

El dispositivo MOSFET

19 de agosto de 2024

Hoy en día, el transistor MOSFET o MOS (Metal-Oxide-Semiconductor Field-Effect Transistor) es el componente fundamental del diseño digital. Desde una perspectiva digital, el MOS se comporta muy bien como switch, es decir, el pasaje entre los estados de corte y saturación se lleva a cabo de manera eficiente con respecto a otros tipos de transistores, además introduce pequeños efectos parásitos. Otra importante ventaja que tienen estos dispositivos es su alta densidad de integración combinada con un proceso de manufactura relativamente "simple", que hace posible crear y producir circuitos digitales complejos a gran escala en forma económica.

1. Descripción del dispositivo

En la Fig. 1 se observan un PMOS y un NMOS en un proceso de pozo simple (N-well). Obsérvese que el transistor NMOS se implementa directamente sobre el sustrato de tipo P.

En la Fig. 2 se muestra una estructura NMOS, donde el body es tipo P. Al terminal de sustrato lo llamaremos *body* (B). Si se aplica una tensión negativa en el *gate* (G) respecto al body, sobre este terminal se acumularán cargas negativas. Luego, los portadores mayoritarios del material tipo P (huecos, h^+) serán atraídos a la región que se encuentra por debajo del gate. Este proceso es conocido como "*modo de acumulación*". En este modo, la estructura se comporta como un capacitor existiendo un campo eléctrico a través del óxido que mantiene los portadores acumulados alrededor del dieléctrico.

En cambio, si se aplica una tensión positiva en el gate respecto del body, el gate se cargará positivamente y los huecos del material tipo P serán repelidos de la región por debajo del gate por efecto de los primeros. Como consecuencia, por debajo del gate se formará una zona libre de portadores, más conocida como *zona de vaciamiento*.

Si la tensión $V_{GB} > 0$ aplicada supera cierto valor umbral que llamaremos V_T , el gate acumulará mayor cantidad de cargas positivas, repeliendo aún más a los portadores mayoritarios positivos. Asimismo, algunos electrones libres en el body serán atraídos a la zona de vaciamiento. Esta capa de conducción en el body tipo P es llamada *inversion layer*.

El transistor NMOS consiste de una estructura MOS de body tipo P con dos difusiones de tipo N+ rodeando la zona del gate. Dichas difusiones se nombran arbitrariamente *drain* (D) y *source* (S).

Obsérvese la Fig. 3. En a), el transistor se encuentra en modo de corte *cutoff*, debido a que no existe un canal conductivo entre S y D. La juntura DB está polarizada en inversa pudiendo circular una corriente de fuga solamente a través de ella. En el caso b), se ha formado un canal entre D y S pero al no haber campo eléctrico entre ellos, los electrones no circulan. En c), la corriente entre los terminales D y S (I_D) aumenta gradualmente conforme lo hace la tensión V_{DS} . Este modo de operación se llama *modo*

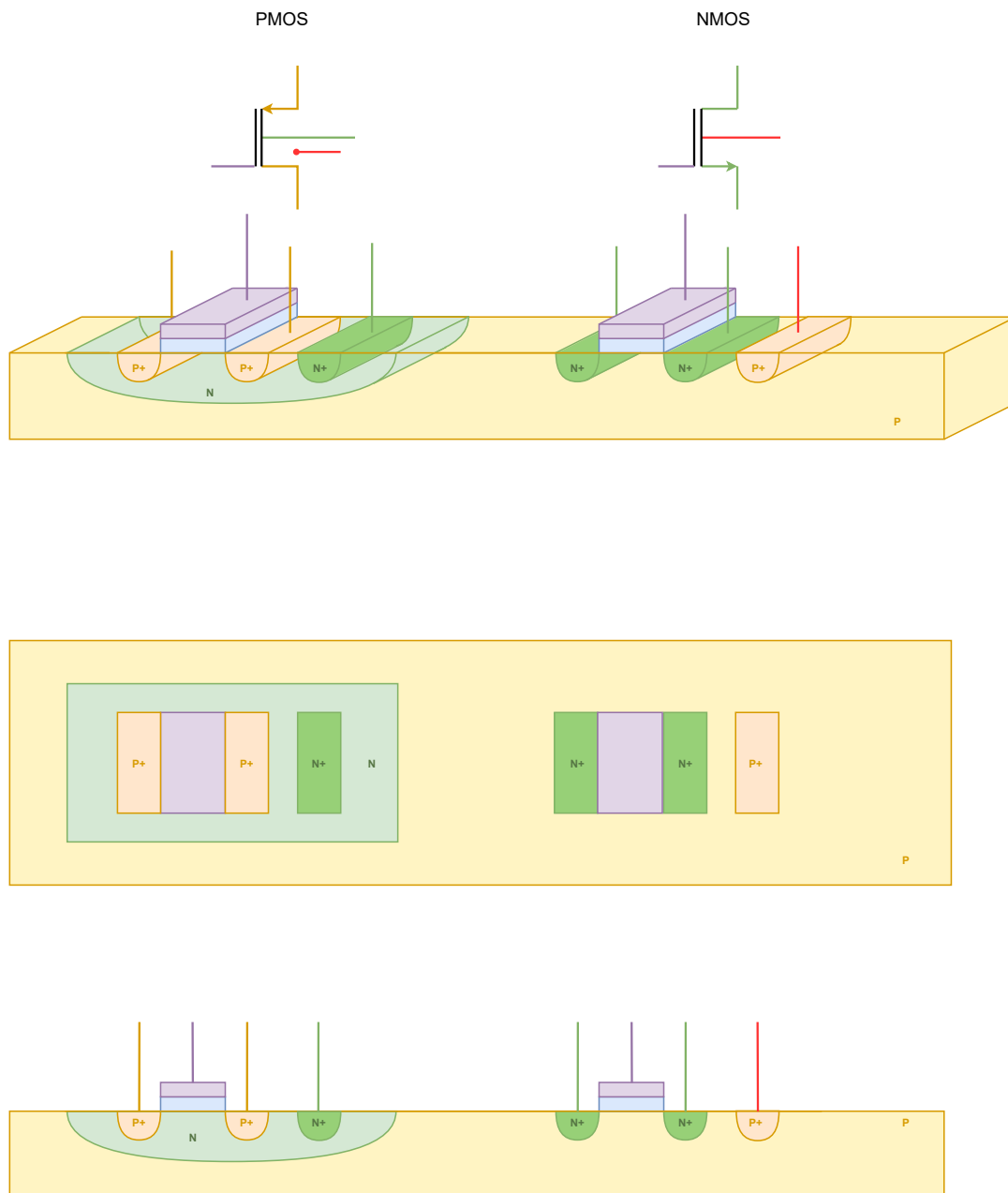


Figura 1: Estructura MOS.

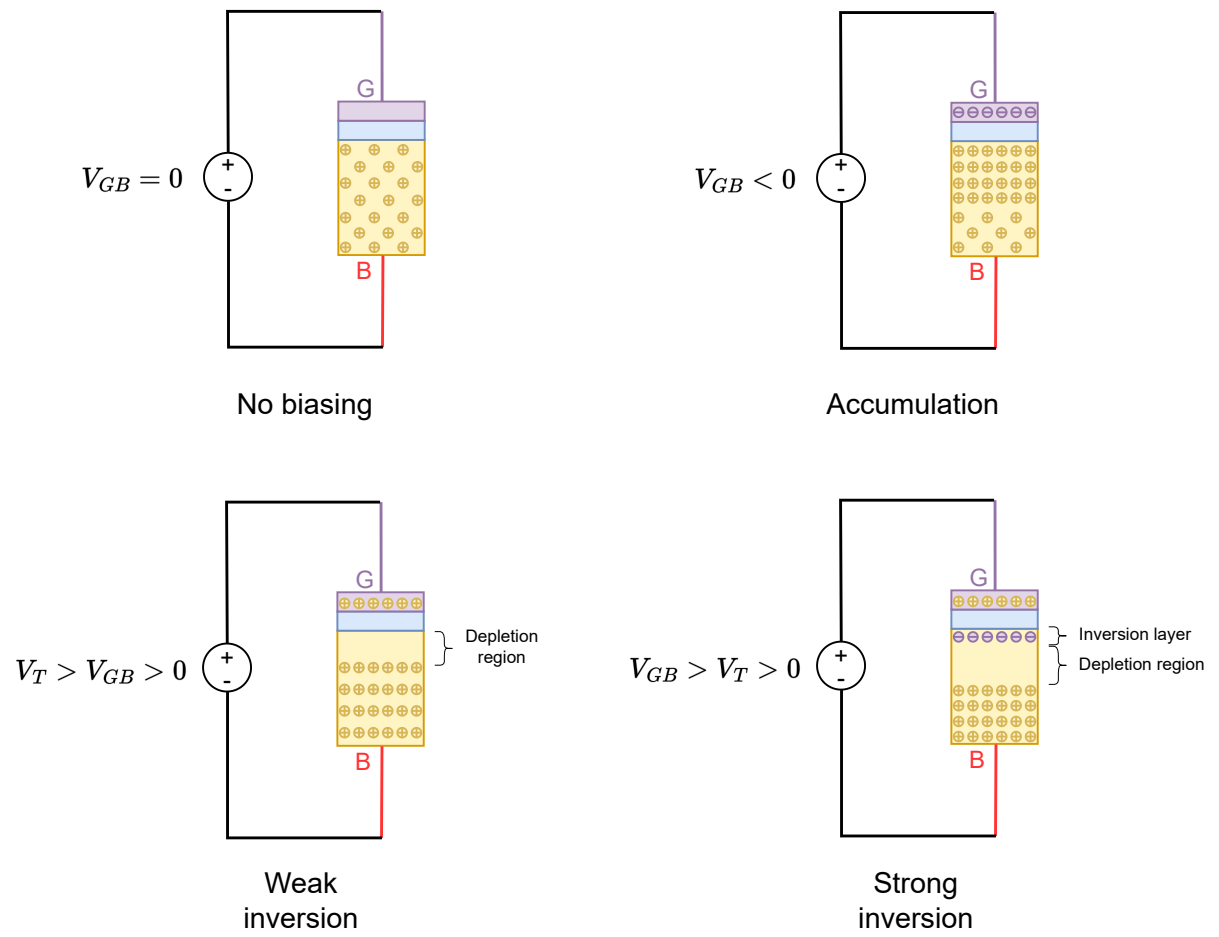


Figura 2: Estructura NMOS con diferentes polarizaciones V_{GB} .

lineal o modo *modo triodo* ya que el comportamiento se asemeja al de una resistencia cuyo valor depende de $V_{GDS} > V_T$. En d), cuando el valor de $V_{DS} > V_{GS} - V_T$, ya no existe inversión cerca de la zona del drain y se dice que el canal se encuentra *pinched off*. La circulación de corriente se mantiene por arrastre del campo eléctrico que existe entre D y el canal. En este modo se dice que el transistor esta en *modo saturación*. La saturación nunca es perfecta, y la corriente se ve algo modulada por el valor de la tensión V_{DS} ya que la longitud efectiva del canal se ve afectada por V_{DS} una vez alcanzada la condición de pinch off.

La tensión V_{SB} que hasta ahora se ha considerado ser cero para el análisis, tiene la función de modificar el valor del parámetro V_T .

2. Modelos de cálculo del transistor MOS

En la industria el modelo que se ha extendido para la simulación de transistores MOS es el BSIM (Berkeley Short-channel IGFET Model) which refers to a family of MOSFET transistor models for integrated circuit design. It also refers to the BSIM group located in the Department of Electrical Engineering and Computer Sciences (EECS) at the University of California, Berkeley, that develops these models. Accurate transistor models are needed for electronic circuit simulation, which in turn is needed for integrated circuit design. As the devices become smaller each process generation, new models are needed to accurately reflect the transistor's behavior.

Commercial and industrial analog simulators (such as SPICE) have added many other device models as technology advanced and earlier models became inaccurate. To attempt standardization of these models so that a set of model parameters may be used in different simulators, an industry working group was formed, the Compact Model Coalition, to choose, maintain, and promote the use of standard models. BSIM models, developed at UC Berkeley, are one of these standards.

Es especialmente útil para procesos *deep subdivision* ya que incluye en su modelo efectos del canal corto como ser la saturación de la velocidad de los portadores que circulan por el canal.

Se utiliza el modelo BSIM3 para procesos en los cuales $130nm < L_{min} < 1\mu m$ y el modelo BSIM4 para $20nm < L_{min} < 130nm$.

Sin embargo, para el análisis manual, se utilizarán el siguiente conjunto de ecuaciones, que si bien es impreciso con respecto al modelo BSIM, es suficiente para la extracción de conceptos:

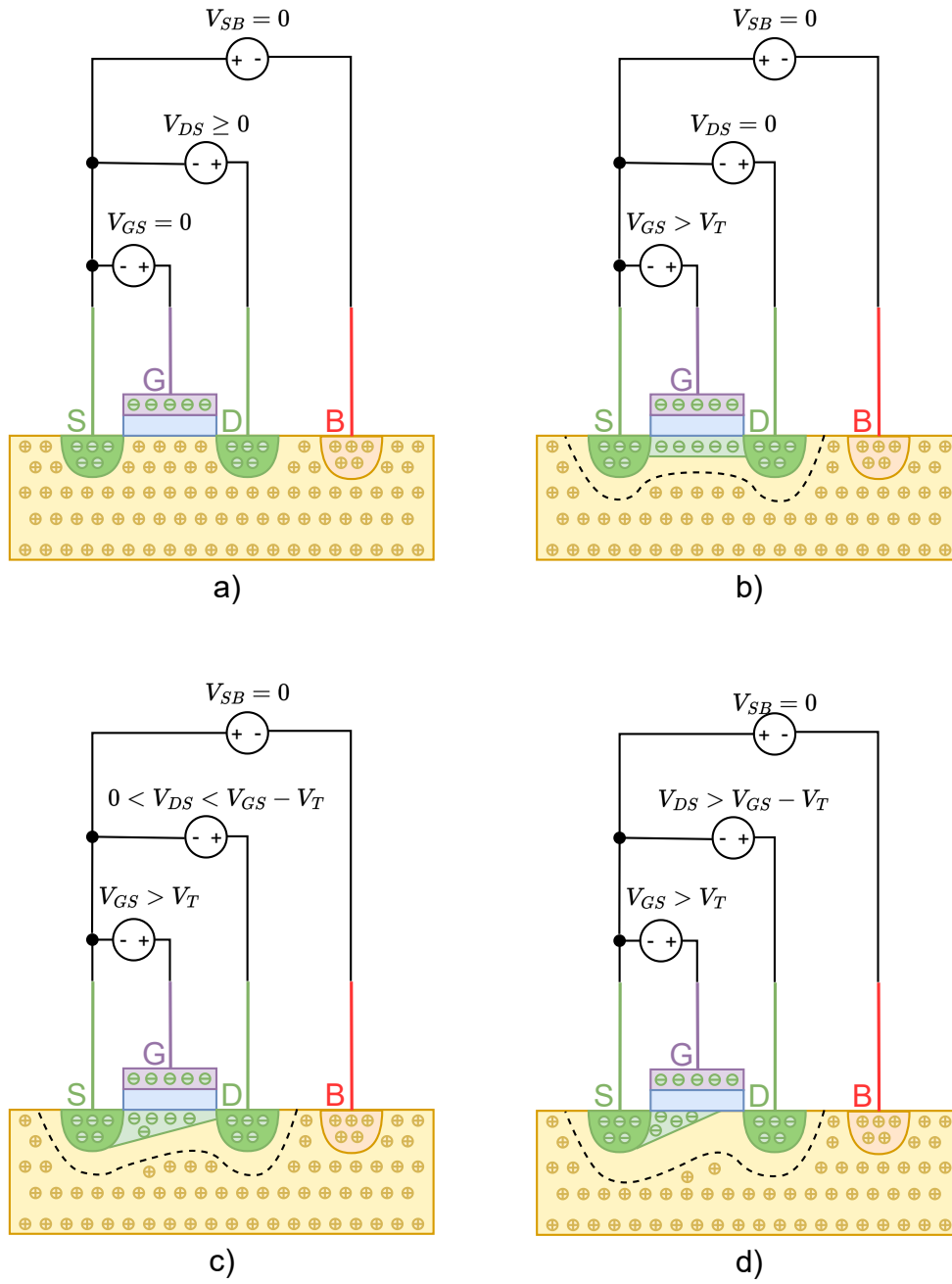


Figura 3: Operación del transistor NMOS.

$$\begin{aligned}
V_T &= V_{T0} + \gamma \sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|2\phi_F|} \\
\phi_F &= -\frac{KT}{q} \ln\left(\frac{N_A}{n_i}\right) \\
V_{GT} &= V_{GS} - V_T \\
I_D &= 0, \quad \text{si } V_{GT} < 0 \\
I_D &= k' \frac{W}{L} \left(V_{GT} V_{min} - \frac{V_{min}^2}{2} \right) (1 + \lambda V_{DS}), \quad \text{si } V_{GT} \geq 0 \\
V_{min} &= \min(V_{GT}, V_{DS}, V_{DSAT})
\end{aligned} \tag{1}$$

donde:

- V_{T0} es la tensión umbral para $V_{SB} = 0$.
- N_A es la concentración de dopaje en el body tipo P.
- n_i es la concentración intrínseca en una muestra pura de material semiconductor sin dopar, para el caso del silicio es aproximadamente $1,5 \times 10^{10} \text{cm}^{-3}$ a 300K.
- KT/q es aproximadamente 25mV a 300K.
- V_{DSAT} es el valor de V_{DS} en el cual se logra la condición de pinch off del canal.
- λ es el efecto de modulación del largo del canal.
- γ es el coeficiente de efecto de body, típicamente $0,4V^{1/2}$.
- $k' = \mu \varepsilon_{ox} / t_{ox}$, donde μ es la movilidad de los portadores en el canal expresada en unidades de $\text{m}^2/(\text{V.s})$, ε_{ox} es la permitividad eléctrica del óxido, típicamente $\varepsilon_{ox} = 4 \times \varepsilon_0 = 3,5 \times 10^{-11} \text{F/m}$, t_{ox} es el espesor del óxido. Por ejemplo, para $t_{ox} = 10 \text{nm} = 100 \text{\AA}$, la capacidad del óxido por unidad de área es $C_{ox} = 3,5 \text{fF}/\mu\text{m}^2$.
- W y L son el ancho (*width*) y largo (*length*) del transistor, tal como se observa en la Fig. 4.

Aclaración: Para un transistor de largo de canal suficientemente grande, resulta $V_{DSAT} \cong V_{GS} - V_T$, y entonces para $V_{GT} \geq 0$ y $V_{DS} > V_{DSAT}$ resulta

$$I_D = \frac{k'}{2} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \tag{2}$$

Se concluye entonces que para análisis manual, es necesario conocer 6 parámetros del proceso CMOS. En la Tabla 1 se observan algunos valores típicos de los mismos.

3. Curva característica

En la Fig. 5 se muestra la curva característica de la dependencia de I_D en función de V_{DS} para distintos valores de V_{GS} y para el caso de $V_{SB} = 0$.

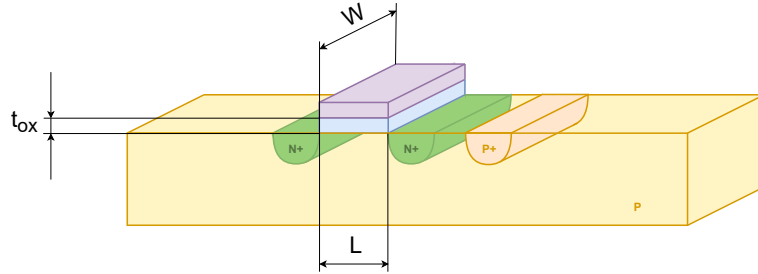


Figura 4: Parámetros geométricos del transistor NMOS.

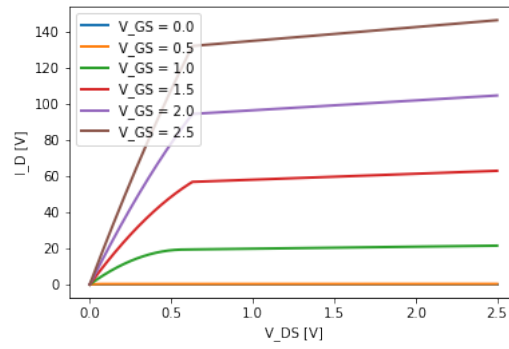


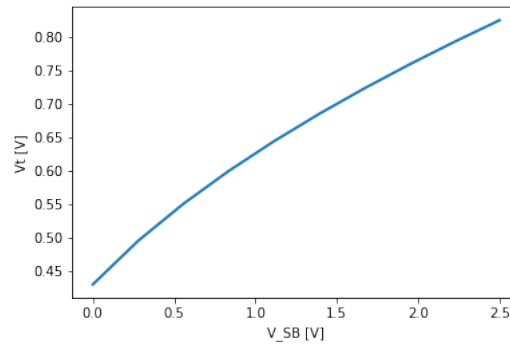
Figura 5: Curva característica de un transistor NMOS de $W=375\text{nm}$, $L=250\text{nm}$, para los parámetros del proceso de la Tabla 1, $V_{SB} = 0$.

	V_{T0} [V]	γ [$V^{1/2}$]	ϕ_F [V]	$V_{D_{SAT}}$ [V]	k' [A/V ²]	λ [1/V]
NMOS	0.43	0.4	-0.3	0.63	115×10^{-6}	0.06
PMOS	-0.4	-0.4	0.3	-1.00	-30×10^{-6}	-0.1

Cuadro 1: Parameters for manual model and typical values.

4. Body bias effect

En la Fig. 6 se muestra la dependencia de V_T en función de V_{SB} . Notar que V_{SB} tiene que ser mayor o igual a 0V para asegurar que la juntura source-body se encuentre siempre en inversa.

Figura 6: Body bias effect on V_T de un transistor NMOS para los parámetros del proceso de la Tabla 1.

5. Resistencia equivalente

Sea el caso que se desea descargar un capacitor C_L a través de un transistor NMOS. Dicho capacitor está cargado inicialmente a un valor de tensión $V_D > V_T$ tal como se observa en la Fig. 7.

Consideremos el caso en el cual V_{GS} es un escalón ideal de tensión de amplitud V_D (Fig. 8). Entonces la descarga de C_L se hará a través de la curva de $I_D(V_{DS})$ para $V_{GS} = V_D$ (Fig. 9). En el instante de tiempo t_0 , el valor de V_{GS} será igual a $V_D > V_T$ por lo cual el transistor empezará a conducir y el capacitor se descargará instantáneamente con un corriente $I_D = I_{D_{A_1}}$. Observemos que el valor de la resistencia equivalente del transistor variará en función del valor instantáneo de V_{DS} :

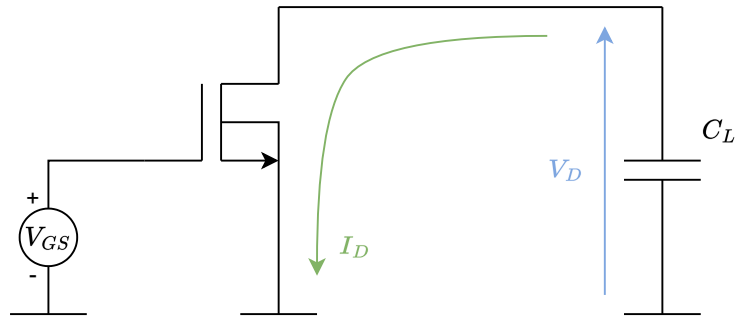


Figura 7: Capacitor descargado a través de un transistor NMOS.

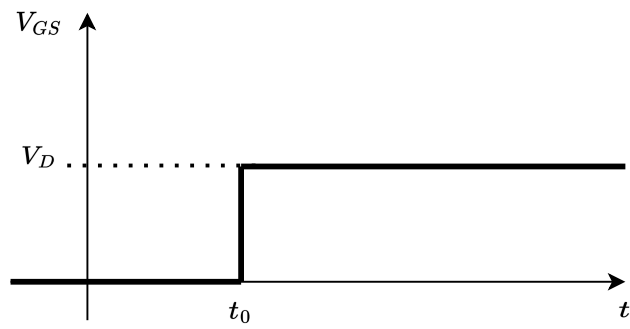


Figura 8: Escalón ideal de amplitud V_D .

$$R_{DS}(V_{DS}) = \left. \frac{V_{DS}}{I_D(V_{DS})} \right|_{V_{GS}=V_D} \quad (3)$$

Inclusive, usando el modelo simplificado para análisis manual dado por la eq. (1), la expresión de R_{DS} sería muy compleja pero además dependiente del punto de operación instantáneo del transistor conforme V_{DS} disminuye mientras el capacitor se descarga. Una opción que permite un análisis cualitativo de la situación es hallar el valor de R_{DS} “promedio”, que lo llamaremos \bar{R}_{DS} . Para ello, se obtiene la pendiente de la recta que une el punto A_i con el origen y luego se promedia dichos valores tal como se observa en la Fig. 9:

$$\bar{R}_{DS} = \frac{1}{N} \sum_{i=1}^N R_{DS_i} \quad (4)$$

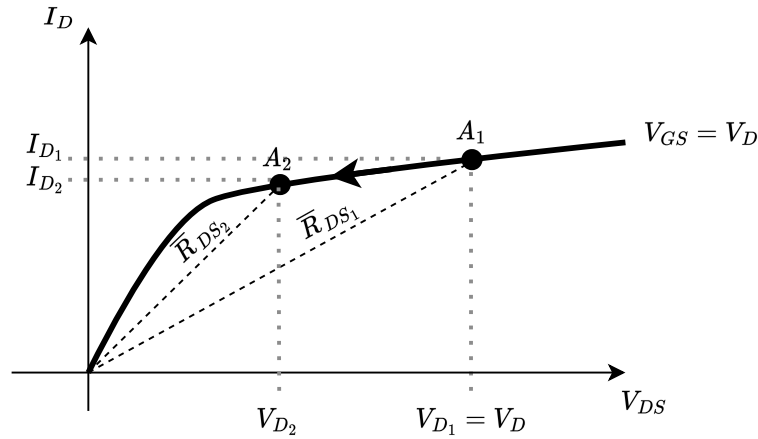


Figura 9: Descarga del capacitor a través de la curva de $I_D(V_{DS})$ para $V_{GS} = V_D$.

Esto puede hacerse a través de una simulación SPICE para cada valor V_{D_i} tal como se observa en la Fig. 10. Promediando los valores de R_{DS} se obtiene $\bar{R}_{DS} = 6457\Omega$.

En la Fig. 11 se muestra la simulación SPICE de un capacitor de 0.1pF el cual es descargado desde el valor inicial de 2.5V a través de un transistor NMOS y una resistencia de 6457 Ω . Observemos la similitud entre ambas curvas lo cual resulta de utilidad para hallar tiempos de subida y bajada donde los valores de tensión instantáneos de la forma de onda no es realmente útil para un circuito digital.

Miremos ahora el valor de resistencia en el punto A_1 cuando $V_{DS} = V_D$ para poder obtener algunas conclusiones útiles. En dicho caso, el modelo de análisis manual (eq. 1) se convierte en:

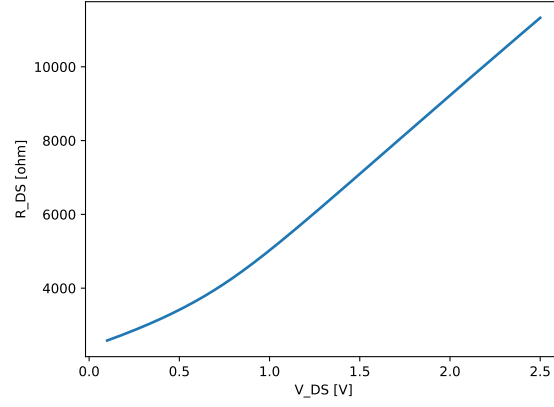


Figura 10: R_{DS} en función de V_{DS} determinados mediante una simulación SPICE para un NMOS de $W=375\text{nm}$, $L=250\text{nm}$.

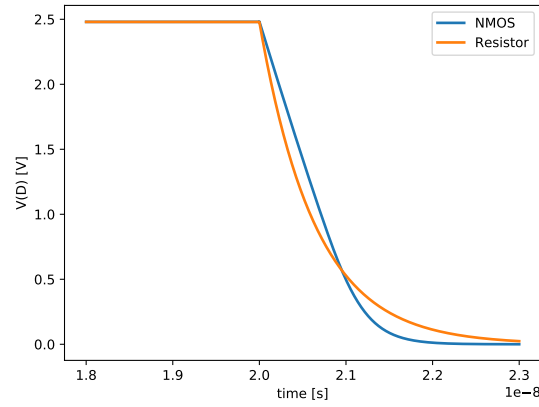


Figura 11: Descarga de un capacitor de 0.1pF a través de un NMOS de $W=375\text{nm}$, $L=250\text{nm}$, y a través de un resistor de 6457Ω .

$$I_{D_1} = \frac{k'}{2} \frac{W}{L} (V_D - V_T)^2 (1 + \lambda V_D) \quad (5)$$

Luego,

$$R_{DS}(V_D) = \frac{V_D}{I_{D_1}} = \frac{2V_D}{k' \frac{W}{L} (V_D - V_T)^2 (1 + \lambda V_D)} \quad (6)$$

De la ec. 6 se observa que aumentar L aumenta proporcionalmente el valor de \bar{R}_{DS} , y al aumentar W disminuye el valor de $R_{DS}(V_D)$ (Fig. 13). En la Fig. 12 se observa la dependencia de $R_{DS}(V_D)$ en función de V_D . Notar que al aumentar el valor de V_D disminuye sustancialmente el valor de $R_{DS}(V_D)$. Este efecto será muy útil en un circuito digital ya que los tiempos de carga y descarga podrán ser modificados dimensionando adecuadamente los transistores (parámetros W y L) y dimensionando adecuadamente la tensión de operación.

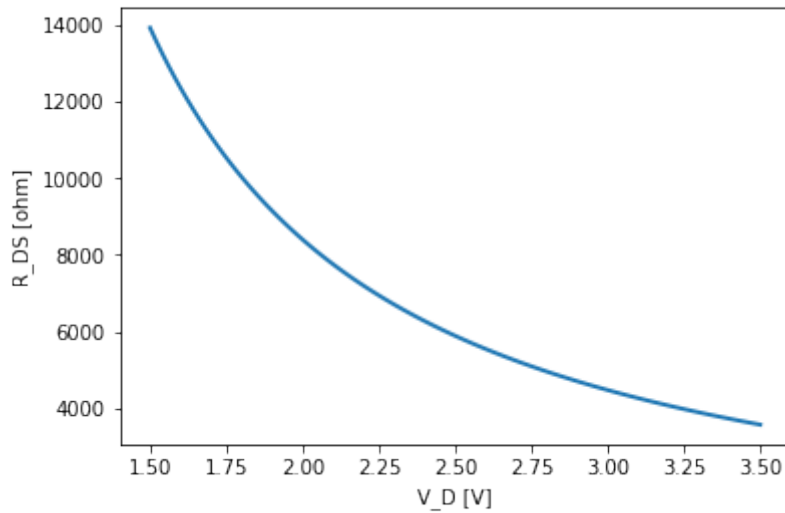


Figura 12: Dependencia de $R_{DS}(V_D)$ en función de V_D , $W = 375\text{nm}$, $L = 250\text{nm}$.

Conclusión:

- Si V_D aumenta, $R_{DS}(V_D)$ disminuye.
- Si $V_D < V_T$ no hay circulación de corriente, $R_{DS}(V_D) = \infty$.
- Si $\frac{W}{L}$ aumenta, $R_{DS}(V_D)$ disminuye.

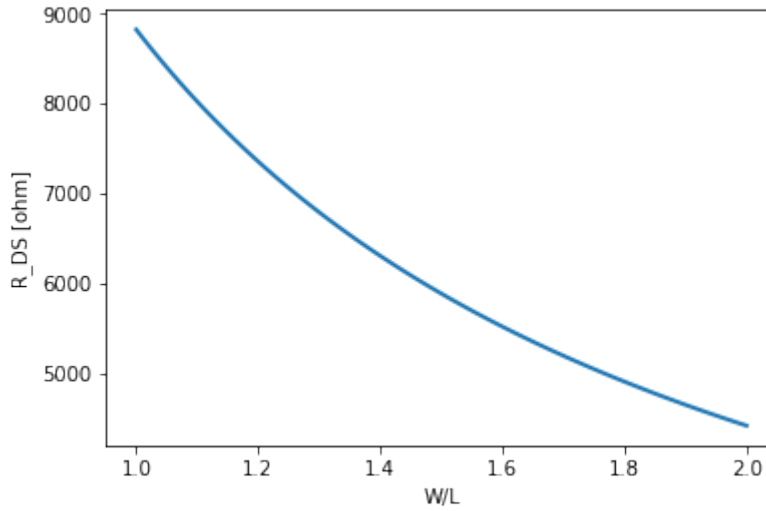


Figura 13: Dependencia de $R_{DS}(V_D)$ en función de W/L , $V_D = 2,5V$.

6. Capacitancia en un MOSFET

El gate del transistor MOS se encuentra separado del canal de conducción por una **capa de óxido**, que tiene una capacitancia por unidad de área dada por $C_{ox} = \epsilon_{ox}/t_{ox}$. La capacitancia total es llamada **Capacitancia del Gate**, C_g , y esta compuesta por dos elementos con distintos comportamientos:

- Una capacidad debida a la estructura topológica del transistor.
- Una capacidad dada por la carga del canal.

6.1. Capacitancia debida a la topología del transistor

Para calcular la capacitancia del MOSFET, observamos la Fig. 14. Podemos ver que **la zona de difusión del source y el drain se extienden por debajo del óxido en una cantidad llamada Difusión Lateral, x_d** . Este solapamiento del gate con las difusiones del source y del drain son **debidas al proceso de fabricación**. Por lo tanto, la **longitud efectiva** del canal será más corta que L_d :

$$L = L_d - 2x_d \quad (7)$$

Esto provoca una capacitancia parásita entre gate-source, y gate-drain, denominada **overlapping capacitance**:

$$C_{GS_o} = C_{GD_o} = C_{ox} \cdot x_d \cdot W = \frac{\varepsilon_{ox}}{t_{ox}} \cdot x_d \cdot W \quad (8)$$

Luego, llamando $C_0 = C_{ox}x_d$, entonces $C_{GS_o} = C_{GD_o} = C_0W$. Observemos que entonces es **proporcional al valor de W** .

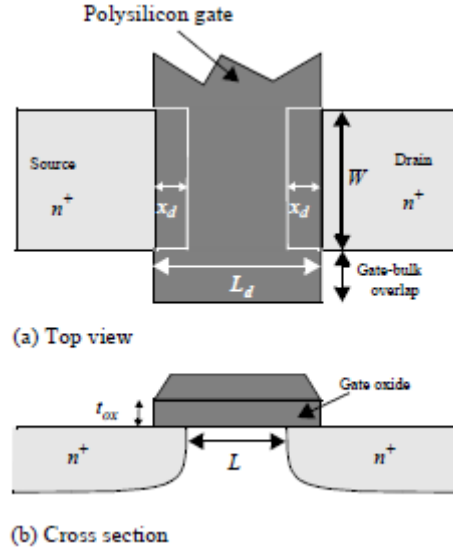


Figura 14: Transistor Mosfet

6.2. Capacitancia del Canal

La capacitancia entre el gate y el canal, C_{GC} se divide en tres componentes:

- $C_{GS_c} \doteq$ Capacitancia Gate-Source.
- $C_{GD_c} \doteq$ Capacitancia Gate-Drain.
- $C_{GB_c} \doteq$ Capacitancia Gate-Body.

Estos componentes son no lineales y varían en los distintos modos de operación del transistor lo cual puede observarse en la Fig. 15. En el modo corte, al no haber canal, toda la capacidad C_{GC} está dada por la capacidad gate-body C_{GB_c} . En el modo resistivo, el body está apantallado por el canal, por lo cual $C_{GB_c} = 0$, y por simetría, $C_{GD_c} = C_{GS_c} = C_{GC}/2$. Finalmente, en modos saturación, el canal está pinchado por lo cual la capacidad gate-drain es casi nula, el body sigue apantallado y entonces $C_{GC} = C_{GS_c}$.

Observemos que siempre la capacidad C_{GC} es proporcional al producto WL ya que el gate contra el canal o el body, forma un capacitor de placas paralelas.

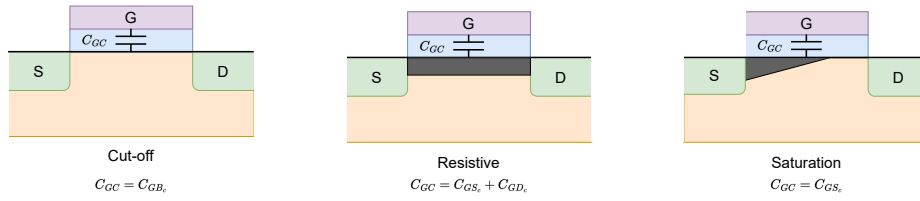


Figura 15: Channel capacitance for the different operating modes.

6.3. Capacitancia de Juntura o de Difusión

Otro componente que aporta a la capacitancia parásita del MOSFET está generado por las junturas PN en inversa formadas por las uniones Source-Body y Drain-Body.

La capacitancia de la zona de vaciamiento a lo largo de dichas junturas es no lineal y decrece a medida que aumenta la corriente inversa.

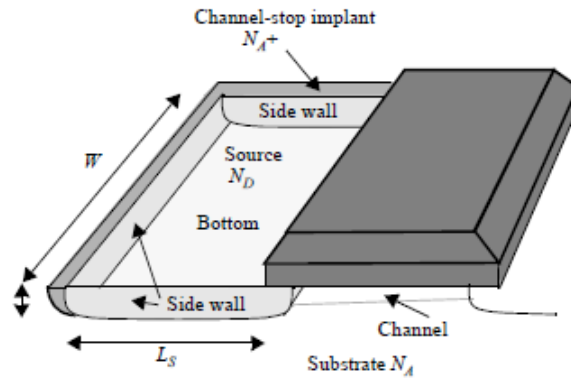


Figura 16: Detalle de las junturas del source.

La Fig. 16 muestra que la juntura esta compuesta por:

- La juntura de la base formada por la región del source, dopado con N_D y el sustrato dopado con N_A . La capacitancia que se genera en esta zona es: $C_{bottom} = C_j \cdot W \cdot L_S$; donde C_j es la capacitancia por unidad de área dada por la capacitancia de juntura del diodo (dependiente de V_D):

$$C_j = \frac{C_{j0}}{\left(1 - \frac{V_D}{\phi_0}\right)^m} \quad (9)$$

- La juntura de la pared lateral, formada por la región del source dopada con N_D

y la zona del límite del canal dopada con N_A^+ . La capacitancia que se genera en esta zona está dada por:

$$C_{sw} = C'_{jsw} \cdot x_j(W + 2 \cdot L_S) \quad (10)$$

Dado que x_j es un parámetro de diseño, conocido como profundidad de la juntura, por lo general se combina con C'_{jsw} para dar una capacitancia por unidad de perímetro, $C_{jsw} = C'_{jsw} \cdot x_j$.

Finalmente, la expresión para la capacitancia total de juntura está dada por:

$$C_{SB_d} = C_{bottom} + C_{sw} = C_j L_S W + C_{jsw}(2L_s + W) \quad (11)$$

siendo $L_s W$ el área, y $2L_s + W$ el perímetro del source. Observemos que resulta proporcional al ancho W del transistor.

Un análisis similar se sigue para el caso capacidad de la juntura drain-body C_{DB_d} .

6.4. Modelo Capacitivo

La ec. 12 muestra todos las componentes a las capacitancias entre terminales de un MOSFET:

$$\begin{aligned} C_{GB} &= C_{GB_c} \\ C_{GD} &= C_{GD_c} + C_{GD_o} \\ C_{GS} &= C_{GS_c} + C_{GS_o} \\ C_{SB} &= C_{SB_d} \\ C_{DB} &= C_{DB_d} \end{aligned} \quad (12)$$

Todas estas capacidades aumentan proporcionalmente con W y en particular, C_{GB_c} , C_{GD_c} , C_{GS_c} lo hacen con WL .

El modelo utilizado para considerar los efectos capacitivos del transistor MOSFET se presenta en la Fig. 17. Se observa que todas las capacidades se encuentran entre un terminal del transistor y VDD o GND con excepción de C_{GD} y C_{GS} .

6.5. Análisis por simulación de la Capacitancia del Gate

Recordemos que la ecuación de la corriente en un capacitor se escribe como:

$$I(t) = C \frac{\partial V(t)}{\partial t} \quad (13)$$

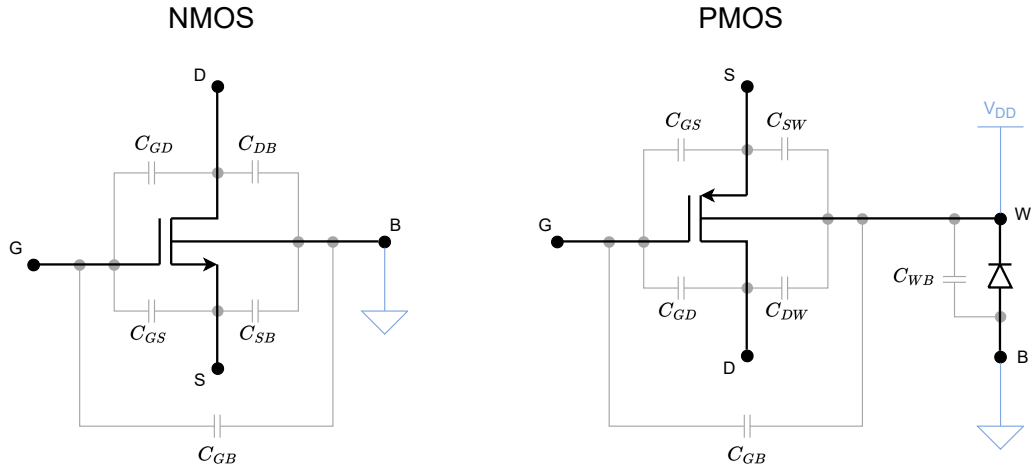


Figura 17: Capacitancia Transistor Mosfet.

Por lo tanto,

$$I = C_G \frac{\partial V_G}{\partial t} \quad (14)$$

Entonces, la capacidad instantánea del gate puede determinarse mediante el circuito de la Fig. 18 utilizando una corriente constante I según:

$$C_G(t) = \frac{I}{\frac{\partial V_G}{\partial t}} \quad (15)$$

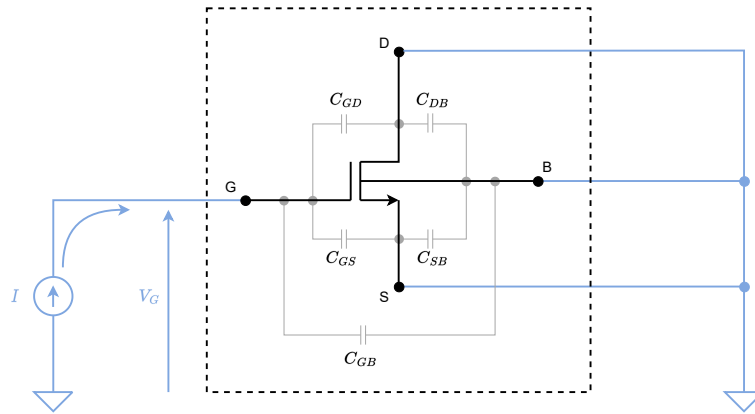


Figura 18: Circuito para determinar la capacidad instantánea del gate.

La Fig. 19 muestra el valor de la capacidad de gate en función de V_G . Se observa la discontinuidad en torno a V_T .

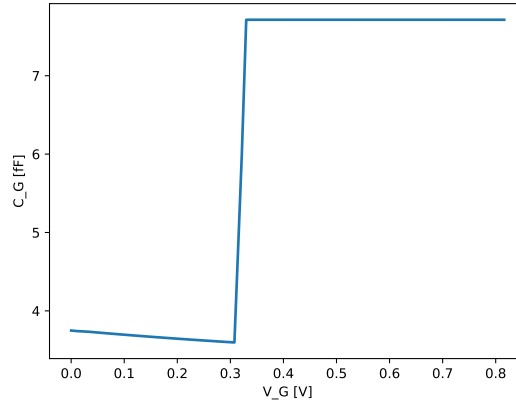


Figura 19: Capacidad de gate C_G en función de V_G , $W = 375\text{nm}$, $L = 250\text{nm}$.

La simulación, tomando a V_G en función del tiempo, muestra como cae la capacitancia cuando V_G se aproxima a V_T conforme aumenta la zona de vaciamiento por debajo del gate, lo que hace que el ancho efectivo del dieléctrico se incremente, y la discontinuidad en torno a V_T cuando el canal empieza a formarse abruptamente.

6.6. Análisis por simulación de la Capacitancia del Drain

La relación entre tensión y corriente en un capacitor dada por la ec. (13) puede ser reescrita como:

$$V(t) = \frac{1}{C} \int_{t_0}^t I(\tau) d\tau + V(t_0) \quad (16)$$

donde $V(t_0) = Q_0/C$ siendo Q_0 la carga inicial del capacitor.

Luego,

$$C = \frac{1}{V(t_{end}) - V(t_0)} \int_{t_0}^{t_{end}} I(\tau) d\tau \quad (17)$$

El circuito de la Fig. 20 se utiliza para determinar la capacidad vista por la fuente V_P . La Fig. 20. En la Fig. 21 se observa la evolución temporal del circuito cuando la capacidad de drain se encuentra inicialmente descargada. En dicha Figura, $t_0 = 10\text{ns}$. Obsérvese que el transistor está cortado y por lo tanto $I_s \cong 0$. Por medio de integración numérica de I_D se determina $C_D = 0,82\text{fF}$. En la Fig. 22 se muestra la evolución de la corriente

de drain del circuito de la Fig. 20 comparada con la corriente de un capacitor de 0,82fF alimentado a través del mismo resistor de 10kΩ.

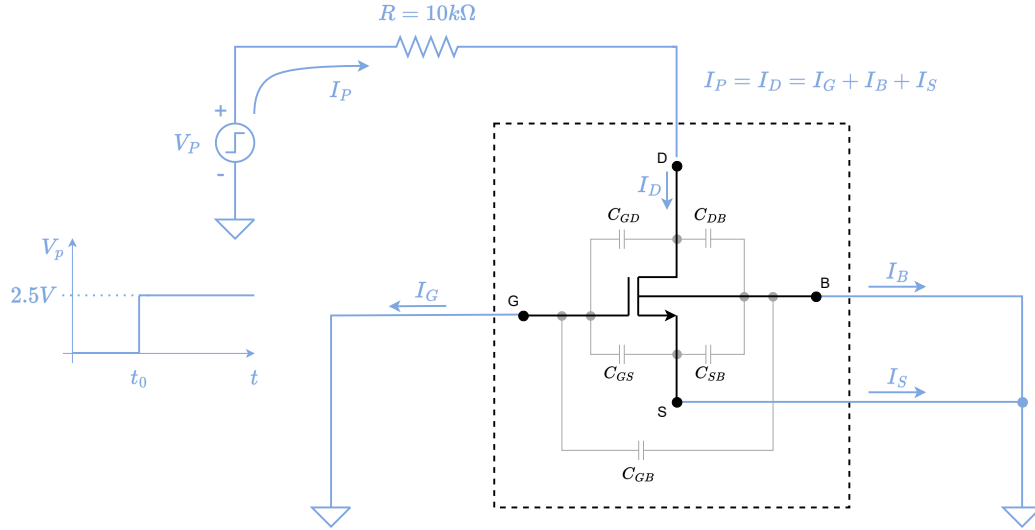


Figura 20: Circuito para determinar la capacidad vista desde el drain, $W = 375\text{nm}$, $L = 250\text{nm}$.

7. Por qué reducir el tamaño del proceso de fabricación aumenta la velocidad

De la ec. (6) podemos resumir que un transistor MOS tiene una $R_{DS} = \alpha_1 \frac{L}{W}$, siendo α_1 un factor de proporcionalidad. Asimismo, posee una capacidad de gate dada por $C_G = \alpha_2 L \cdot W$. Los tiempos de carga y descarga de una capacidad de gate a través de otro transistor MOS serán proporcionales a:

$$\tau_L = R_{DS} C_G = \alpha_1 \frac{L}{W} \alpha_2 L W = \alpha_3 L^2 \quad (18)$$

Por ejemplo, si se reduce a la mitad ambas dimensiones (manteniendo todas las otras variables del proceso constante como ser, dopaje, espesor del óxido, etc) entonces:

$$\tau_{L/2} = R_{DS_2} C_{G_2} = \alpha_1 \frac{L/2}{W/2} \alpha_2 (L/2)(W/2) = \alpha_3 \frac{L^2}{4} \quad (19)$$

En la práctica, muchos de los parámetros del proceso no pueden mantenerse constante al reducir de tamaño, el efecto de saturación de velocidad de portadores en el canal (efecto

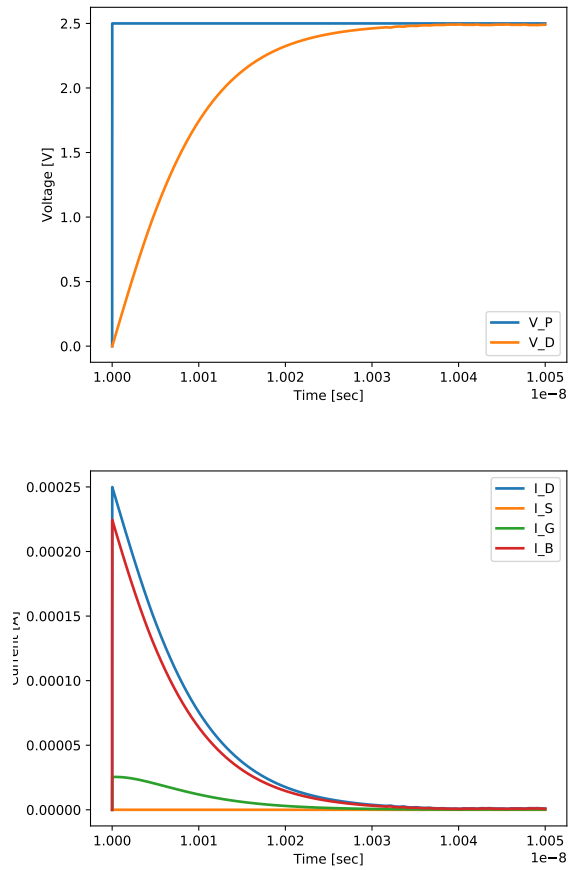


Figura 21: Evolución temporal del circuito de la Fig. 20.

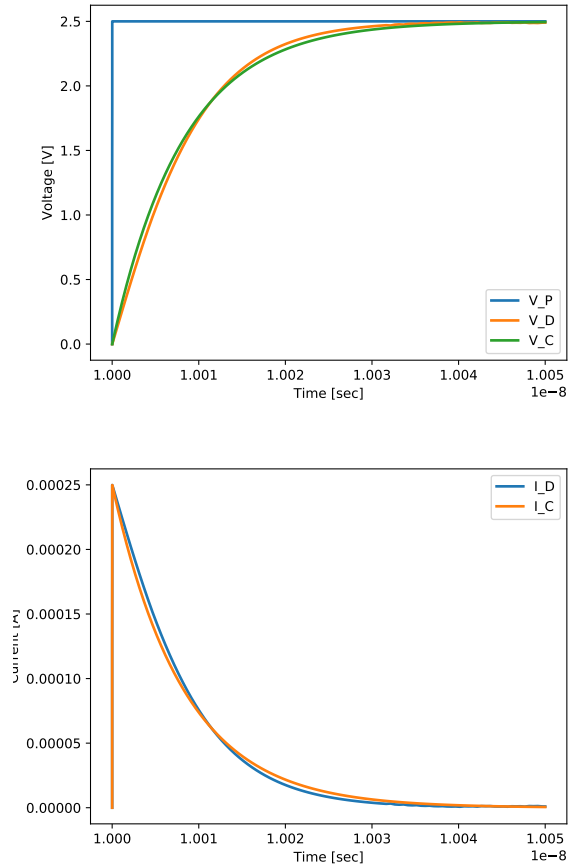


Figura 22: Comparación de la evolución temporal del circuito de la Fig. 20 y un capacitor de 0,82fF alimentado a través de un resistor de 10k Ω .

de canal corte) se hace más pronunciado y por lo tanto los transistores relativamente más resistivos a su tamaño, la tensión de alimentación debe reducirse debido a que un canal más corto tendrá mayor campo eléctrico a través de él si se mantiene el valor de tensión V_{DS} haciendo muy notorio el efecto de canal corto y además aumentando el leakage a través del canal para tensión de V_{GS} por debajo de V_T , etc. Todas estas razones hacen que los tiempos de carga y descarga en la práctica, se ven reducidos al reducir el proceso pero no tanto como hubiese determinado la ec. (19).

8. Condiciones de trabajo (PVT)

Hasta ahora hemos asumido que el modelo de un transistor está determinado por un conjunto de parámetros. Sin embargo, todo transistor está sometido a variaciones del proceso de fabricación como ser diferencias en el dopaje de las difusiones, variaciones en el ancho del óxido, variaciones en la zona de overlapping del drain y el source contra el gate, variaciones en la geometría debido a variaciones en la litografía, etc. Esto puede ocurrir lote a lote, e incluso dentro de un mismo die. Por lo cual, cuando se trabaja con un modelo de transistor, es más conveniente trabajar con un modelo que contemple variaciones aleatorias de los diferentes parámetros del modelo dentro de ciertos límites razonables.

En un circuito digital, las variaciones de proceso en general se traducen en variaciones de retardo de los tiempos de carga y descarga de las capacidades parásitas. Por lo cual, los fabricantes proveen no solo modelos con parámetros aleatorios y los límites de los mismos, sino que además dichos límites de parámetros se combinan de forma tal de proveer el transistor más lento, y el más rápido que puede ser encontrado. Conjuntamente se provee el modelo del transistor típico. Por lo cual, en general se proveen cuatro modelos de transistor.

- Típico: parámetros promedio.
- Slow: límites de los parámetros de forma tal que sus capacidades parásitas y su resistencia de salida sea la más alta posible.
- Fast: límites de los parámetros de forma tal que sus capacidades parásitas y su resistencia de salida sea la más baja posible.
- Random: se especifican los parámetros del transistor en función de su distribución (típicamente gaussianos).

En general, cuando los parámetros tienen distribución gaussiana, los límites máximo y mínimo de cada uno se especifican $\pm 3\sigma$ alrededor del valor medio, lo que significa un 99.7% de probabilidad de encontrar el valor verdadero de dicho parámetro entre esos límites.

Por otro lado, puede suceder que en cierto instante de tiempo la tensión de alimentación pueda sufrir variaciones. Una de las razones de dichas fluctuaciones es causada por la circulación de corriente por las resistencias parásitas de los conductores que forman la red de alimentación (*power grid*).

Esto es debido a dos factores principales:

- Cuando los circuitos lógicos conmutan de estado cargando y descargando capacidades, el consumo de corriente al ser pulsado en el tiempo, produce una caída de tensión pulsda sobre la resistencia de los conductores que conforman la power grid. Esto logra variaciones de tensión de alimentación para cada transistor a lo largo del die.
- Otro factor importante que produce variaciones en la tensión es el causado por inductancias parásitas de la power grid provocando fluctuaciones en la tensión.

Estos dos efectos logran lo que se conoce como *power-ground bouncing*. Como hemos mencionado anteriormente, la resistencia de un transistor MOSFET está directamente relacionada a la tensión de alimentación. El power-ground bouncing causará por lo tanto variaciones en los tiempos de carga y descarga que se producen en los cambios de estado lógico.

Finalmente, la temperatura de operación puede ser en un amplio rango. Muchos parámetros del transistor tienen dependencias marcadas con la temperatura como ser las corrientes de fuga de las junturas polarizadas en inversa, la tensión de umbral V_T , etc. Nuevamente, el amplio rango de operación de temperatura causa variaciones en los tiempos de carga y descarga.

En resumen:

- Velocidad de carga y descarga:
 - Slow: $V_{DD} - 10\%$, T_{max} , Slow Transistor
 - Typ: V_{DD} , 25C (room temperature), Typical Transistor
 - Fast: $V_{DD} + 10\%$, T_{min} , Fast Transistor
- Potencia:
 - Max: $V_{DD} + 10\%$, T_{max} , Fast Transistor
 - Typ: V_{DD} , 25C (room temperature), Typical Transistor
 - Min: $V_{DD} - 10\%$, T_{min} , Slow Transistor

Note: the margin $\pm 10\%$ is the most common case but any other margin could be used like $\pm 5\%$, $\pm 15\%$, $\pm 20\%$, etc.

En la Fig. 23 se observa que un transistor tomado al azar construido con cierto proceso de fabricación se encontrará operando dentro del cubo. Notése que el eje P indica el proceso pero es adimensional es sólo qualitativo con límites fijados entre el *slow* y el *fast* transistor.

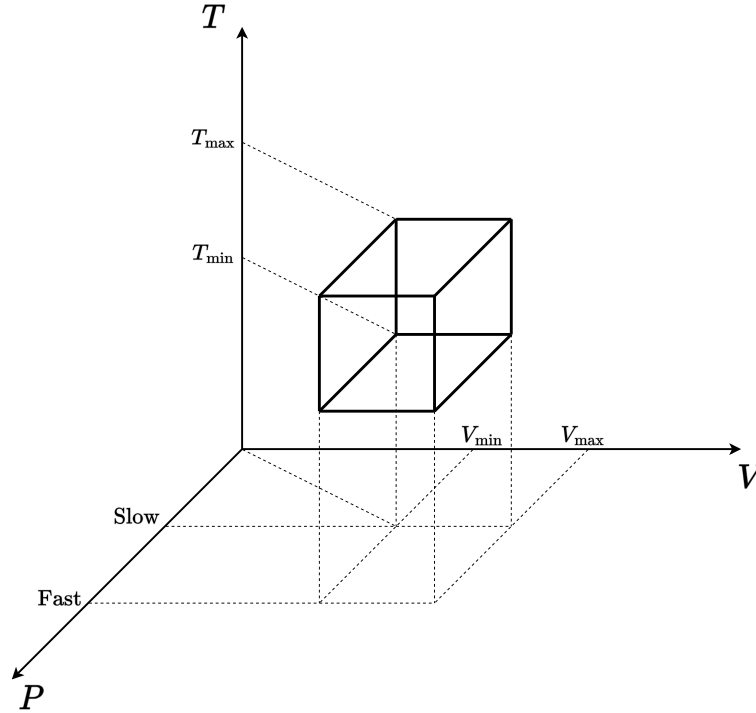
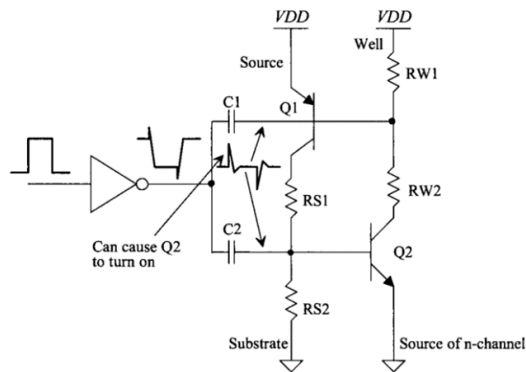


Figura 23: Condiciones PVT.

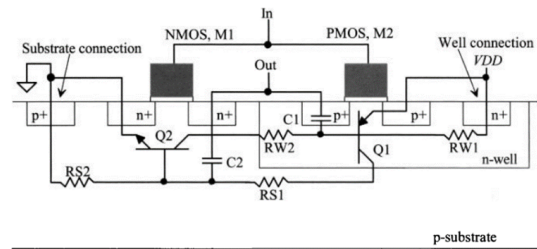
9. Latch-up in CMOS technology

El latch-up es un estado de falla indeseado en un dispositivo CMOS. Se produce debido a un par de transistores bipolares parásitos (NPN y PNP), propios de la estructura CMOS, que forman un tiristor (estructura PNPN). El estado de latch-up se puede disparar por picos de sobretensión en la Fuente de alimentación, pulsos de ruido o radiaciones ionizantes externas.

Cuando el dispositivo entra en el estado de latch-up queda bloqueado y se mantiene en ese estado hasta que se retira la alimentación o se destruye el dispositivo por exceso de corriente.



(a) Schematic used to describe latch-up



(b) Cross-sectional view of an inverter showing parasitic bipolar transistors and resistors.

9.1. Solución por diseño

- Hacer más lentos los tiempos de subida (rise time) y bajada (fall time) de las señales digitales, esto tiene un efecto negativo en circuitos digitales debido al aumento del consumo y la disminución de la velocidad de operación.
- Reducir el área del Drain de los transistores M1 y M2 reduce el valor de C1 y C2 obteniendo el mismo efecto.
- Minimizar las resistencias parásitas RW1 (well) y RS2 (sustrato) para que Q1 y Q2 nunca se disparen.
- Acortar la distancia entre los contactos de well y sustrato en el diseño del layout o incluir varios en paralelo.
- Utilizar una oblea SOI(silicon on Insulator) elimina totalmente el efecto de latch-up debido a la capa de óxido enterrado entre la capa superior e inferior de la oblea.
- Al utilizar una oblea epitaxial (EPI silicon wafer) se reduce la resistencia del bulk al tener una capa epitaxial ligeramente dopada y un sustrato altamente dopado.