

Profesor: Dr.-Ing. Paola Vega Castillo

Capítulo 5. Aplicaciones del MOSFET

5.1 El MOSFET como interruptor

En la lógica digital se representan solamente dos estados, correspondientes al 1 y 0 en el sistema binario. El nivel uno corresponde a una tensión de V_{DD} , mientras que el nivel cero corresponde a una tensión de 0 V.

Una de las aplicaciones básicas del MOSFET es la implementación de circuitos digitales, en los cuales una vez transcurrido el transiente del circuito, puede modelarse el transistor como un interruptor abierto o un interruptor cerrado, según la tensión aplicada a su compuerta. En el caso del interruptor abierto, el transistor opera en la región de subumbral, por lo cual su comportamiento es bastante cercano al ideal. En el caso del interruptor cerrado, deben analizarse dos casos: la transmisión de un 1 lógico y la de un 0 lógico.

Consideremos primeramente la transmisión de un 0 lógico. Iniciemos con el caso de un transistor NMOS, con ayuda del circuito de la figura 5.1, el cual tiene un capacitor de carga C_L conectado a la salida. El capacitor de carga incluye la capacitancia parásita de las interconexiones, las capacitancias del transistor MOSFET y las capacitancias de cualquier otro circuito que esté conectado a terminal de salida (no mostrado en la figura 5.1). Aplicaremos una tensión de V_{DD} (1 lógico) a la compuerta de transistor con el fin de formar el canal. Asuma que la tensión de entrada pasa de V_{DD} a 0 en tiempo cero, y que el capacitor inicialmente se encuentra cargado a V_{DD} . Ahora se desea fijar un 0 lógico a la salida del circuito, es decir, debemos descargar el capacitor C_L .

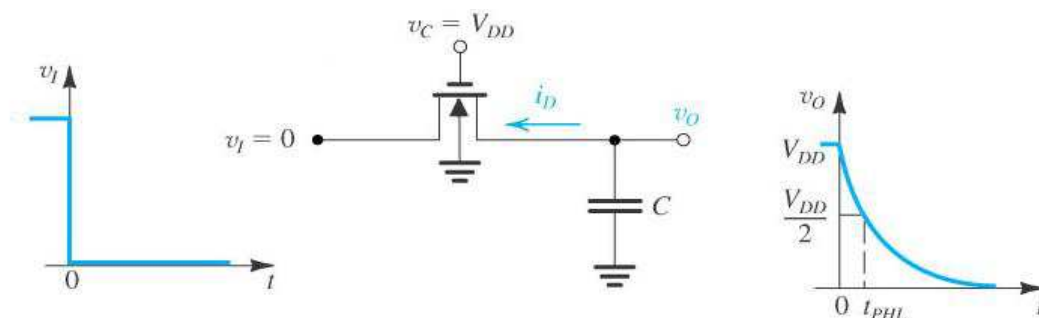


Figura 5.1 Transmisión de un cero lógico utilizando un transistor NMOS

Debido a la tensión V_{DD} aplicada a la compuerta del transistor, en $t < 0$ la corriente en el circuito es cero, puesto que el capacitor estaba cargado a V_{DD} y además $V_{in} = V_{DD}$, por lo que el transistor opera en la región de corte con $V_{GS} =$

0. En $t > 0$, puede ahora distinguirse entre drenador y surtidor, debido a que la tensión aplicada a cada terminal es distinta.

De acuerdo con la Figura 5.1, $V_{GS}=V_{DD}=V_{DS}$ y al inicio de la descarga el transistor opera en saturación. Ahora, puesto que V_G es constante y V_S es constante, el capacitor puede descargarse completamente y al final de la descarga el transistor opera en la región lineal, es decir, el canal está formado, sin embargo, con $V_C=0$, $V_{DS}=0$ y no fluye corriente en el transistor, dado que el capacitor ya no entrega carga. De esta forma, el transistor NMOS fija de forma óptima un cero lógico a la salida del circuito.

Considere ahora el mismo caso, pero sustituyendo el transistor NMOS por un transistor PMOS, como se muestra en la Figura 5.2, en el cual se aplica una tensión de 0V en la compuerta del transistor.

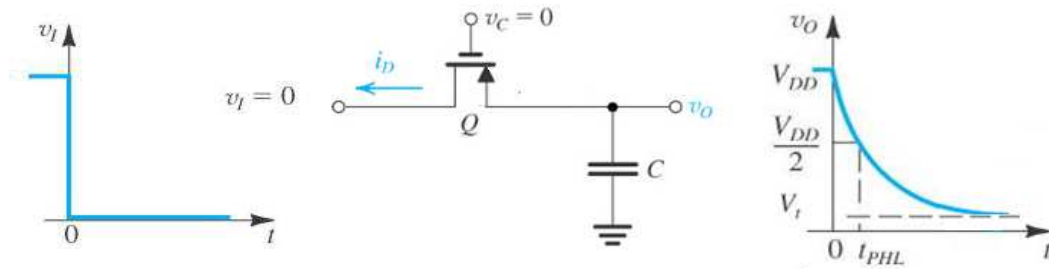


Figura 5.2 Transmisión de un cero lógico utilizando un transistor PMOS

En $t < 0$, con $V_{in}=V_{DD}$ aún no puede distinguirse entre drenador y surtidor del transistor. Sin embargo, puede verse que $V_{SG}=V_{DD}$ y el transistor opera en la región lineal. No fluye corriente en el transistor debido a que $V_{SD}=0$. En $t=0^+$, dado que $V_{in}=0$, puede distinguirse entre drenador y surtidor, con $V_{SG}=V_{DD}$ y $V_{SD}=V_{DD}$.

Observamos también que $V_S=V_C(t)$ y por lo tanto $V_{SG}=V_C(t)-0$. Así cuando el capacitor se descarga hasta alcanzar $V_C=V_{TH}$, el transistor está aún encendido y en el umbral de conducción, pero si la tensión disminuye un poco más, el transistor se apaga. Por ello el transistor PMOS no transmite (fija) de manera óptima un 0 lógico a la salida, dado que la tensión más baja a la que puede descargar el capacitor es $V_{C,min}=V_{TH}$.

Una ventaja adicional del transistor NMOS para la transmisión de un cero lógico es el hecho de que durante dicha operación no sufre de efecto de cuerpo, a diferencia del transistor PMOS.

Veamos ahora la transmisión de un 1 lógico, es decir, deseamos fijar un 1 lógico a la salida del circuito, y para eso debemos cargar el capacitor C_L . Iniciemos con el caso de un transistor NMOS, con ayuda del circuito de la Figura 5.3. La tensión inicial en el capacitor de carga es 0V. Se aplica una tensión de V_{DD} (1 lógico) a la compuerta de transistor con el fin de formar el canal. Asuma que la tensión de entrada pasa de 0 a V_{DD} en tiempo cero.

Debido a la tensión aplicada a la compuerta del transistor, en $t < 0$ la corriente en el circuito es cero, puesto que el capacitor estaba cargado a 0 y además $V_{in} = 0$, por lo que el transistor opera en la región de lineal; en este caso el canal está formado pero no fluye corriente porque $V_{DS} = 0$. En $t > 0$, puede ahora distinguirse entre drenador y surtidor.

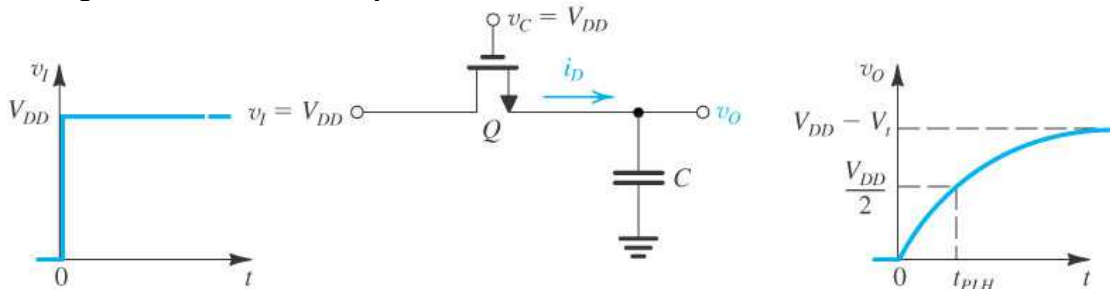


Figura 5.3 Transmisión de un uno lógico utilizando un transistor NMOS

De acuerdo con la Figura 5.3, $V_{GS} = V_{DD} = V_{DS}$ y al inicio de la carga el transistor opera en saturación. Ahora, puesto que V_G es constante y $V_S = V_C(t)$, el capacitor se empieza a cargar. Sin embargo, cuando el capacitor alcanza $V_C(t) = V_{DD} - V_{TH}$, la tensión V_{GS} correspondiente es V_{TH} . Si la tensión del capacitor aumenta un poco más, el transistor NMOS entra a operar en la región de corte. De esta forma, el transistor NMOS no puede fijar de forma óptima un uno lógico a la salida del circuito, puesto que la tensión máxima a la que puede cargar el capacitor es $V_{DD} - V_{TH}$.

Considere ahora el mismo caso, pero sustituyendo el transistor NMOS por un transistor PMOS, como se muestra en la Figura 5.4, en el cual se aplica una tensión de 0V en la compuerta del transistor.

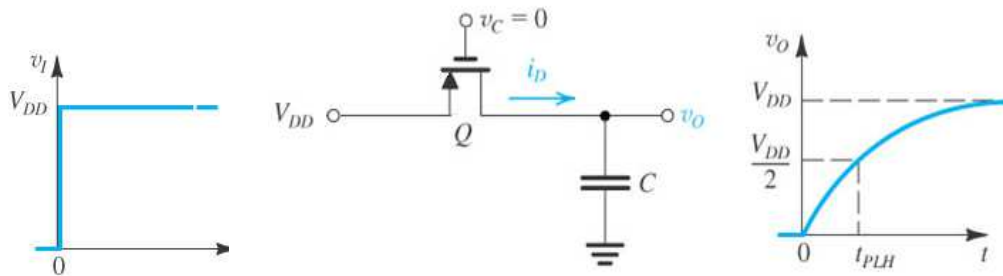


Figura 5.4 Transmisión de un uno lógico utilizando un transistor PMOS

En $t < 0$, con $V_{in} = 0$, $V_{SG} = 0$ y el transistor opera en la región de corte. En $t = 0^+$, dado que $V_{in} = V_{DD}$, puede distinguirse entre drenador y surtidor. Aquí $V_{SG} = V_{DD}$ y $V_{SD} = V_{DD}$ y el transistor opera en la región de saturación.

Observamos también que $V_S = V_{DD}$ y por lo tanto V_{SG} es constante, de forma que el transistor puede permanecer encendido y puede completarse la carga del capacitor hasta el valor de V_{DD} . Por ello el transistor PMOS transmite (fija) de manera óptima un 1 lógico a la salida.

Una ventaja adicional del transistor PMOS para la transmisión de un uno lógico es que, a diferencia del transistor NMOS, no sufre de efecto de cuerpo durante esta operación.

5.2 Compuertas lógicas básicas

5.2.1 El inversor CMOS

El inversor CMOS es la compuerta lógica básica más sencilla que puede implementarse. La Figura 5.5 muestra el esquema de un inversor CMOS, es decir, un inversor implementado con transistores MOSFET complementarios, ya que consta de un transistor PMOS y un transistor NMOS. Observe que debido al comportamiento de los transistores PMOS y NMOS como interruptores, el transistor PMOS se conecta entre la V_{DD} y la salida para transmitir un 1 lógico a la salida, mientras que el transistor NMOS se conecta entre la salida y tierra para transmitir un 0 lógico a la salida.

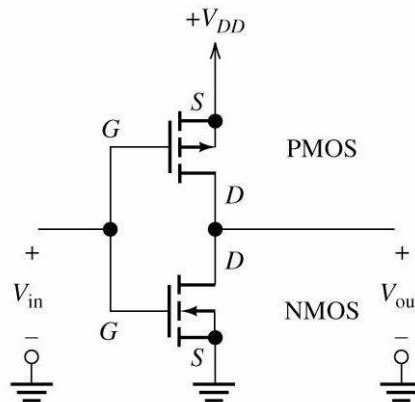


Figura 5.5. Inversor CMOS

La tabla de verdad de un circuito lógico muestra el comportamiento del circuito, es decir, sus valores de salida, ante las diferentes combinaciones de señales de entrada. Los valores mostrados en las tablas de verdad son binarios. La Tabla 5.1 muestra la tabla de verdad de un inversor.

Tabla 5.1
Tabla de verdad de un inversor lógico

Entrada	Salida
0	1
1	0

Observe que la función de inversor es transmitir a la salida el valor contrario al aplicado a la entrada, es decir, invertir la señal. Desde el punto de vista lógico, esta función se conoce también como **negación**, es decir, NOT.

Para obtener la tabla de verdad de un circuito lógico debe considerarse el valor final de la salida, es decir, el valor en estado estable. Al analizar el circuito debe observarse si la tensión de entrada contribuye a activar o inactivar los transistores a los cuales está conectada. Es importante recordar que, desde el punto de vista de un circuito digital, las señales solamente puede tomar dos valores: 0 lógico (tierra) o 1 lógico (V_{DD}). Además, la tensión de alimentación V_{DD} es mayor que la magnitud de los voltajes de umbral de los transistores.

Con estas consideraciones en mente, procederemos a analizar el comportamiento del inversor CMOS. Para ello, en el estado estable representaremos a un transistor en corte como un interruptor abierto, y a un transistor activo como un interruptor cerrado.

Con $V_{in} = 0$ lógico, el transistor NMOS presenta una tensión V_{GS} de 0V, por lo cual opera en la región de corte, mientras que la tensión V_{SG} del transistor PMOS es $V_S - V_G = V_{DD}$, por lo que el transistor PMOS se activa. De esta forma, en el estado estable, podemos representar al transistor NMOS como un interruptor abierto y al PMOS como un interruptor cerrado. De esta forma, el transistor PMOS conecta la salida V_{out} a V_{DD} , definiendo el valor de la salida como un 1 lógico. Dado que en estado estable la tensión V_{SD} es de 0V, el transistor PMOS mantiene el canal formado pero no fluye corriente entre sus terminales, y el transistor opera en la región lineal.

Si ahora $V_{in} = 1$ lógico, el transistor PMOS presenta una tensión V_{SG} de 0V, por lo cual opera en la región de corte, mientras que la tensión V_{GS} del transistor NMOS es $V_G - V_S = V_{DD}$, por lo que el transistor NMOS se activa. Representamos entonces al transistor NMOS como un interruptor cerrado y al PMOS como un interruptor abierto. De esta forma, el transistor NMOS conecta la salida V_{out} a tierra, definiendo el valor de la salida como un 0 lógico. En estado estable la tensión V_{DS} del transistor NMOS es de 0V, por lo que el transistor mantiene el canal formado pero no fluye corriente entre sus terminales, y el transistor opera en la región lineal.

El comportamiento del inversor CMOS para ambos casos de la tensión de entrada se muestra en la Figura 5.6. El símbolo lógico de un inversor se muestra en la Figura 5.7. El triángulo representa la transmisión de la señal de la entrada a la salida, mientras que el círculo representa inversión.

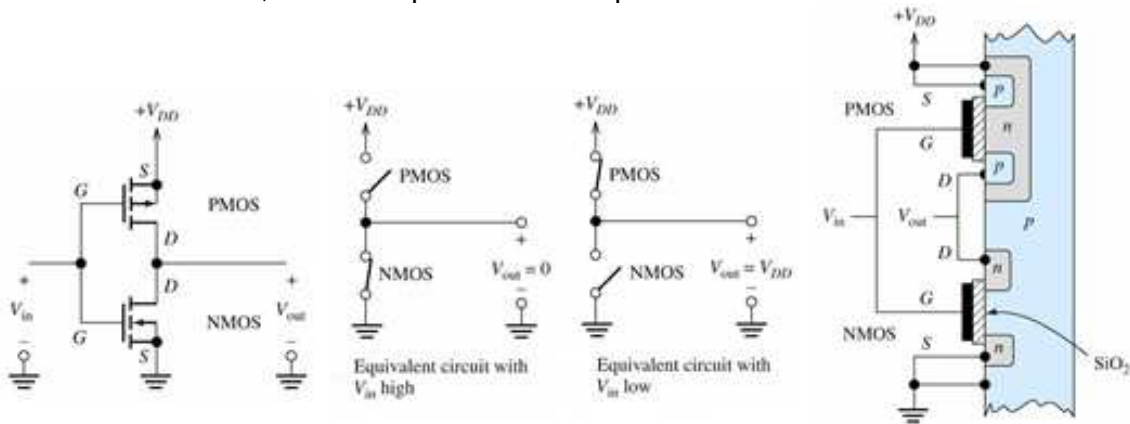


Figura 5.6. Comportamiento del inversor CMOS

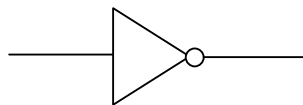


Figura 5.7. Símbolo lógico de un inversor

En un circuito, el comportamiento de la tensión de salida como función de la tensión de entrada se conoce como curva de transferencia (función de

transferencia). La función de transferencia del inversor CMOS se muestra en la Figura 5.8; se observa que existe un rango de tensiones de entrada para las cuales ambos transistores se encuentran activos, y por lo tanto existe un flujo de corriente entre V_{DD} y tierra a través de ambos transistores. Esta corriente se conoce como corriente de corto circuito, pues los transistores forman una red de baja impedancia entre V_{DD} y tierra.

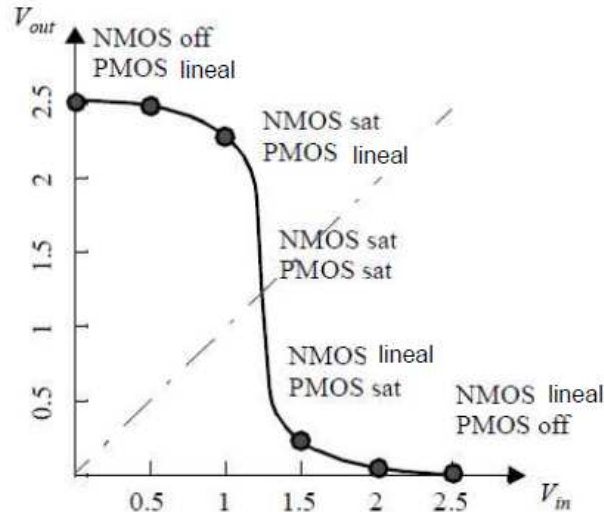


Figura 5.8. Curva de transferencia de un inversor CMOS

En la curva característica existe un valor de tensión de entrada tal que $V_{in}=V_{out}$. Este valor se conoce como punto de disparo V_{SP} (switching point) de una compuerta lógica. Cuando el punto de disparo de un inversor es $V_{DD}/2$, se dice que el inversor es simétrico.

La ecuación del punto de disparo de un inversor CMOS es

$$V_{SP} = \frac{V_{TH,N} + \sqrt{\frac{K_P}{K_N}}(V_{DD} - |V_{TH,P}|)}{1 + \sqrt{\frac{K_P}{K_N}}}$$

Donde $V_{TH,N}$: tensión de umbral del transistor NMOS, $V_{TH,P}$: tensión de umbral del transistor PMOS, K_N : parámetro de transconductancia del transistor NMOS, K_P : parámetro de transconductancia del transistor PMOS.

Como puede observarse de la ecuación del punto de disparo, para que un inversor sea simétrico es necesario que $V_{TH,N} = |V_{TH,P}|$ y $K_P/K_N=1$. Esto implica que la diferencia de movilidades de huecos y electrones debe compensarse ajustando el ancho del transistor PMOS de forma que $K_P = K_N$.

Así, cuando un inversor es simétrico, cumple las siguientes características:

- $V_{TH,N} = |V_{TH,P}|$
- $K_P = K_N$
- $V_{SP} = V_{DD}/2$

- Su tiempo de subida t_{rise} y su tiempo de bajada t_{fall} son iguales

Es importante considerar que aunque analizamos las salidas de los circuitos digitales en estado estable, en realidad las señales de entrada y salida tienen un tiempo de transición diferente de cero. El tiempo en el cual una señal pasa de 0 a 1 lógico se conoce como tiempo de subida t_{rise} , y el tiempo en el cual pasa de 1 a 0 se conoce como tiempo de bajada t_{fall} .

5.2.2 Compuerta NAND

La compuerta NAND CMOS de dos entradas se compone de cuatro transistores: dos PMOS y dos NMOS, como se muestra en la Figura 5.9. Es importante recalcar que, dado que los transistores PMOS se utilizan para transmitir el valor de 1 lógico a la salida, están conectados entre V_{DD} y la salida, mientras que los NMOS deben estar conectados entre la salida y a tierra, porque se utilizan para transmitir el valor de 0 lógico a la salida.

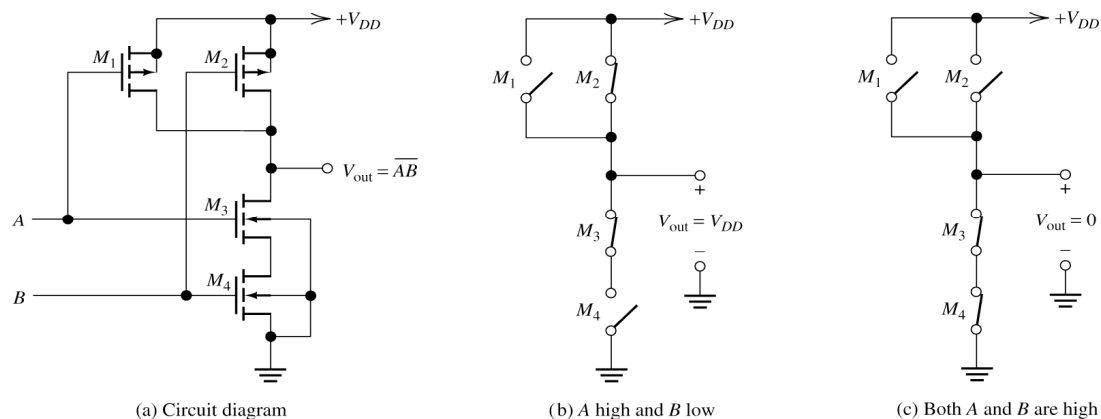


Figura 5.9. Compuerta NAND CMOS de dos entradas

Como se observa en la Figura 5.9, en una compuerta NAND CMOS se necesitan dos transistores por cada entrada, uno NMOS y otro PMOS.

La función lógica NAND corresponde a una función AND negada. Para comprender esto, definamos primero la función lógica AND, cuyo comportamiento se muestra en la Tabla 5.2.

Tabla 5.2
Comportamiento de una compuerta AND

Entrada lógica A	Entrada lógica B	Salida lógica
0	0	0
0	1	0
1	0	0
1	1	1

Como podemos observar en la Tabla 5.2, la salida tendrá un valor de 1 lógico si y solo si **A y B** son 1, de allí el nombre AND. La compuerta NAND es una AND seguida de un inversor (NAND= NOT AND), como se resume en la Tabla 5.3. Las primeras tres columnas de la Tabla 5.3 corresponden a la tabla de verdad de la compuerta NAND, mientras que las últimas dos muestran los valores de tensión para cada fila de la tabla de verdad. Los símbolos lógicos de las compuertas AND y NAND se muestran en la Figura 5.10.

Tabla 5.3
Tabla de verdad de una compuerta NAND

Entrada lógica A	Entrada lógica B	Salida lógica	Voltaje de entrada	Voltaje de salida
0	0	0	$V_A=0, V_B=0$	V_{DD}
0	1	0	$V_A=0, V_B=V_{DD}$	V_{DD}
1	0	0	$V_A=V_{DD}, V_B=0$	V_{DD}
1	1	1	$V_A=V_{DD}, V_B=V_{DD}$	0

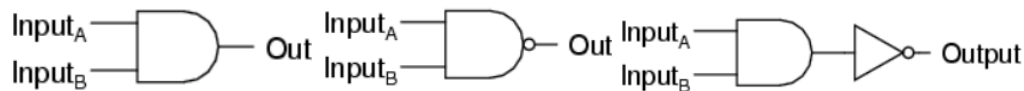


Figura 5.10. Símbolos lógicos de las compuertas AND, NAND y equivalencia de la compuerta NAND

5.2.3 Compuerta NOR

Al igual que la compuerta NAND CMOS, la compuerta NOR CMOS de dos entradas se compone de cuatro transistores, dos por cada entrada: dos PMOS y dos NMOS, como se muestra en la Figura 5.11.

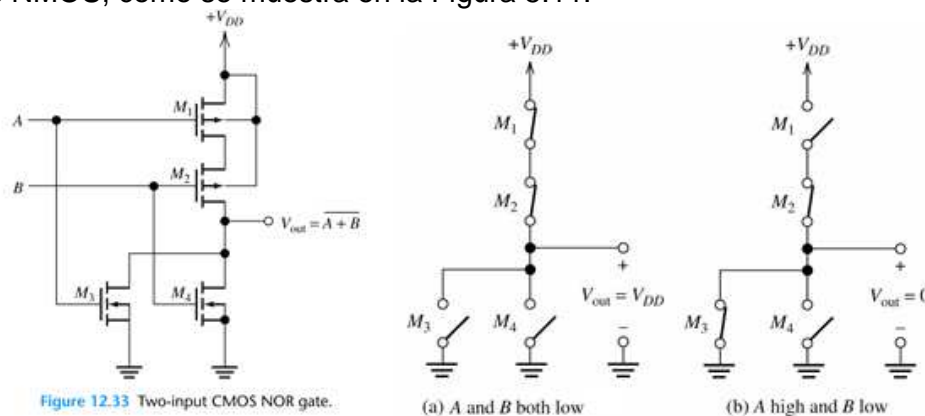


Figura 5.11. Compuerta NOR CMOS de dos entradas

La función lógica NOR corresponde a una función OR negada. Para comprender esto, definamos primero la función lógica OR, cuyo comportamiento se muestra en la Tabla 5.4.

Tabla 5.4
Tabla de verdad de una compuerta OR

Entrada lógica A	Entrada lógica B	Salida lógica
0	0	0
0	1	1
1	0	1
1	1	1

Como podemos observar en la Tabla 5.4, la salida tendrá un valor de 1 lógico si **A o B** son 1, de allí el nombre OR. La compuerta NOR es una OR seguida de un inversor (NOR= NOT OR), como se resume en la Tabla 5.5. Las primeras tres columnas de la Tabla 5.5 corresponden a la tabla de verdad de la compuerta NOR, mientras que las últimas dos muestran los valores de tensión para cada fila de la tabla de verdad. Los símbolos lógicos de las compuertas OR y NOR se muestran en la Figura 5.12.

Tabla 5.5
Tabla de verdad de una compuerta NOR

Entrada lógica A	Entrada lógica B	Salida lógica	Voltaje de entrada	Voltaje de salida
0	0	1	$V_A=0, V_B=0$	V_{DD}
0	1	0	$V_A=0, V_B=V_{DD}$	0
1	0	0	$V_A=V_{DD}, V_B=0$	0
1	1	0	$V_A=V_{DD}, V_B=V_{DD}$	0

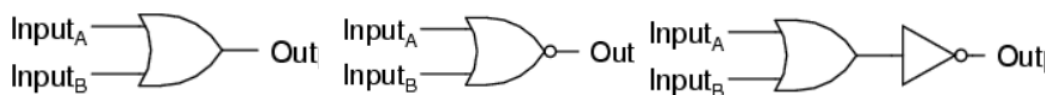


Figura 5.12. Símbolos lógicos de las compuertas OR, NOR y equivalencia de la compuerta NOR

5.3 Arreglos de transistores

Todo transistor MOSFET tiene una resistencia asociada, la cual se refleja en las regiones lineal y de saturación por medio del parámetro de transconductancia y la resistencia de salida del transistor. Cuando se conectan transistores MOSFET en serie, estos pueden reemplazarse por un único transistor equivalente si su parámetro de transconductancia da como resultado la misma corriente que todos los transistores sustituidos juntos. El transistor equivalente tiene la misma tensión de umbral que la de los transistores a los que sustituye. En el caso de transistores conectados en paralelo, podemos encontrar el parámetro de transconductancia del transistor equivalente la suma de los parámetros de transconductancias individuales. Para transistores conectados en serie, el parámetro de transconductancia del transistor equivalente es el inverso de la suma de los inversos de los parámetros de transconductancia individuales. Dicho de otra forma, se utilizan las fórmulas para el cálculo de una conductancia equivalente, tomando para ello los valores de parámetro de transconductancia.

5.4 Consumo de potencia

El consumo de potencia de un circuito es uno de las especificaciones de diseño más importantes en la Electrónica. Para un circuito dado, debe considerarse tanto la potencia en estado activo como inactivo. De acuerdo con esto, existen dos tipos de consumo de potencia:

- a) Potencia dinámica, que es la potencia consumida por los circuitos en estado activo, por ejemplo, cuando una compuerta lógica cambia de estado
- b) Potencia estática, que es la potencia consumida por los circuitos en estado inactivo, por ejemplo, un dispositivo en “stand by” o una compuerta lógica en espera de un cambio en su señal de entrada

La potencia estática se consideró despreciable durante mucho tiempo, pero con el advenimiento de los dispositivos portátiles y el aumento en la densidad de integración de los circuitos microelectrónicos, su contribución no puede despreciarse, ya que determina en buena parte la duración de la carga en las baterías de computadoras portátiles, teléfonos celulares, y en general, de dispositivos móviles.

5.4.1 Potencia estática

La potencia estática está causada por:

- Corriente de fuga de compuerta
- Corriente de reversa de juntas PN
- Corriente de subumbral

La corriente de fuga de compuerta es aquella corriente que fluye entrando o saliendo de la compuerta del transistor. Se debe a que el material dieléctrico

de la compuerta no ofrece condiciones de aislamiento 100% ideales, aunque se aproximan mucho a esto. Además de que el dieléctrico de la compuerta tiene una conductividad diferente de cero, la reducción del espesor del dieléctrico depositado contribuye a aumentar la corriente de fuga de compuerta por efecto túnel.

La corriente de reversa de las juntas PN difusión-substrato de los transistores MOSFET son el siguiente factor causante de la potencia estática. Durante la operación de los transistores MOSFET el sustrato se polariza de forma que estas juntas PN se mantengan operando en la región de corte, donde presentan una corriente que, aunque muy baja, es diferente de cero, y por lo tanto, disipan potencia.

La corriente de subumbral es la causa más importante de la potencia estática, y ocurre cuando un transistor opera en la región de corte (región de subumbral). Su valor está determinado tanto por la tensión aplicada a los transistores que estén operando en la región de corte, como por el valor de la tensión de umbral y la pendiente de subumbral. La pendiente de subumbral, a su vez, depende de la temperatura, de forma que cuanto mayor temperatura, tanto mayor será la corriente de subumbral.

Para el cálculo exacto de la potencia estática, se requieren las tres corrientes de fuga: de compuerta, de reversa de las juntas PN y de subumbral. Sin embargo, dado que la corriente de subumbral es la mayor de todas, la potencia estática puede aproximarse utilizando para el cálculo solamente la corriente de subumbral.

5.4.2 Potencia dinámica

La potencia dinámica se presenta cuando existen transiciones de un estado lógico a otro, es decir, cuando la salida de un circuito digital cambia en respuesta a una tensión aplicada a la entrada.

La potencia dinámica está causada por:

- Corriente de corto circuito
- Corriente debido a cargas capacitivas

Para comprender estas componentes de la potencia dinámica, así como obtener ecuaciones para calcularlas, considere un inversor CMOS.

Cuando cambia la tensión aplicada a la entrada del inversor, el cambio no se produce instantáneamente, sino que la tensión de entrada presenta una pendiente, llamada flanco de subida (para transiciones de 0 a 1 lógico) o flanco de bajada (para transiciones de 1 a 0 lógico). Esto origina un consumo de potencia dinámica durante las transiciones.

5.4.2.1 Potencia dinámica de corto circuito

Durante la transición de la salida de un estado lógico a otro, la tensión de entrada toma por un tiempo determinado valores de voltaje entre 0 y V_{DD} que hacen que el transistor PMOS y NMOS se encuentran activos simultáneamente, causando una trayectoria de baja impedancia entre V_{DD} y tierra (GND) y con ello un alto consumo de corriente en el inversor. La corriente que fluye en esta condición se conoce como corriente de corto circuito, y es la causa de la potencia dinámica de corto circuito.

El valor máximo de la corriente de corto circuito se obtiene en la condición $V_{in}=V_{out}$, es decir, en la tensión de disparo del inversor. La potencia dinámica de corto circuito para un inversor simétrico se calcula como:

$$P_{SC} = I_{SC} \cdot V_{DD} = \frac{2}{3} \cdot K \cdot \frac{t_r}{T} \left(\frac{V_{DD}}{2} - V_{TH} \right)^3$$

Donde P_{SC} : potencia de corto circuito

I_{SC} : corriente de corto circuito

K : parámetro del transconductancia de los transistores del inversor simétrico

t_r : tiempo de subida

T : período de la señal de entrada

Esta fórmula es aplicable bajo las siguientes condiciones:

- El inversor es simétrico
- El factor de actividad es 1
- El tiempo de subida t_{rise} y el tiempo de bajada t_{fall} son iguales

5.4.2.1 Potencia dinámica de carga capacitiva

La potencia dinámica de carga capacitiva se debe a las capacitancias parásitas presentes en el circuito, y que pueden representarse como una sola capacitancia de carga C_L , como se muestra en la Figura 5.13. Esta capacitancia de carga se debe a la suma de:

- La capacitancia de compuerta C_{OX} de las compuertas lógicas conectadas a la salida
- La capacitancia de compuerta C_{OX} propia de la compuerta lógica
- La capacitancia parásita C_W de las interconexiones conectadas a la salida

La potencia dinámica de carga capacitiva se debe a la potencia consumida por la compuerta lógica en el proceso de carga y descarga del capacitor al cambiar el estado lógico de la salida. Esta potencia se calcula como:

$$P_L = A \cdot f \cdot C_L \cdot V_{DD}^2$$

Donde f : frecuencia de conmutación, C_L : capacitancia de carga, A : factor de actividad.

El factor de actividad A es la probabilidad de conmutación de la salida (cambio de estado) debido a cambios en la entrada, en un período. En el caso de un inversor al que se aplica una señal cuadrada en la entrada, el factor de actividad es 1.

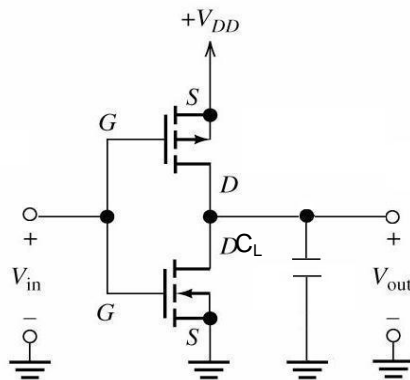


Figura 5.13. Capacitancia de carga en un inversor CMOS

5.5 Memorias semiconductoras

Las memorias electrónicas comerciales actuales se basan principalmente en el uso de transistores MOSFET, debido a la alta densidad de integración que puede alcanzarse con la tecnología CMOS.

Las memorias semiconductoras pueden clasificarse en memorias volátiles y no volátiles, como se muestra en la Figura 5.14. Las memorias volátiles no pueden almacenar los datos una vez que se ha removido la alimentación del circuito. Las memorias no volátiles, por el contrario, pueden retenerlos hasta por 10 años, de acuerdo con los estándares comerciales.

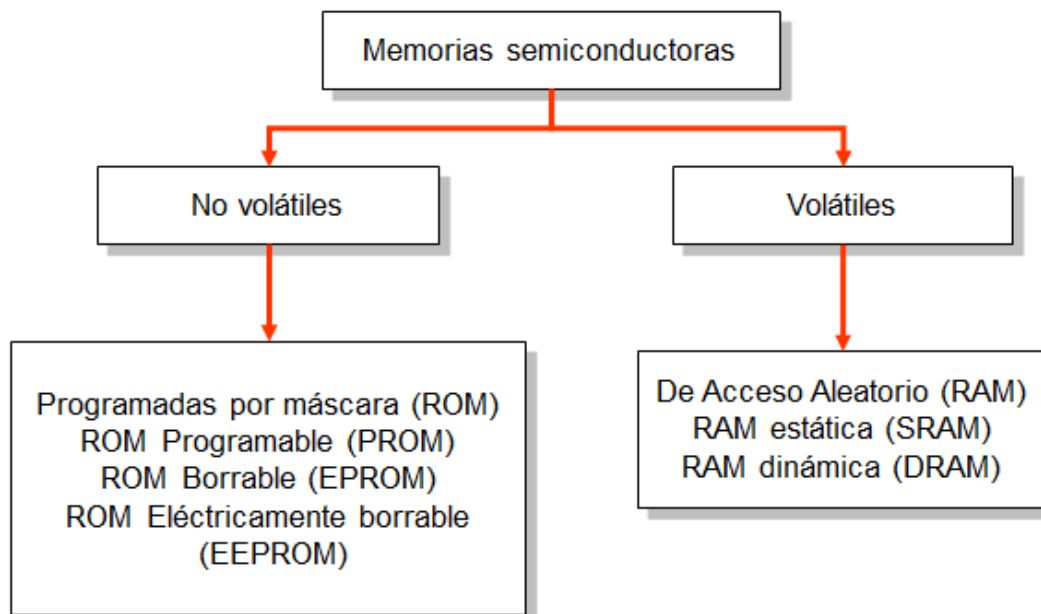


Figura 5.14. Clasificación de memorias semiconductoras

5.5.1. Organización de las memorias

Para almacenar los datos, las memorias se organizan como una matriz, como muestra la Figura 5.15. Cada bit se almacena en la intersección de una fila y una columna, es decir, en cada intersección se localiza un elemento de almacenamiento de datos. A las columnas se les conoce como *bitlines* (líneas de bit) y a las filas se les conoce como *wordlines* (líneas de palabra). Para leer un bit individual debe indicarse la posición del dato utilizando la fila y la columna correspondiente. Para leer una palabra (un conjunto de bits), debe indicarse la fila correspondiente y activar todas las líneas de bit de esa fila. Así, la dirección completa de un dato en una memoria se compone de bits de fila y bits de columna. Para formar palabras, se utiliza un conjunto de matrices y de cada matriz se lee un bit de la palabra, como muestra la Figura 5.16.

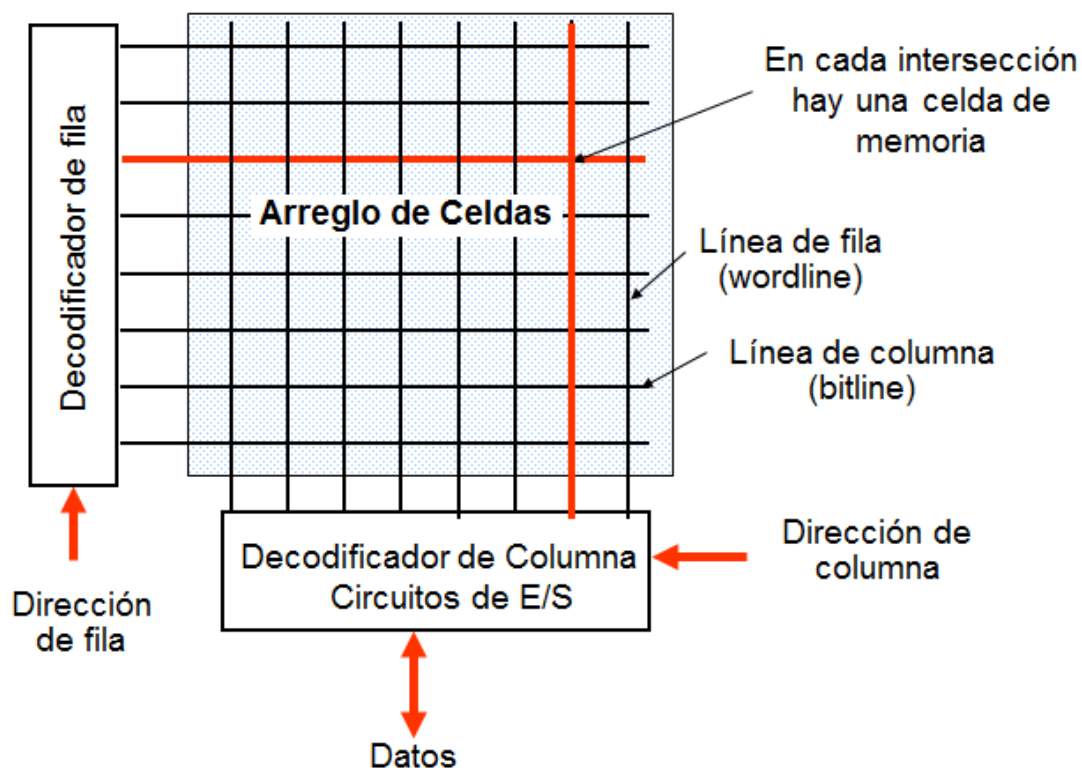


Figura 5.15. Organización de un arreglo de memoria

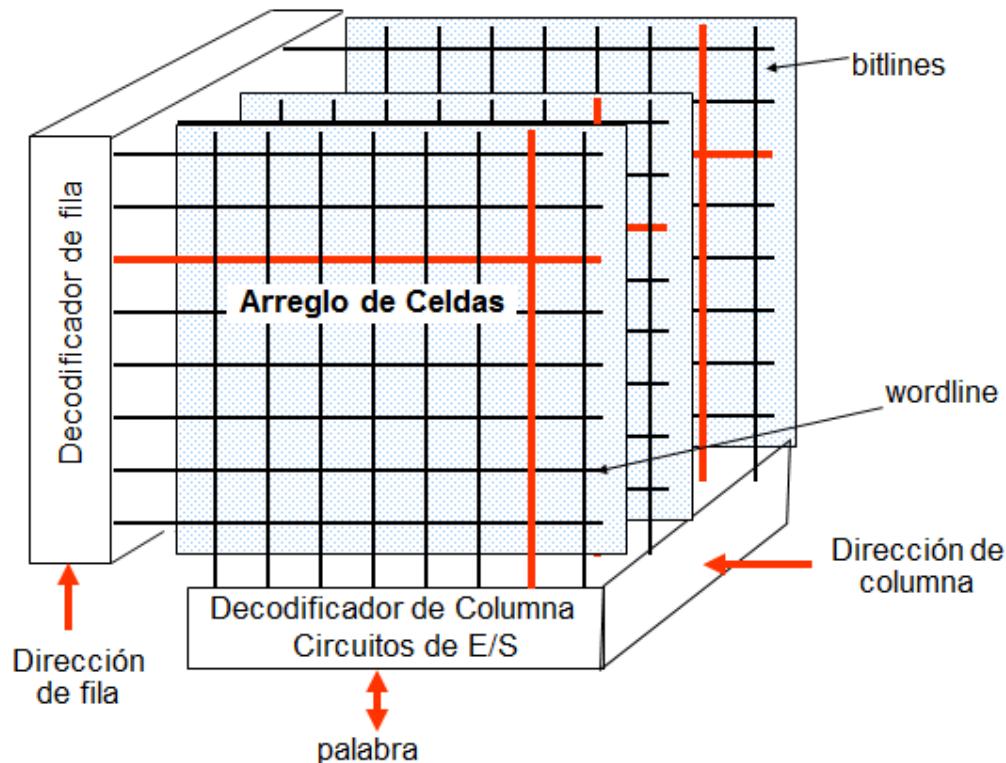


Figura 5.16. Organización de una memoria para formar una palabra

5.5.2. Memorias no volátiles

Existen varios tipos de memorias no volátiles, de acuerdo con las posibilidades de programación y borrado. Las memorias no volátiles se identificaron inicialmente con el acrónimo ROM (read only memory), pues el usuario final no podía borrarlas ni escribirlas, sino sólo leerlas. Con el tiempo se implementaron mecanismos de escritura eléctricos y borrado eléctrico u óptico de estas memorias, aunque el acrónimo ROM se mantiene como parte del nombre de este tipo de memorias para indicar su característica de no volatilidad. Estas son las memorias que se utilizan en USB/memory sticks (llaves mayas), memorias de cámaras y otros aparatos electrónicos, almacenamiento de BIOS y similares.

Las memorias no volátiles se clasifican de acuerdo con sus métodos de programación y borrado, como sigue:

- Memorias programadas por máscara (ROM)
- Memorias ROM programables eléctricamente (PROM)
- Memorias ROM programables eléctricamente y borrables ópticamente (EPROM)
- Memorias ROM programables y borrables eléctricamente (EEPROM)

Las memorias programadas por máscara se programan durante el proceso de fabricación. Una forma de diferenciar entre los estados de cero y uno lógico consiste en fabricar transistores MOSFET de canal N, uno de mayor tensión de

umbral para uno de los estados y otro con menor tensión de umbral para el otro estado. Esto puede lograrse creando transistores con diferente espesor del aislante de la compuerta o bien ajustando las tensiones de umbral por medio de implantación iónica. En el primer caso, un transistor con aislante más grueso presenta una mayor tensión de umbral que un transistor de aislante delgado. En el segundo caso, se utilizan transistores de enriquecimiento y de empobrecimiento para representar los estados lógicos. Para crear el transistor de empobrecimiento, se dopa la región bajo la compuerta para crear el canal y con ello obtener una tensión de umbral negativa.

Otra forma de obtener memorias no volátiles programadas por máscara es contar con una matriz de transistores MOSFET en los cuales los dos estados lógicos se representan por la presencia/ausencia de una interconexión al drenador del transistor, permitiendo o impidiendo el flujo de corriente, respectivamente.

En las memorias programables eléctricamente se cuenta con transistores MOSFET de canal N con doble compuerta a los cuales se inyectan huecos o electrones los cuales quedan atrapados en el óxido de compuerta. La inyección de huecos permite disminuir la tensión de umbral y eventualmente volverla negativa, mientras que la inyección de electrones aumenta la tensión de umbral. Así puede distinguirse entre dos estados lógicos. La inyección de portadores de carga en la compuerta se lleva a cabo aplicando una diferencia de tensión de diferencia compuerta y drenador/surtidor que genere un campo eléctrico suficientemente alto para observar el efecto cuántico llamado efecto túnel (Fowler-Norheim o directo). Dado que los transistores tienen doble compuerta, una de ellas es flotante y permite que los portadores de carga queden atrapados en vez de fluir por el transistor como una corriente de compuerta.

Sin embargo, cada vez que se inyectan portadores de carga en el óxido de la compuerta flotante, dicho óxido sufre daños. Estos daños limitan el número de veces que la memoria se puede programar y borrar (ciclos de programación-borrado), así como el tiempo durante el cual los transistores pueden almacenar los datos. Para toda memoria programable y borrrable los fabricantes deben garantizar, al menos, 100 mil ciclos de programación-borrado y una permanencia de datos en la memoria de al menos 10 años. Estas son las memorias PROM, EPROM, EEPROM y Flash, la cual es una memoria EEPROM borrrable por sectores.

Para borrar los datos de una memoria no volátil puede recurrirse a luz ultravioleta o a métodos eléctricos. En caso de las memorias EPROM se utiliza luz ultravioleta para excitar los electrones inyectados en la compuerta flotante de forma que puedan regresar al sustrato del transistor; este proceso de borrado dura algunos minutos (hasta 20 minutos) y consiste en exponer la memoria a luz ultravioleta, al cual radia sobre el transistor a través de una ventana de vidrio en el encapsulado de la memoria. Cuanto mayor sea el número de ciclos de escritura-borrado que haya experimentado la memoria, tanto más tiempo se requiere para completar un proceso de borrado. En las memorias borrrables eléctricamente, se aplica un campo eléctrico entre

compuerta y drenador/surtidor que extraiga los electrones de la compuerta flotante por medio de efecto túnel.

5.5.2.1 Lectura de datos en memorias no volátiles

Las memorias no volátiles pueden presentar dos configuraciones: NOR y NAND. La más utilizada es la NAND debido a que permite implementar memorias más densas. En la configuración NAND, los transistores que almacenan la información están conectados en serie. Para leer el dato de un transistor en particular, se aplica a todos los demás transistores de la rama en serie una tensión $V_G = V_{DD}$, mientras que al transistor por leer se le aplica 0V en la compuerta. Para determinar el dato almacenado se sensa la corriente por la rama: si el transistor es de enriquecimiento, la corriente en la rama será cero, pues $V_G=0V$ no será suficiente para encender el transistor; por el contrario, si el transistor es de empobrecimiento se podrá sensar una corriente diferente de cero en la rama. De esta forma puede determinarse la presencia de un cero o un uno lógico almacenado.

5.5.3 Memoria SRAM

La memoria SRAM es volátil. La figura 5.17 muestra una celda de memoria SRAM. Esta consiste en dos inversores mutuamente acoplados (cerrojo) cuyas salidas están conectadas a transistores de acceso, para un total de 6 transistores. Las compuertas de los transistores de acceso están conectadas a la línea de palabra (wordline), y permiten acceder a los nodos internos de la memoria donde se leen los datos, es decir, las salidas de los inversores, conectando estos nodos a las líneas de bit (bitlines).

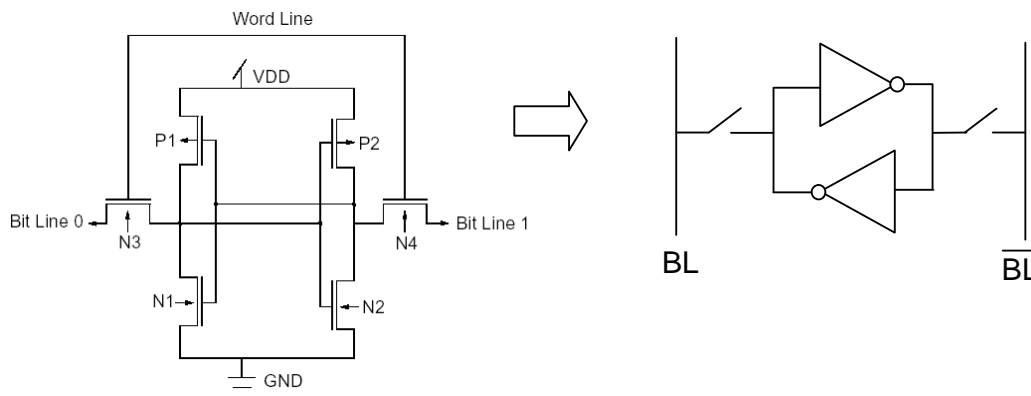


Figura 5.17. Celda de memoria SRAM

5.5.3.1 Lectura de un dato de una celda de memoria SRAM

¿Cómo se lee un dato de una celda de memoria SRAM? Para este ejemplo, asumamos que tenemos un '1' lógico almacenado en el lado izquierdo de la celda y un '0' al lado derecho, de acuerdo con la Figura 5.18.

El proceso es el siguiente:

a) Preparación para la lectura

Los transistores de acceso de la celda SRAM están en corte ($WL=0$) para mantener los nodos internos de la celda aislados de la línea de bit. Se precargan las líneas de bit a V_{DD} (y con ello los capacitores parásitos de bitline). Para esto se activan los transistores de precarga para que las líneas de bitline se conecten a V_{DD} . Se activa el transistor de ecualización para igualar las tensiones de la línea bit. Pasado el tiempo de ecualización, se desactivan los transistores de precarga y ecualización.

b) Lectura de dato

Se activan los transistores de acceso de la celda SRAM. Se espera un tiempo determinado para que las tensiones de línea bit cambien debido a la tensión de los nodos internos de la celda, obteniendo una diferencia de tensión de 100mV.

Puesto que hay un '0' en uno de los nodos de la celda, una línea bit empezará a descargarse. El transistor M_2 está activo y empieza a descargar el capacitor de la de la línea de bit derecha. En el lado izquierdo de la celda, tanto en el nodo interno de la celda como la línea de bit están a V_{DD} , pues M_5 está encendido y en saturación. En este lado la tensión no cambia. Después de un tiempo determinado, la diferencia de tensión entre la línea de bit izquierda y derecha alcanza 100mV, como sigue:

Lado izquierdo = $V_{DD} - 100\text{mV}$

Lado derecho = V_{DD}

Cuando esto sucede se desactivan los transistores de acceso de la celda SRAM.

c) Amplificación del dato

En esta etapa se activa el transistor de activación de sensado de la red NMOS del amplificador de sensado (Activación de sensado = 1). En el lado derecho del amplificador tiene una tensión de $V_{DD} - 100\text{mV}$, y el lado izquierdo a V_{DD} .

El transistor M_8 está activo a ($V_{GS8} = V_{DD}$) y empieza a descargar la capacitancia de bitline del lado derecho a través de M_8 . Conforme la tensión del lado derecho disminuye, M_7 se empieza a desactivar. Ahora se activa el transistor de activación de sensado de la red PMOS del amplificador de sensado (Activación de sensado = 0) y M_9 se empieza a activar.

Esto mantiene la tensión del lado izquierdo en V_{DD} , y con ello M_8 se mantiene activo completando la descarga del lado derecho. Esto completa la lectura del dato, y a partir de aquí se activan circuitos de salida (no mostrados) que trasladan el dato a las patillas del circuito integrado.

Siempre se traslada sólo un dato (0 o 1) a la salida de la memoria, es decir, el dato que está conectado a Bitline (no a Bitline). El diseñador debe leer siempre del mismo lado en todas las celdas.

El mismo proceso se sigue si los voltajes están intercambiados (V_{DD} a la derecha y '0' a la izquierda), con la diferencia de que los transistores que descargan la línea de bit son los del lado izquierdo.

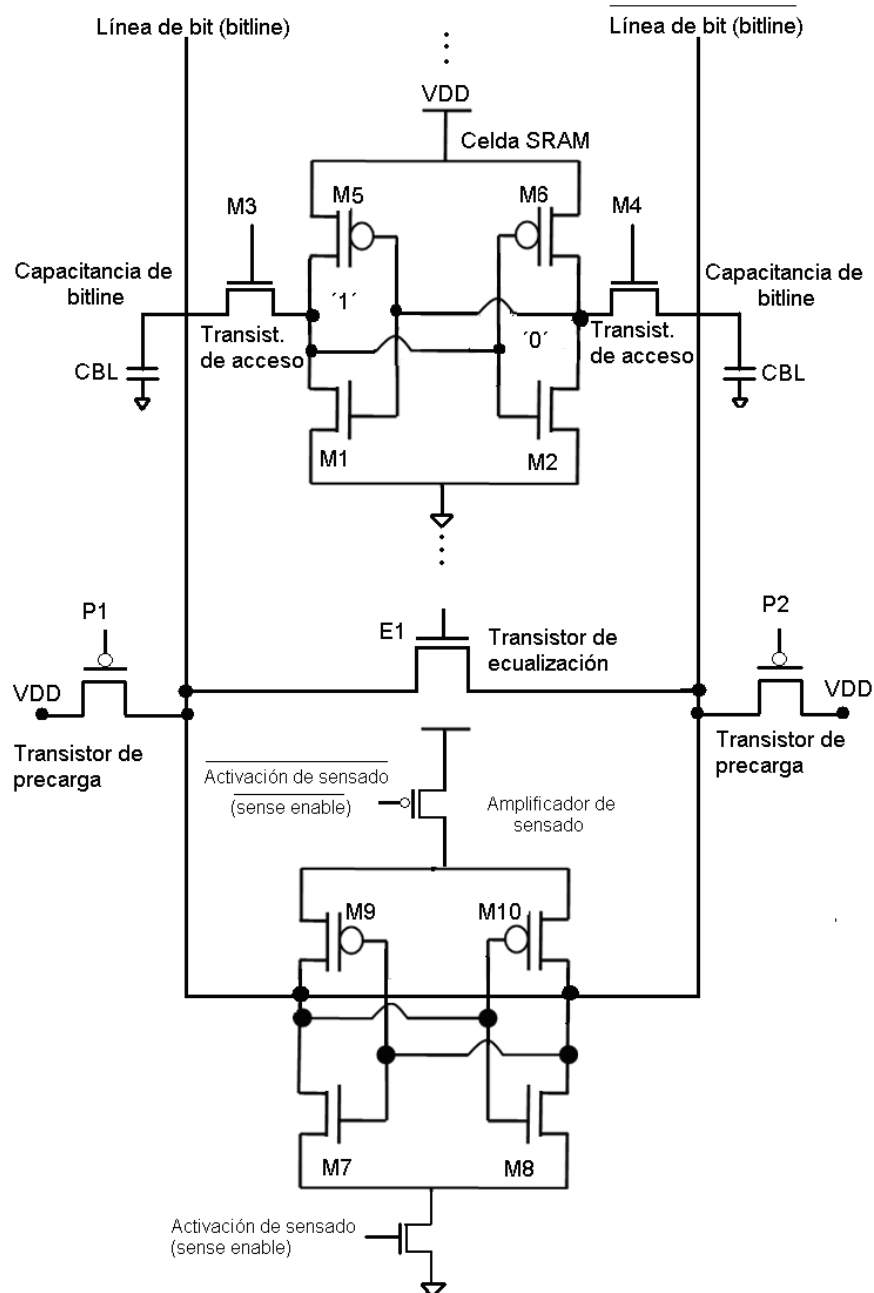


Figura 5.18. Lectura de una memoria SRAM

5.5.4 Memoria DRAM

La memoria DRAM es volátil. La figura 5.19 muestra una celda de memoria DRAM. Esta consiste en un transistor de acceso y un capacitor de almacenamiento C_{st} . La compuerta del transistor de acceso está conectada a la línea de palabra (wordline), y permite acceder al capacitor de almacenamiento, conectando este capacitor a la línea de bit (bitline).

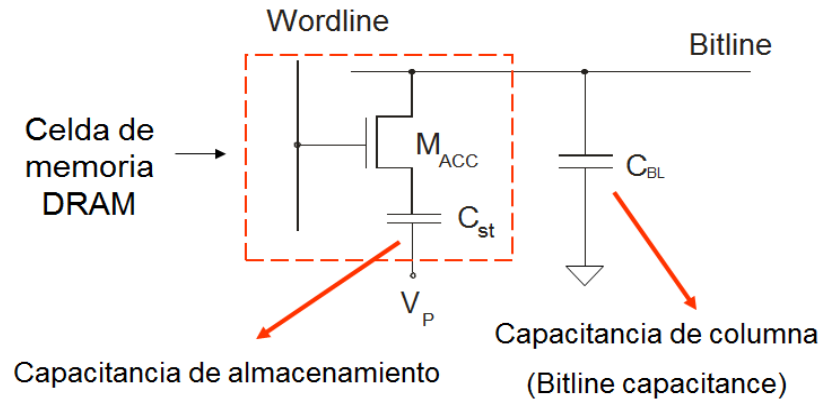


Figura 5.19. Celda de memoria DRAM

5.5.4.1 Lectura de un dato de una celda de memoria DRAM

¿Cómo se lee un dato en una memoria DRAM? Para este ejemplo, asumamos que tenemos un '1' lógico almacenado en la celda de memoria conectada en Bitline, de acuerdo con la Figura 5.20. La operación de lectura ocurre en tres etapas, explicadas a continuación.

a) Preparación para la lectura

En esta etapa se inicia con la activación de los transistores de precarga P1 y P2. Con esto, las líneas de bit se precargan a $V_{DD}/2$. Después de esto se desactivan los transistores de precarga y se activa el transistor de ecualización, para igualar ambas tensiones de la línea de bit a $V_{DD}/2$.

b) Lectura del dato

Se activa solamente la línea de palabra de la celda de memoria que contiene el dato a leer. La otra línea de palabra (Wordline) se deja en 0 para que el transistor de acceso de esa celda opere en corte, pero que la tensión de precarga en Bitline sirva como tensión de comparación para el amplificador de sentido.

Al activar la línea de palabra para leer el dato, el transistor de acceso se activa, conectando la línea de bit precargada a $V_{DD}/2$ con el capacitor de almacenamiento C_{ST} . Si el transistor de acceso activo se considera como un corto circuito, inmediatamente después de activar el transistor de acceso se

obtiene el circuito de la Figura 5.21. Si el dato almacenado era un cero lógico, el capacitor de bitline se descarga 100mV, en el caso de un uno lógico se carga 100mV.

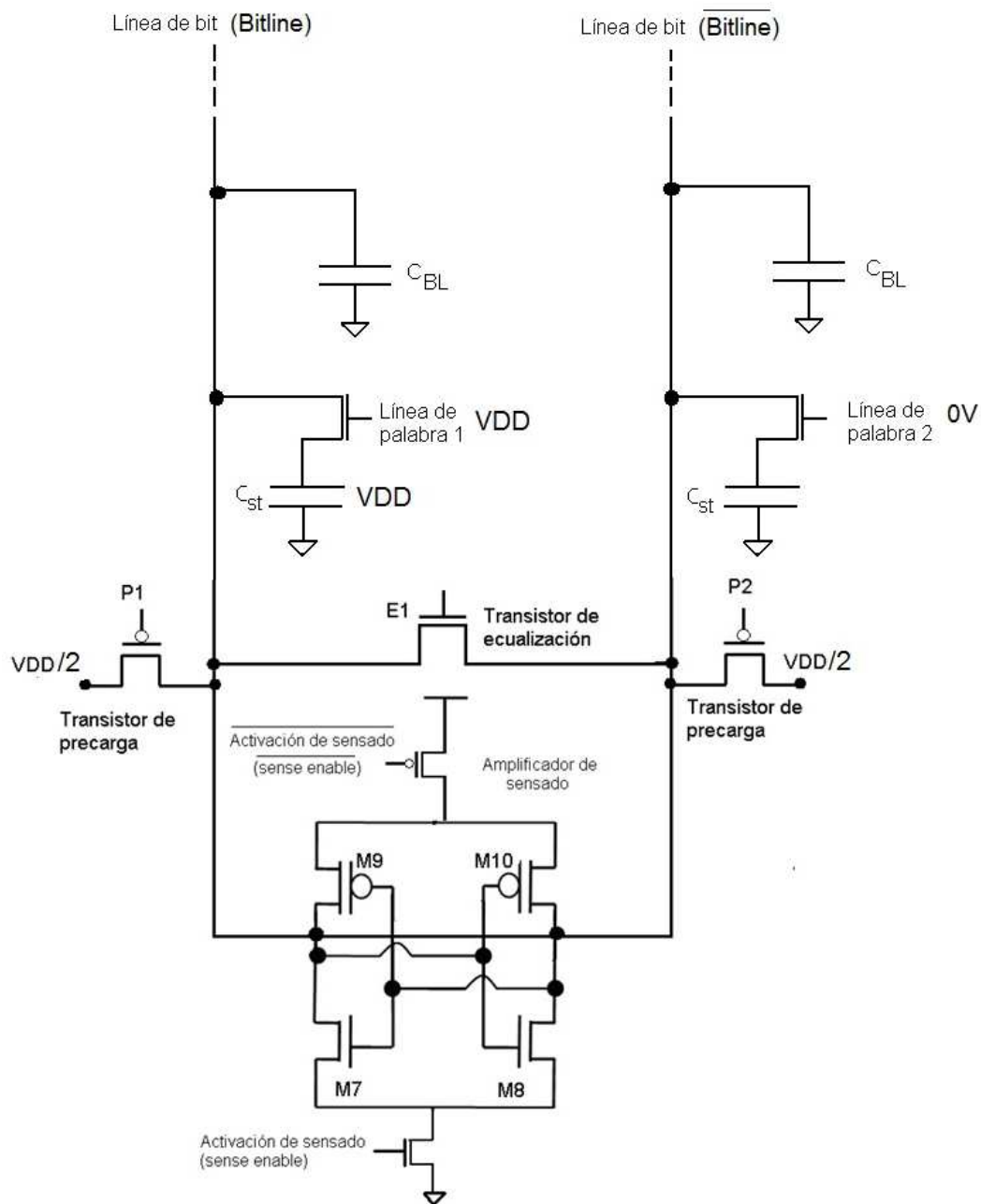


Figura 5.20. Lectura de una memoria DRAM

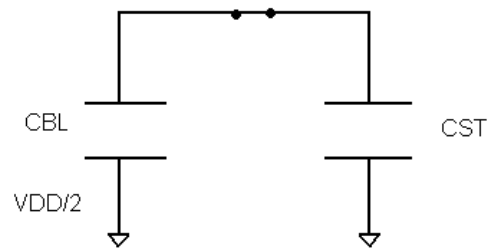


Figura 5.21 Circuito equivalente para la lectura de un dato de una celda de memoria DRAM

$$\begin{aligned}
 Q_{inicial} &= Q_{final} \\
 Q_{final} &= (C_{ST} + C_{BL}) \cdot V_{ST} \\
 Q_{inicial} &= C_{BL} \cdot V_{precharge} \\
 Q_{final} &= C_{BL} \cdot V_{BL} = C_{BL} \left(\frac{V_{DD}}{2} \pm 100mV \right)
 \end{aligned}$$

c) Amplificación del dato

En esta etapa se activa el transistor de activación de sensado de la red NMOS del amplificador de sensado (Activación de sensado = 1). En un lado del amplificador se tiene una tensión de $V_{DD} \pm 100mV$, y en el otro $V_{DD}/2$. Para continuar esta explicación, asuma que la celda que se está leyendo está conectada al lado izquierdo del amplificador de sensado y que el dato almacenado es un 1 lógico.

La compuerta del transistor M_8 tiene entonces una tensión de $V_{DD}/2 + 100mV$ ($V_{GS8} = V_{DD}/2 + 100mV$) y la capacitancia de Bitline se descarga más rápido la que la capacitancia de Bitline. Conforme la tensión del lado derecho disminuye, M_7 se empieza a desactivar. Ahora se activa el transistor de activación de sensado de la red PMOS del amplificador de sensado, es decir, (Activación de sensado = 0) y M_9 conduce cada vez más.

Esto empieza a aumentar la tensión del lado izquierdo hacia V_{DD} , y con ello M_8 se mantiene activo completando la descarga del lado derecho, mientras que M_9 completa la carga del lado izquierdo. Esto completa la lectura del dato, y a partir de aquí se activan circuitos de salida (no mostrados) que trasladan el dato de Bitline a las patillas del circuito integrado.

El mismo proceso se sigue si los voltajes están intercambiados (V_{DD} a la derecha y '0' a la izquierda), con la diferencia de que los transistores que descargan la línea de bit son los del lado izquierdo y los que la cargan son los del lado derecho.