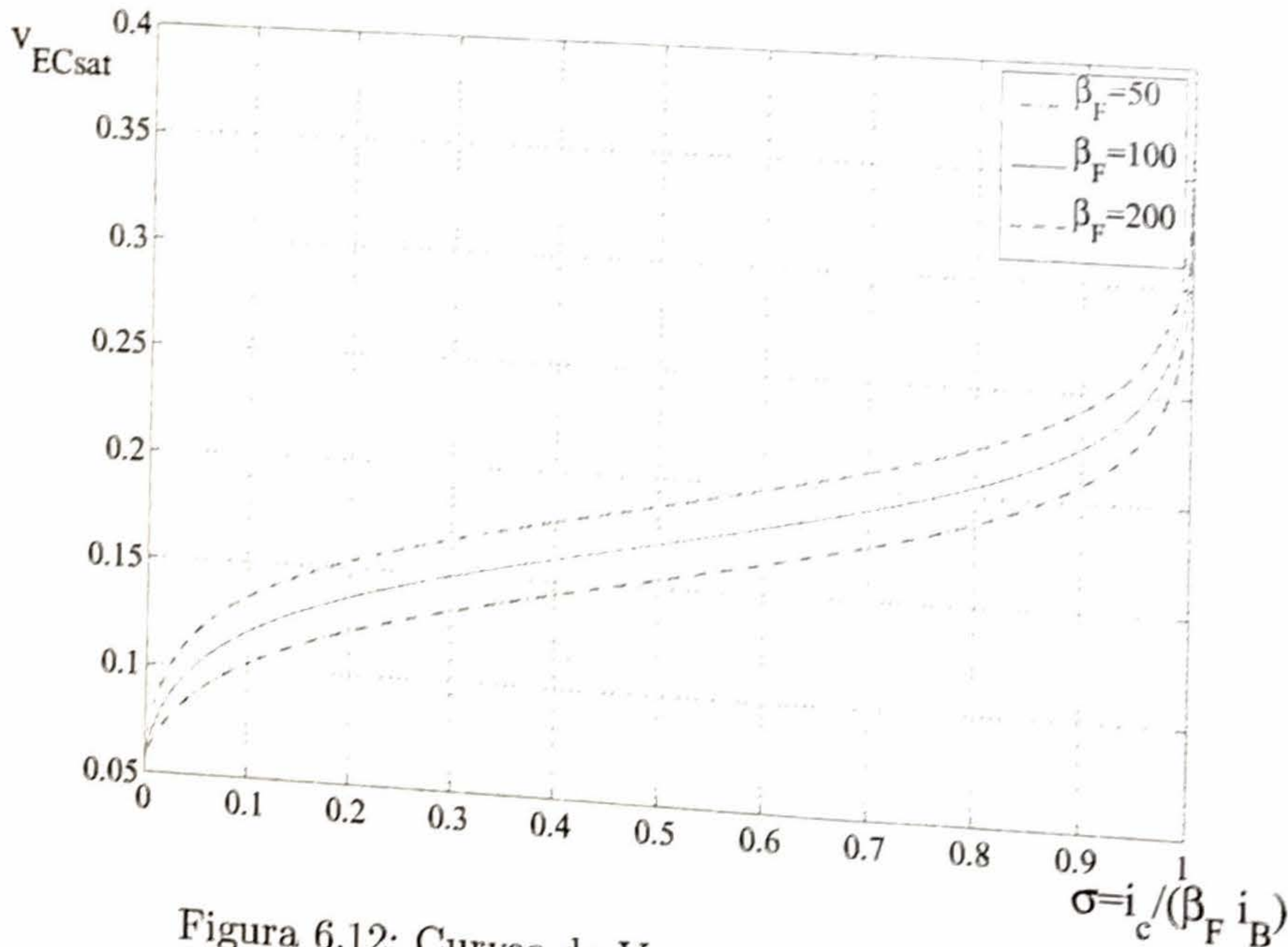


Figura 6.12: Curvas de $V_{EC(sat)}$ en función de σ .

Ejemplo 25 Si se supone $\alpha_R = 0,1$, el siguiente código de Matlab produce las curvas de $V_{CE(sat)}$ en función de σ , que se muestran en la Fig. 6.12:

```
UT=25e-3;
alfaR=0.1;
```


6.2. Modelo de DC del transistor PNP

```

sigma=[0:0.001:1];
betaF1=50;
VECsat1=UT * log((1+(1-alfaR)*betaF1*sigma)./((1-sigma)*alfaR));
betaF2=100;
VECsat2=UT * log((1+(1-alfaR)*betaF2*sigma)./((1-sigma)*alfaR));
betaF3=200;
VECsat3=UT * log((1+(1-alfaR)*betaF3*sigma)./((1-sigma)*alfaR));
plot(sigma,VECsat1,sigma,VECsat2,sigma,VECsat3)

```

La corriente de colector es proporcional a la corriente de base en la región de activa directa, cuando $\sigma = 1$. En saturación, a medida que el cociente de las corrientes se achica, lo cual sucede para $v_{CE} > -0,35V$, la tensión $V_{CE(sat)}$ se mantiene aproximadamente constante con un valor medio de $V_{CE(sat)} \approx -0,2V$. En esta situación, dado que ambas junturas están en conducción directa, $V_{EB(sat)} \approx 0,7V$ y $V_{CB(sat)} \approx 0,5V$.

La Fig. 6.13 muestra las curvas experimentales de salida de un transistor PNP en la región de saturación.

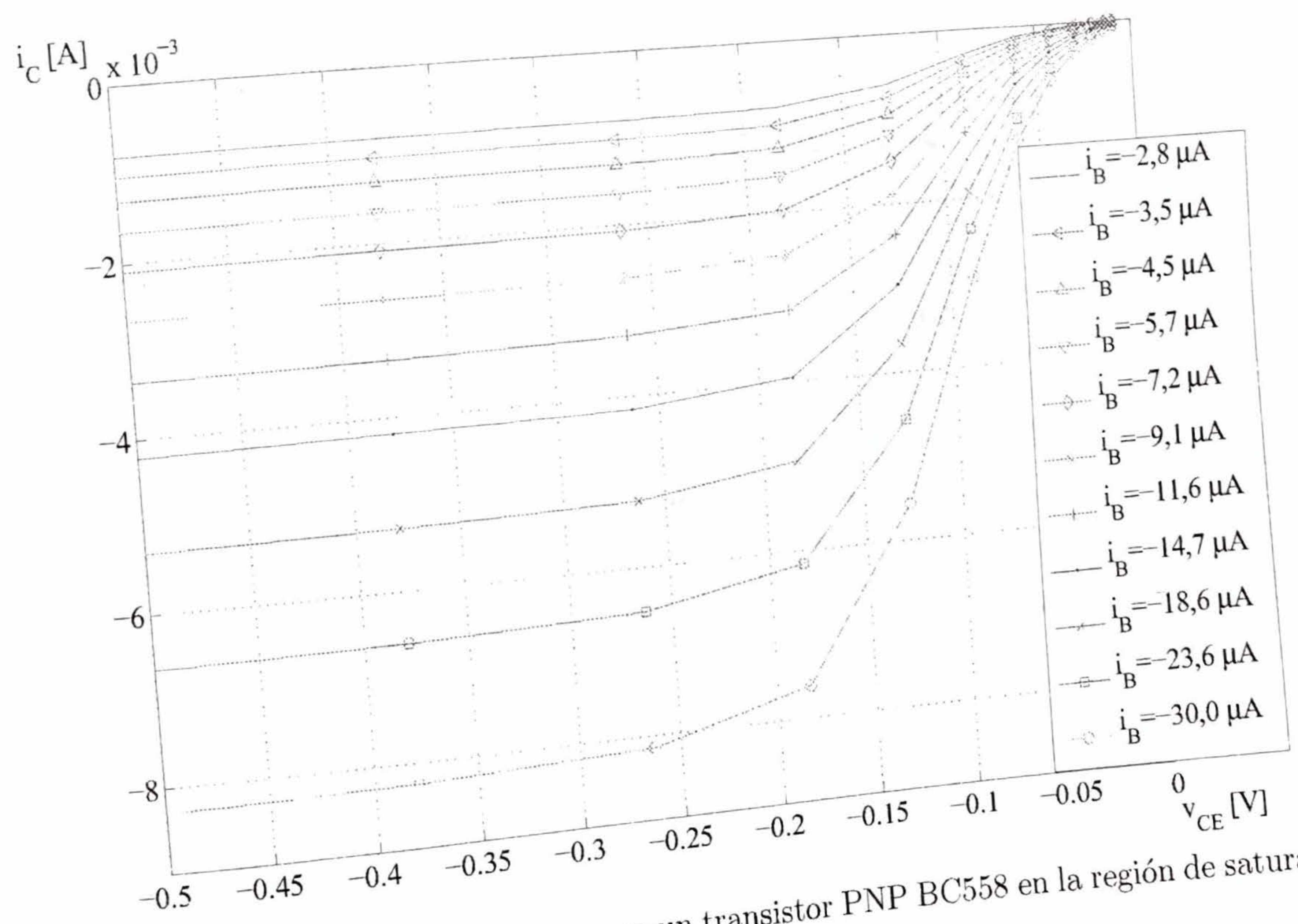


Figura 6.13: Curvas experimentales de salida para un transistor PNP BC558 en la región de saturación.

6.2.4. Modelos Simplificados

El modelo de Ebers-Moll puede utilizarse, y de hecho, es utilizado por simuladores, para resolver circuitos con transistores bipolares. Sin embargo, para cálculos manuales, es conveniente adaptarlo a expresiones más sencillas. En la zona activa directa, $v_{CB} \ll 0$, con lo cual las ecuaciones se reducen a:

$$\begin{aligned}
 i_E &= I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \\
 i_C &= -\alpha_F I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \approx -\alpha_F i_E \\
 i_B &= \frac{1 - \alpha_F}{\alpha_F} i_C = \frac{1}{\beta_F} i_C
 \end{aligned}
 \tag{6.81}$$

Estas ecuaciones corresponden a los circuitos de la Fig. 6.14, compuestos por un diodo entre el emisor y la base, y una FCCT entre colector y emisor.

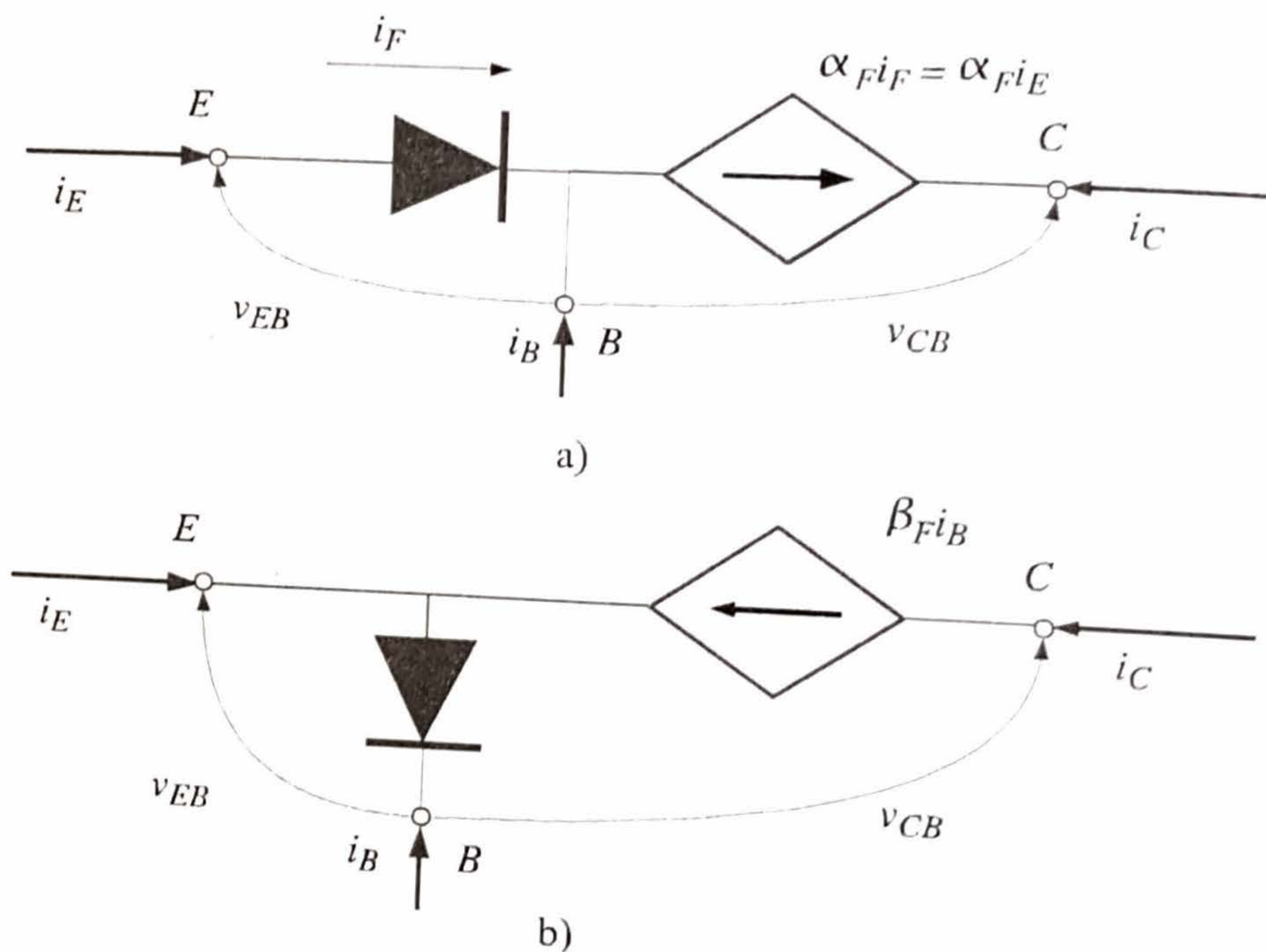


Figura 6.14: Circuito simplificado de Ebers-Moll en conducción directa activa para un transistor PNP: a) referido a la base; b) referido al emisor.

En la zona inversa activa, $v_{EB} \ll 0$, con lo cual el modelo resulta de la siguiente forma:

$$\begin{aligned}
 i_C &= I_{CS} \left(e^{v_{CB}/U_T} - 1 \right) \\
 i_E &= -\alpha_R I_{CS} \left(e^{v_{CB}/U_T} - 1 \right) \approx -\alpha_R i_C \\
 i_B &= \frac{1 - \alpha_R}{\alpha_R} i_E = \frac{1}{\beta_R} i_E
 \end{aligned}
 \tag{6.82}$$

Estas ecuaciones corresponden a un diodo entre el colector y la base, y una FCCT entre colector y emisor.

Ejemplo 26 Los circuitos de las Figs. 6.15-a y 6.15-d utilizan un transistor con $\alpha_F = 0,99$, es decir $\beta_F = 100$ y la fuente de alimentación tiene un valor $V_{CC} = 5V$. Cuanto debe valer la resistencia R_B para que la corriente de colector sea $I_C = -1mA$?

Comenzando por el circuito de la Fig. 6.15-a, si se reemplaza el transistor por el modelo equivalente de la Fig. 6.14-a, resulta el circuito de la Fig. 6.15-b. La corriente de base debe ser

$$I_B = \frac{1 - \alpha_F}{\alpha_F} I_C = \frac{-1mA}{100} = -10\mu A \quad (6.83)$$

Planteando la malla exterior resulta:

$$V_{CC} = 0,7V + (-I_B) \times R_B \quad (6.84)$$

Con lo cual R_B resulta:

$$R_B = \frac{-V_{CC} + 0,6V}{I_B} = \frac{4,4V}{10\mu A} = 440K\Omega \quad (6.85)$$

La Fig. 6.15-c ilustra las curvas de salida, el punto de funcionamiento, dado por $V_{CE} = -5V$ y $I_C = -1mA$, y la recta de carga (que es la ecuación que la fuente de alimentación le impone al transistor), dada por:

$$v_{CE} = -V_{CC} \quad (6.86)$$

La Fig. 6.15-d agrega una resistencia de colector $R_C = 2K$. Este caso se resuelve utilizando el modelo equivalente de la Fig. 6.14-b, para ilustrar la equivalencia de ambos modelos. El circuito resultante se ilustra en la Fig. 6.14-e. La corriente de base resulta sin cambios, dado que la malla externa no presenta modificaciones. La corriente de colector, suponiendo que el transistor está en conducción activa directa, es igual a:

$$I_C = \beta_F I_B = 100 \times -10\mu A = -1mA \quad (6.87)$$

Con esta corriente, la caída de tensión en la resistencia R_C es de $V_{R_C} = (-I_C) \times R_C = 2V$. La tensión V_{CE} puede calcularse mediante la malla de salida:

$$V_{CC} = -V_{CE} + (-I_C)R_C \quad (6.88)$$

de donde resulta:

$$V_{CE} = -V_{CC} - I_C R_C = -5V + 2V = -3V \quad (6.89)$$

La Fig. 6.15-f ilustra las curvas de salida, el punto de funcionamiento, y la recta de carga (que es la ecuación que la fuente de alimentación y la resistencia de colector R_C le imponen al transistor), dada por:

$$V_{CE} = -I_C R_C - V_{CC} \quad (6.90)$$

El hecho que resulte $V_{CB} = V_C - V_B = 2V - (5V - 0,6V) = -2,4V < 0$ confirma que la juntura CB está en inversa y el transistor está en la zona activa directa.

En la región de saturación, ambas junturas están en conducción directa, con lo cual pueden reemplazarse por dos fuentes de tensión de valores:

$$V_{EB(sat)} = 0,7V \quad (6.91)$$

$$V_{CB(sat)} = 0,5V \quad (6.92)$$

como se muestra en la Fig. 6.16-a. De manera equivalente (desde los terminales externos), la segunda fuente de tensión puede colocarse entre el colector y el emisor, de manera que el circuito resultante posea dos fuentes de tensión de valor:

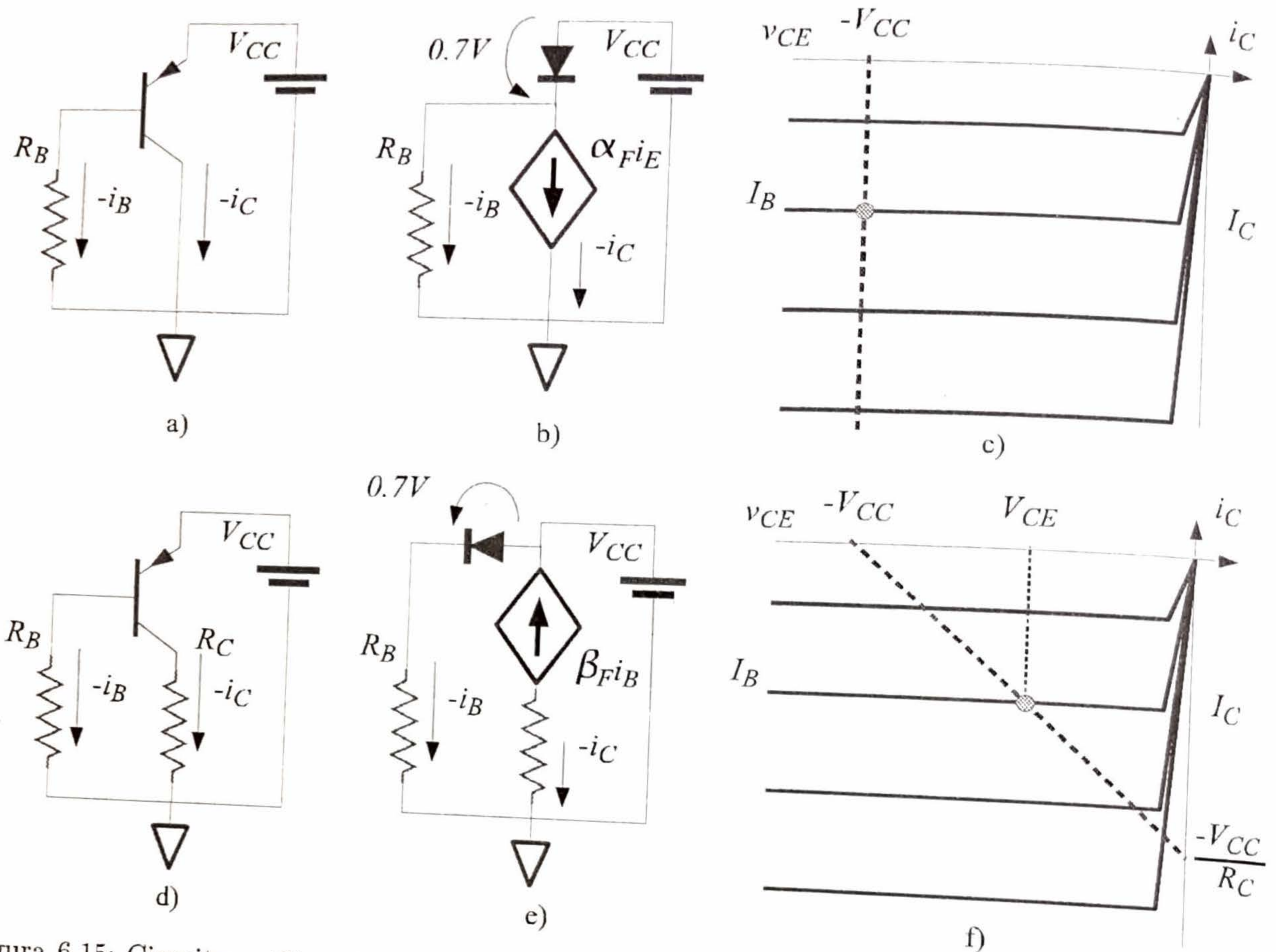


Figura 6.15: Circuitos utilizando transistores PNP: a) circuito sin resistencia de colector; b) circuito equivalente; c) curvas de salida, recta de carga y punto de funcionamiento; d) circuito con resistencia de colector; e) circuito equivalente; f) curvas de salida, recta de carga y punto de funcionamiento;

$$V_{EB(sat)} = 0,7V \tag{6.93}$$

$$V_{EC(sat)} = 0,2V \tag{6.94}$$

como se muestra en la Fig. 6.16-b.

Ejemplo 27 Si se considera el circuito de la Fig. 6.15-d, con $R_B = 440K$ y $R_C = 7K$, cuanto vale la corriente de colector ?

Con una resistencia de base $R_B = 440K$, de acuerdo a los cálculos realizados en el ejemplo anterior, la corriente de base es $-I_B = 10\mu A$. Si se supone que el transistor está en la región de conducción activa, y se utiliza el modelo de la Fig. 6.15-e, tal como se ilustra en la Fig. 6.17-a, la corriente de colector resulta $I_C = \beta_F I_B = -1mA$. La caída de tensión en la resistencia R_C resulta $V_{R_C} = -I_C \times R_C = 7V$. De acuerdo a la malla de salida:

$$V_{CC} = -V_{CE} + V_{R_C} \tag{6.95}$$

con lo cual resulta:

$$V_{CE} = -V_{CC} + V_{R_C} = -5V + 7V = +2V \tag{6.96}$$

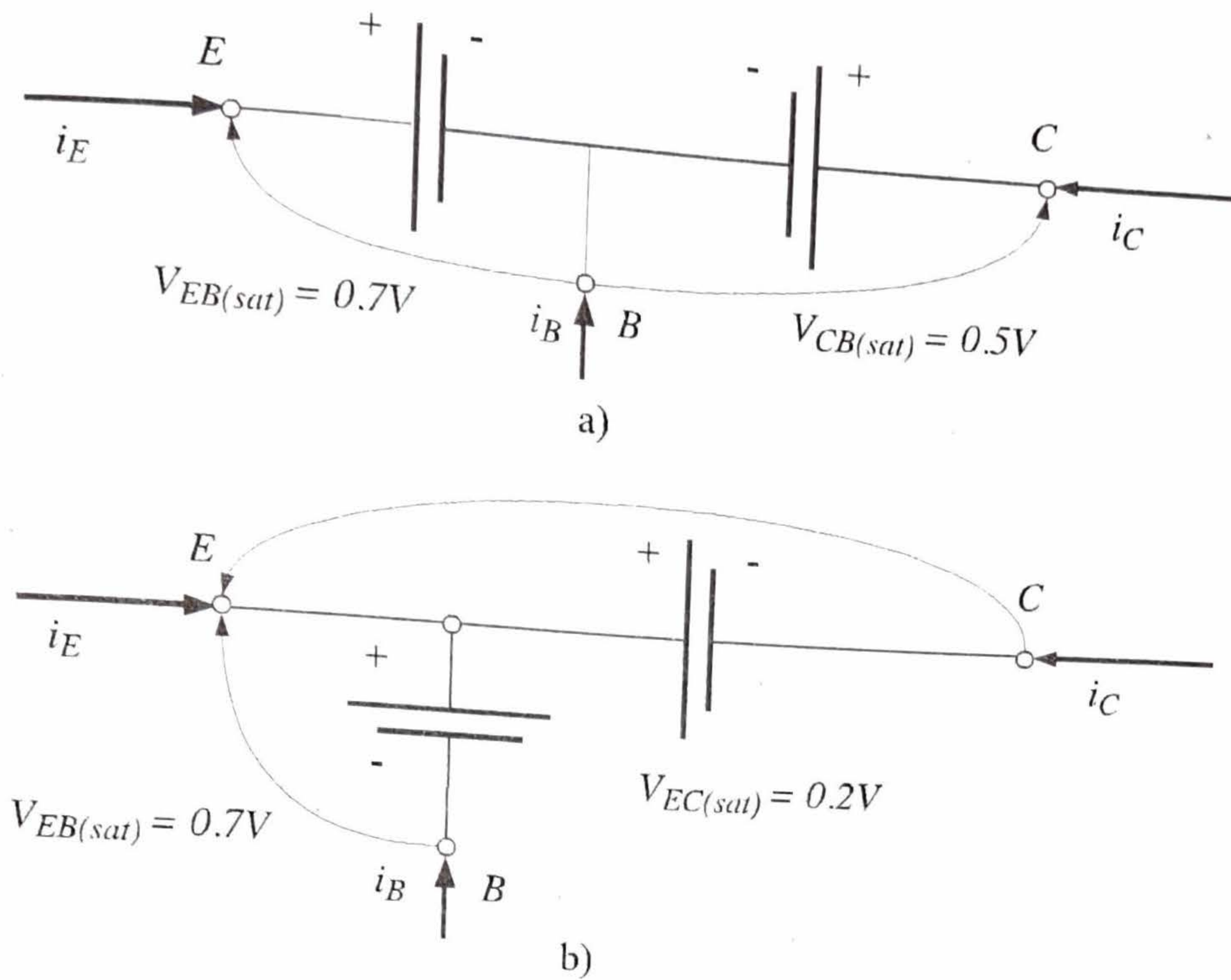


Figura 6.16: Circuito simplificado de Ebers-Moll en saturación para un transistor PNP: a) referido a la base; b) referido al emisor.

De acuerdo a esto, la tensión de la base es de $V_B = 5V - 0,7V = 4,3V$ y la tensión del colector es de $V_C = 5V + 2V = 7V$, con lo cual la juntura CB tiene una tensión

$$V_{CB} = 7V - 4,3V = 2,7V \quad (6.97)$$

y está en directa. Esto se opone a la suposición hecha, que el transistor estaba en la región de conducción activa directa. Si se grafica la recta de carga, como se muestra en la Fig. 6.17-b, el punto de funcionamiento con $V_{CE} > 0$ quedaría fuera de la zona de funcionamiento del transistor (lo cual es incorrecto).

Se debe suponer entonces que el transistor está en otra región de funcionamiento: la región de saturación. Para ello, se utiliza el circuito equivalente de la Fig. 6.16-b, como se ilustra en la Fig. 6.17-c. En este caso, la corriente de base resulta sin cambios. La corriente de colector resulta de evaluar la malla de salida:

$$V_{CC} = -V_{CE(sat)} + V_{RC} = 0,2V + (-I_C)R_C \quad (6.98)$$

De aquí surge el valor de corriente de colector:

$$I_C = -\frac{5V - 0,2V}{7K} = -0,68mA \quad (6.99)$$

El índice de saturación es $\sigma = 0,68mA/1mA = 0,68$.

La Fig. 6.17-d muestra la recta de carga y el punto de funcionamiento en la región de saturación.

Un paso más en el análisis permite hallar el valor de resistencia de colector que coloca al transistor en el límite de las regiones de conducción activa directa y saturación. En este punto se deben verificar dos condiciones: para que el transistor esté en saturación debe suceder $v_{CE} = V_{CE(sat)}$; y para que esté en la región de conducción directa debe suceder $i_C = \beta_F i_B$. Con estas dos condiciones se puede plantear la malla de salida como:

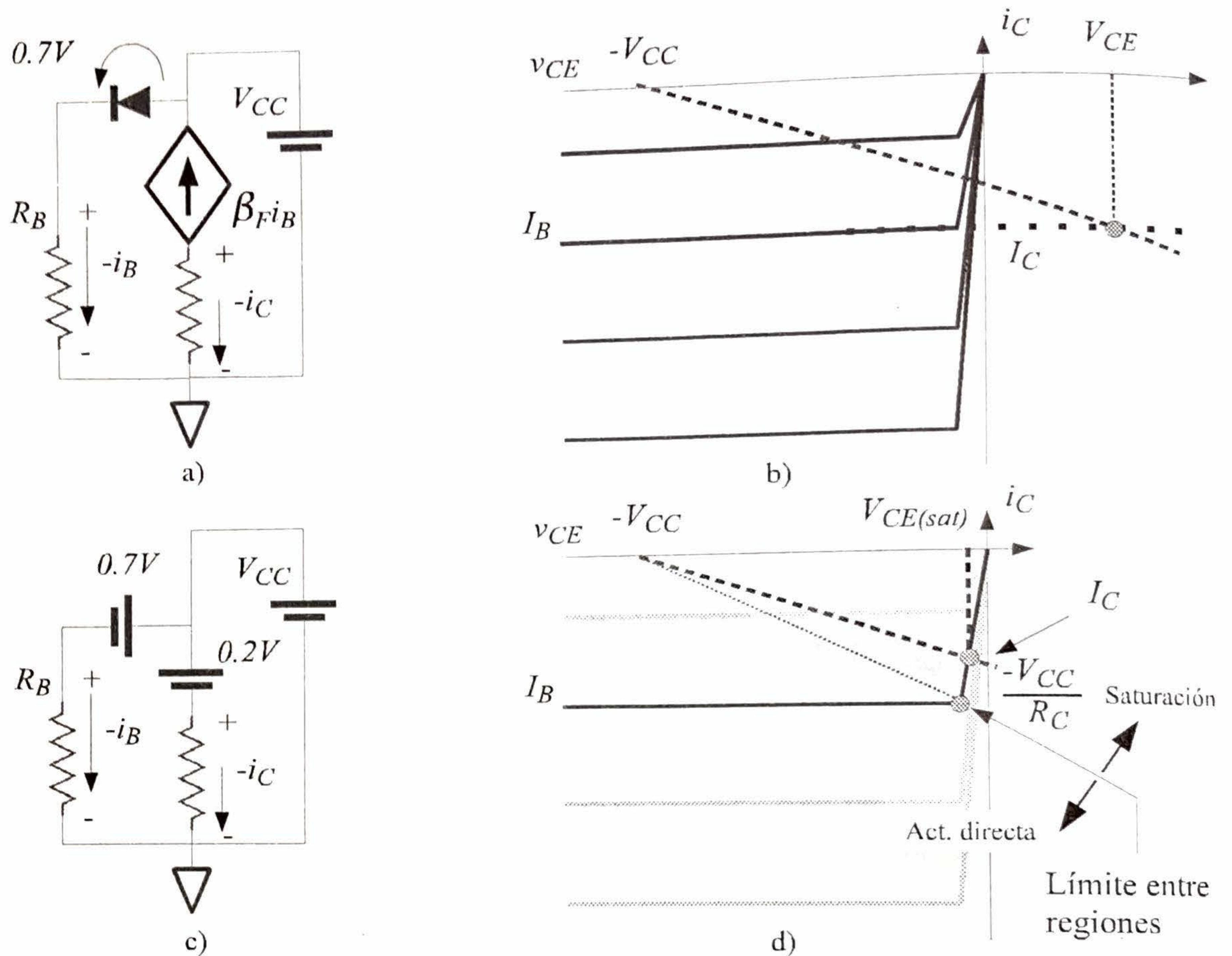


Figura 6.17: Circuito con un transistor PNP en saturación: a) circuito suponiendo el transistor en conducción directa; b) recta de carga correspondiente (incorrecta); c) circuito suponiendo el transistor en saturación; d) recta de carga correspondiente, y recta de carga para el caso límite entre saturación y conducción directa.

$$V_{CC} = -V_{CE(sat)} + \beta_F(-I_B)R_C = 0,2V + 1mA \times R_C \quad (6.100)$$

De esta ecuación surge

$$R_C = \frac{V_{CC} - 0,2V}{1mA} = \frac{5V - 0,2V}{1mA} = 4,8K \quad (6.101)$$

Este punto se indica en la Fig. 6.17-d, donde se puede ver la pertenencia simultánea a las regiones de saturación y conducción directa. Todo valor de resistencia superior a $R_C = 4,8K$ (manteniendo $I_B = -10\mu A$) hace que el circuito esté en saturación con valores cada vez más pequeños de corriente de colector.

6.2.5. Desviaciones del comportamiento ideal

Corriente de pérdida

Cuando el transistor bipolar funciona en la zona activa directa, en la corriente de colector se puede advertir el término i_R que corresponde a la corriente inversa, o de pérdida, de la juntura CB, que fuera despreciado en la Sección 6.2.1. Como se ve a continuación, esta corriente puede tener un efecto considerable con el aumento de la temperatura.

Se comienza el análisis a partir de la expresión de la corriente de base del modelo de Ebers-Moll:

$$i_B = -(1 - \alpha_F)i_F - (1 - \alpha_R)i_R \quad (6.102)$$

Dado que el efecto de la componente de corriente i_R sobre la corriente de emisor es despreciable, se puede simplificar:

$$i_E = i_F - \alpha_R i_R \approx i_F \quad (6.103)$$

Por otro lado, en conducción directa, $v_{CB} \ll 0$, con lo cual:

$$i_R = I_{CS} \left(e^{v_{CB}/U_T} - 1 \right) \approx -I_{CS} \quad (6.104)$$

Reemplazando en (6.102) se obtiene:

$$i_B = -(1 - \alpha_F)i_E + (1 - \alpha_R)I_{CS} \quad (6.105)$$

Ahora, se define la corriente de pérdida:

$$I_{co} \triangleq (1 - \alpha_R)I_{CS} \quad (6.106)$$

con lo cual, (6.105) se puede escribir como:

$$i_B = -(1 - \alpha_F)i_E + I_{co} = -\frac{1}{\beta_F + 1}i_E + I_{co} \quad (6.107)$$

La corriente de colector en conducción directa refleja una amplificación de esta corriente:

$$i_C = \beta_F i_B = -\frac{\beta_F}{\beta_F + 1}i_E + \beta_F I_{co} \quad (6.108)$$

Dada la fuerte dependencia de I_{CS} con la temperatura, para temperaturas elevadas, el término $\beta_F I_{co}$ tiene efectos notables. La corriente I_{CS} es la corriente en inversa de la juntura CB, compuesta por electrones que van de colector a base, y por huecos que van de base a colector, tal como se ilustra en la Fig. 6.18. La corriente I_{co} es la componente de I_{CS} que va hacia la base, es decir, es la corriente inversa de electrones de la juntura CB.

Es interesante notar, a partir de (6.107), que en el caso en que $i_B = 0$, es decir, cuando la base está desconectada, la corriente que circula por el dispositivo es $i_E = (\beta_F + 1)I_{co}$. Por otro lado, si se desconecta el emisor, haciendo $i_E = 0$, la corriente resultante, de acuerdo a (6.107), es $i_B = I_{co}$. En un diodo real, se mide una corriente I_{CBO} que resulta mucho mayor que I_{co} , por varias razones, entre ellas: pérdidas adicionales por defectos en la unión metalúrgica, generación térmica adicional en la zona de vaciamiento, y multiplicación por efecto avalancha. Estos tres términos son dependientes de la tensión inversa, y en menor medida de la temperatura. I_{CBO} se duplica aproximadamente cada 10°C para diodos de Silicio a temperatura ambiente.

Ganancia de corriente

En un dispositivo ideal, los portadores del emisor sufren recombinaciones al entrar a la base. En un transistor PNP, el emisor inyecta huecos, los cuales son portadores minoritarios en la base, que por ser tipo N, tiene portadores mayoritarios negativos. Debido a la recombinación de portadores, la corriente de colector se reduce en un factor α_T , denominado "factor de transporte de la base":

$$\alpha_T = \frac{\text{Huecos que alcanzan el colector}}{\text{Huecos inyectados en la base}} \quad (6.109)$$

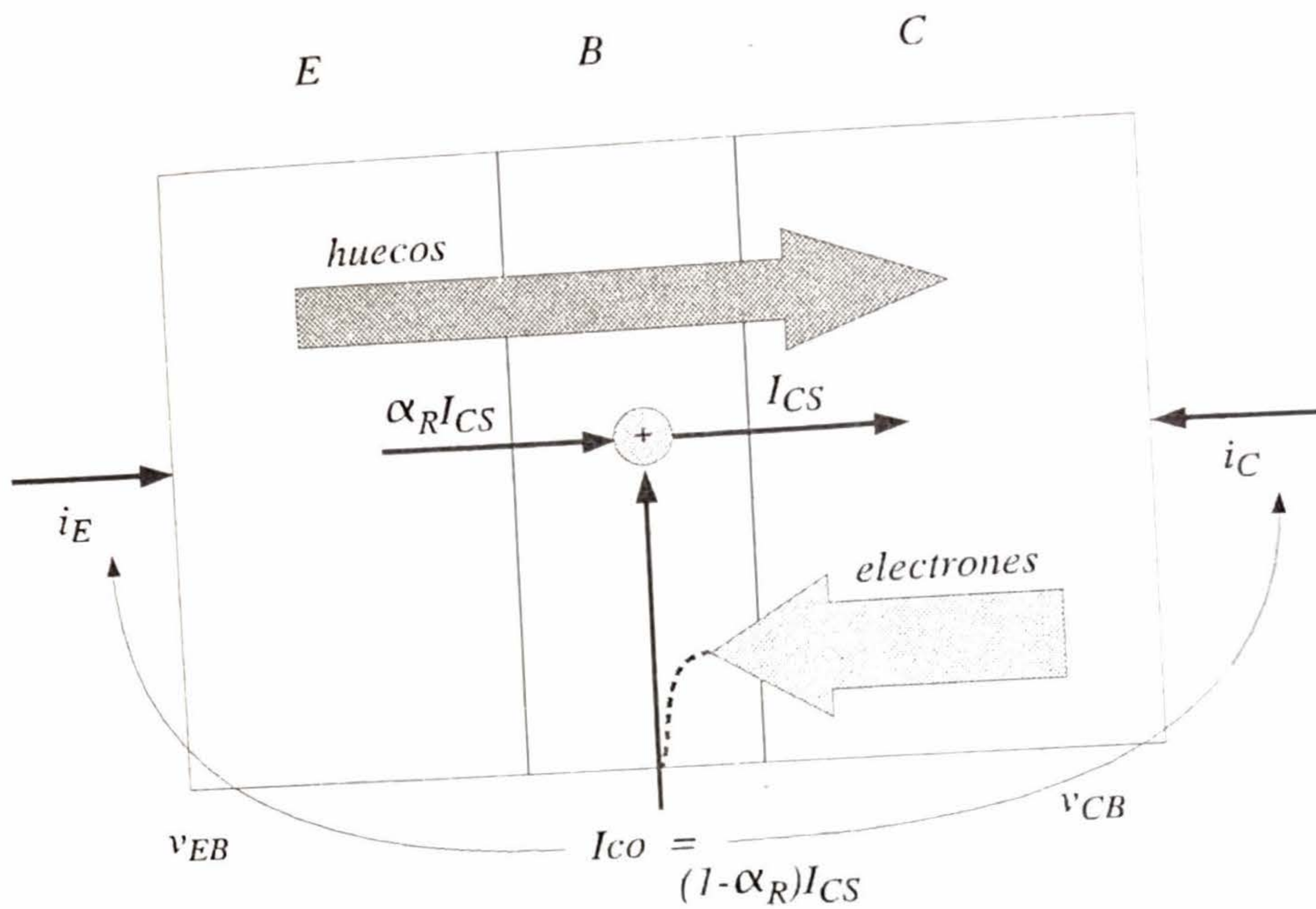


Figura 6.18: Corriente de pérdida I_{co} en un transistor bipolar PNP. Las flechas delgadas indican corrientes, y las gruesas, flujo de portadores.

El factor de transporte de la base puede llevarse prácticamente a uno (99% - 99.8%) mediante varias estrategias: reduciendo la longitud de la base; diseñando el dispositivo con una geometría que haga que los huecos puedan alcanzar el colector de manera directa; y, dopando menos la base, de manera que haya menos electrones para recombinación.

Si ahora se define la eficiencia de emisor, γ , como:

$$\gamma = \frac{\text{Corriente de huecos inyectada en la base}}{\text{Corriente total de emisor}} \tag{6.110}$$

la ganancia de corriente entre emisor y colector⁵, resulta:

$$\alpha_F \triangleq \alpha_T \times \gamma \tag{6.111}$$

Mediante el dopado mayor del emisor, el término γ se hace usualmente cercano a 99,6%. Con estos valores, es común encontrar valores de α_F alrededor de 99,5%. Dada la cercanía de α_F a 1, el parámetro β_F , cuyo denominador es $1 - \alpha_F$ presenta gran dispersión en los dispositivos comerciales, pudiendo variar entre 100 y 300.

La relación entre corriente de colector y corriente de base, es además, función de la corriente de colector empleada. Para bajas corrientes, la corriente de base está dominada por las corrientes de recombinación y resulta $I_C \sim \exp(v_{EB}/2U_T)$. Luego, sigue una región de comportamiento ideal, donde las corrientes se relacionan mediante el β_F ideal. A continuación, hay una región de inyección moderada donde la caída de tensión en la resistencia de base es apreciable; y por último, hay una región de alta inyección donde $I_C \sim \exp(v_{EB}/2U_T)$. La característica resultante para β_F se muestra en la Fig. 6.19.

⁵La ganancia de corriente también está afectada por el término M , introducido en la Sección 3.6.1, Ec. (3.124), que se denomina el factor de multiplicación del colector. Este factor se define como la relación entre la corriente de huecos inyectada en el colector y la corriente total de colector, y para tensiones menores que las tensiones de avalancha es $M = 1$.

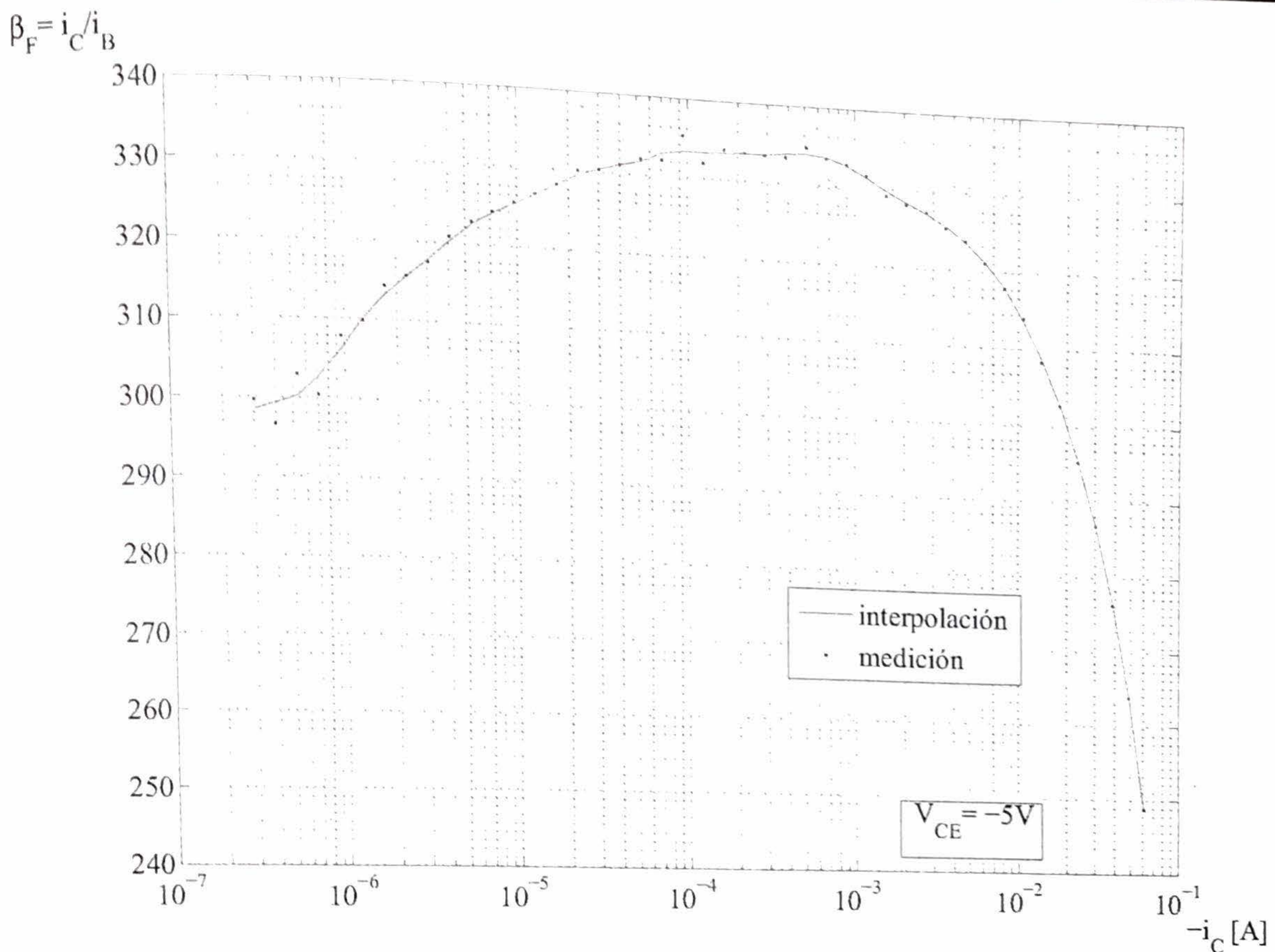


Figura 6.19: Relación β_F , entre la corriente de colector y la corriente de base, en función de la corriente de colector, $-i_C$, para un transistor bipolar PNP BC558.

6.3. Modelo de DC del transistor NPN

En esta sección se realizan, de manera sintética, los desarrollos para el caso de un transistor bipolar NPN. Constructivamente, un transistor bipolar NPN se logra haciendo una implantación profunda tipo N sobre un sustrato de Silicio, para obtener un colector de buena conductividad. A continuación, mediante un proceso epitaxial se hace crecer Silicio, y sobre éste se realiza el implante de base tipo P, y el implante de emisor tipo N fuertemente dopado. La situación se ilustra en la Fig. 6.20.

6.3.1. Región de conducción activa directa

Para este análisis, se utilizará la estructura que se muestra en la Fig. 6.21, donde la longitud del emisor es W_E , la de la base W_B , y la del colector W_C . En el caso de conducción activa directa (forward active), el potencial del emisor es el menor de los tres. El potencial de la base es mayor que el del emisor, de tal manera que la juntura BE esté polarizada en conducción directa. A su vez, el potencial del colector es mayor que el de la base, de tal manera que la juntura BC esté en conducción inversa. Se verifica entonces:

$$v_{BE} > 0 \quad v_{BC} < 0 \quad (6.112)$$

De igual manera que con el análisis de la juntura, se utilizará la aproximación de juntura corta, es decir, se supondrá que W_E , W_B , y W_C , son de una longitud lo suficientemente pequeña como para que los portadores se trasladen sin sufrir recombinación. Asimismo, para simplificar la notación, se despreciarán las longitudes de la zona de vaciamiento. Del análisis de la juntura se puede ver que debido a la polarización en directa de la juntura BE, los niveles de portadores a ambos lados de

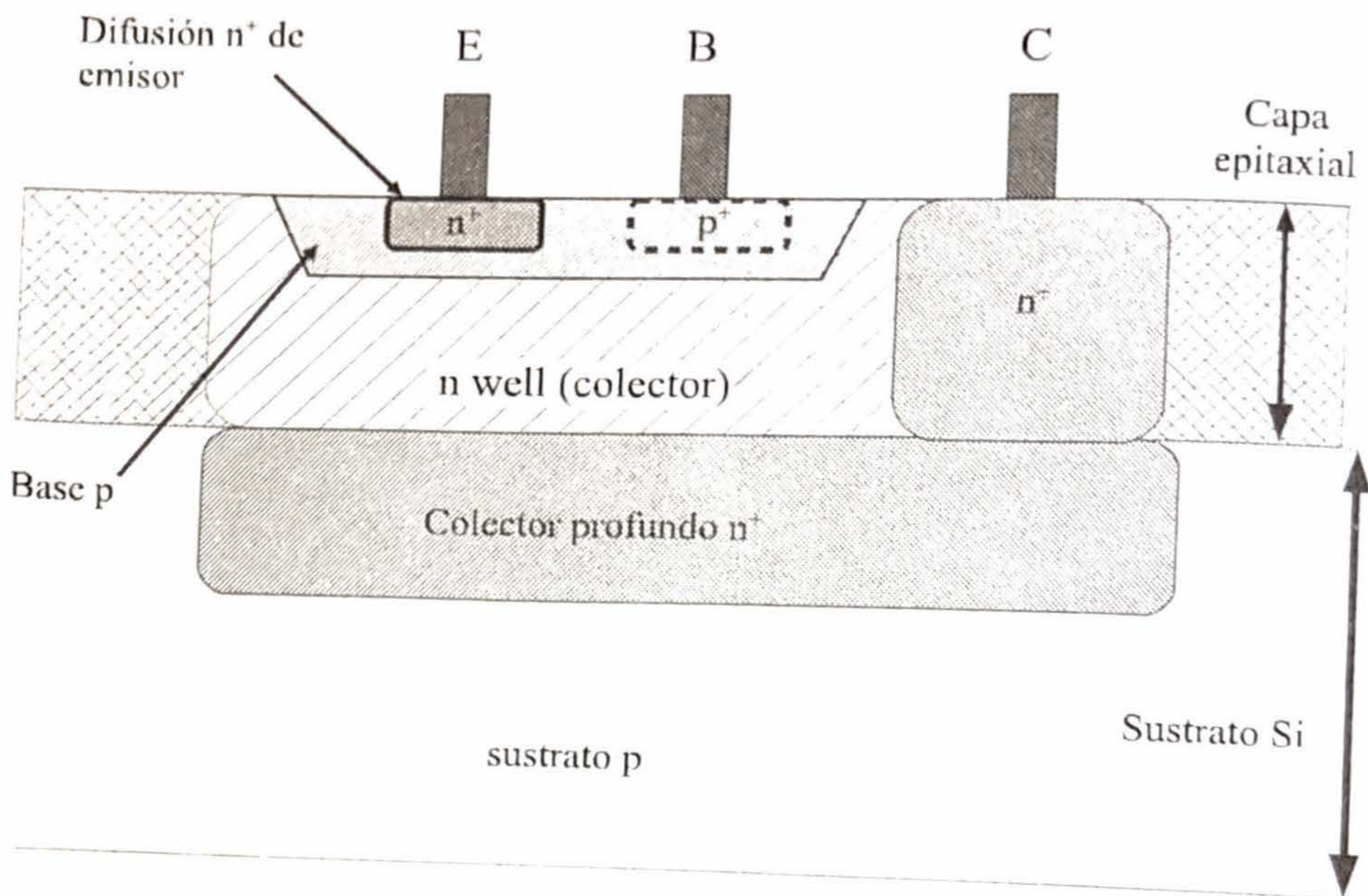


Figura 6.20: Esquema de un transistor NPN integrado. La implantación de colector profundo requiere un paso especial de fabricación.

esta zona de vaciamiento se incrementan en forma exponencial con la tensión v_{BE} . Por otro lado, la polarización en inversa de la juntura BC hace que los niveles portadores a ambos lados de la zona de vaciamiento de esta juntura se reduzcan a cero. Esta situación se ilustra en la Fig. 6.22.

La generación de corriente se origina en la juntura en conducción directa, BE, mientras que la juntura CB, que está en conducción inversa, se limita a transportar los portadores emitidos por la primera.

La corriente de difusión de electrones J_1 en el emisor se puede escribir como:

$$J_1 = qD_n \frac{n(W_B) - n(0)}{W_B} \tag{6.113}$$

donde:

$$n(W_B) = 0 \tag{6.114}$$

y

$$n(0) = N_{dE} e^{-\phi_{JBE}/U_T} = n_{PB} e^{v_{BE}/U_T} \tag{6.115}$$

dato que $\phi_{JBE} = \phi_{BBE} - v_{BE}$ y $N_{dE} e^{-\phi_{JBE}/U_T} = n_{PB}$. Utilizando estas dos ecuaciones en (6.113) resulta:

$$J_1 = -qD_n \frac{n_{PB} e^{v_{BE}/U_T}}{W_B} \tag{6.116}$$

La corriente de difusión de huecos J_2 está dada por:

$$J_2 = -qD_p \frac{p(0) - p(-W_E)}{W_E} \tag{6.117}$$

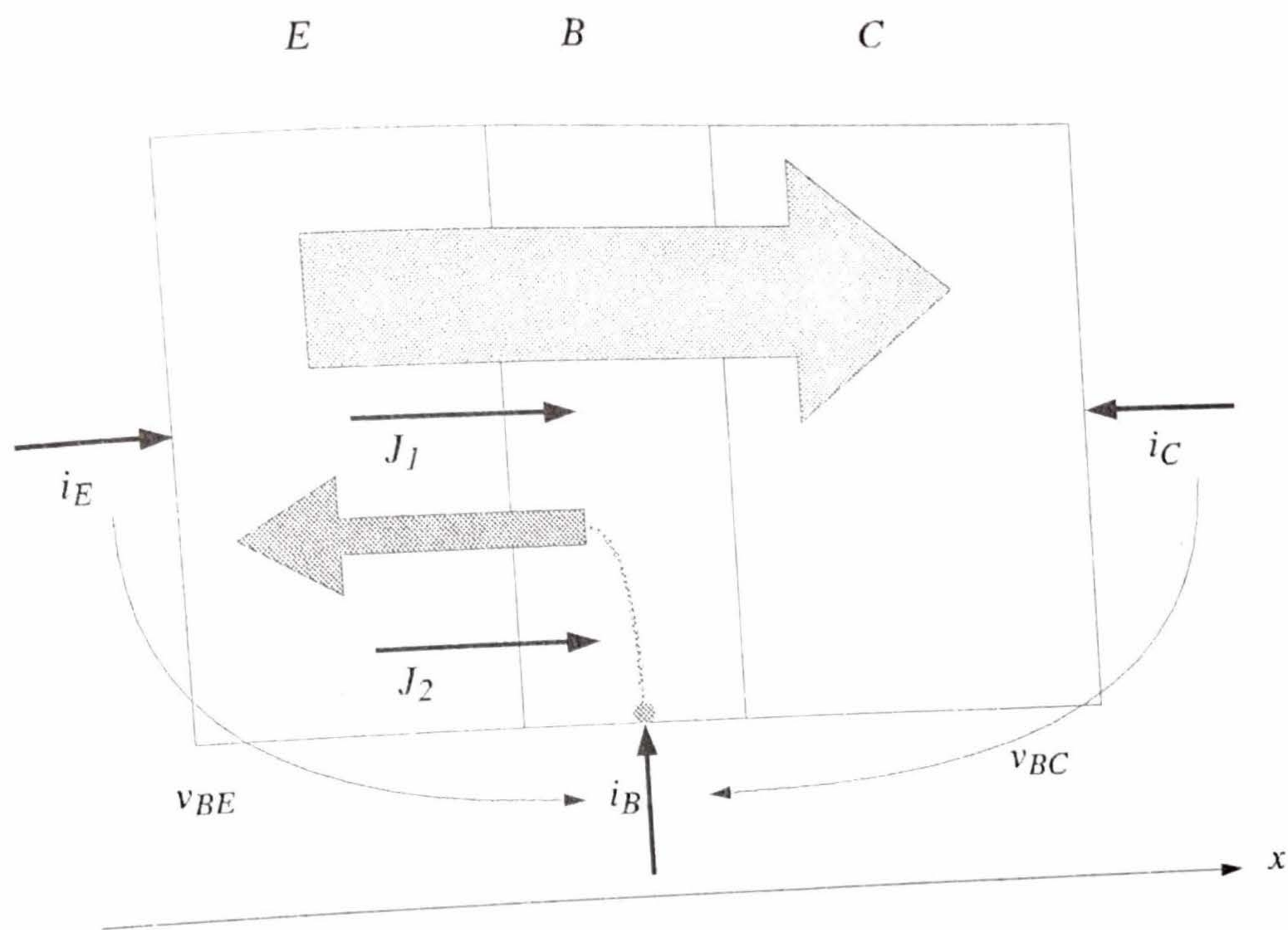


Figura 6.21: Esquema de un transistor bipolar NPN y las corrientes en el modo de conducción activa directa.

donde:

$$p(0) = N_{AB} e^{-\phi_{JBE}/U_T} = p_{NE} e^{v_{BE}/U_T} \quad (6.118)$$

y

$$p(-W_E) = p_{NE} \quad (6.119)$$

La corriente J_2 resulta de la siguiente manera:

$$J_2 = -qD_p \frac{p_{NE} e^{v_{BE}/U_T} - p_{NE}}{W_E} = -qD_p \frac{p_{NE} (e^{v_{BE}/U_T} - 1)}{W_E} \quad (6.120)$$

En base a estas dos corrientes, es posible resolver las corrientes de los terminales del transistor. De hecho, planteando nodos resulta:

$$\begin{cases} i_E = A(J_1 + J_2) \\ i_B = -AJ_2 \\ i_C = -AJ_1 \end{cases} \quad (6.121)$$

A partir de (6.121), se pueden calcular las expresiones de las corrientes individuales. La corriente de emisor es la suma de las contribuciones de los electrones, que van de emisor a colector, y de los huecos, que van de base a emisor, razón por la cual es la mayor de las tres en magnitud:

$$i_E = -AqD_n \frac{n_{PB} e^{v_{BE}/U_T}}{W_B} - AqD_p \frac{p_{NE} (e^{v_{BE}/U_T} - 1)}{W_E} \quad (6.122)$$

Considerando que, en conducción directa, el término -1 es muy pequeño con respecto al término exponencial, (6.122) puede escribirse como:

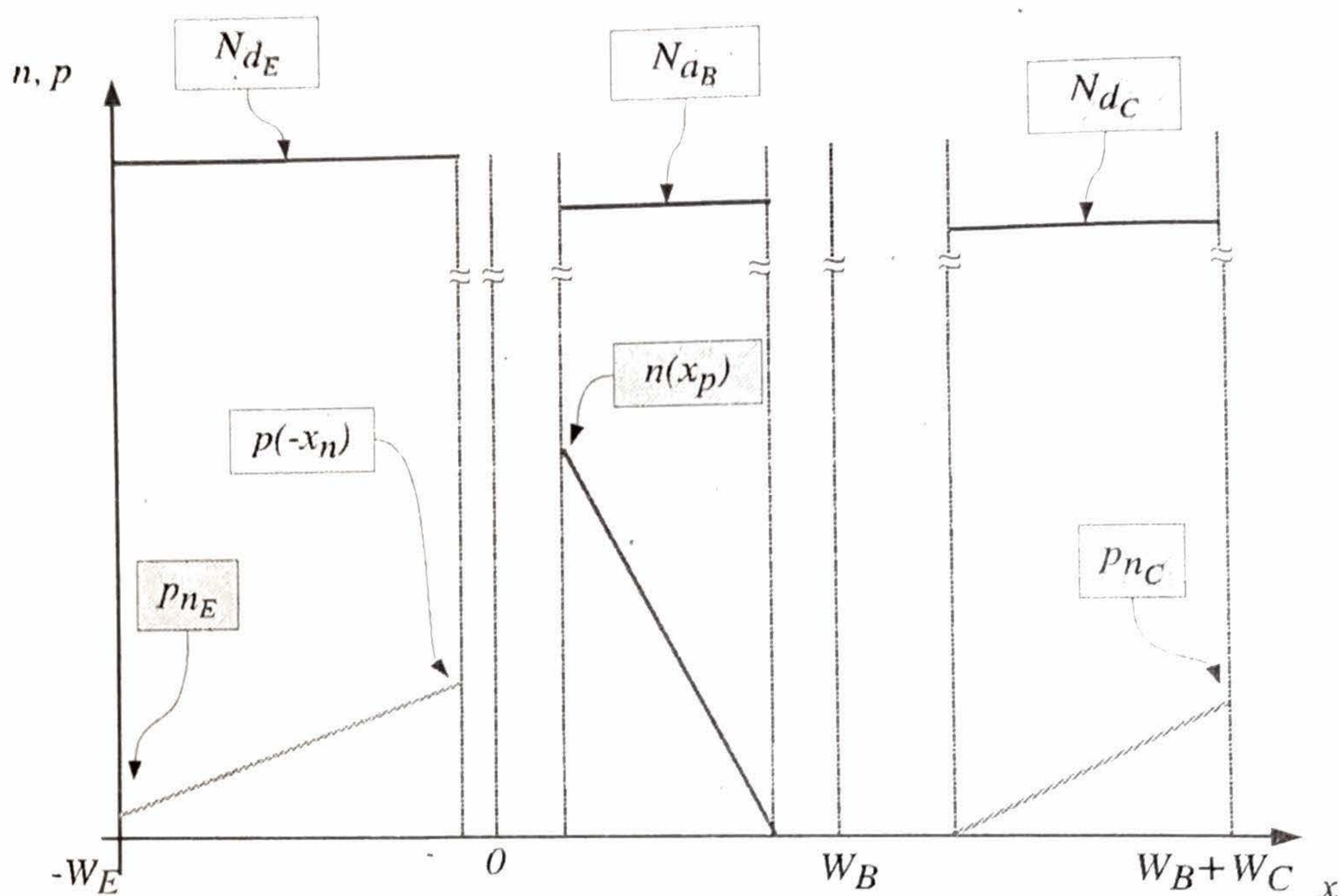


Figura 6.22: Niveles de portadores en un transistor NPN en el modo de conducción activo directo.

$$i_E = -A \left(\frac{qD_n n_{PB}}{W_B} + \frac{qD_p p_{NE}}{W_E} \right) e^{v_{BE}/U_T} \quad (6.123)$$

Por la convención elegida, en conducción directa, es decir con $v_{BE} > 0$, resulta $i_E < 0$.

La corriente de base está compuesta solamente por huecos, que circulan desde la base hacia el emisor, y retornan a la base, gracias a la fuente de alimentación. La expresión de esta corriente es:

$$i_B = A \frac{qD_p p_{NE}}{W_E} \left(e^{v_{BE}/U_T} - 1 \right) \quad (6.124)$$

Por la convención elegida, la corriente de base resulta positiva, $i_B > 0$.

Por último, la corriente de colector está compuesta solo por los electrones que salen del emisor, atraviesan la base, y llegan al colector. La diferencia entre la corriente de colector y la de emisor, es la corriente de huecos que circula por la base; por ello, la corriente de colector es ligeramente menor en magnitud que la corriente de emisor:

$$i_C = A \frac{qD_n n_{PB}}{W_B} e^{v_{BE}/U_T} \quad (6.125)$$

Esta corriente también resulta positiva con la convención elegida, $i_C > 0$.

El cociente α_F , entre la corriente de colector y la corriente de emisor, en activa directa, es:

$$\alpha_F = \frac{1}{1 + \frac{D_p N_{aB} W_B}{D_n N_{dE} W_E}} \quad (6.126)$$

Por ejemplo, si se consideran valores típicos, $N_{aE} = 10^{19} \text{ cm}^{-3}$, $N_{dB} = 2 \times 10^{17} \text{ cm}^{-3}$, y $W_B \approx W_E$, se obtiene $D_n \approx 8,5D_p$, y el valor de α_F es:

$$\alpha_F = \frac{1}{1 + \frac{1}{8,5 \times 50 \times 1}} = 0,9977 \quad (6.127)$$

El cociente β_F , entre la corriente de colector y la corriente de base, es:

$$\beta_F = \frac{D_n N_{dE} W_E}{D_p N_{aB} W_B} \quad (6.128)$$

Para los mismos valores típicos, el valor de β_F resulta:

$$\beta_F = 8,5 \times 50 = 425 \quad (6.129)$$

En este caso, para lograr un β_F grande, se debe dopar el emisor en mayor medida que la base:

$$N_{dE} \gg N_{aB} \quad (6.130)$$

De esta manera, el flujo de electrones del emisor a la base será mucho mayor que el flujo de huecos de la base al emisor, y la corriente de base será pequeña con respecto a i_E e i_C . También en este caso, conviene hacer la longitud de la base reducida ($W_B \ll W_E$), para que el perfil de portadores en la base sea más abrupto, provocando una mayor corriente de difusión. Por otro lado, al hacer la base corta, se evita que los electrones provenientes del emisor se recombinen con los huecos de la base.

En resumen, de (6.125), se deduce que en la zona de funcionamiento activa directa, la corriente de colector depende de la corriente de base, o de la tensión v_{BE} , y no depende de la tensión v_{BC} . En esta región de funcionamiento, el transistor puede considerarse como una fuente de corriente, controlada por la corriente de base i_B , o controlada por la tensión v_{BE} . La Ec. (6.125), representa la curva de transferencia, que relaciona la corriente de salida vs. la tensión de entrada, como se ilustra en la Fig. 6.23. Si en esta última ecuación, se reemplaza la corriente de colector por la de base, se obtiene la curva de entrada del transistor bipolar:

$$i_B = \frac{1}{\beta_F} \frac{AqD_n n_{PB}}{W_B} \left(e^{v_{BE}/U_T} - 1 \right) \quad (6.131)$$

la cual es igual a la curva de transferencia, salvo por el factor de escala β .

Comúnmente, se toma como tensión de salida del transistor a la tensión v_{CE} (en lugar de la tensión v_{BC}). Las curvas de salida ideales del dispositivo, i_C versus v_{CE} , son rectas horizontales, parametrizadas en función de la tensión v_{BE} , o de la corriente de base i_B . La Fig. 6.24 muestra las curvas experimentales de salida de un transistor NPN⁶. Para valores menores (hacia la izquierda) que $v_{CB} = 0V$, o $v_{CE} = 0,6V$, la juntura BC comienza a estar en zona directa, y el transistor pasa a la región de saturación.

Al hacerse la tensión de colector más positiva con respecto a la base, la zona de vaciamiento de la juntura BC aumenta, reduciendo la longitud efectiva de la base. Esto hace que la pendiente del perfil de portadores sea mayor, lo cual implica un aumento de la corriente. Esta variación de la corriente, en función de v_{BC} , se modela utilizando la tensión de Early, $V_A > 0$:

⁶En algunos casos, los fabricantes pueden especificar las curvas de corriente i_C versus v_{BC} . En estos casos, si se plantea la ecuación de mallas en el dispositivo, resulta:

$$v_{CE} = v_{BE} + v_{CB} \quad (6.132)$$

Nótese que los puntos de corriente para los cuales $v_{CB} = 0$ se trasladan a $v_{CE} = v_{BE}$. Dado que en activa directa, la juntura BE está en directa, su valor de tensión será aproximadamente $V_{BE} = 600mV$. Debido a esto, $v_{CE} \approx v_{CB} + 0,6V$ y la curva de salida en función de v_{CE} es una versión, desplazada hacia la derecha, de la curva versus v_{CB} .

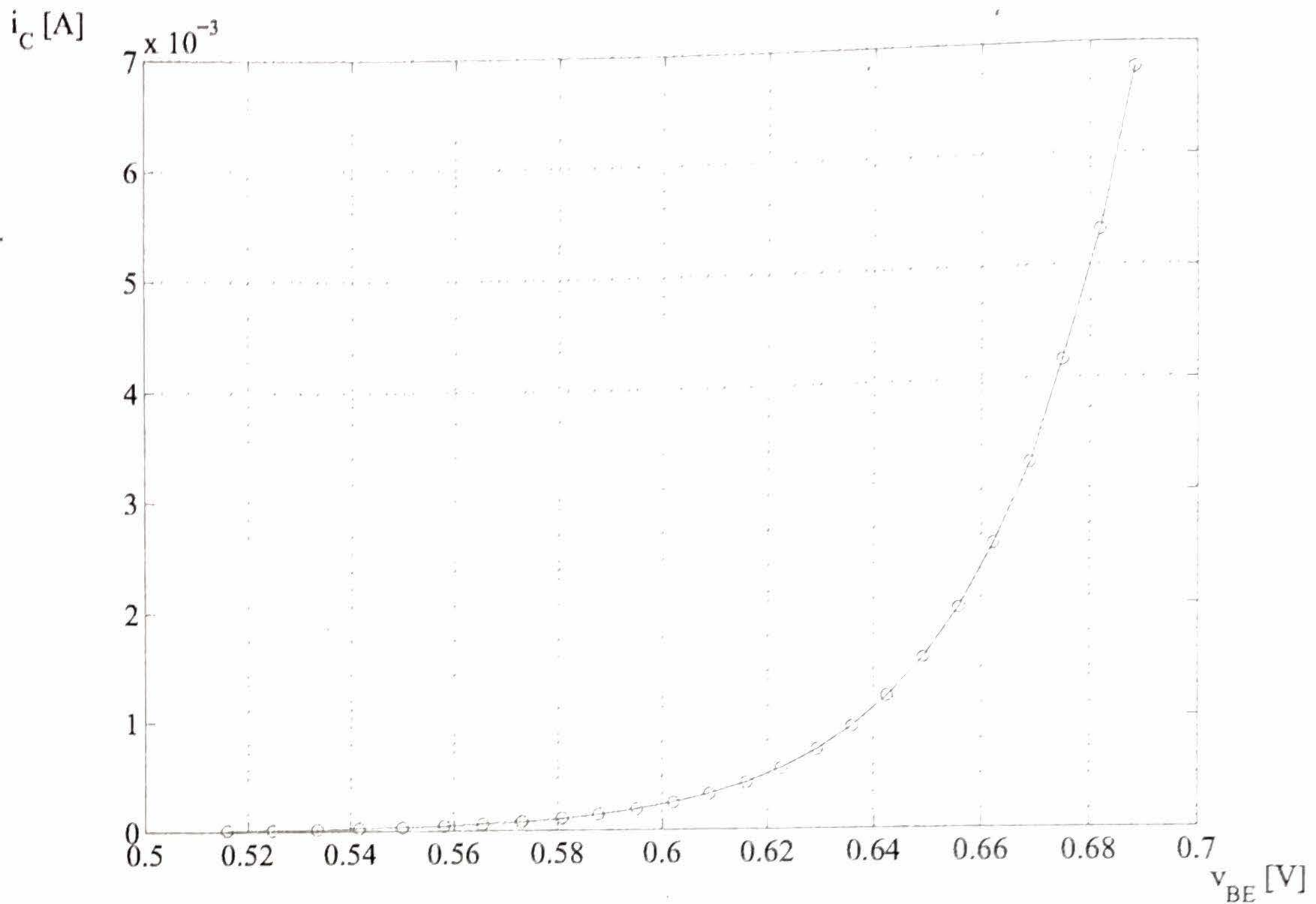


Figura 6.23: Curva experimental de transferencia de un transistor NPN BC548, para $v_{CE} = 380\text{mV}$.

$$i_C = \beta_F i_B \left(1 + \frac{v_{CB}}{V_A} \right) \quad (6.133)$$

o

$$i_C = A \frac{q D_n n_{PB}}{W_B} \left(e^{v_{BE}/U_T} - 1 \right) \left(1 + \frac{v_{CB}}{V_A} \right) \quad (6.134)$$

Similares expresiones son válidas, en caso de utilizar v_{CE} :

$$i_C = \beta_F i_B \left(1 + \frac{v_{CE} - V_{CE(sat)}}{V_A} \right) \quad (6.135)$$

o

$$i_C = A \frac{q D_n n_{PB}}{W_B} \left(e^{v_{BE}/U_T} - 1 \right) \left(1 + \frac{v_{CE} - V_{CE(sat)}}{V_A} \right) \quad (6.136)$$

donde $V_{CE(sat)}$ es el valor de v_{CE} a partir del cual el dispositivo está en activa directa.

El efecto de la tensión v_{BC} sobre la zona de vaciamiento de la juntura BC se puede ver con el auxilio de las Ecs. (3.27) y (3.28).

La zona de vaciamiento de la juntura BC, en la región de la base, está dada por:

$$x_{pB} = \sqrt{2q^{-1}\epsilon S i(\phi_B - v_{BC})} \sqrt{\frac{N_{dC}}{N_{aB}(N_{dC} + N_{aB})}} \quad (6.137)$$

mientras que en la región del colector:

$$x_{nC} = \sqrt{2q^{-1}\epsilon S i(\phi_B - v_{BC})} \sqrt{\frac{N_{aB}}{N_{dC}(N_{dC} + N_{aB})}} \quad (6.138)$$

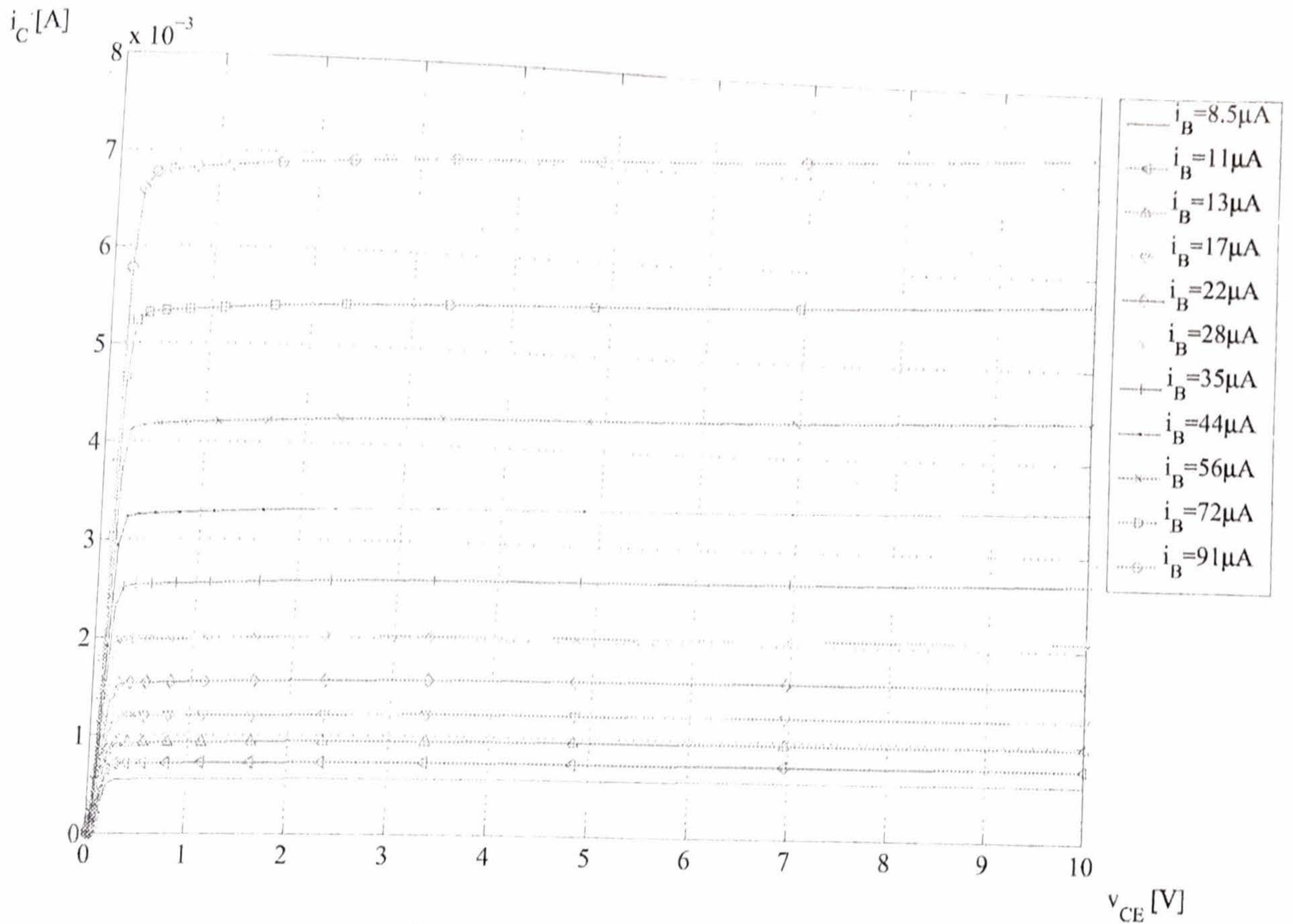


Figura 6.24: Curvas experimentales de salida de un transistor NPN BC548, para varios valores de corriente de base.

La relación entre ambas es:

$$\frac{x_{pB}}{x_{nC}} = \sqrt{\frac{N_{dC}^2}{N_{aB}^2}} = \frac{N_{dC}}{N_{aB}} \quad (6.139)$$

Para que la longitud de la zona de vaciamiento de la base, x_{pB} , sea mínima, es conveniente que se cumpla:

$$N_{dC} \ll N_{aB} \quad (6.140)$$

En consecuencia, para los transistores bipolares NPN, la relación de dopado será descendente a medida que se recorre el dispositivo, desde el emisor hacia el colector:

$$N_{dE} \gg N_{aB} \gg N_{dC} \quad (6.141)$$

6.3.2. Región de conducción activa inversa

En el caso del transistor NPN, la región activa inversa se caracteriza por tener la juntura BE en conducción inversa, y la juntura BC en conducción directa, es decir:

$$v_{BE} < 0 \quad v_{BC} > 0 \quad (6.142)$$

Para el desarrollo de este caso, se utilizará la estructura de la Fig. 6.25, donde se ve una conducción de electrones del colector hacia la base, que son arrastrados por el campo eléctrico de la zona de vaciamiento BE hacia el emisor. Por otro lado, hay también una conducción de huecos desde la base hacia el colector. El perfil resultante de portadores se ilustra en la Fig. 6.26.

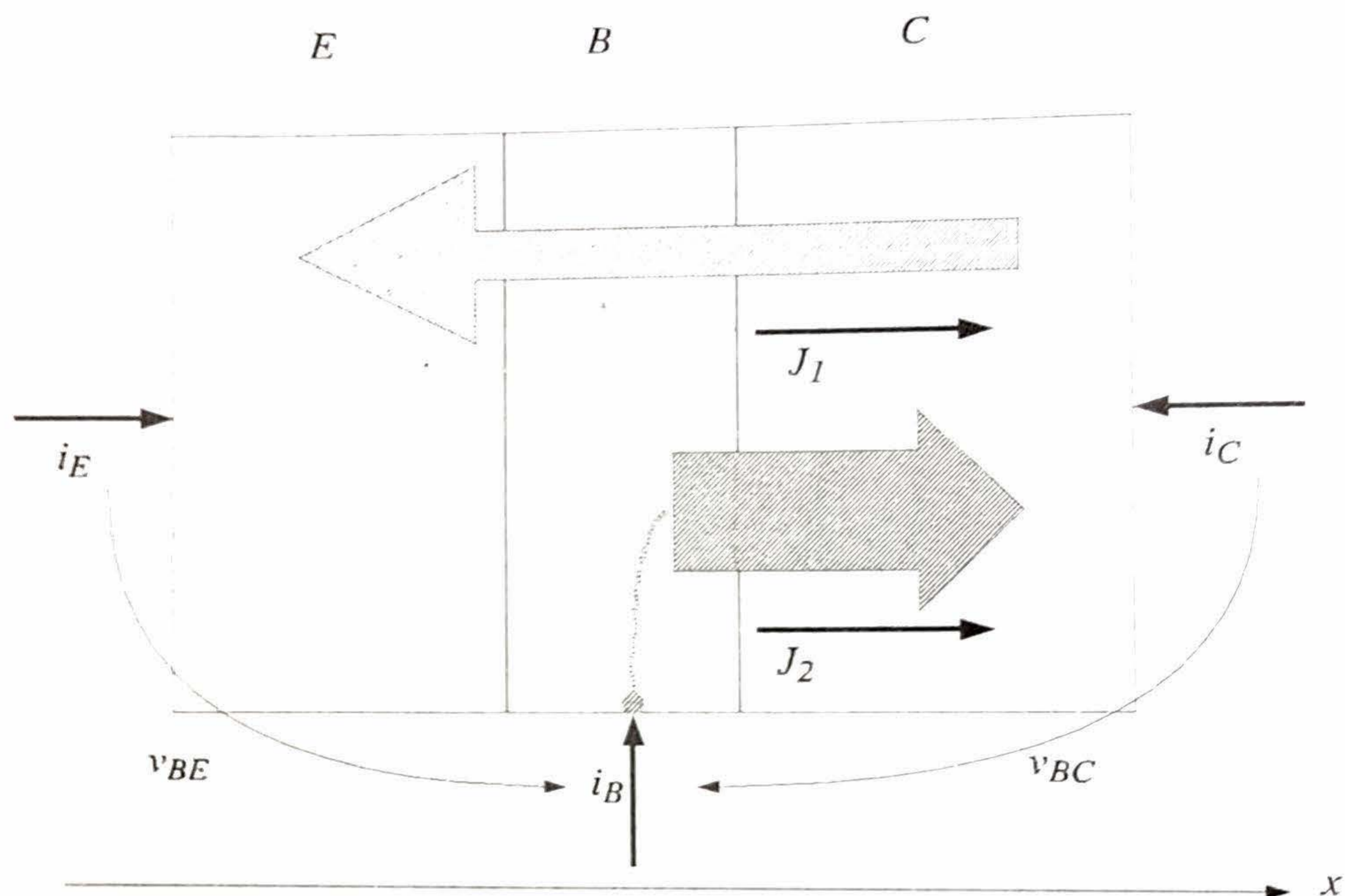


Figura 6.25: Esquema de un transistor bipolar NPN y las corrientes en el modo de conducción activo inverso.

La corriente de electrones que se dirigen del colector al emisor se calcula como:

$$J_1 = qD_n \frac{n(W_B) - n(0)}{W_B} \quad (6.143)$$

donde ahora:

$$n(W_B^-) = N_{dC} e^{-\phi_{JBC}/U_T} = n_{PB} e^{v_{BC}/U_T} \quad (6.144)$$

y

$$n(0) = 0 \quad (6.145)$$

luego de haber reemplazado $\phi_{JBC} = \phi_{BBC} - v_{BC}$, y $N_{dC} e^{-\phi_{JBC}/U_T} = n_{PB}$. Utilizando estas dos ecuaciones en (6.143), resulta:

$$J_1 = qD_n \frac{n_{PB} e^{v_{BC}/U_T}}{W_B} \quad (6.146)$$

La corriente de difusión de huecos, que se dirigen de la base al colector, se puede escribir como:

$$J_2 = -qD_p \frac{p(W_B + W_C) - p(W_B)}{W_C} \quad (6.147)$$

donde:

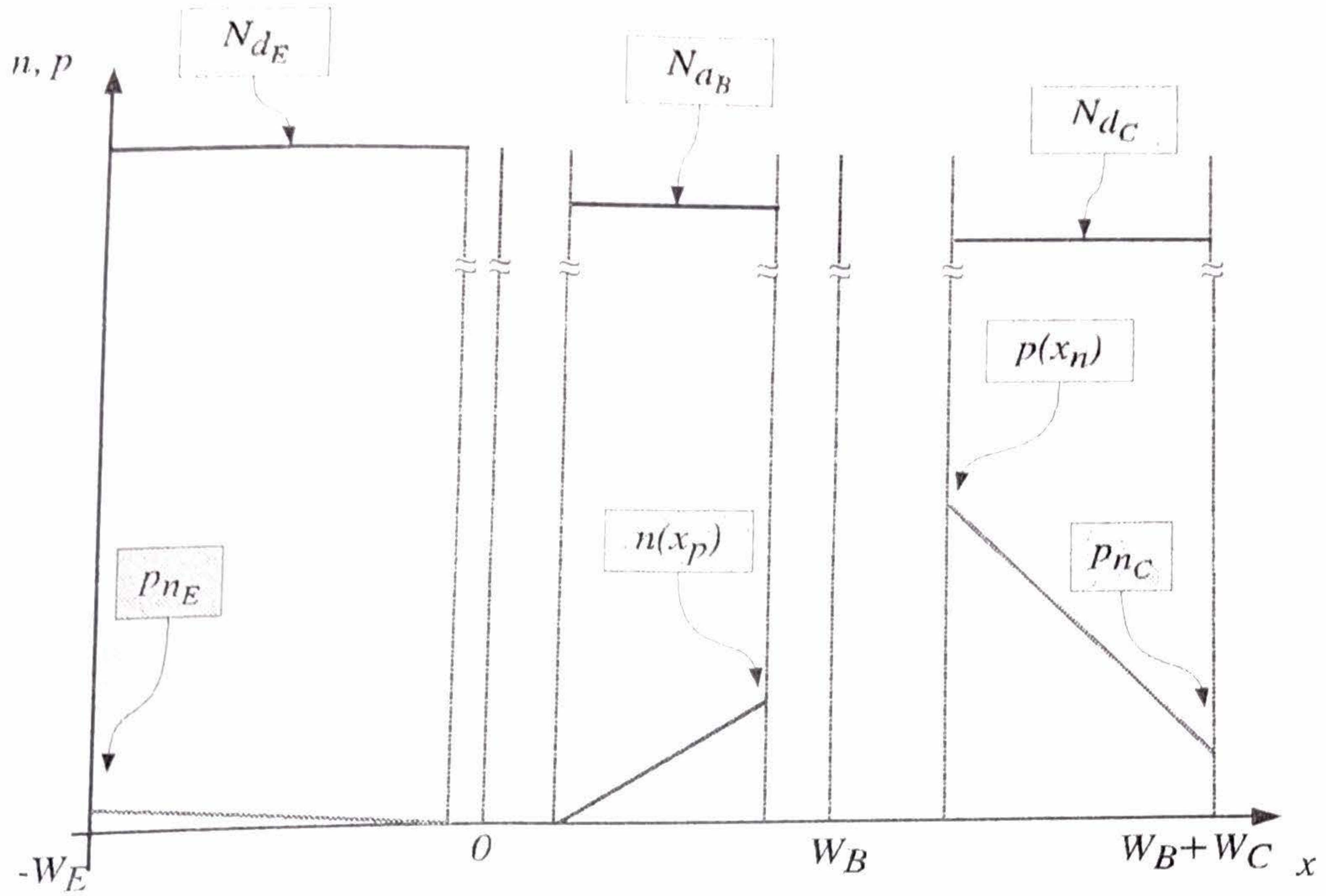


Figura 6.26: Niveles de portadores en un transistor NPN en el modo de conducción activo inverso.

$$p(W_B) = N_{dB} e^{-\phi_{JBC}/U_T} = p_{nC} e^{v_{BC}/U_T} \quad (6.148)$$

y

$$p(W_B + W_C) = p_{nC} \quad (6.149)$$

La corriente J_2 resulta de la forma:

$$J_2 = qD_p \frac{p_{nC} (e^{v_{BC}/U_T} - 1)}{W_C} \quad (6.150)$$

La corriente de colector está compuesta por los electrones, que van de colector a emisor, y los huecos, que van de base a colector, razón por la cual será la mayor de las tres en magnitud:

$$i_C = -AqD_n \frac{n_{pB} e^{v_{BC}/U_T}}{W_B} - AqD_p \frac{p_{nC} (e^{v_{BC}/U_T} - 1)}{W_C} \quad (6.151)$$

Si se considera que, en conducción directa, el término -1 es muy pequeño con respecto al término exponencial, (6.151) puede escribirse como:

$$i_C = -A \left(\frac{qD_n n_{pB}}{W_B} + \frac{qD_p p_{nC}}{W_C} \right) (e^{v_{BC}/U_T} - 1) \quad (6.152)$$

Nótese que en conducción inversa resulta $i_C < 0$.

La corriente de base está compuesta solamente por los huecos que circulan hacia el colector. Estos huecos circulan de base a colector y retornan a la base gracias a la fuente de alimentación. La expresión de esta corriente es:

$$i_B = A \frac{qD_p p_{nC}}{W_C} (e^{v_{BC}/U_T} - 1) \quad (6.153)$$

La corriente de base en este caso sigue siendo positiva, $i_B > 0$.

Por último, la corriente de emisor está compuesta solo por los electrones que se originan en el colector, atraviesan la base, y llegan al emisor, y es ligeramente menor en magnitud que la corriente de colector:

$$i_E = A \frac{qD_n n_{PB}}{W_B} \left(e^{v_{BC}/U_T} - 1 \right) \quad (6.154)$$

Esta corriente también resulta positiva con la convención elegida, $i_E > 0$.

El cociente α_R se define de la siguiente manera:

$$\alpha_R \triangleq -\frac{i_E}{i_C} \quad (6.155)$$

En función de los niveles de dopado, resulta:

$$\alpha_R = \frac{1}{1 + \frac{D_p N_{dC} W_B}{D_n N_{aB} W_C}} \quad (6.156)$$

El cociente entre la corriente de emisor y la corriente de base en activa inversa se define de la siguiente manera:

$$\beta_R \triangleq \frac{i_E}{i_B} \quad (6.157)$$

En función de los niveles de dopado, el parámetro β_R , se puede expresar como:

$$\beta_R = \frac{D_n N_{dC} W_C}{D_p N_{aB} W_B} \quad (6.158)$$

En este caso, dado que $N_{dC} \ll N_{aB}$, también resulta $\beta_R \ll \beta_F$.

Las curvas de salida expresan la variación de i_E versus v_{EB} ó v_{CE} , y se ven afectadas por una tensión de Early mayor, producto que el dopado del emisor es mayor que el de la base:

$$i_E = \beta_R i_B \left(1 + \frac{v_{EB}}{V_A} \right) \quad (6.159)$$

o

$$i_E = A \frac{qD_n n_{PB}}{W_B} \left(e^{v_{BC}/U_T} - 1 \right) \left(1 + \frac{v_{EB}}{V_A} \right) \quad (6.160)$$

Similares expresiones son válidas, en caso de utilizar v_{EC} :

$$i_C = \beta_R i_B \left(1 + \frac{v_{EC} - V_{EC(sat)}}{V_A} \right) \quad (6.161)$$

o

$$i_C = A \frac{qD_n n_{PB}}{W_B} \left(e^{v_{BC}/U_T} - 1 \right) \left(1 + \frac{v_{EC} - V_{EC(sat)}}{V_A} \right) \quad (6.162)$$

donde $V_{EC(sat)}$, es el valor de v_{EC} a partir del cual el dispositivo está en activa inversa.

6.3.3. Región de saturación y el Modelo de Ebers-Moll

En el caso en que la juntura BC esté en conducción directa, el dispositivo está en la zona de funcionamiento denominada saturación. Esta región está caracterizada por:

$$v_{BE} > 0 \quad v_{BC} > 0 \quad (6.163)$$

En la práctica, la conducción directa de la juntura BC comienza a apreciarse cuando $v_{CE} \approx 0,2V$ y $v_{BC} \approx -0,5V$

Para el análisis de esta región, se utilizará la estructura de la Fig. 6.27. Aquí se puede apreciar la presencia de dos corrientes adicionales J'_1 y J'_2 . Estas dos corrientes son producto de la conducción directa de la juntura BC. En la juntura BC el campo eléctrico arrastra los electrones inyectados por la juntura BE (que determinan la corriente J_1) hacia el colector. En la juntura BE, el campo eléctrico arrastra los electrones inyectados por la juntura BC (que determinan la corriente J'_1) hacia el emisor.

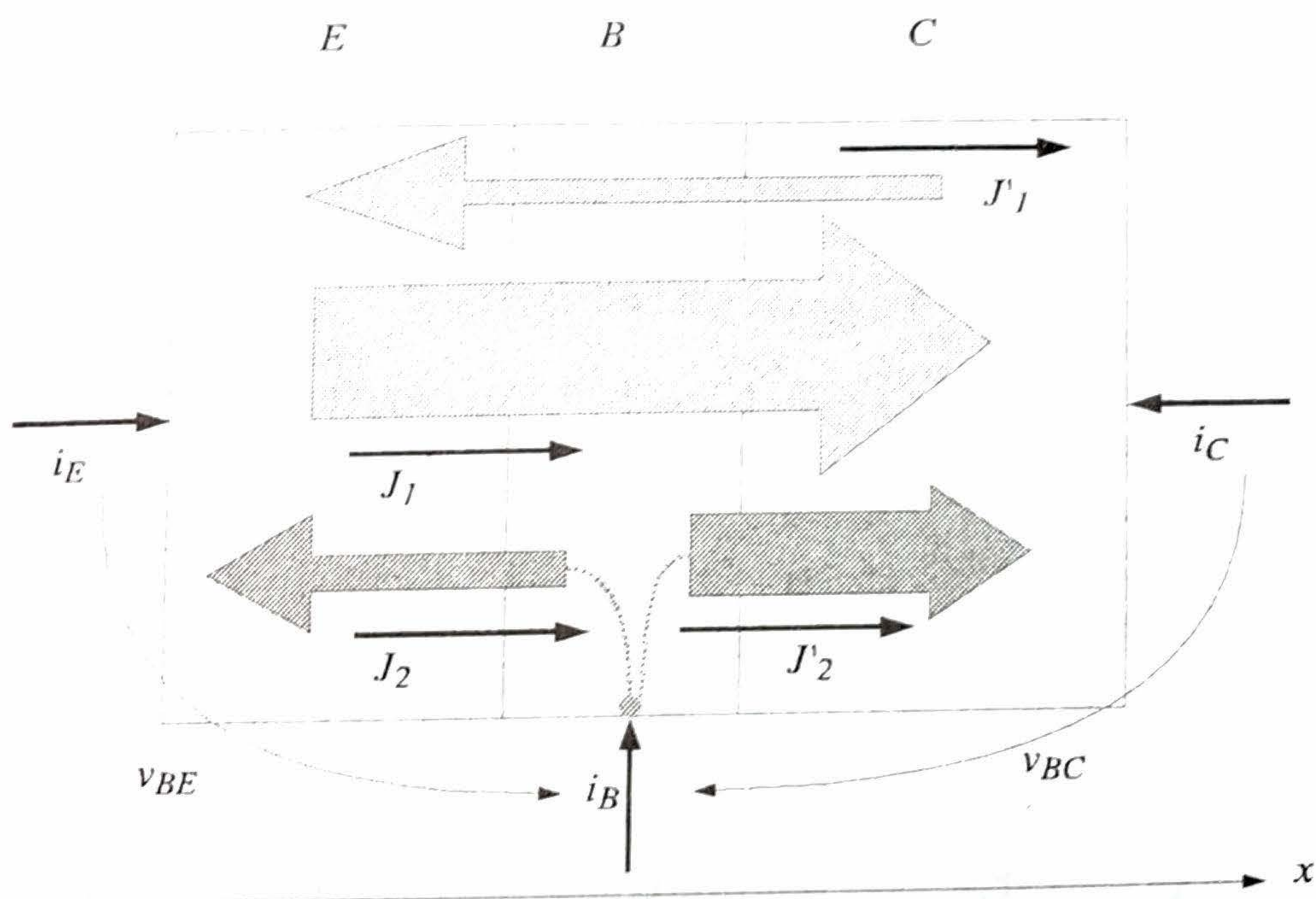


Figura 6.27: Esquema de un transistor bipolar NPN y las corrientes en el modo de saturación.

En este modo de funcionamiento, las corrientes en el dispositivo pueden hallarse haciendo una superposición de los casos de directa activa e inversa activa, es decir, considerando las corrientes individuales J_1 , J_2 , J'_1 y J'_2 , de (6.116), (6.120), (6.146) y (6.147), respectivamente.

Planteando las ecuaciones de nodos en el emisor y el colector:

$$i_E = A_E(J_1 + J_2 + J'_1) \quad (6.164)$$

$$i_C = -A_C(J_1 + J'_1 + J'_2) \quad (6.165)$$

Reemplazando las expresiones (6.116), (6.120), (6.146) y (6.147), las corrientes de emisor y colector resultan:

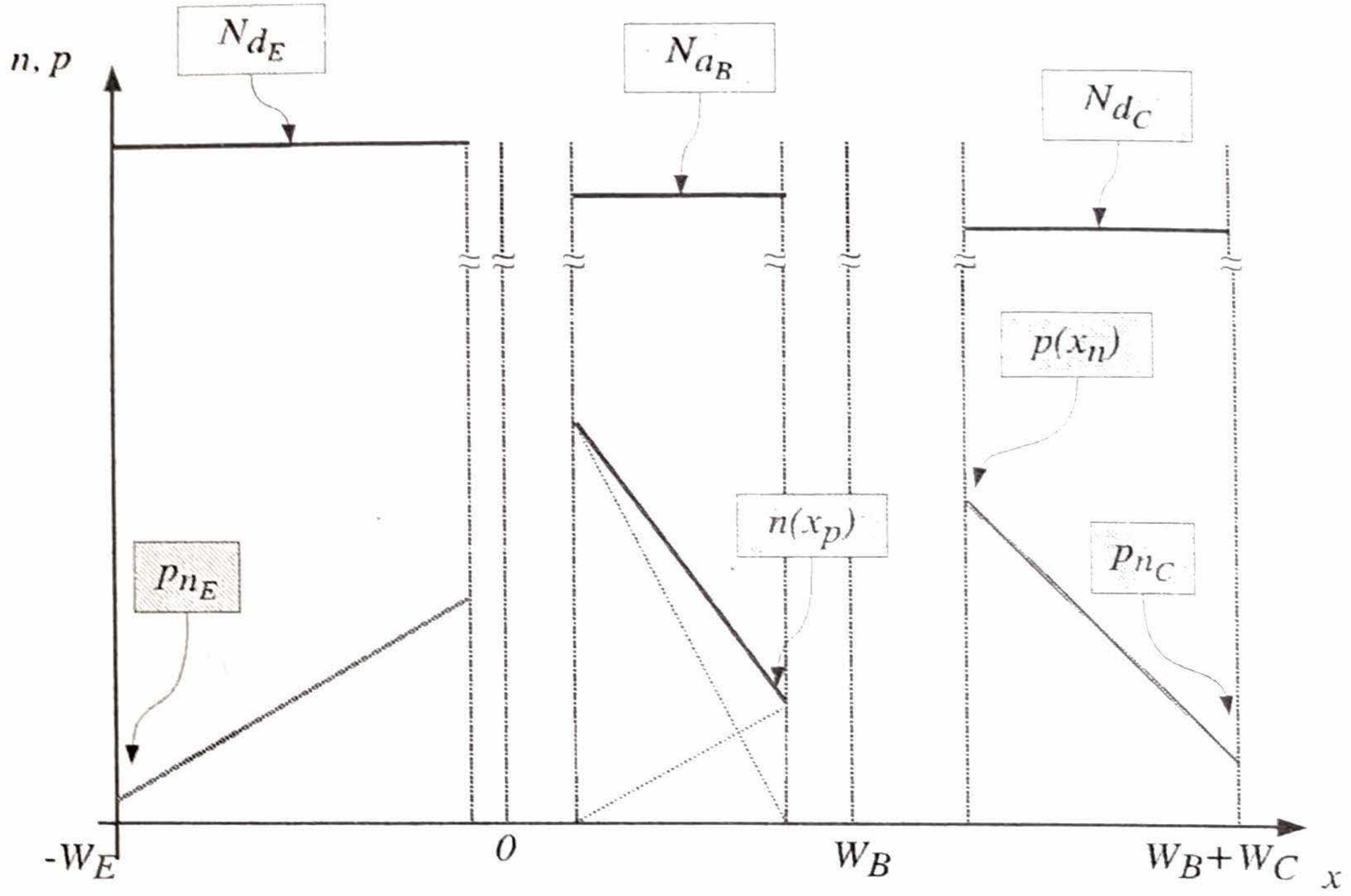


Figura 6.28: Niveles de portadores en un transistor NPN en el modo de saturación.

$$\begin{aligned}
 i_E &= -qA_E \left(\frac{D_n n_{PB} e^{v_{BE}/U_T}}{W_B} + \frac{D_p p_{NE} (e^{v_{BE}/U_T} - 1)}{W_E} - \frac{D_n n_{PB} e^{v_{BC}/U_T}}{W_B} \right) \\
 &= -qA_E \left(\frac{D_n n_{PB} (e^{v_{BE}/U_T} - 1)}{W_B} + \frac{D_p p_{NE} (e^{v_{BE}/U_T} - 1)}{W_E} - \frac{D_n n_{PB} (e^{v_{BC}/U_T} - 1)}{W_B} \right) \\
 &= -qA_E \left(\left(\frac{D_n n_{PB}}{W_B} + \frac{D_p p_{NE}}{W_E} \right) (e^{v_{BE}/U_T} - 1) - \frac{D_n n_{PB}}{W_B} (e^{v_{BC}/U_T} - 1) \right) \quad (6.166)
 \end{aligned}$$

$$\begin{aligned}
 i_C &= -qA_C \left(-\frac{D_n n_{PB} e^{v_{BE}/U_T}}{W_B} + \frac{D_n n_{PB} e^{v_{BC}/U_T}}{W_B} + \frac{D_p p_{NC} (e^{v_{BC}/U_T} - 1)}{W_C} \right) \\
 &= -qA_C \left(-\frac{D_n n_{PB} (e^{v_{BE}/U_T} - 1)}{W_B} + \frac{D_n n_{PB} (e^{v_{BC}/U_T} - 1)}{W_B} + \frac{D_p p_{NC} (e^{v_{BC}/U_T} - 1)}{W_C} \right) \\
 &= -qA_C \left(\left(\frac{D_n n_{PB}}{W_B} + \frac{D_p p_{NC}}{W_C} \right) (e^{v_{BC}/U_T} - 1) - \frac{D_n n_{PB}}{W_B} (e^{v_{BE}/U_T} - 1) \right) \quad (6.167)
 \end{aligned}$$

donde, en ambas ecuaciones, se ha sumado y restado el factor $D_n n_{PB}/W_B$, para completar los términos exponenciales con el factor -1 .

Si se considera $A_E = A_C = A$, luego de definir:

$$I_{ES} \triangleq qA \left(\frac{D_n n_{PB}}{W_B} + \frac{D_p p_{NE}}{W_E} \right) \quad (6.168)$$

$$I_{CS} \triangleq qA \left(\frac{D_n n_{PB}}{W_B} + \frac{D_p p_{NC}}{W_C} \right) \quad (6.169)$$

y sacando factor común, las corrientes de emisor y colector pueden re-escribirse como:

$$i_E = -I_{ES} \left(e^{v_{BE}/U_T} - 1 \right) + \alpha_R I_{CS} \left(e^{v_{BC}/U_T} - 1 \right) \quad (6.170)$$

$$i_C = -I_{CS} \left(e^{v_{BC}/U_T} - 1 \right) + \alpha_F I_{ES} \left(e^{v_{BE}/U_T} - 1 \right) \quad (6.171)$$

En este caso, también se verifica, de (6.166) y (6.167), que

$$I_S = \alpha_R I_{CS} = \alpha_F I_{ES} = \frac{D_n n_{PB}}{W_B} \quad (6.172)$$

Si se define una corriente directa i_F y una corriente inversa i_R que circulan por dos junturas, situadas entre BE y BC, respectivamente:

$$i_F \triangleq I_{ES} \left(e^{v_{EB}/U_T} - 1 \right) \quad (6.173)$$

$$i_R \triangleq I_{CS} \left(e^{v_{CB}/U_T} - 1 \right) \quad (6.174)$$

el modelo de Ebers-Moll puede expresarse como:

$$\begin{aligned} i_E &= -i_F(v_{BE}) + \alpha_R i_R(v_{BC}) \\ i_C &= -i_R(v_{BC}) + \alpha_F i_F(v_{BE}) \end{aligned} \quad (6.175)$$

El circuito correspondiente al modelo de Ebers-Moll se muestra en la Fig. 6.29. El mismo está compuesto por dos junturas con corrientes i_F e i_R , y por dos FCCC de valores α_F y α_R . Nótese que los sentidos de los diodos y las fuentes de corriente son opuestos al caso del transistor PNP.

La corriente de base puede obtenerse de la resta de estas últimas dos ecuaciones como:

$$i_B = (1 - \alpha_F) i_F(v_{EB}) + (1 - \alpha_R) i_R(v_{CB}) \quad (6.176)$$

La saturación comienza cuando la tensión de colector (positiva) ha disminuido lo suficiente como para que sea $v_{BC} = 0$. En ese momento, se verifica:

$$\begin{aligned} i_E &= -i_F = -I_{ES} \left(e^{v_{BE}/U_T} - 1 \right) \\ i_C &= \alpha_F i_F = \alpha_F I_{ES} \left(e^{v_{BE}/U_T} - 1 \right) \\ i_B &= (1 - \alpha_F) i_F \end{aligned} \quad (6.177)$$

A partir de este punto, si se produce un incremento en v_{BC} como para producir un aumento de la corriente Δi_R , o una disminución de la corriente de colector Δi_C , similar (pero menor) a i_F , se ve que se produce un ligero aumento de i_E en un factor $\alpha_R i_R$, y un aumento de i_B casi igual a i_R dado por $(1 - \alpha_R) i_R$. Si se continúa aumentando la tensión v_{BC} , eventualmente, la componente de corriente de colector debida a la corriente directa de la juntura CB, dada por i_R , igualará en magnitud la inyección de electrones de la juntura BE dada por $\alpha_F i_F$. En este momento la corriente de colector es nula, por lo cual el dispositivo perdió por completo su ganancia de corriente i_C/i_B , y la corriente de emisor completa su recorrido por la base, es decir:

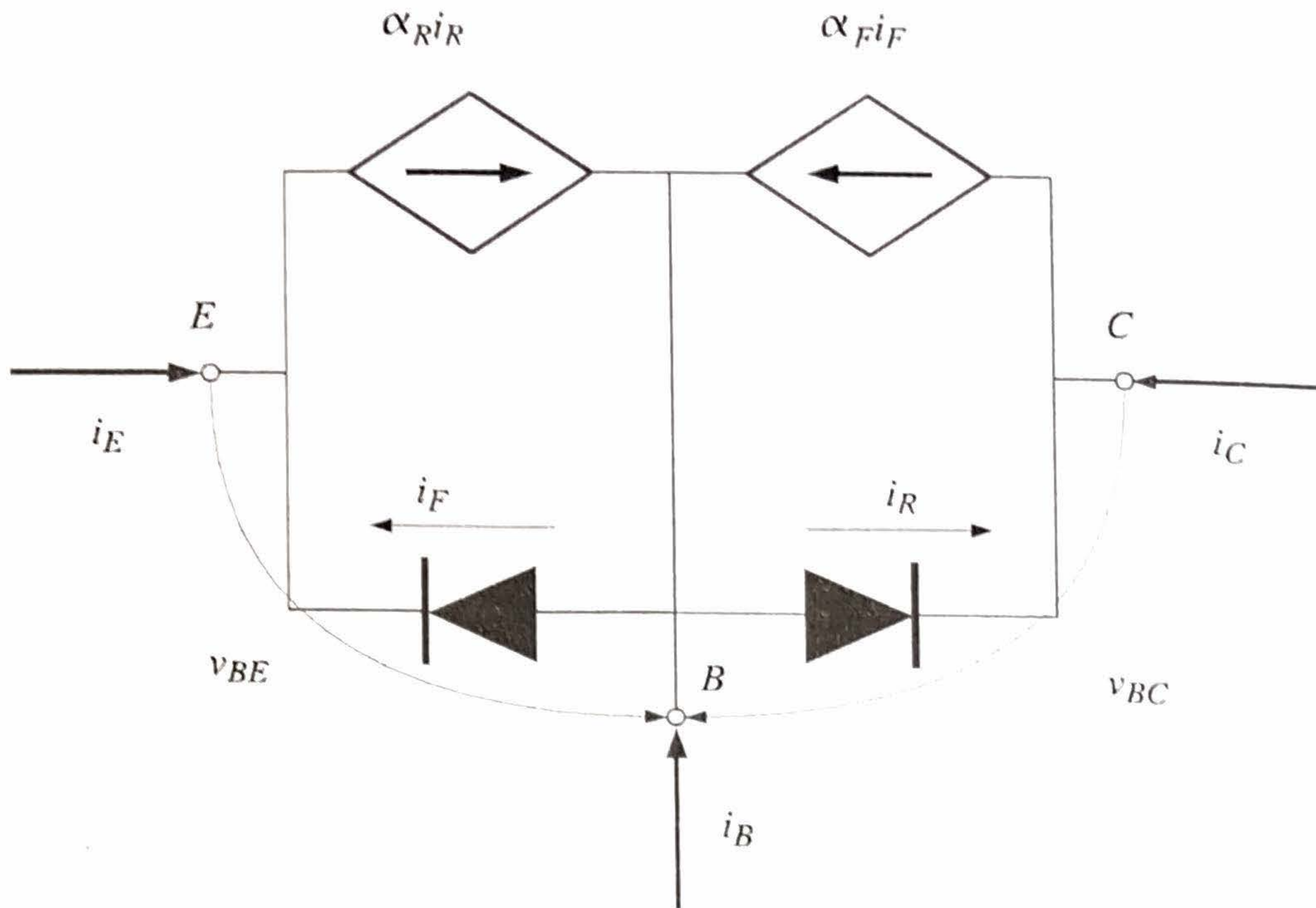


Figura 6.29: Circuito correspondiente al modelo de Ebers-Moll para un transistor NPN.

$$\begin{aligned}
 i_E &= -(1 - \alpha_F \alpha_R) I_{ES} \left(e^{v_{BE}/U_T} - 1 \right) \\
 i_C &= 0 \\
 i_B &= -i_E
 \end{aligned}
 \tag{6.178}$$

La zona de saturación comprende los valores v_{BC} , tal que $0 \leq v_{BC} \leq v_{BC(sat)}$; es decir, desde que la corriente de colector comienza su reducción (situación descrita por la ecuaciones en (6.177)), hasta que se anula (situación descrita por la ecuaciones en (6.178)).

La Fig. 6.12 para el transistor PNP puede utilizarse aquí (con la sustitución de V_{EC} por V_{EC}) para visualizar la situación en saturación. A medida que el cociente de las corrientes se achica, lo cual sucede para $v_{CE} < 0,35V$, la tensión $V_{CE(sat)}$ se mantiene aproximadamente constante con un valor medio de $V_{CE(sat)} \approx 0,2V$. En esta situación, dado que ambas junturas están en conducción directa, $V_{BE(sat)} \approx 0,7V$ y $V_{BC(sat)} \approx 0,5V$.

La Fig. 6.30 muestra las curvas experimentales de salida de un transistor NPN en la región de saturación.

6.3.4. Modelos Simplificados

A continuación, se resumen los modelos simplificados para el transistor NPN.

En la zona activa directa, $v_{BC} \ll 0$, con lo cual las ecuaciones del modelo son las siguientes:

$$\begin{aligned}
 i_E &= -I_{ES} \left(e^{v_{BE}/U_T} - 1 \right) \\
 i_C &= \alpha_F I_{ES} \left(e^{v_{BE}/U_T} - 1 \right) \approx -\alpha_F i_E \\
 i_B &= \frac{1 - \alpha_F}{\alpha_F} i_C = \frac{1}{\beta_F} i_C
 \end{aligned}
 \tag{6.179}$$

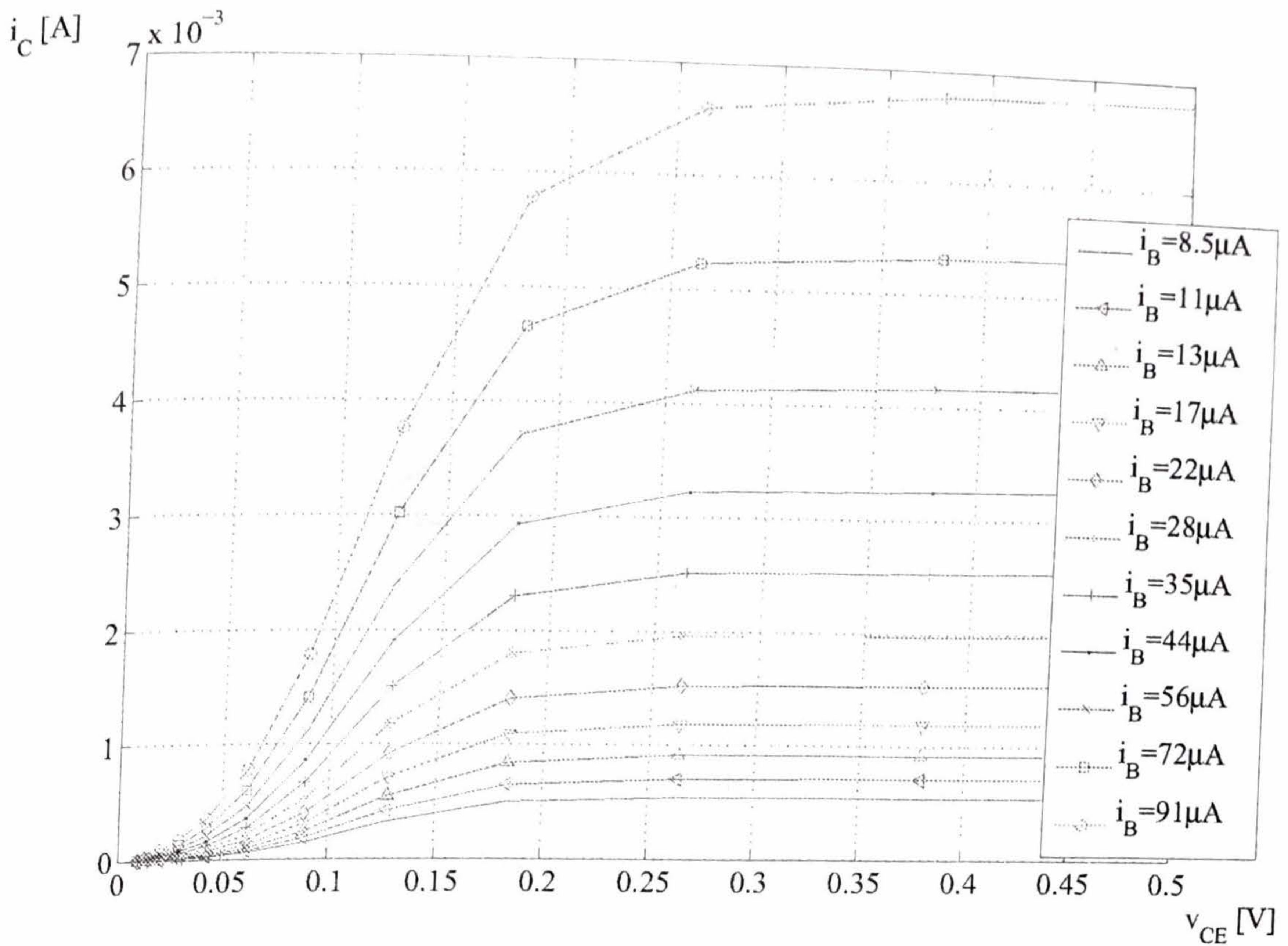


Figura 6.30: Curvas experimentales de salida para un transistor NPN BC548 en la región de saturación.

Estas ecuaciones corresponden a los circuitos de la Fig. 6.31, compuestos por un diodo entre base y emisor, y una FCCT entre colector y emisor.

En la zona reversa activa, $v_{BE} \ll 0$, con lo cual las ecuaciones del modelo son las siguientes:

$$\begin{aligned}
 i_C &= -I_{CS} \left(e^{v_{BC}/U_T} - 1 \right) \\
 i_E &= \alpha_R I_{CS} \left(e^{v_{BC}/U_T} - 1 \right) \approx -\alpha_R i_C \\
 i_B &= \frac{1 - \alpha_R}{\alpha_R} i_E = \frac{1}{\beta_R} i_E
 \end{aligned} \tag{6.180}$$

Estas ecuaciones corresponden a un diodo entre base y colector, y una FCCT entre colector y emisor.

En saturación, ambas junturas están en conducción directa, con lo cual pueden reemplazarse por dos fuentes de tensión, de valores:

$$V_{BE(sat)} = 0,7V \tag{6.181}$$

$$V_{BC(sat)} = 0,5V \tag{6.182}$$

como se muestra en la Fig. 6.32-a.

De manera equivalente, la segunda fuente de tensión puede colocarse entre el colector y el emisor, de manera que el circuito resultante posea dos fuentes de tensión de valor:

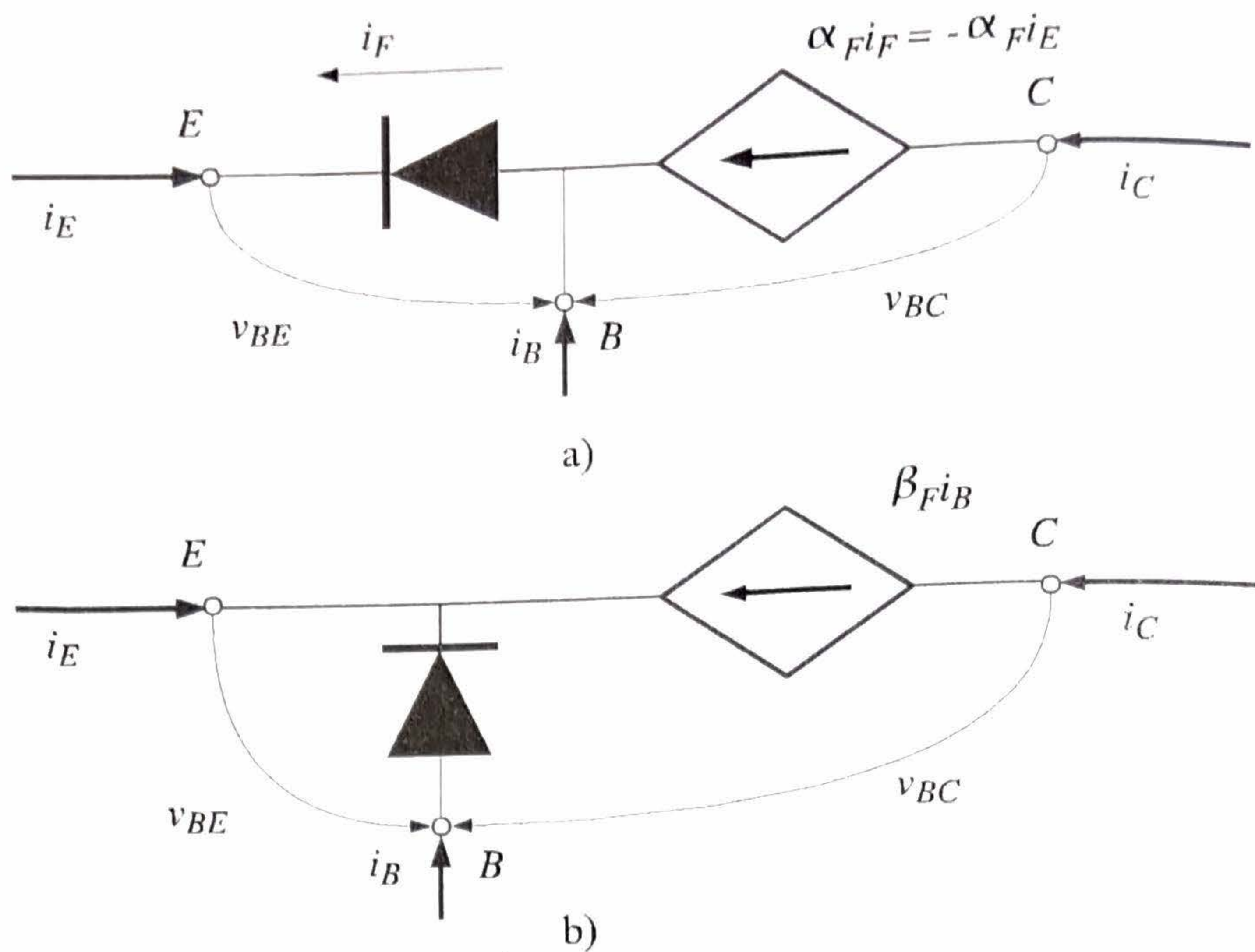


Figura 6.31: Circuito simplificado de Ebers-Moll en conducción directa activa para un transistor NPN: a) referido a la base; b) referido al emisor.

$$V_{BE(sat)} = 0,7V \quad (6.183)$$

$$V_{CE(sat)} = 0,2V \quad (6.184)$$

como se muestra en la Fig. 6.32-b.

6.4. Modelo Lineal Incremental (MLI)

Para el desarrollo del MLI del transistor bipolar, se parte del modelo simplificado de Ebers-Moll, de un transistor PNP para el caso de conducción activa directa, explicado en la Sección 6.2.4, que se repite por conveniencia:

$$\begin{aligned} i_E &= I_{ES} e^{v_{EB}/U_T} \\ i_C &= -\alpha_F i_E = \beta_F i_B \\ i_B &= -\frac{1}{\beta_F + 1} i_E \end{aligned} \quad (6.185)$$

Entre base y emisor hay una juntura en modo de conducción directa, y por lo tanto, su MLI es una conductancia/resistencia dinámica, tal como en el caso del diodo. La única diferencia, en este caso, es que es posible utilizar i_E o i_B como corrientes de entrada. En el caso que se tome i_E como corriente de entrada, la conductancia dinámica estará dada por:

$$g_e \triangleq \left. \frac{di_E}{dv_{EB}} \right|_Q = \frac{i_e}{v_{eb}} = \frac{I_{ES}}{U_T} e^{v_{EB}/U_T} = \frac{I_E}{U_T} \quad (6.186)$$

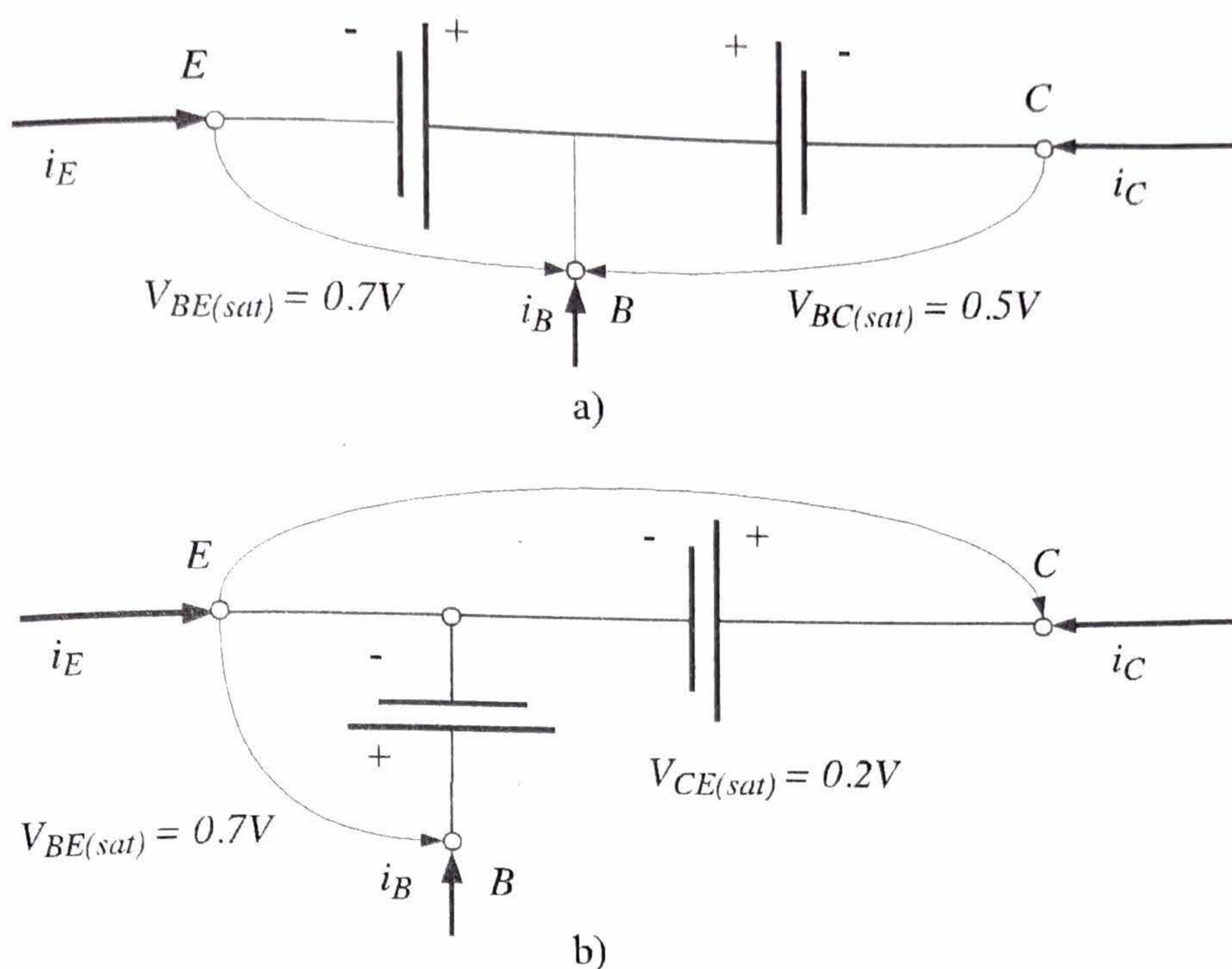


Figura 6.32: Circuito simplificado de Ebers-Moll en saturación para un transistor NPN: a) referido a la base; b) referido al emisor.

En el caso que se tome i_B como corriente de entrada, la conductancia dinámica estará dada por:

$$g_b \triangleq \left. \frac{di_B}{dv_{BE}} \right|_Q = \frac{i_b}{v_{be}} = \frac{I_{ES}}{(\beta_F + 1)U_T} e^{-v_{BE}/U_T} = \frac{I_E}{(\beta_F + 1)U_T} = -\frac{I_B}{U_T} \quad (6.187)$$

Debe notarse aquí que I_B es negativa, por lo tanto, la conductancia tiene valor positivo. Las resistencias dinámicas correspondientes son:

$$r_e \triangleq g_e^{-1} = \frac{U_T}{I_E} \quad r_b \triangleq g_b^{-1} = \frac{(\beta_F + 1)U_T}{I_E} \quad (6.188)$$

La resistencia dinámica que se observa desde la base es $\beta_F + 1$ veces más grande que la vista desde el emisor, dado que desde la base se ve a la juntura BE con una conducción de corriente $\beta_F + 1$ mas pequeña⁷.

Si se toma la corriente de colector como variable de salida, ésta puede interpretarse como una función de la corriente de base, o como una función de la tensión desarrollada sobre la juntura BE. En caso que se interprete como una función de la corriente de base, se debe definir una ganancia de corriente:

$$\beta_f \triangleq \left. \frac{di_C}{di_B} \right|_Q = \frac{i_c}{i_b} = \beta_F \quad (6.189)$$

En caso que se interprete como una función de la tensión de la juntura EB, se debe definir una transconductancia:

⁷Como se puede ver de (6.187), para obtener el valor de la resistencia dinámica vista desde la base, se debe utilizar v_{be} en lugar de v_{eb}

$$g_m \triangleq \left. \frac{di_C}{dv_{BE}} \right|_Q = \frac{i_c}{v_{be}} = \frac{\beta_F I_{ES}}{(\beta_F + 1)U_T} e^{-v_{BE}/U_T} = -\frac{I_C}{U_T} \quad (6.190)$$

La transconductancia se define como la relación entre la variación de la corriente de colector y la variación de la tensión de base, con respecto a la de emisor, y su valor es positivo; es decir, que un incremento de la tensión de base produce un incremento de la corriente de colector. Si se analiza la variación de corriente de colector con respecto a la variación de la tensión de emisor, la transconductancia tiene la misma magnitud, pero signo contrario; es decir, $i_c/v_{eb} = -i_c/v_{be}$. El significado de esto es que un aumento de la tensión de emisor produce una disminución de la corriente de colector.

Por último, si se analiza la salida del dispositivo, de acuerdo a (6.135), y teniendo en cuenta el efecto de Early, la conductancia de salida es:

$$g_o \triangleq \left. \frac{di_C}{dv_{CE}} \right|_Q = \frac{i_c}{v_{ce}} = \frac{\beta_F I_B}{V_A} = \frac{I_C}{V_A} \quad (6.191)$$

El modelo híbrido- π

El modelo híbrido- π es una de las representaciones más frecuentes del transistor bipolar. En esta representación se modela el transistor de tres terminales como una red de dos puertos, con su terminal de emisor como terminal de referencia (ver Fig. 6.33-a). El modelo se denomina híbrido, porque considera como variables una tensión de entrada, v_{be} , y una corriente de salida, i_c . La representación es la siguiente:

$$\begin{bmatrix} i_c \\ v_b \end{bmatrix} = \begin{bmatrix} h_{fe} & h_{oe} \\ h_{ie} & h_{re} \end{bmatrix} \times \begin{bmatrix} i_b \\ v_c \end{bmatrix} \quad (6.192)$$

Los términos de la matriz tienen dos subíndices. El segundo subíndice indica el terminal que se toma como referencia. En este caso, "e" indica que el emisor se toma como terminal de referencia.

En la ecuación de salida, el subíndice "f" indica "forward", y relaciona el efecto de una variable de entrada (la corriente i_b), sobre una variable de salida, i_c (a través del término h_{fe}). En esta misma ecuación, el subíndice "o", indica "output", y relaciona el efecto de una variable de salida (la tensión v_c), sobre otra variable de salida, i_c (a través del término h_{oe}).

En la ecuación de entrada, el subíndice "i", indica "input", y relaciona el efecto de una variable de entrada (la corriente i_b), sobre una variable de entrada, v_b (a través del término h_{ie}). En esta misma ecuación, el subíndice "r", indica "reverse", y relaciona el efecto de una variable de salida (la tensión v_c), sobre una variable de entrada, v_b (a través del término h_{re}).

El lector puede verificar de (6.192), y del modelo de la Sección 6.4, que:

$$h_{fe} = \left. \frac{i_c}{i_b} \right|_{v_c=0} = \beta_f \quad (6.193)$$

$$h_{oe} = \left. \frac{i_c}{v_c} \right|_{i_b=0} = g_o = \frac{I_C}{V_A} \quad (6.194)$$

$$h_{ie} = \left. \frac{v_b}{i_b} \right|_{v_c=0} = r_b = (\beta_F + 1) \frac{U_T}{I_E} \quad (6.195)$$

El último término

$$h_{re} = \left. \frac{v_b}{v_c} \right|_{i_b=0} = \frac{v_b}{i_c} \times \frac{i_c}{v_c} = \frac{U_T}{I_C} \frac{I_C}{V_A} = \frac{U_T}{V_A} \quad (6.196)$$

6.4. Modelo Lineal Incremental (MLI)

es la fracción de la variación de tensión de salida que aparece en la entrada, y para valores típicos de $V_A \approx 100V$, resulta en $h_{re} \approx 2,5 \times 10^{-4}$, con lo cual, en general se adopta $h_{re} = 0$. De esta manera, el modelo híbrido- π queda determinado como se muestra en la Fig. 6.33-b.

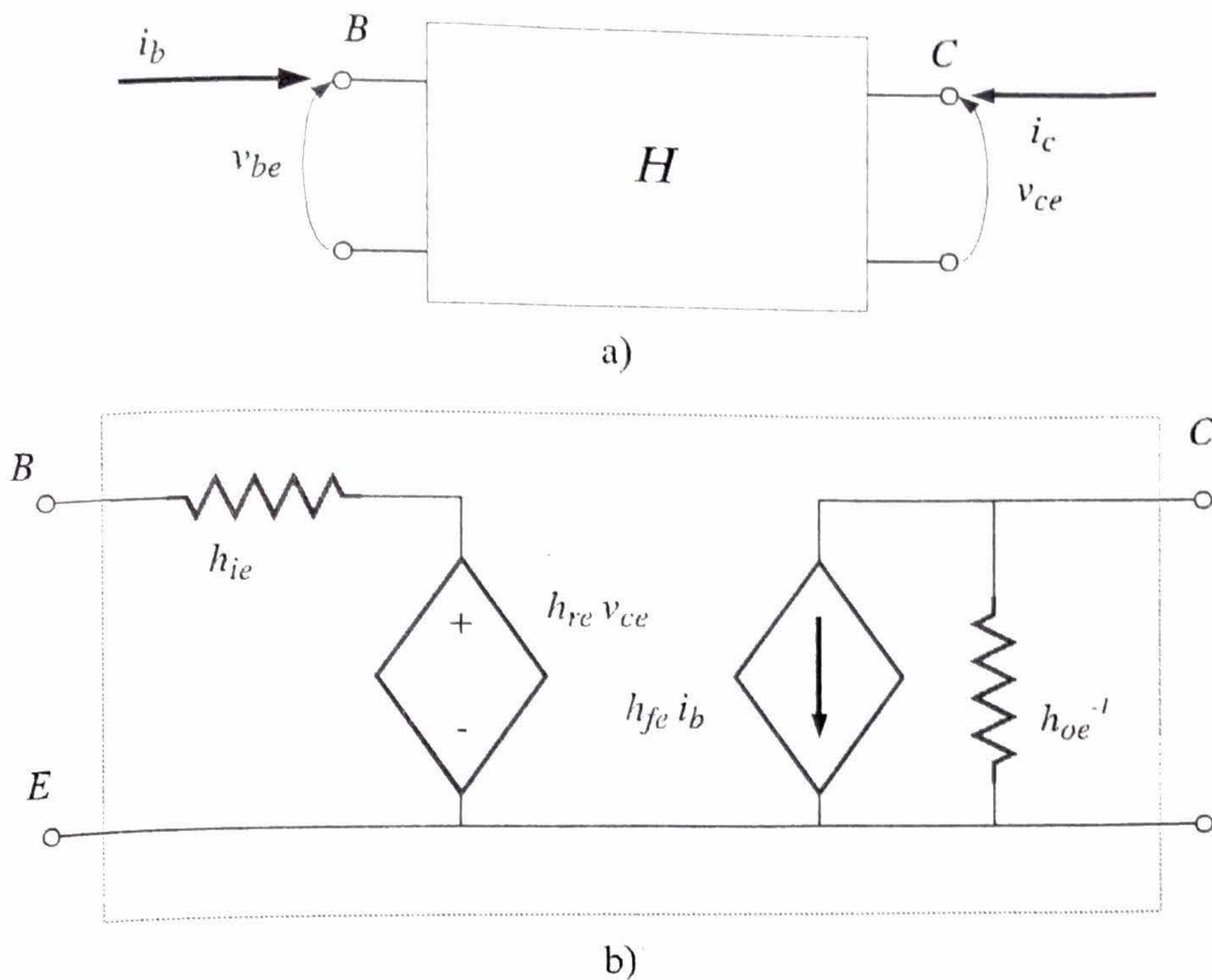


Figura 6.33: Modelo lineal incremental híbrido- π de un transistor bipolar.

Ejemplo 28 Dado el circuito del Ejemplo 26, correspondiente a la Fig. 6.15-d, con $V_A = -100V$. Qué valores tienen las resistencias de entrada y salida? Si se aplica una tensión de señal, $v_b = 2,5mV$, a la base del transistor, cuáles son los valores de la corriente y tensión de salida de señal resultante?

El dispositivo funciona con los siguientes valores de DC: $I_B = -10\mu A$, $I_C = -1mA$, $V_{RC} = -2V$, $V_{CE} = -3V$. Para estos valores de DC, el transistor está en la región de conducción activa directa.

La resistencia de entrada (tomando la base como entrada) es:

$$r_b = h_{ie} = \frac{(\beta_F + 1)U_T}{I_E} \approx \frac{101 \times 25mV}{1mA} = 2,525K\Omega \quad (6.197)$$

donde se supuso $\beta_F = \beta_f$ y $I_E \approx -I_C$. La conductancia de salida es:

$$g_o = \frac{-1mA}{-100V} = 10 \frac{1}{M\Omega} \quad (6.198)$$

Esta conductancia corresponde a una resistencia de salida $r_o = 100K\Omega$. Si la tensión de señal en la base es $v_b = 2,5mV$, entonces la corriente de base (de señal) será:

$$i_b = \frac{v_b}{r_b} = \frac{2,5mV}{2,525K\Omega} = 0,99\mu A \quad (6.199)$$

La corriente de salida, en este caso, la corriente de colector, será $i_c = 99\mu A$. Como se ve en el circuito de pequeña señal de la Fig. 6.34, la resistencia de salida total, R_o , es la combinación paralelo de la resistencia de salida del transistor $r_o = 100K\Omega$ y la resistencia de colector $R_C = 2K\Omega$, con lo cual $R_o = 1,96K\Omega$. La tensión de señal resultante sobre esta resistencia es:

$$v_o = v_{ce} = -i_c \times R_o = -99\mu A \times 1,96K\Omega = -194mV \quad (6.200)$$

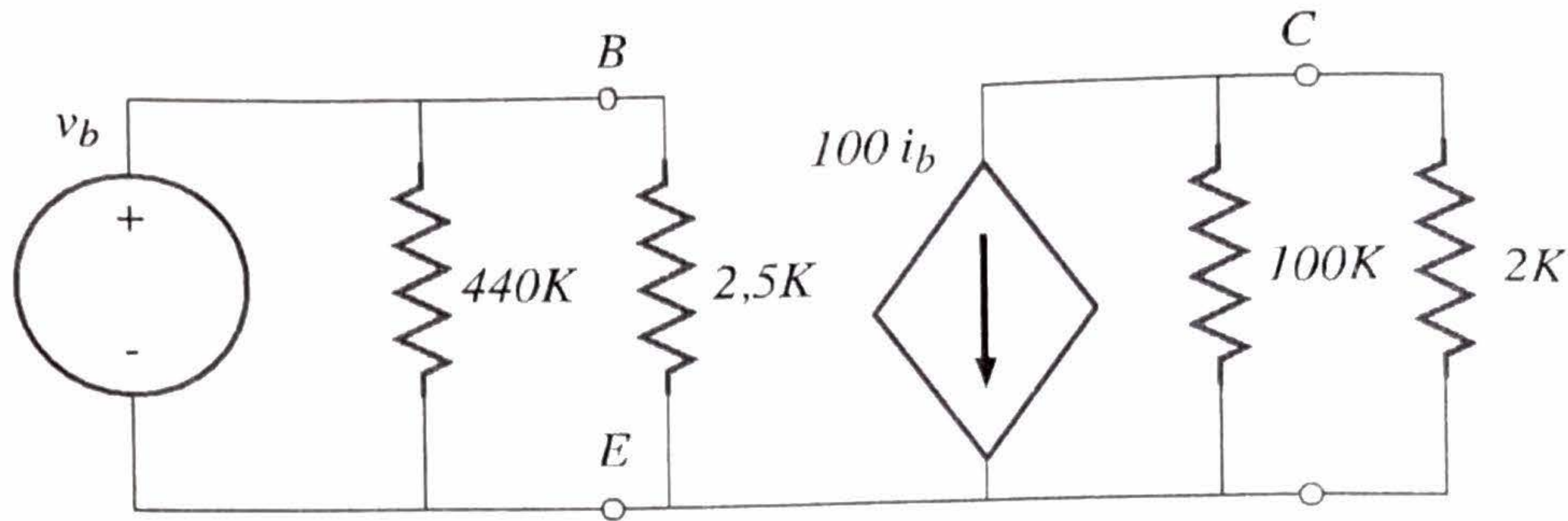


Figura 6.34: Modelo de pequeña señal del circuito del Ejemplo 28.

6.5. Modelo de AC

6.5.1. Capacidad de vaciamiento

Dado que el transistor bipolar está compuesto por dos junturas, una en directa (EB) y otra en inversa (CB), surgen dos capacidades que siguen los lineamientos descritos en el Cap. 3. La componente de vaciamiento de la capacidad entre emisor y base, está dada por (3.109):

$$C_{veb} = \sqrt{2}C_{jeb0} \quad (6.201)$$

donde C_{jeb0} , es la capacidad, sin tensión aplicada, de la juntura EB.

Por otro lado, la capacidad de vaciamiento de la juntura de CB tendrá un valor:

$$C_{vcb} = C_{jcb0} \frac{1}{\left(1 - \frac{V_{CB}}{\phi_{B_{cb}}}\right)^{m_{cb}}} \quad (6.202)$$

donde C_{jcb0} es la capacidad, sin tensión aplicada, de la juntura CB, $\phi_{B_{cb}}$ es el potencial intrínseco de la juntura CB y m_{cb} su coeficiente asociado.

En caso que el colector esté dispuesto sobre un sustrato, habrá otra capacidad entre el terminal de colector y el sustrato, dada por:

$$C_{vcs} = C_{jcs0} \frac{1}{\left(1 - \frac{V_{CS}}{\phi_{B_{cs}}}\right)^{m_{cs}}} \quad (6.203)$$

donde C_{jcs0} es la capacidad, sin tensión aplicada, de la juntura CS, $\phi_{B_{cs}}$ es el potencial intrínseco de la juntura CS y m_{cs} su coeficiente asociado.

6.5.2. Capacidad de carga de la base

De la misma manera que en el caso del diodo, cuando el transistor se halla en conducción directa, su juntura EB se halla en directa, y se produce una acumulación de portadores minoritarios a ambos lados de la misma. Dado que el dopado del emisor es mucho mayor que el de la base, solo es necesario

considerar la carga acumulada en la base (huecos provenientes del emisor). Una variación de la tensión v_{EB} provocará un cambio en la carga acumulada, dando lugar a un efecto capacitivo.

La carga acumulada en la base es de la forma:

$$\begin{aligned} Q_p &= qAW_B \frac{(p(W_B) + p(0))}{2} \\ &= \frac{1}{2} qAW_{BP} N_B e^{v_{EB}/U_T} \end{aligned} \quad (6.204)$$

La capacidad incremental es:

$$C_B = \left. \frac{Q_p}{dv_{cb}} \right|_Q = \frac{qAW_{BP} N_B e^{v_{EB}/U_T}}{2U_T} \quad (6.205)$$

Si se define el tiempo de tránsito de la base:

$$\tau_F \triangleq \frac{W_B^2}{2D_{pB}} \quad (6.206)$$

y se multiplica y divide (6.204) por W_B/D_{pB} , resulta:

$$\begin{aligned} C_B &= \frac{qAD_{pB} p_{NB} e^{v_{EB}/U_T}}{W_B} \times \frac{W_B^2}{2U_T D_{pB}} \\ &= \frac{I_C}{U_T} \times \tau_F = g_m \tau_F \end{aligned} \quad (6.207)$$

El tiempo de tránsito τ_F es el tiempo promedio que tardan los huecos en recorrer la base, y es un parámetro de gran importancia en los transistores bipolares, dado que determina la máxima velocidad de operación. De acuerdo a [10], un valor típico puede obtenerse teniendo en cuenta $W_B = 300\text{nm}$ y $D_{pB} = 20\text{cm}^2/\text{s}$. Esto arroja un valor de $\tau_F = 90\text{ps}$.

El circuito correspondiente al MLI de AC se ilustra en la Fig. 6.35-a donde se han definido dos capacidades equivalentes:

$$C_\pi \triangleq C_{VEB} + C_B \quad (6.208)$$

$$C_\mu \triangleq C_{VCB} \quad (6.209)$$

6.5.3. Límite de validez del modelo de AC

De manera similar a lo descrito para el caso del transistor MOS, conforme se eleva la frecuencia de operación, la ganancia de corriente decrece y es necesario utilizar otros modelos. En el caso del transistor bipolar, también se define la frecuencia de transición f_T , como aquella frecuencia para la cual la ganancia de corriente se hace igual a 1. Desde el punto de vista del circuito, lo que sucede es que las capacidades C_π y C_μ disminuyen su impedancia con la frecuencia, y toman más corriente de la entrada. Debido a ello, la corriente de base, y por ende, la corriente de colector (que es la corriente de salida), disminuye. Para el cálculo de f_T se utiliza el circuito de la Fig. 6.35-b, y su circuito equivalente de pequeña señal de la Fig. 6.35-c. En la base del transistor se coloca una fuente de corriente de señal, de valor:

$$i(t) = I_o \sin(2\pi ft) \quad (6.210)$$

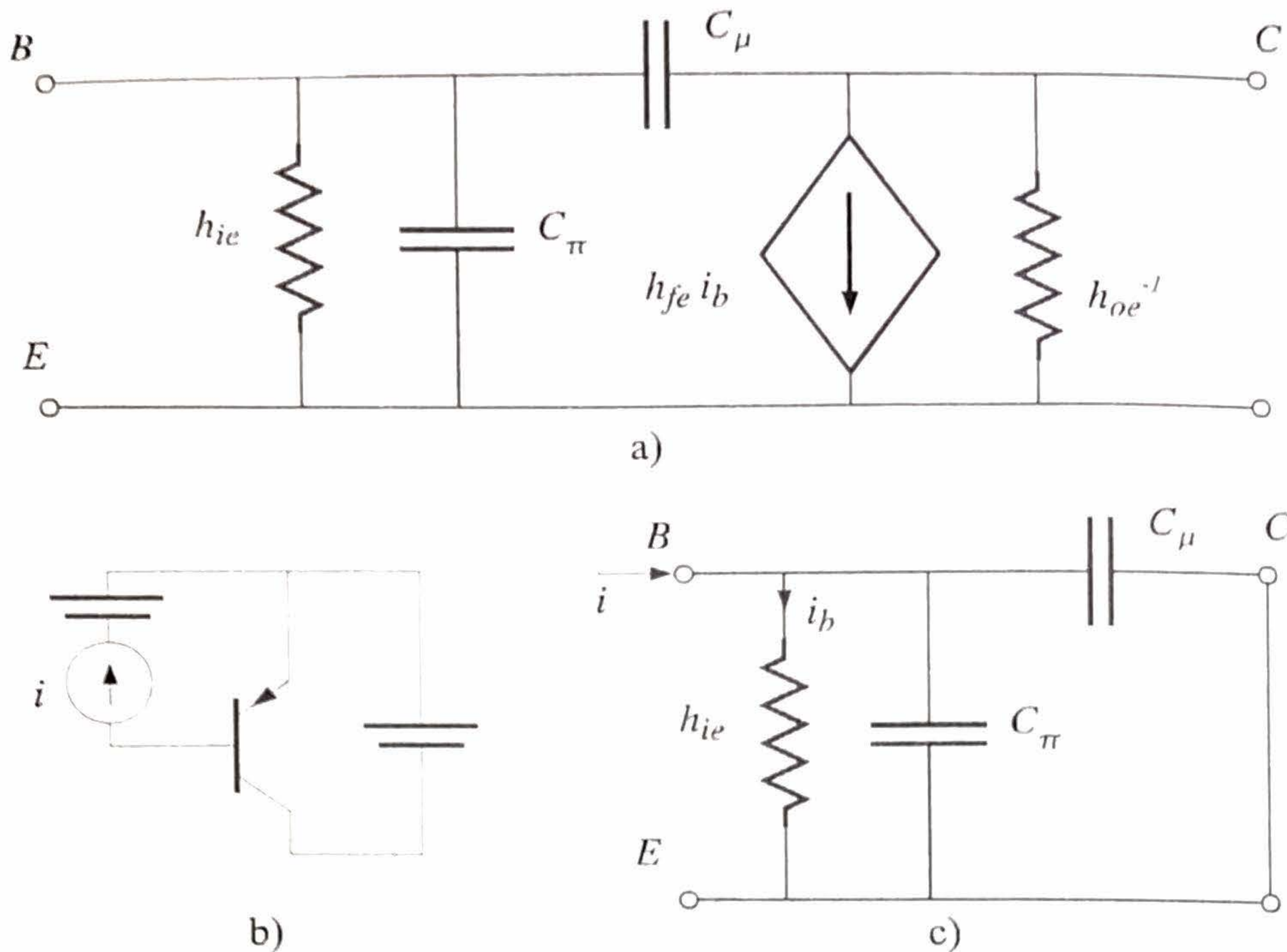


Figura 6.35: a) Modelo lineal incremental híbrido- π de AC de un transistor bipolar; b) circuito para determinar la frecuencia de transición; c) circuito de pequeña señal correspondiente al circuito en b).

La tensión v_{be} satisface :

$$v_{be} = i_b h_{ie} = (i - i_b) \frac{1}{j2\pi f(C_\pi + C_\mu)} \quad (6.211)$$

y de esta igualdad es posible hallar la relación entre i_b e i :

$$i_b = \frac{1}{j2\pi f h_{ie}(C_\pi + C_\mu) + 1} \times i \quad (6.212)$$

Dado que $i_c = h_{fe} i_b$, y despreciando el término 1 frente a $2\pi f h_{ie}(C_\pi + C_\mu)$, la ganancia de corriente (en módulo) es igual a:

$$A_i = \frac{h_{fe}}{2\pi f h_{ie}(C_\pi + C_\mu)} \quad (6.213)$$

Considerando que $h_{ie} = h_{fe} U_T / I$ y que $C_B = \tau_F I / U_T$, donde $I = I_E \approx I_C$, si se iguala $A_i(f_T) = 1$ y se despeja f_T , se obtiene:

$$f_T = \frac{1}{2\pi \left(\tau_F + \frac{U_T}{I} (C_{VEB} + C_\mu) \right)} \quad (6.214)$$

De esta ecuación se ve que el incremento de la corriente produce un aumento de la frecuencia de transición, permitiendo que el dispositivo funcione a mayores frecuencias. Por otro lado, el aumento de la tensión de la juntura BC, hace que C_μ disminuya, permitiendo un incremento de la frecuencia de operación. Tal como se menciona en [15], para valores significativos de corriente, τ_F comienza a aumentar, y f_T cae; debido a esto, el punto de operación se debe elegir para tener valores razonables, tanto de h_{fe} , como de velocidad.

6.5. Modelo de AC

Ejemplo 29 Para el circuito del Ejemplo 28, suponga los siguientes parámetros⁸: $C_{jeb0} = 11fF$, $\phi_{B_{cb}} = 0,9V$, $m_{cb} = 0,49$, $C_{jcb0} = 2,23fF$, $\phi_{B_{cb}} = 0,9V$, $m_{cb} = 0,29$, $\tau_F = 5ns$. Determine el modelo de pequeña señal y dé una aproximación al valor de f_T .

La capacidad de vaciamiento de la juntura BE se puede aproximar como:

$$C_{veb} = \sqrt{2} \times 11fF = 15,5fF \quad (6.215)$$

En el peor caso, la capacidad de vaciamiento de la juntura BC será:

$$C_{vcb} = 2,23fF \quad (6.216)$$

La capacidad de carga de la base es:

$$C_B = \frac{1mA}{25mV} \times 5ns = 200pF \quad (6.217)$$

Las capacidades resultantes del modelo son:

$$C_\pi \triangleq 15,5fF + 200pF \approx 200pF \quad (6.218)$$

$$C_\mu \triangleq 2,23fF \quad (6.219)$$

El modelo circuital se muestra en la Fig. 6.36.

De acuerdo a (6.214), una aproximación a la frecuencia de transición del transistor puede hallarse de la siguiente manera:

$$f_T = \frac{1}{2\pi \left(5ns + \frac{25mV}{1mA} (15,5fF + 2,23fF) \right)} = 31MHz \quad (6.220)$$

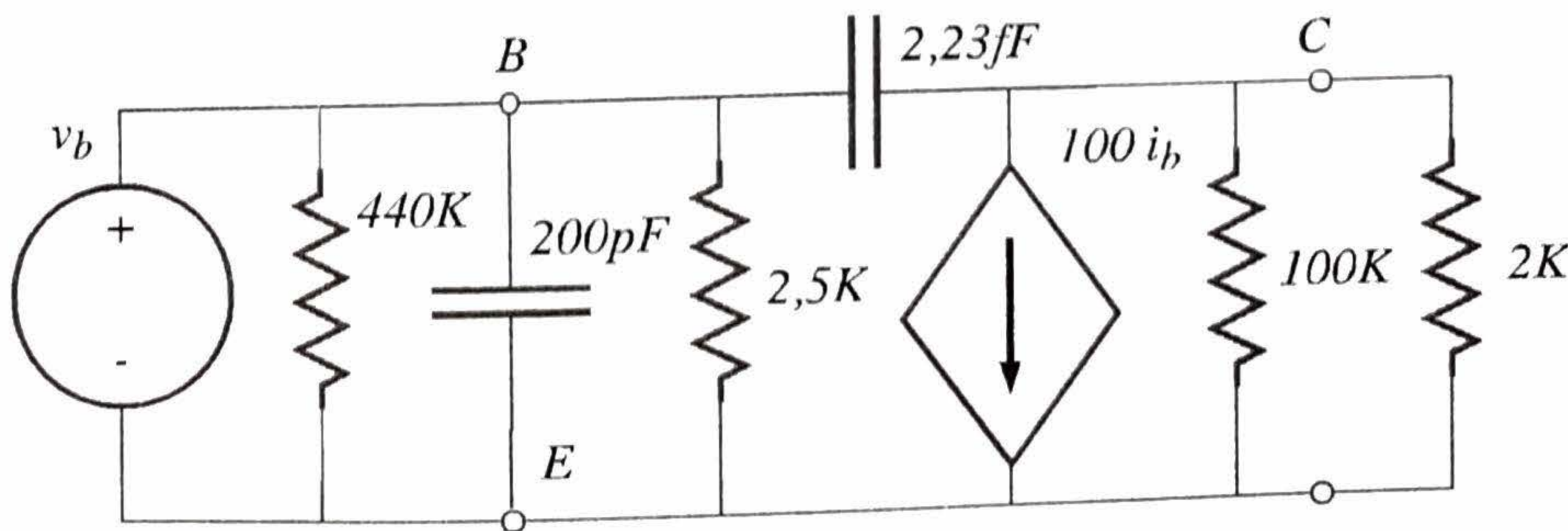


Figura 6.36: Modelo de pequeña señal de AC del circuito del Ejemplo 29.

⁸Parámetros correspondientes a un transistor bipolar en un proceso CMOS de $0,5\mu m$. Este proceso optimiza el desempeño de los transistores MOS; los transistores bipolares son un sub-producto, por lo cual su desempeño no está optimizado.