

INTRODUCCIÓN A LOS TRANSISTORES

Dr. Ing. Eduardo A. Romero

EL TRANSISTOR BIPOLAR

Los transistores bipolares se construyen con una fina capa de material semiconductor de tipo P entre dos capas de material tipo N. A esta estructura se le denomina N-P-N. Si bien existen diferentes geometrías posibles para la implementación de transistores bipolares, el estudio de los mismos puede hacerse mediante una estructura simplificada, como la que puede verse en la Figura 1. A cada una de las regiones semiconductoras utilizadas se les conecta un conductor para que sea posible su conexión a otros componentes de circuito como componentes activos, resistores, capacitores, inductores, etc.

Se denominan a cada uno de los conductores (también denominados terminales) y a las regiones semiconductoras a las que están conectados como colector, base y emisor.

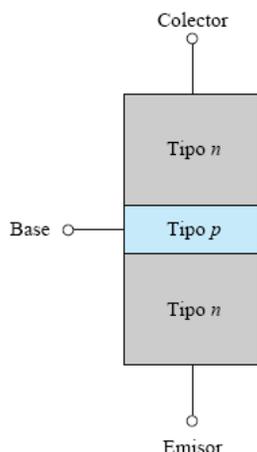


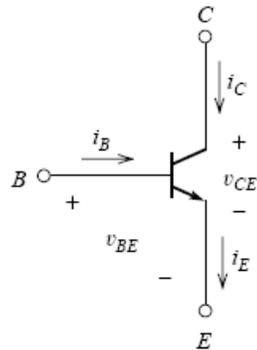
Figura 1. Estructura física simplificada del transistor bipolar

En este punto es conveniente remarcar que esta estructura no se obtiene de materiales independientes que son unidos mediante algún procedimiento. Por el contrario, la estructura observada es el resultado de un proceso metalúrgico en el que se cambia el dopado del silicio empleado como base para la construcción del dispositivo. El procedimiento seguido para la construcción es complejo, ya que son necesarios una serie de pasos físicos y químicos de gran precisión. El estudio de estos procesos se encuentra más allá de lo que se pretende en el presente curso.

Debido a que la utilización de la estructura de la Figura 1 en circuitos complejos resulta sumamente engorrosa, se hace necesaria la adopción de un símbolo de circuito (tal como los usados en resistencias, bobinas o condensadores), que permita el trazado de circuitos de manera sencilla. Por este motivo se adopta el símbolo de circuito para el transistor bipolar que se muestra en la Figura 2. En esta figura pueden verse también las corrientes y tensiones que se definen para el transistor:

i_c : corriente de colector
 i_b : corriente de base
 i_e : corriente de emisor

v_{be} : diferencia de potencial (tensión) existente entre la base y el emisor
 v_{ce} : diferencia de potencial (tensión) existente entre el colector y el emisor



(b) Símbolo esquemático

Figura 2. Símbolo simplificado del transistor bipolar

La operación del transistor NPN puede explicarse considerando al mismo como dos junturas PN yuxtapuestas. Consideremos la juntura base-emisor, suponiendo por ahora que el colector se encuentra a circuito abierto. Supongamos también que existe una fuente de polarización de continua (una batería) que polariza en forma directa la juntura semiconductor, tal como se observa en la Figura 3. Bajo estas condiciones, el comportamiento de la juntura BE es similar a la de un diodo. Consecuentemente, circularán las corrientes de electrones y huecos que se muestran en la figura. Debido a que el dopado del emisor es mucho mayor que el de la base, la corriente de electrones es mucho mayor que la de huecos. La coexistencia de estas dos corrientes hace que se encuentren presentes en la región de base pares electrón-huecos, los cuales tienden a recombinarse. Los portadores remanentes darán una corriente neta desde la base al emisor.

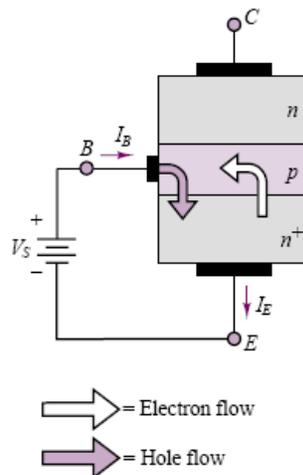


Figura 3. Comportamiento de la juntura BE con el colector abierto.

Consideremos ahora que se agrega otra fuente que polariza la juntura colector-base en forma inversa, tal como puede observarse en la Figura 4. Esta juntura también se comporta como un diodo, pero inversamente polarizado.

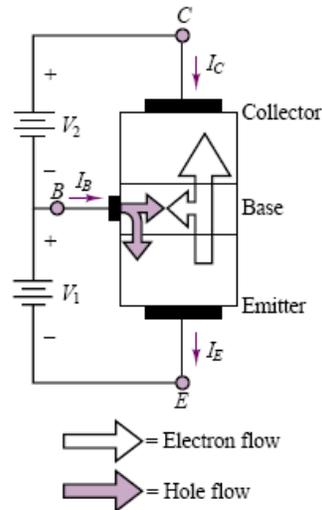


Figura 4. Efecto de la polarización de la juntura colector-base

Bajo estas condiciones, un fenómeno interesante tiene lugar: los electrones emitidos por el emisor, bajo el efecto de la polarización directa de la juntura base-emisor, alcanzan la estrecha región de base. En esta región (debido a las dimensiones del dispositivo) existe una baja probabilidad de recombinación, razón por la cual la mayoría de los electrones emitidos alcanzan las cercanías de la juntura colector-base. El fuerte campo eléctrico presente en esta región hace que los portadores sean barridos hacia el colector y colectados en el terminal correspondiente. Tal como veremos más adelante, la corriente circulante entre el colector y el emisor es mucho más grande que la corriente circulante entre la base y el emisor.

ECUACIONES DEL DISPOSITIVO

Debido a que la juntura base emisor se comporta como un diodo polarizado en sentido directo, es posible utilizar la ecuación del diodo para modelar su comportamiento:

$$i_E = I_{ES} \left[\exp\left(\frac{v_{BE}}{V_T}\right) - 1 \right]$$

Aplicando la ley de Kirchoff, se tiene que:

$$i_E = i_C + i_B$$

Se define el parámetro α como la relación entre la corriente de colector y la de emisor

$$\alpha = \frac{i_C}{i_E}$$

Este parámetro es prácticamente igual a la unidad debido a la forma en la que se fabrican los transistores. Valores típicos para este parámetro están en el rango de 0,99 a 0,999.

Utilizando la primera expresión se puede obtener:

$$i_C = \alpha I_{ES} \left[\exp\left(\frac{v_{BE}}{V_T}\right) - 1 \right]$$

Definiendo la corriente de escala como:

$$I_s = \alpha I_{ES}$$

y considerando que trabajamos con tensiones V_{be} muy superiores a V_t (lo cual permite despreciar el 1 del corchete), se obtiene:

$$i_C \cong I_s \exp\left(\frac{v_{BE}}{V_T}\right)$$

Considerando las ecuaciones 2 y 3 se obtiene:

$$i_B = (1 - \alpha)i_E$$

Reemplazando i_E por su expresión:

$$i_B = (1 - \alpha)I_{ES} \left[\exp\left(\frac{v_{BE}}{V_T}\right) - 1 \right]$$

Reescribiendo la expresión de i_E :

$$i_C = \alpha I_{ES} \left[\exp\left(\frac{v_{BE}}{V_T}\right) - 1 \right]$$

Relacionando las últimas dos expresiones se obtiene:

$$\beta = \frac{i_C}{i_B} = \frac{\alpha}{1 - \alpha}$$

Dados los valores de α usuales, se deduce que $\beta \gg 1$ o en otros términos, que la corriente de colector es mucho más grande que la de la base. Adicionalmente, note que ambas corrientes están relacionadas por una constante. Esta es la relación básica sobre la que se apoya el fenómeno de amplificación en transistores bipolares: la corriente en el colector es β veces más grande que la corriente de base. Por lo tanto, una corriente pequeña en la base puede controlar una corriente grande entre el colector y el emisor.

CURVAS CARACTERÍSTICAS DEL TRANSISTOR BIPOLAR

La caracterización de dispositivos de dos terminales (como el caso del diodo o resistencias) se realiza haciendo un barrido de tensiones entre sus terminales y midiendo las corrientes que circulan. En el caso de componentes de tres terminales, esta caracterización se hace de a pares de terminales. Para el transistor bipolar son importantes dos curvas: las de la juntura base emisor y las de la juntura colector emisor. El resultado de la caracterización experimental de un dado transistor puede verse en la Figura 5.

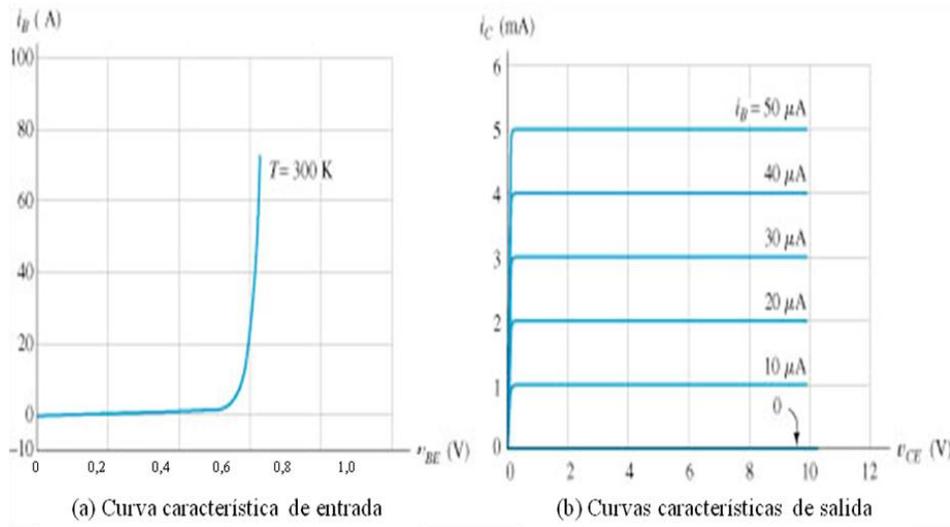


Figura 5. Curvas características del transistor bipolar

EL FENÓMENO DE AMPLIFICACIÓN EN EL TRANSISTOR BIPOLAR

Consideremos el circuito de la Figura 6, en la que se presenta un transistor bipolar con una fuente de continua que polariza la juntura emisor colector y otra haciendo lo propio en la juntura base emisor. Superpuesta a esta última, se conecta una fuente de alterna. En la figura también se presentan las ecuaciones de las dos mallas del circuito.

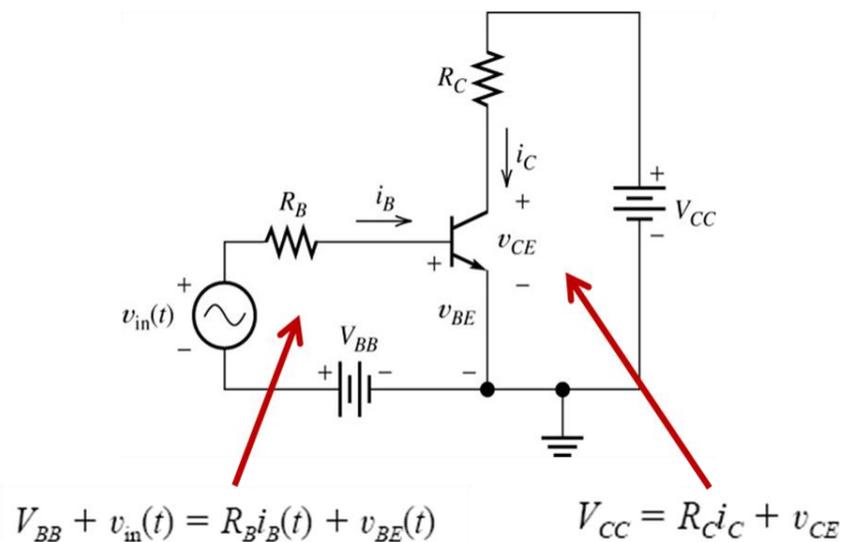


Figura 6. Circuito típico y sus ecuaciones.

Es usual emplear una combinación gráfica analítica para el análisis del circuito, lo que evita el tratamiento de sistemas de ecuaciones no lineales relativamente tediosas. En la malla base emisor se tiene que cumplir la ley de suma de tensiones pero al mismo tiempo debe cumplirse la característica corriente tensión de esa malla (Figura 5). Una situación idéntica se da en la malla colector emisor. Si se grafican sobre los mismos pares de ejes las ecuaciones de malla y las características se obtienen las gráficas de la Figura 7.

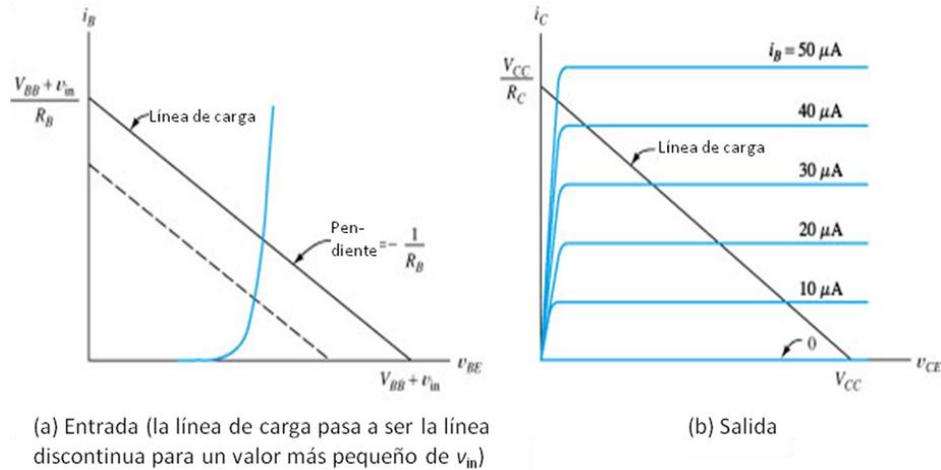


Figura 7. Rectas de carga y características de las junturas.

Si se supone que $V_{in}=0$, se tiene una recta como la mostrada en líneas de trazos en la Figura 7-a. Si V_{in} tiene algún valor positivo, se obtiene una recta como la de trazo continuo en la misma figura. Cuando la señal de entrada varía de manera continua, se da en la malla base-emisor el fenómeno ilustrado en la Figura 8. La señal sinusoidal, desplaza la recta de carga en forma dinámica, provocando variaciones en la corriente de base del transistor. En el momento que $V_{in}=0$, se dice que se tienen condiciones de no señal, también denominadas de reposo. En esta situación se tiene una corriente circulando por la base que se obtiene del punto de intersección de la recta y la característica base-emisor del transistor (señalado como punto Q). Cuando la señal de entrada adopta su valor máximo, la recta se desplaza paralelamente, haciendo que la corriente circulando por la base aumente (el punto de intersección se desplaza hacia arriba). Una situación inversa se da cuando la señal adopta su valor mínimo. Debido a que la señal de entrada es una función continua del tiempo, adopta infinitos valores entre los señalados en la figura. Por este motivo hay infinitas rectas de carga, lo que hace que la corriente de base varíe también de manera continua con el tiempo.

Bajo las condiciones anteriores, y dado que la corriente en el colector es β veces la corriente de base, se produce un cambio continuo de curva paramétrica de salida. Esta situación se ilustra en la Figura 9. Dado que la variación de la corriente de base es continua y que existen también infinitas curvas paramétricas, la variación de la corriente de colector es también continua. Esto hace que el punto de intersección de la recta de carga de salida con la característica colector emisor del transistor varíe de manera continua. Esto hace que la tensión colector-emisor varíe en el rango indicado en la figura. Como puede verse de relacionar las figuras, la tensión de salida tiene un valor de pico de 2V, mientras que la entrada sólo de 0,4V, de lo cual se deduce que existe amplificación de tensión entre la señal entre C y E y la de entrada.

señal de entrada

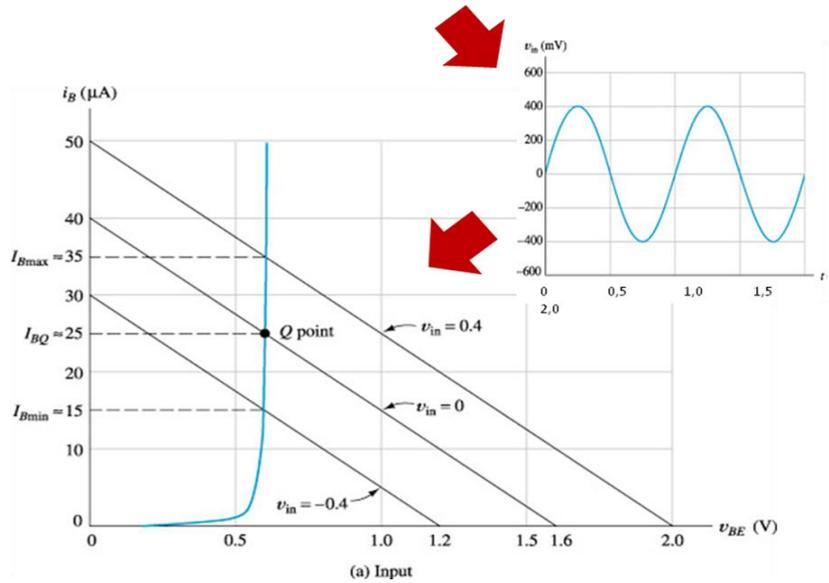


Figura 8. Comportamiento de la malla de entrada bajo excitación sinusoidal.

Las variaciones de la corriente de base provocan

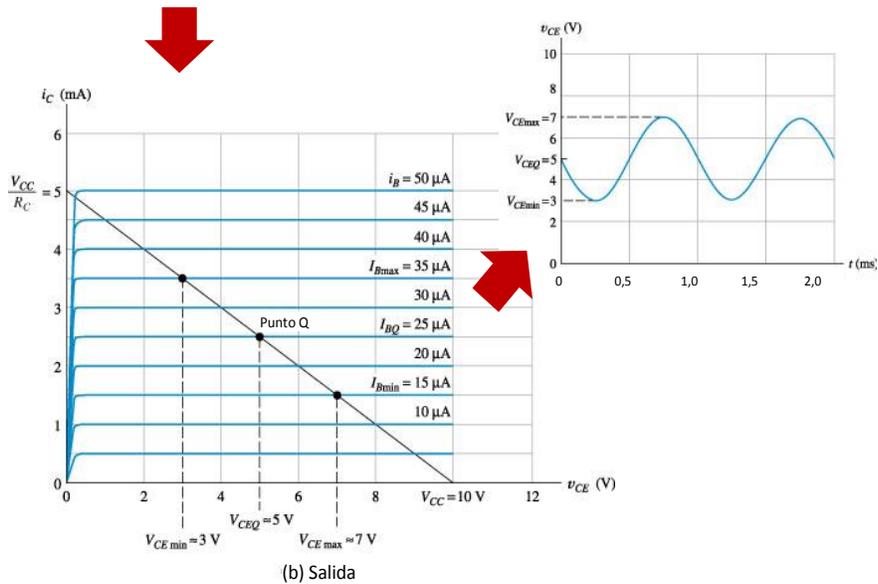


Figura 9. Comportamiento de la malla de salida bajo excitación sinusoidal.

EL TRANSISTOR DE EFECTO DE CAMPO MOS

La estructura del transistor de efecto de campo de compuerta aislada (también denominado MOS, Metal-Óxido-Semiconductor) puede verse en la Figura 10. El transistor se construye sobre un sus-

trato de silicio tipo p, sobre el cual se implementan dos regiones de tipo n, denominadas drenador y surtidor. Sobre el sustrato y levemente solapada con las regiones de drenador y surtidor, se hace crecer una capa de óxido que se comporta como un aislador casi perfecto. Sobre esta capa se practica un depósito metálico (aluminio) al que se denomina compuerta. Se conectan a las regiones de fuente, compuerta, drenador y sustrato conductores que se denominan terminales que son los que se conectan al circuito externo.

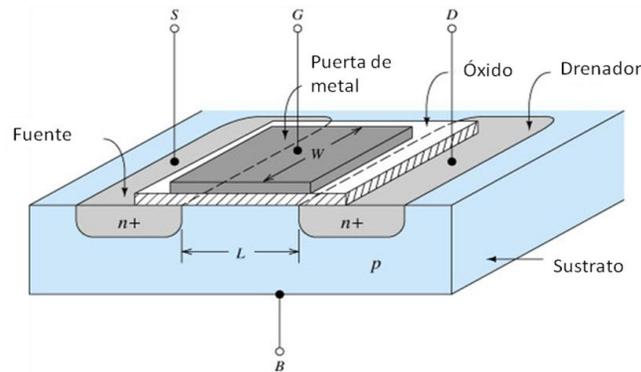


Figura 10. Estructura física del transistor MOS

El símbolo esquemático utilizado para representar el transistor en circuitos complejos se muestra en la Figura 11.

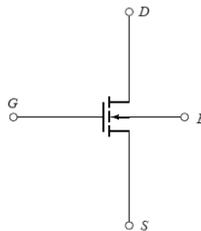


Figura 11. Símbolo esquemático del transistor MOS

Para comprender el funcionamiento del transistor, supongamos el circuito de la

Figura 12. En este esquema, una fuente de continua denominada V_{gs} establece la diferencia de potencial entre los terminales G y S. Se hace notar que los terminales S y B se encuentran unidos, una situación usual en este tipo de transistores. Por otro lado, otra fuente de tensión de continua establece la diferencia de tensión entre los terminales D y S. En el estudio que sigue se supondrá que estas fuentes pueden ser variadas (fuentes variables de tensión continua).

Supongamos para comenzar nuestro estudio que la fuente V_{gs} es igual a 0V y V_{ds} tiene algún valor positivo. Bajo estas condiciones, y con las polaridades mostradas en la figura, las junturas existentes entre D y B y entre B y S se encuentran polarizadas en sentido inverso. Por lo tanto estas regiones se comportan como dos diodos polarizados en sentido inverso, lo que hace imposible la circulación de corriente entre D y S. Es por este motivo que, bajo las condiciones anteriores, se dice que el transistor se encuentra en estado de corte.

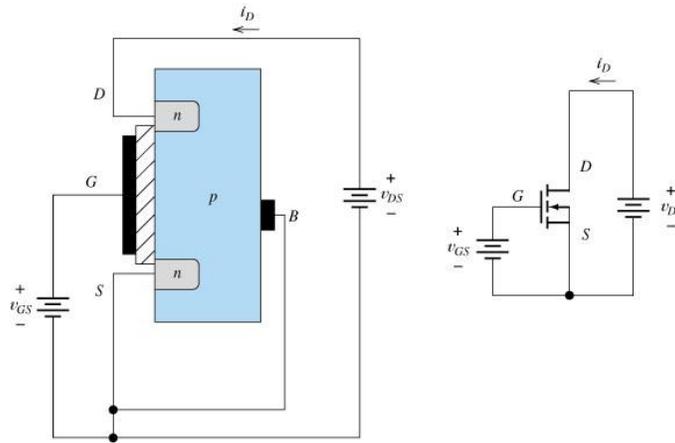


Figura 12. Funcionamiento en la región de corte

Si se procede a incrementar el valor de V_{gs} , los portadores minoritarios de la región P (electrones) son atraídos por el campo eléctrico impuesto por la fuente. Recordemos que esta fuente establece la diferencia de potencial entre G y B. Bajo el influjo del campo eléctrico, los electrones tienden a acumularse debajo del óxido. Al mismo tiempo, los portadores mayoritarios son repelidos al extremo opuesto. La cantidad de electrones acumulados aumenta mientras mayor sea el campo eléctrico o, equivalentemente, la diferencia de potencial impuesta por V_{gs} . Cuando se alcanza una diferencia de potencial dada, la cantidad de portadores acumulados hace la región inmediatamente por debajo de la compuerta deje de comportarse como material P y pase a comportarse como material tipo N. Bajo estas condiciones se dice que aparece un canal inducido. Este cambio se da solo localmente y bajo la acción de la fuente V_{gs} . Si desaparece la polarización, desaparece el canal de tipo N. La situación descrita se muestra en la Figura 13. Adicionalmente, debe señalarse que es necesario que se supere un determinado valor de tensión V_{gs} para que se induzca el canal. A esta tensión se le denomina tensión de umbral (V_t) y depende de las características constructivas del transistor.

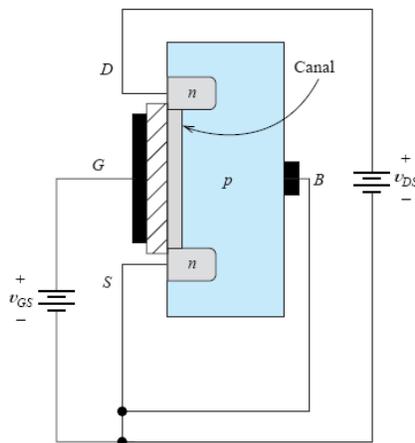


Figura 13. Funcionamiento en la región resistiva.

Si bajo las condiciones anteriores se deja fija la fuente V_{ds} y se varía la tensión V_{gs} , la cantidad de portadores en el canal variará en consecuencia. Por lo tanto, la resistividad de este canal disminuye a medida que aumenta esta tensión. Si se suponen ahora variaciones pequeñas en la fuente V_{ds} (en

un entorno del origen) es posible graficar la relación entre la corriente de drenador y V_{ds} . Esto puede verse en la Figura 14. Variaciones en la fuente V_{gs} cambian la pendiente de la gráfica, lo cual significa un cambio en la resistencia del canal. Es por este motivo que se dice que en esta región el transistor se comporta como un resistor controlado por tensión.

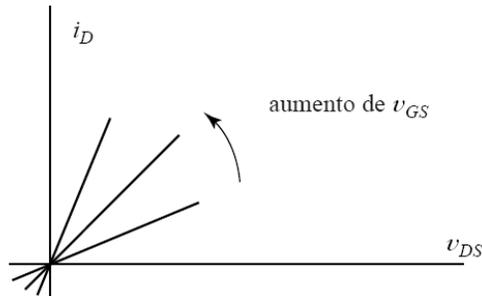


Figura 14. Comportamiento resistivo.

Veamos que sucede si V_{gs} se mantiene constante (pero superior al V_t) y se aumenta progresivamente V_{ds} . El flujo de corriente por el canal del transistor produce que las tensiones de los puntos del canal respecto del surtidor sean progresivamente mayores a medida que nos movemos hacia el drenador. En consecuencia, las tensiones entre la compuerta y regiones del canal cercanas al drenador son menores que las existentes entre la primera y regiones cercanas al surtidor. Es por esto que el campo eléctrico que sostiene el canal inducido es menor en cercanías del drenador. Esto provoca que el canal sea más delgado en esta región, lo que provoca un aumento de la resistividad. Este fenómeno es más marcado mientras más grande es V_{ds} . Esta situación puede observarse en la Figura 15.

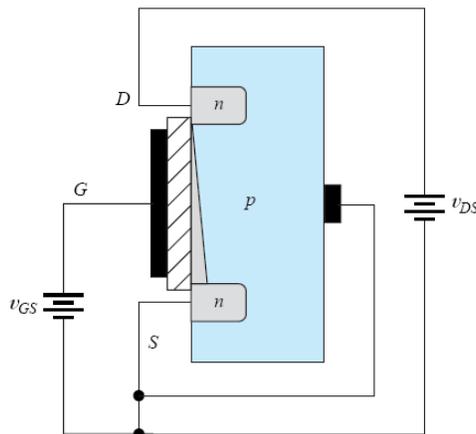


Figura 15. Estrangulamiento del canal.

Si se aumenta aún más la tensión V_{ds} no se obtienen incrementos en la corriente de drenador, tal como puede observarse en la Figura 16. A la región de corriente constante se la denomina región de saturación.

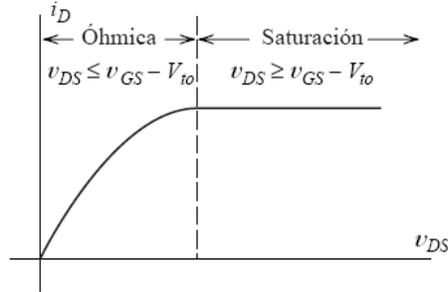


Figura 16. Corriente de drenador vs. Tensión Vds para Vgs constante.

La curva anterior ha sido obtenida fijando una tensión Vgs y realizando un barrido de tensión Vds. Si ahora se repite el experimento para diferentes valores de tensión Vgs, se obtiene la familia de curvas paramétricas de la Figura 17.

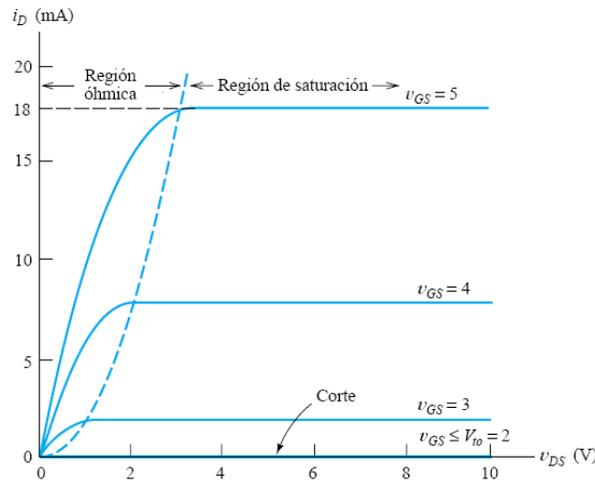


Figura 17. Curvas características de salida.

EL FENÓMENO DE AMPLIFICACIÓN EN EL TRANSISTOR MOS

Consideremos el circuito de la Figura 18Figura 6, en la que se presenta un MOS con una fuente de continua que polariza la juntura drenador-surtidor y otra haciendo lo propio en la malla compuerta-surtidor. Superpuesta a esta última, se conecta una fuente de alterna. Se aplicará aquí el mismo procedimiento de análisis utilizado previamente para el transistor bipolar. Las ecuaciones de malla pueden plantearse como:

$$v_{GS}(t) = v_{in}(t) + V_{GG}$$

$$V_{DD} = R_D i_D(t) + v_{DS}(t)$$

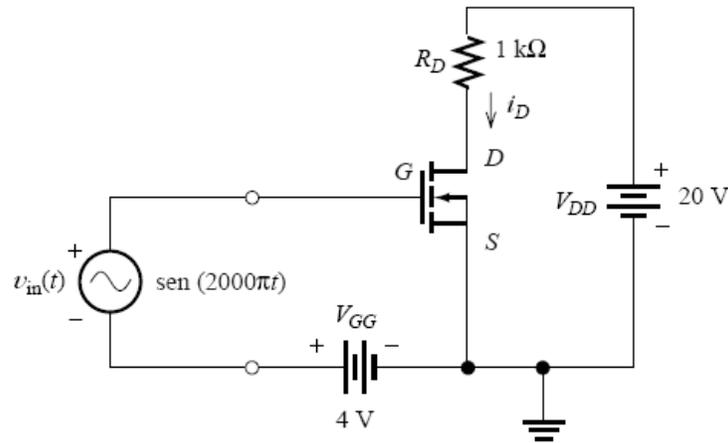


Figura 18. Amplificador MOS

A título de ejemplo, consideraremos que la señal de entrada es sinusoidal con 1V de pico y frecuencia de 1KHz. Los demás parámetros circuitales están indicados en la figura.

Como puede verse, la tensión de entrada es la superposición de una fuente de continua y una de alterna. Si consideramos solamente la fuente de continua, y debido a que no hay circulación de corriente hacia la compuerta del transistor MOS (debido a la capa de óxido debajo de la compuerta), se tiene que la tensión V_{gs} es 4V. Sabiendo esto, puede buscarse la curva paramétrica correspondiente en las características de drenador de la Figura 19. La intersección de la recta de carga (dada por la suma de tensiones en la malla de salida) con la curva paramétrica nos da el punto de reposo (corriente y tensión bajo condiciones de continua) del transistor.

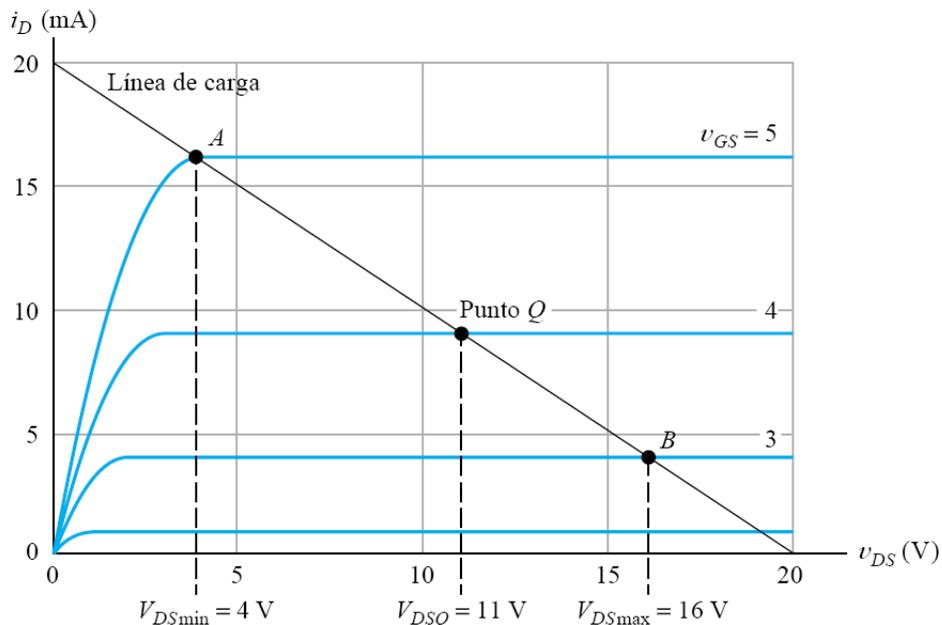


Figura 19. Características del drenador y recta de carga.

Por otro lado, la presencia de la fuente de señal hará que la recta de carga se desplace paralelamente a sí misma hacia arriba y hacia abajo, cambiando el punto de trabajo del transistor (debido a que al desplazarse corta a otra curva paramétrica). El comportamiento es totalmente análogo al del transistor bipolar ya explicado con anterioridad. Para los valores de los parámetros del circuito, la fuente de señal hace que el punto de reposo se desplace instantáneamente 1V hacia abajo y hacia arriba, determinando los puntos B y A respectivamente. Como puede observarse en la Figura 19, esto hace que la tensiones mínima y máxima del drenador sean respectivamente 4V y 16 V. La tensión de alterna del drenador es la que se muestra en la Figura 20. Esta señal tiene una amplitud pico a pico de 12V, mientras que la entrada es de 2V (también pico a pico). Consecuentemente, la señal de salida es 6 veces más grande que la de entrada, lo cual expone las características de amplificación del circuito tomado como ejemplo.

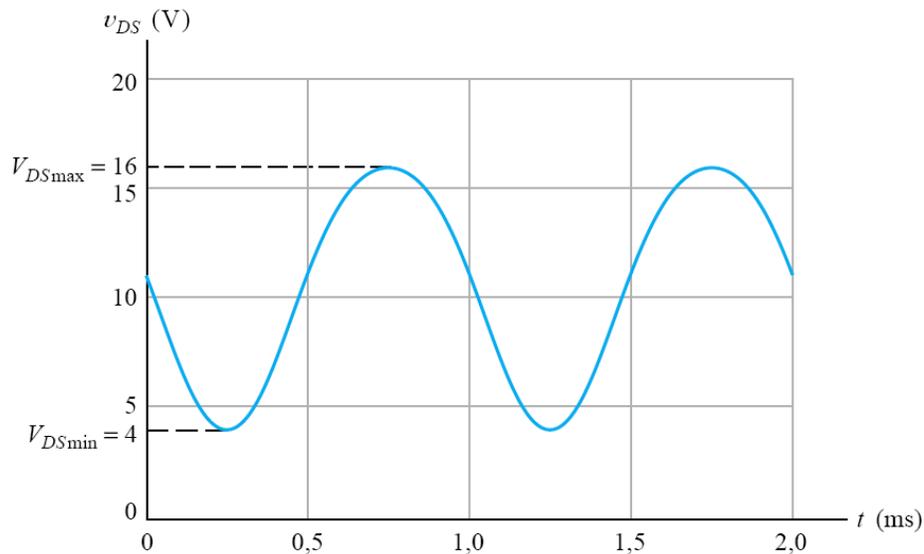


Figura 20. Forma de onda de salida.