

T7 ESTUDIO EN DETALLE DEL INVERSOR CMOS

T7.1. El inversor CMOS

T7.2. Tiempo de propagación y mejora del mismo

T7.3. Disipación de potencia

T7.4. Puertas CMOS

Interesa estudiar en detalle el comportamiento de las puertas CMOS y ese detalle se refiere a tensiones, intensidades, tiempos y consumos: su función de transferencia (tensión de salida en relación con la tensión de entrada), la intensidad suministrable por la salida, los tiempos de conmutación, la disipación de potencia,... El inversor, como puerta más simple, de una sola entrada, resulta sumamente adecuado para desarrollar este estudio.

En el capítulo anterior, se introdujeron los «diagramas de Memelink» como forma de representar gráficamente la distribución de los portadores de carga en el canal de los transistores MOS y de facilitar los cálculos relativos a tensiones e intensidades. En el caso de un inversor, los «diagramas de Memelink» de los dos transistores que lo forman pueden representarse conjuntamente (en un mismo diagrama), de manera tal que puede calcularse la tensión de salida y la intensidad a través del inversor, para cualquier valor de su tensión de entrada.

De esta forma, en el presente capítulo se calcula la tensión de conmutación del inversor y la intensidad que consume en tal situación y se obtienen, también, las expresiones algebraicas que corresponden a los diversos «tramos» de la función de transferencia y, a partir de ella, los márgenes de ruido.

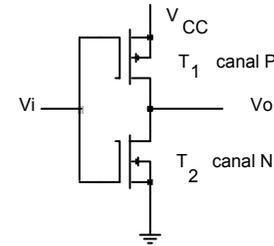
Asimismo es de gran interés comprender los factores que afectan a los tiempos de propagación y cómo dichos tiempos pueden mejorarse mucho, si la carga capacitiva es alta, utilizando una serie de inversores suplementarios, «en escalera» (cada uno de ellos de anchura superior al anterior).

La disipación de potencia en una puerta lógica, su consumo, tiene importancia en dos aspectos complementarios: la fuente de alimentación que ha de proporcionar suficiente intensidad y el calentamiento del circuito que aumenta sus tiempos de propagación y, en ocasiones, necesita adecuada refrigeración.

Por último, se recuerdan las características funcionales de las puertas CMOS y se destaca la limitación del número de entradas de las mismas, para evitar que un alto número de transistores en serie empeore en gran medida esas características.

T7.1. Inversor CMOS

El inversor CMOS utiliza un transistor de canal N y otro de canal P:



Supuestos $V_{CC} = 5\text{ V}$ y $V_{TO,NMOS} = 1\text{ V}$ y $V_{TO,PMOS} = -1\text{ V}$

Para $V_i < 1\text{ V}$ T_1 conduce y T_2 no: $V_o = V_{CC} = V(1)$

Para $V_i > 4\text{ V}$ T_1 no conduce y T_2 sí: $V_o = 0\text{ V} = V(0)$

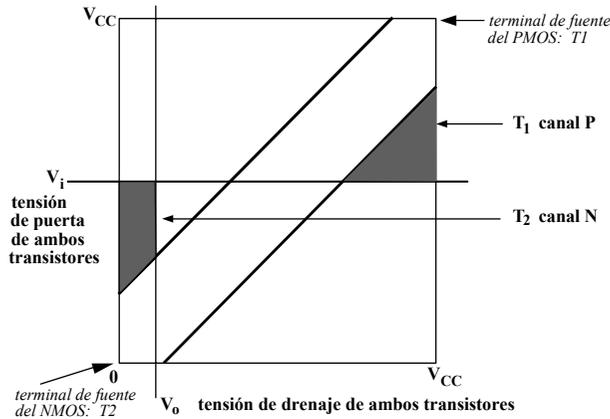
Las propias tensiones umbrales de los transistores aseguran las tensiones de salida booleanas $V(0)$ y $V(1)$ al mantener uno de los dos transistores en corte; en ambas situaciones booleanas, el consumo es nulo pues el transistor cortado impide el paso de corriente a través de ambos (que se encuentran en serie).

En la figura se ha representado también la polarización de los substratos: en el diseño físico han de configurarse las conexiones del substrato P (del transistor NMOS) a 0 V y del substrato N (pozo) a V_{CC} . Tales conexiones de polarización aseguran que todas las uniones PN que forman los transistores con el substrato se encuentran polarizadas inversamente, aislando a cada transistor respecto de su substrato.

El transistor NMOS tiene su terminal de fuente a 0 V y las tensiones de entrada V_{GS} y de salida V_{DS} son positivas y su referencia es 0 V. En cambio, el transistor PMOS tiene su terminal de fuente a V_{CC} (se encuentra colocado «boca abajo») y las tensiones de entrada V_{GS} y de salida V_{DS} son negativas, ambas con referencia a V_{CC} .

Para analizar la función de transferencia del inversor ($V_o - V_i$) puede utilizarse un diagrama de Memelink conjunto de ambos transistores. En dicho diagrama el origen de coordenadas del transistor NMOS (punto 0,0) no coincide con el del transistor PMOS (punto V_{CC}, V_{CC}) ya que la tensión de referencia, tensión de fuente, de ambos transistores es diferente. En cambio, la tensión de entrada V_i (tensión de puerta) de ambos es común y actúa sobre el eje Y , y también lo es su tensión de salida V_o (tensión de drenaje) que actúa sobre el eje X .

En una situación estática, la intensidad que circula por ambos transistores es la misma $I_{T1} = I_{T2}$ ($I_P = I_N$) ya que ambos están en serie; para lo cual tiene que cumplirse que $\alpha_P \cdot A_P = \alpha_N \cdot A_N$, siendo $\alpha = \mu_{COX} \cdot W/L$ el parámetro que incluye los aspectos tecnológicos ($\mu_{COX} = K_p$) y geométricos ($W/L = ff$) y A el efecto de las tensiones aplicadas sobre el transistor, que viene dado (en el diagrama de Memelink) por el área limitada por las cuatro rectas $x = V_S$, $x = V_D = V_o$, $y = V_G = V_i$ e $y = V_{TO} + V$.



$$I_{NMOS} = I_{PMOS} \quad \alpha_N \cdot A_N = \alpha_P \cdot A_P \quad \mu_{electrones} \cdot C_{ox} \cdot W_N/L_N \cdot A_N = \mu_{huecos} \cdot C_{ox} \cdot W_P/L_P \cdot A_P$$

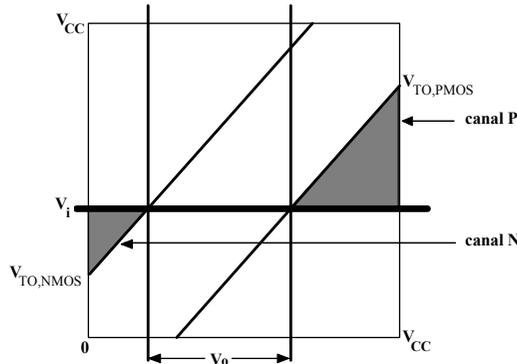
$$\frac{A_N}{A_P} = \frac{\alpha_P}{\alpha_N} = \frac{\mu_{huecos}}{\mu_{electrones}} \frac{W_P/L_P}{W_N/L_N}; \quad \frac{\mu_{huecos}}{\mu_{electrones}} = \mu'; \quad \frac{A_N}{A_P} = \mu' \frac{ff_P}{ff_N} = \mu' \cdot \beta$$

donde $\mu' = \mu_{huecos} / \mu_{electrones}$, relación entre las movilidades de huecos (transistor PMOS) y electrones (NMOS), es del orden de 1/3 y $\beta = (W_1/L_1) / (W_2/L_2) = ff_1 / ff_2$ es la relación de geometrías entre ambos transistores.

Para cualquier tensión de entrada V_i la tensión de salida V_o es la que corresponde a la siguiente relación entre áreas en el diagrama de Memelink: $A_N = \mu' \cdot \beta \cdot A_P$.

Tensión de conmutación

La tensión de conmutación corresponde a aquella tensión de entrada V_i en la que las áreas de conducción de ambos transistores en el diagrama de Memelink son triangulares: en tal situación, según se aprecia en la figura, la tensión de salida V_o no es un valor único, sino un amplio intervalo de tensiones; dicho intervalo corresponde, precisamente, a la conmutación de la salida.



La tensión de entrada V_i a la cual se produce la conmutación es aquella para la cual las áreas de los dos triángulos rectángulos e isosceles (formados por V_i con $V_{TO} + V$) cumplen la relación anterior, ya que para dicha tensión de entrada existe todo un intervalo de tensiones de salida posibles, intervalo que corresponde al salto brusco de la conmutación.

$$\left. \begin{aligned} A_P &= \frac{1}{2} (V_{CC} - |V_{TO,PMOS}| - V_i)^2 \\ A_N &= \frac{1}{2} (V_i - V_{TO,NMOS})^2 \end{aligned} \right\} A_N = \mu' \cdot \beta \cdot A_P$$

Sea $V_{uP} = |V_{TO,PMOS}|$ el valor absoluto de la tensión umbral del transistor PMOS (dicha tensión es negativa), $V_{uN} = V_{TO,NMOS}$ la tensión umbral del transistor NMOS, V_{conm} la tensión de conmutación del inversor y $k' = \sqrt{\mu' \cdot \beta}$.

$$\left. \begin{aligned} A_P &= \frac{1}{2} (V_{CC} - V_{uP} - V_{conm})^2 \\ A_N &= \frac{1}{2} (V_{conm} - V_{uN})^2 \end{aligned} \right\} A_N = \mu' \cdot \beta \cdot A_P$$

$$\rightarrow (V_{conm} - V_{uN}) = k' \cdot (V_{CC} - V_{uP} - V_{conm}) \rightarrow V_{conm} = \frac{V_{uN} + k' \cdot (V_{CC} - V_{uP})}{1 + k'}$$

La tensión de conmutación del inversor puede calcularse mediante la expresión

$$V_{conm} = \frac{V_{uN} + k' \cdot (V_{CC} - V_{uP})}{1 + k'} \quad \text{donde } k' = \sqrt{\mu' \cdot \beta}.$$

Si las tensiones umbrales son iguales en valor absoluto ($V_{uN} = V_{uP}$) y los transistores son iguales en dimensiones (relación de geometrías $\beta = 1$), la función de transferencia no es simétrica ($V_{conm} \neq V_{CC}/2$) sino que la conmutación está desplazada hacia valores inferiores de V_i , debido a la menor movilidad de los huecos que repercute en una menor conductividad del transistor PMOS.

En la conmutación, la salida puede adoptar cualquier tensión dentro del intervalo señalado en la figura que viene delimitado por los vértices de ambos triángulos. Teniendo en cuenta que la recta $V_{TO} + V$ tiene una pendiente de 45°, los dos catetos de dichos triángulos son iguales y su valor es respectivamente:

NMOS: $V_{conm} - V_{uN};$ $V_o(\text{vértice del triángulo}) = V_{conm} - V_{uN}$

PMOS: $V_{CC} - V_{uP} - V_{conm};$
 $V_o(\text{vértice del triángulo}) = V_{CC} - (V_{CC} - V_{uP} - V_{conm}) = V_{conm} + V_{uP}$

Es decir, la conmutación se produce con una variación vertical de la tensión de salida entre $V_{conm} - V_{uN}$ y $V_{conm} + V_{uP}$.

La intensidad que circula por el inversor en la situación de conmutación será:

$$I_{\text{conm}} = \alpha \cdot \mathbf{A}_N = K_p \cdot \frac{W}{L} \cdot \mathbf{A}_N = K_p \cdot \frac{W}{L} \cdot \frac{1}{2} (V_{\text{conm}} - V_{uN})^2.$$

Para la tecnología Ψ :

$$V_{CC} = 5 \text{ V}; \quad V_{uN} = 1 \text{ V}; \quad V_{uP} = 1 \text{ V};$$

$$\mu_{\text{electrones}} = 600 \text{ cm}^2/\text{Vs}; \quad \mu_{\text{huecos}} = 200 \text{ cm}^2/\text{Vs}; \quad \mu' = 1/3;$$

supuestos ambos transistores iguales en dimensiones,

por ejemplo, ambos de tamaño mínimo: $ff_P/ff_N = \beta = 1$

$$k' = \sqrt{\mu' \cdot \beta} = \sqrt{1/3} \approx 0,58$$

la tensión de conmutación será:

$$V_{\text{conm}} = \frac{V_{uN} + k' \cdot (V_{CC} - V_{uP})}{1 + k'} = \frac{1 + 0,58 \cdot (5 - 1)}{1 + 0,58} = 2,1 \text{ V}$$

La tensión de salida conmutará entre los siguientes valores:

$$\text{NMOS:} \quad V_o(\text{vértice del triángulo}) = V_{\text{conm}} - V_{uN} = 1,1 \text{ V}$$

$$\text{PMOS:} \quad V_o(\text{vértice del triángulo}) = V_{\text{conm}} + V_{uP} = 3,1 \text{ V}$$

y la intensidad en dicha situación es:

$$K_p (\text{NMOS}) \sim 40 \mu\text{A}/\text{V}^2; \quad W/L = 1,5$$

$$I_{\text{conm}} = K_p \cdot \frac{W}{L} \cdot \frac{1}{2} (V_{\text{conm}} - V_{uN})^2 = 40 \frac{\mu\text{A}}{\text{V}^2} \cdot 1,5 \cdot 0,5 \cdot (2,1 - 1)^2 \text{ V}^2 = 36 \mu\text{A}.$$

La función de transferencia no es simétrica ($V_{\text{conm}} = 2,1 \text{ V} \neq 2,5 \text{ V}$) sino que la conmutación está desplazada hacia valores inferiores de V_i , debido a la menor movilidad de los huecos que repercute en una menor conductividad del transistor PMOS.

Es posible conseguir una función de transferencia simétrica, compensando, a través de β , la relación entre las movilidades de los dos transistores: en este caso en que ambas tensiones umbrales son iguales y $\mu' = 1/3$, para $\beta = 3$ resulta $k' = \sqrt{\mu' \cdot \beta} = 1$, $\mathbf{A}_P = \mathbf{A}_N$ y $V_i = 2,5 \text{ V}$.

Para ello es necesario hacer el transistor PMOS de anchura triple, de manera que ocupará mayor superficie de integración y, además, aumentarán los tiempos de propagación de la puerta anterior: los transistores PMOS dejan de ser mínimos y aumentan sus dimensiones y su capacidad de puerta. No es, por ello, una buena solución y, en la práctica, no suele hacerse este «equilibrado» de los transistores PMOS.

Habida cuenta de que la precisión del proceso de fabricación es limitada, cabe preguntarse en qué medida una desviación en las dimensiones de los transistores afecta a la tensión de conmutación.

Supongamos que, por alguna razón, se modifican las dimensiones de los transistores y no resultan exactamente iguales; expresando la tensión de conmutación V_{conm} en función de la relación de geometrías β :

$$V_{\text{conm}} = \frac{V_{uN} + k' \cdot (V_{CC} - V_{uP})}{1 + k'} = \frac{V_{uN} + \sqrt{\mu' \cdot \beta} \cdot (V_{CC} - V_{uP})}{1 + \sqrt{\mu' \cdot \beta}}$$

Para la tecnología Ψ :

$$V_{\text{conm}} = \frac{V_{uN} + \sqrt{\mu' \cdot \beta} \cdot (V_{CC} - V_{uP})}{1 + \sqrt{\mu' \cdot \beta}} = \frac{1 + 0,58 \cdot \sqrt{\beta} \cdot (5 - 1)}{1 + 0,58 \cdot \sqrt{\beta}} = \frac{1 + 2,3 \cdot \sqrt{\beta}}{1 + 0,58 \cdot \sqrt{\beta}}$$

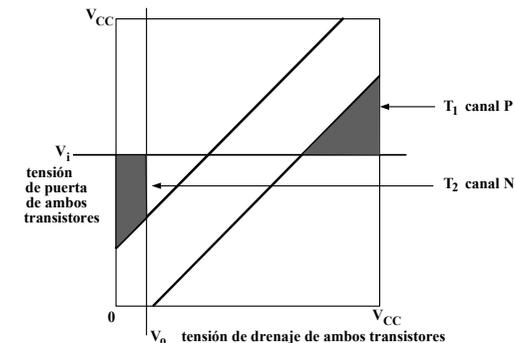
- si β en lugar de 1 pasa a valor 2 (modificación muy amplia, del 100 %, que equivale a un factor de forma del transistor PMOS doble) entonces $V_i = 2,34 \text{ V}$ (desviación en la tensión de conmutación de $0,24 \text{ V} < 12 \%$);

- en cambio, si β pasa a 0,5 (factor de forma del NMOS doble), $V_i = 1,86$ (desviación análoga a la anterior, $0,24 \text{ V} < 12 \%$).

Es decir, el efecto de una desviación en las dimensiones de los transistores sobre la tensión de conmutación del inversor es muy pequeño.

Función de transferencia

El *diagrama de Memelink* del inversor permite obtener la curva de transferencia ($V_o - V_i$) punto a punto: dado un valor de V_i , calcular el correspondiente de V_o ;



La relación entre la tensión de salida y la de entrada se obtiene a partir de la igualdad de intensidades en ambos transistores, $\mathbf{A}_N = \mu' \cdot \beta \cdot \mathbf{A}_P$, expresando el área de conducción de cada transistor en función de las tensiones de entrada V_i y de salida V_o .

Puede efectuarse un análisis por tramos, obteniendo las expresiones algebraicas que relacionan V_o con V_i en cada tramo.

Los tramos diferenciados son los siguientes:

- I $V_i < V_{TO,NMOS}$ el transistor N no conduce y $V_o = V_{CC}$
- II $V_i < V_{conmutación}$ el área de conducción del transistor N es triangular y, en cambio, el área correspondiente al P es trapezoidal
- III $V_i = V_{conmutación}$ el área de conducción del transistor N es triangular y el área correspondiente al P también es triangular
- IV $V_i > V_{conmutación}$ el área de conducción del transistor N es trapezoidal y, en cambio, el área correspondiente al P es triangular
- V $V_i > V_{CC} - |V_{TO,PMOS}|$ el transistor P no conduce y $V_o = 0$

Para la tecnología Ψ :

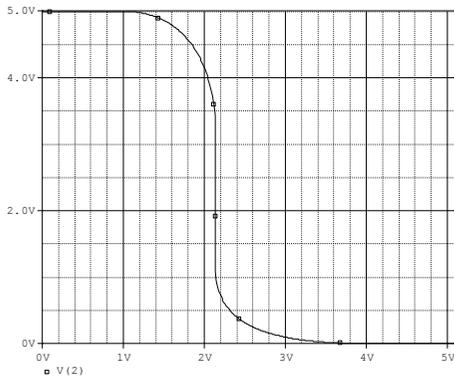
- I $V_i < 1\text{ V}$ $V_o = 5\text{ V}$
- II $1\text{ V} < V_i < 2,1\text{ V}$ $V_o = V_i + 1 + \sqrt{13 - 2V_i - 2V_i^2}$
- III $V_i = 2,1\text{ V}$ $V_o = [1,1 ; 3,1]$
- IV $2,1\text{ V} > V_i < 4\text{ V}$ $V_o = V_i - 1 - \sqrt{(2V_i^2 + 2V_i - 13) / 3}$
- V $V_i > 4\text{ V}$ $V_o = 0\text{ V}$

En conmutación, $V_{conm} = 2,1\text{ V}$, las expresiones de los tramos II y IV proporcionan los valores de 3,1 V y 1,1 V entre los cuales conmuta el inversor:

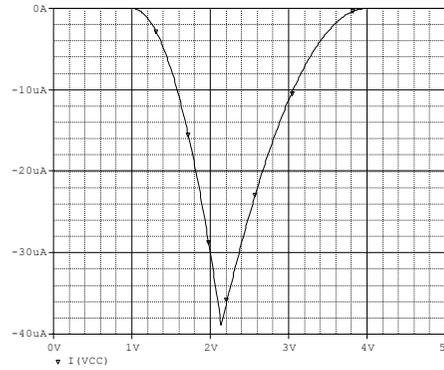
$$V_i = V_{conm} \quad V_o = [3,1 ; 1,1]\text{ V}$$

También puede obtenerse la función de transferencia mediante simulación SPICE: las siguientes gráficas muestran, respectivamente, la función de transferencia $V_o - V_i$ y el consumo de intensidad de un inversor CMOS de la tecnología Ψ :

$L = 1\ \mu\text{m}$, $W = 1,5\ \mu\text{m}$, $|V_{TO}| = 1\text{ V}$, $K_{p,NMOS} = 40\ \mu\text{A}/\text{V}^2$ y $K_{p,PMOS} = 15\ \mu\text{A}/\text{V}^2$.



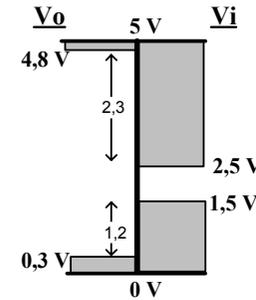
Función de transferencia $V_o - V_i$ CMOS



Consumo de intensidad I_{CC} del inversor

En la anterior función de transferencia se aprecia que, en relación con el «ruido» (para determinar mayores márgenes respecto al ruido), es razonable extender la tensión de entrada correspondiente al **0** hasta 1,5 V (para dicho intervalo, $V_o(\mathbf{1}) > 4,8\text{ V}$) y la tensión de entrada para valor **1** puede tomarse desde 2,5 V (intervalo para el cual $V_o(\mathbf{0}) < 0,3\text{ V}$):

$$\text{para } V_i < 1,5\text{ V}, \quad V_o > 4,8\text{ V} \quad \text{y} \quad \text{para } V_i > 2,5\text{ V}, \quad V_o < 0,3\text{ V}.$$



Márgenes de ruido en tensión:

$$\Delta V(\mathbf{0}) = V_{i,\text{máx}}(\mathbf{0}) - V_{o,\text{máx}}(\mathbf{0}) = 1,5 - 0,3 = 1,2\text{ V}$$

$$\Delta V(\mathbf{1}) = V_{o,\text{mín}}(\mathbf{1}) - V_{i,\text{mín}}(\mathbf{1}) = 4,8 - 2,5 = 2,3\text{ V}$$

Es mayor el margen correspondiente al **1** a causa del desplazamiento de la función de transferencia hacia tensiones bajas, debido a la menor movilidad de los huecos.

Margen de ruido en potencia: $\Delta P = (\Delta V)^2 / R_o$. El margen de ruido en potencia (o sea, en energía por unidad de tiempo) es más significativo que el margen en tensión, ya que el ruido actúa como energía perturbativa sobre el circuito digital.

Resistencia de salida

La resistencia de salida es un parámetro indicativo del comportamiento de las puertas lógicas (una referencia de calidad de las mismas), en cuanto a intensidad disponible en la salida, inmunidad frente al ruido y tiempos de propagación, pues afecta fuertemente a estos tres aspectos: en principio, cuanto menor sea la resistencia de salida mayores serán la intensidad suministrable por la misma, el margen frente al ruido y la velocidad de trabajo.

También interesa resistencia de salida baja en relación con el acoplo en tensión, pero dicho acoplo ya viene garantizado por el altísimo valor de la resistencia de entrada (que es cuasi-infinita).

Las situaciones booleanas corresponden a un transistor en zona lineal: para salida **0** el transistor NMOS se encontrará conduciendo en su zona lineal, mientras que para salida **1** será el transistor PMOS el que se encuentre en zona óhmica; de forma que $R_o(\mathbf{0})$ y $R_o(\mathbf{1})$ corresponden, respectivamente, a las resistencias que presentan los transistores NMOS y PMOS en zona lineal.

Valores de las resistencias de salida (tecnología Ψ):

Para transistores de tamaño mínimo ($W=1,5L$; $ff=1,5$) con tensión umbral de 1 V y tensión de puerta de 5 V los valores de la resistencia equivalente en zona lineal serán:

NMOS	PMOS
$K_p \approx 40 \mu\text{A}/\text{V}^2$	$K_p \approx 15 \mu\text{A}/\text{V}^2$
$R_{\text{eq}} = \frac{1}{K_p \cdot W/L \cdot (V_G - V_{T0})}$	
$R_{\text{eq}} = 1 / (40 \cdot 10^{-6} \cdot 1,5 \cdot (5-1)) \approx 4 \text{ k}\Omega$	$R_{\text{eq}} = 1 / (15 \cdot 10^{-6} \cdot 1,5 \cdot (5-1)) \approx 11 \text{ k}\Omega$

De manera que las resistencias de salida en situación booleana serán:

$R_o(0) \sim 4 \text{ K}\Omega$	$R_o(1) \sim 11 \text{ K}\Omega$
---------------------------------	----------------------------------

Los valores de las resistencias de salida de los inversores de tamaño mínimo son del orden de varios $\text{K}\Omega$, $\sim 10^3$ ohmios; no son valores pequeños, pero son adecuados para el comportamiento de las puertas booleanas en el interior del circuito integrado. Téngase en cuenta que las capacidades de entrada de dichas puertas son pequeñas (del orden del femtofaradio, 10^{-15} F) y, por ello, también lo son las intensidades que se requieren y las constantes de tiempo que se generan ($10^3 \cdot 10^{-15} \sim \text{ps}$); asimismo, habida cuenta de que las dimensiones de las conexiones son muy reducidas, el efecto del ruido es muy pequeño.

En los terminales de salida de un circuito integrado interesan resistencias de salida inferiores a 100 ohmios (pues en este caso, se requieren intensidades del orden de 10 mA, las capacidades de carga son del orden de 10 pF y las longitudes de las conexiones van en centímetros); en cambio, en las puertas interiores son adecuadas resistencias del orden de 10 K (por las razones antedichas: intensidades del orden de 0,1 mA, capacidades de 1 fF y longitud de las conexiones en micras).

En todo caso, puede disminuirse el valor de la resistencia de salida aumentando en igual medida la anchura W del correspondiente transistor, el PMOS para $R_o(1)$ y el NMOS para $R_o(0)$; esto es lo que se hace, precisamente, en los adaptadores de las salidas del circuito integrado.

La intensidad suministrable por el inversor, en cada uno de sus dos estados, depende de la caída de tensión que admitamos en la salida: $I_o = \Delta V_o / R_o$. Esta intensidad es relativamente pequeña y, en ocasiones, no es suficiente la correspondiente a transistores de tamaño mínimo. Por ejemplo, en los citados adaptadores de las salidas, que precisan de una capacidad de intensidad (hacia el exterior del circuito integrado) del orden de 10 mA. En tales casos, basta aumentar adecuadamente el factor de forma de los transistores ($ff = W / L$), es decir, es necesario hacer $W \gg L$.

Para la tecnología Ψ :

NMOS	PMOS
$R_o(0) \sim 4 \text{ K}\Omega$	$R_o(1) \sim 11 \text{ K}\Omega$

Intensidad suministrable por el inversor: $I_o = \Delta V_o / R_o$

Admitiendo $\Delta V(0) = 0,2$ V y $\Delta V(1) = 0,6$ V que son valores aceptables (cuya diferencia va en relación a la asimetría de la función de transferencia y a la diferencia de márgenes de ruido):

$I_o(0) \sim 0,2 / 4\text{K} \approx 50 \mu\text{A}$	$I_o(1) \sim 0,6 / 11\text{K} \approx 50 \mu\text{A}$
--	---

Haciendo $W > L$ se consiguen mayores intensidades de salida:

para $W = 15L$	$I_o \approx 0,5 \text{ mA}$
----------------	------------------------------

para $W = 50L$	$I_o \approx 1,5 \text{ mA}$
----------------	------------------------------

para $W = 300L$	$I_o \approx 10 \text{ mA}$
-----------------	-----------------------------

Margen de ruido en potencia del inversor con transistores de tamaño mínimo:

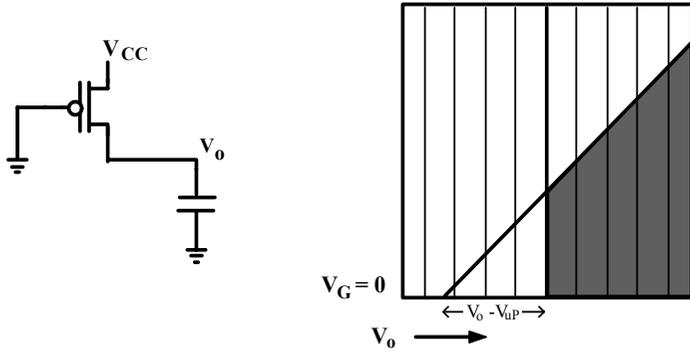
$\Delta P(0) \sim 1,2^2 / 4 \text{ K} \approx 0,36 \text{ mW}$	$\Delta P(1) \sim 2,3^2 / 11 \text{ K} \approx 0,48 \text{ mW}$
--	---

El ruido actúa efectivamente como potencia: energía perturbativa por unidad de tiempo. En los terminales de un circuito integrado se requieren márgenes de ruido superiores de 20 mW para ambientes «normales» (viviendas o lugares públicos) y a 100 mW en ambientes «industriales»; en el interior de los circuitos integrados (en donde la captación y efecto del ruido es muy inferior, ya que las conexiones son «milimétricas»), son suficientes márgenes del orden de 0,1 mW.

T7.2. Tiempo de propagación y mejora del mismo

T7.2.1. Tiempo de conmutación de una puerta CMOS

La salida de una puerta booleana se encontrará conectada a la entrada de otra u otras puertas (u otros componentes del circuito global), cuyo efecto equivalente es el de una capacidad que ha de cargarse y descargarse en la conmutación, originando unos tiempos de retraso hasta que el proceso de carga o descarga alcanza los valores de tensión apropiados para el 0 y 1 booleanos.



Proceso de carga a través de un transmisor PMOS: conmutación de la salida de 0 a 1

Estudiemos el proceso de carga, expresando los tiempos de retraso en función de la intensidad con canal saturado $I_{SAT} = I$ y de la capacidad de carga $C_L = C$ que soporta:

$I = I_{SAT} = \alpha \cdot A_{canal\ saturado}$, siendo A el área del triángulo de conducción en el diagrama de Memelink;

En el proceso de carga, existen dos tiempos diferenciados:

t_1 : V_o de 0 a $V_{uP} = |V_{TO,PMOS}|$ canal saturado intensidad constante $I_1 = I$,

$$\Delta Q = I \cdot t = C \cdot \Delta V \quad \Delta V = V_{uP} \quad t_1 = C \cdot V_{uP} / I$$

t_2 : V_o de V_{uP} a V_{CC} canal no saturado $I_2 < I$ y decrece al aumentar V_o ,

$$dQ = I(t) dt = C dV$$

el área de conducción corresponde ahora a la diferencia entre dos triángulos: el propio de la saturación de canal del transistor y el que determina la tensión de salida:

$$A = A_{canal\ saturado} - 1/2 \cdot (V - V_{uP})^2 \quad I(t) = I - \alpha \cdot (V - V_{uP})^2 / 2$$

expresando α en función de I : $I = I_{SAT} = \alpha \cdot A_{canal\ saturado} = \alpha \cdot 1/2 \cdot (V_{CC} - V_{uP})^2$;

$$\alpha = 2 \cdot I / (V_{CC} - V_{uP})^2 \quad I(t) = I - I \cdot (V - V_{uP})^2 / (V_{CC} - V_{uP})^2$$

$$I(t) dt = C dV \quad (I - I \cdot (V - V_{uP})^2 / (V_{CC} - V_{uP})^2) dt = C dV$$

$$\frac{I \cdot (V_{CC} - V_{uP})^2 - I \cdot (V - V_{uP})^2}{(V_{CC} - V_{uP})^2} dt = C dV$$

$$\frac{I}{C \cdot (V_{CC} - V_{uP})^2} dt = \frac{dV}{(V_{CC} - V_{uP})^2 - (V - V_{uP})^2}$$

Integrando ambos miembros de la ecuación diferencial anterior:

$$\frac{I}{C \cdot (V_{CC} - V_{uP})^2} t = \frac{1}{(V_{CC} - V_{uP})} \text{arc tgh} \frac{(V - V_{uP})}{(V_{CC} - V_{uP})}$$

$$\frac{I}{C \cdot (V_{CC} - V_{uP})} t = \text{arc tgh} \frac{(V - V_{uP})}{(V_{CC} - V_{uP})}$$

$$V(t) = V_{uP} + (V_{CC} - V_{uP}) \cdot \text{tgh} \frac{I}{(V_{CC} - V_{uP}) \cdot C} t$$

La función tangente hiperbólica $\text{tgh}(x)$ es fuertemente creciente (del tipo $1 - e^{-x}$, con pendiente aún mayor): tiende a 1 asintóticamente y alcanza el valor 0.76 para $x = 1$. Para dicho valor ($x = 1$), la salida habrá recorrido, aproximadamente, el 80% de la tensión de alimentación:

$$V_0 = V_{uP} + (V_{CC} - V_{uP}) \cdot 0,76 = 0,76 \cdot V_{CC} + 0,24 \cdot V_{uP} \approx 0,8 V_{CC}$$

(habida cuenta de que V_{uP} se encontrará, generalmente, entre $0,2 \cdot V_{CC}$ y $0,4 \cdot V_{CC}$).

La conmutación de la salida hacia V_{CC} alcanzará $0,8V_{CC}$ en un tiempo aproximado:

$$\frac{I}{(V_{CC} - V_{uP}) \cdot C} t_2 = 1 \quad t_2 = \frac{(V_{CC} - V_{uP}) \cdot C}{I}$$

$$t = t_1 + t_2 = \frac{V_{uP} \cdot C}{I} + \frac{(V_{CC} - V_{uP}) \cdot C}{I} = \frac{V_{CC} \cdot C}{I} \quad t_{pLH} = t_1 + t_2 = V_{CC} \cdot C / I$$

donde I es la intensidad de canal saturado del transistor PMOS.

La determinación del tiempo de descarga a través del transistor NMOS hasta disminuir a $0,2V_{CC}$ es análoga y conduce a la misma expresión $t_{pHL} = V_{CC} \cdot C / I$:

$$t_{pLH} = V_{CC} \cdot C / I(\text{PMOS})$$

$$t_{pHL} = V_{CC} \cdot C / I(\text{NMOS}).$$

La intensidad de canal saturado del transistor PMOS $I(\text{PMOS})$ suele ser menor que la del NMOS $I(\text{NMOS})$ como consecuencia de la menor movilidad de los huecos que conforman el canal del transistor PMOS y, consiguientemente, $t_{pLH} > t_{pHL}$: el tiempo de subida es superior al de bajada. Esta desigualdad entre ambos tiempos de propagación podría evitarse dimensionando adecuadamente el transistor PMOS, pero implicaría mayor área de integración y, a la vez, mayor capacidad de carga de cada entrada lo cual redundaría en empeorar los tiempos de conmutación de la puerta anterior. Por ello, no interesa dimensionar las puertas lógicas para conseguir su simetría funcional.

Las expresiones de los tiempos de propagación parecen indicar que dichos tiempos disminuyen al hacerlo la tensión de alimentación V_{CC} , pero sucede al revés ya que la intensidad de canal saturado también depende de la tensión de alimentación y disminuye fuertemente con ella: $I_{SAT} = \alpha \cdot A_{canal\ saturado} = K_p \cdot W/L \cdot 1/2 \cdot (V_{CC} - V_{uP})^2$.

Para la tecnología Ψ :

Para transistores de tamaño mínimo ($L = 1 \mu\text{m}$, $W = 1,5 \mu\text{m}$) el valor de la capacidad de puerta será: $C_G = C_{ox} \cdot W \cdot L = 0,7 \text{ fF} / \mu\text{m}^2 \cdot 1,5 \mu\text{m} \cdot 1 \mu\text{m} \approx 1 \text{ fF}$

y como cada entrada se encuentra conectada a dos transistores, NMOS y PMOS, la capacidad de entrada será: $C_i \approx 2 \text{ fF}$.

Para $V_{CC} = 5 \text{ V}$ y $V_{TO} = 1 \text{ V}$: $I(P) = I_{sat}(\text{canal P}) \approx 0,2 \text{ mA}$

$I(N) = I_{sat}(\text{canal N}) \approx 0,5 \text{ mA}$

y con un $\text{fan-out} = 1$ (una sola entrada conectada a la salida de un inversor mínimo):

$t_{pLH} \approx 5 \text{ V} \cdot 2 \text{ fF} / 0,2 \text{ mA} \sim 0,05 \text{ ns}$ $t_{pHL} \approx 5 \text{ V} \cdot 2 \text{ fF} / 0,5 \text{ mA} \sim 0,02 \text{ ns}$

Nos encontramos con tiempos de conmutación por debajo del nanosegundo, que dependen linealmente de la capacidad de carga y, por ello, del fan-out y de las dimensiones de los transistores:

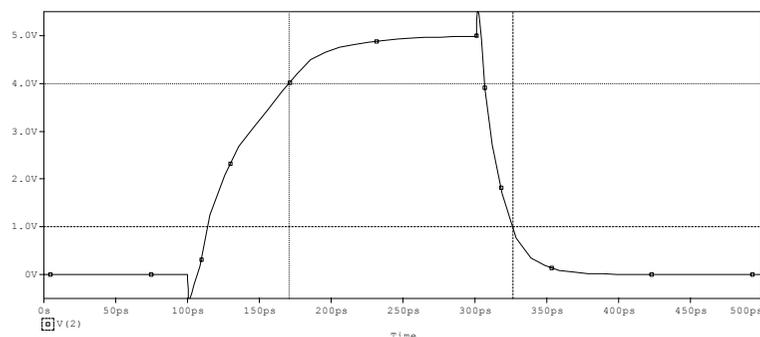
- al aumentar el fan-out , los tiempos de conmutación crecen en la misma proporción
- al disminuir las dimensiones de la tecnología los tiempos de conmutación se reducen cuadráticamente (C_i disminuye con L y con W).

Para una tecnología de 0,5 micras: $L_{\text{mín}} = 0,5 \mu\text{m}$, $W = 0,75 \mu\text{m}$, $C_i \sim 0,5 \text{ fF}$
con la misma alimentación $V_{CC} = 5 \text{ V}$, tensión umbral $V_{TO} = 1 \text{ V}$ y $\text{fan-out} = 1$:

$t_{pLH} \approx 5 \text{ V} \cdot 0,5 \text{ fF} / 0,2 \text{ mA} \sim 0,015 \text{ ns}$ $t_{pHL} \approx 5 \text{ V} \cdot 0,5 \text{ fF} / 0,5 \text{ mA} \sim 0,005 \text{ ns}$

La siguiente gráfica muestra la conmutación de un inversor con un fan-out de 1, es decir, a su salida se encuentra conectado otro inversor análogo; la simulación SPICE se ha efectuado con los datos de la tecnología Ψ y transistores de tamaño mínimo:

$L = 1 \mu\text{m}$, $W = 1,5 \mu\text{m}$, $|V_{TO}| = 1 \text{ V}$, $K_{p,NMOS} = 40 \mu\text{A}/\text{V}^2$ y $K_{p,PMOS} = 15 \mu\text{A}/\text{V}^2$.



Onda de conmutación de un inversor: $V(2) = V_o$

Los tiempos medidos en esta gráfica son: $t_{pLH} \approx 0,071 \text{ ns}$ y $t_{pHL} \approx 0,026 \text{ ns}$.

El retardo que introduce un inversor es algo superior al tiempo de conmutación calculado en este apartado: dicho tiempo de conmutación ha sido determinado en relación con la capacidad de carga C_L (fan-out) y, en la práctica, influyen también otros efectos capacitivos internos del propio inversor.

Sin ser exactos, las formulas y cálculos anteriores ofrecen una buena estimación del orden de magnitud y una buena aproximación al valor real de los tiempos de propagación. La simulación SPICE de los mismos puede proporcionar aún mayor precisión.

Además, habrá que tener en cuenta que cada conexión entre la salida de una puerta y la entrada de la siguiente introduce una línea conductora con sendos efectos resistivo y capacitivo y, con ello, un nuevo sumando al retardo de propagación de las señales; en tecnologías submicrónicas no es extraño que el retardo introducido por las líneas de conexión sea superior al tiempo de propagación de la propia puerta booleana.

Por otra parte, los tiempos de conmutación dependen fuertemente de la temperatura, ya que al aumentar ésta disminuye fuertemente la movilidad de los portadores (aumentan sus choques con los núcleos de la red cristalina) y, en consecuencia, se reduce K_p y la intensidad de saturación de los transistores. De ahí, el alto interés de evitar el calentamiento de los circuitos integrados, disipando adecuadamente el calor generado por la potencia consumida en ellos, si se desea aprovechar la velocidad máxima de trabajo.

«Equilibrado» de transistores (tecnología Ψ):

Si «equilibramos» ambos transistores para que su transconductancia K_p sea la misma:

$$W_{PMOS} = 3 \cdot W_{\text{mín}}$$

las intensidades de canal saturado se igualan: $I(P) = I(N) \approx 0,5 \text{ mA}$

la capacidad de entrada se duplica:

$$C_i = C_{ox} \cdot (W_{NMOS} \cdot L_{NMOS} + W_{PMOS} \cdot L_{PMOS}) = C_{ox} \cdot (4 \cdot W_{\text{mín}} \cdot L_{\text{mín}}) \approx 4 \text{ fF}$$

y también se igualan los tiempos de propagación:

$$t_{pLH} \approx 5 \text{ V} \cdot 4 \text{ fF} / 0,5 \text{ mA} \sim 0,04 \text{ ns} \quad t_{pHL} \approx 5 \text{ V} \cdot 4 \text{ fF} / 0,5 \text{ mA} \sim 0,04 \text{ ns}$$

Pero, calculando el retardo que introducen dos inversores seguidos, resultaría:

$$t_2 \text{ inversores} = t_{pLH} + t_{pHL} \sim 0,08 \text{ ns}$$

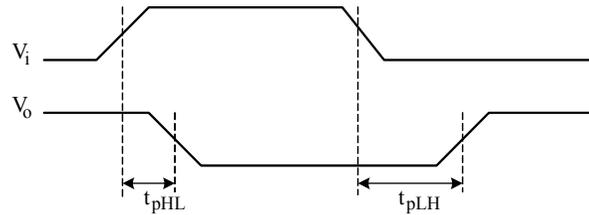
un tiempo de propagación superior al que resulta con dos inversores de tamaño mínimo:

$$t_2 \text{ inversores mínimos} = t_{pLH} + t_{pHL} \sim 0,05 + 0,02 \text{ ns} = 0,07 \text{ ns}$$

O sea que, «equilibrando» el transistor PMOS de forma que conduzca igual que el NMOS y sus tiempos de propagación sean iguales, ocupamos mayor área de silicio y la velocidad de trabajo (considerada en conjuntos de puertas sucesivas) resulta inferior: en principio, no es una buena solución y, en la práctica, no suele hacerse.

Al conectar varios inversores seguidos, el tiempo de propagación del conjunto es mayor que los tiempos de conmutación individuales pero es inferior a la suma de ellos, ya que existe solapamiento entre ellos. Es decir, los tiempos de conmutación no son linealmente acumulativos ya que cada puerta inicia su conmutación antes de que la anterior complete la suya.

Para facilitar la suma de tiempos, cada tiempo de propagación suele medirse por el retraso entre el punto medio de la conmutación (tensión $V_{cc}/2$) de la onda de entrada y el punto medio de la conmutación de la señal de salida.



T7.2.2. Mejora de los tiempos de propagación

En ocasiones las capacidades de carga son altas; por ejemplo, en los adaptadores de las salidas de los circuitos integrados (que han de ser capaces de soportar su conexión a cargas equivalentes de decenas de picofaradios) o en puertas que deban soportar a su salida un *fan-out* alto (por ejemplo, las que transmiten a los diversos biestables la señal de reloj) o largas longitudes de polisilicio. En tales casos, los tiempos de propagación se elevan en demasía y es preciso reducirlos mediante adaptadores de intensidad.

Para la tecnología Ψ :

Para una capacidad de carga de 50 pF (25.000 veces superior a C_i -*fan-out* 1-):

$$t_{pLH} = V_{CC} \cdot C / I(P) \approx 5 \text{ V} \cdot 50 \text{ pF} / 0,2 \text{ mA} \approx 1,5 \mu\text{s}$$

$$t_{pHL} = V_{CC} \cdot C / I(N) \approx 5 \text{ V} \cdot 50 \text{ pF} / 0,5 \text{ mA} \approx 0,5 \mu\text{s}$$

tiempos que resultan relativamente altos y limitan la velocidad por debajo del MHz.

Para disminuir los tiempos de propagación causados por altas capacidades de carga o, lo que es lo mismo, por la necesidad de intensidades de salida altas, se utilizan esquemas de amplificación en cascada con inversores cuya anchura de transistor es progresivamente creciente.

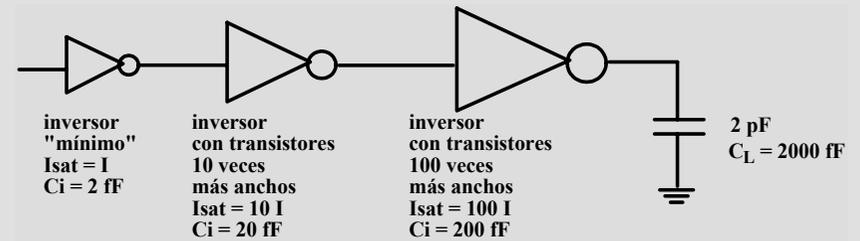
Ejemplo desarrollado con la tecnología Ψ :

Supongamos una carga de 2 pF que resulta ser unas 1.000 mayor que C_i (capacidad de carga que corresponde a *fan-out* 1):

$$t_{pLH} = V_{CC} \cdot C / I(P) \approx 5 \text{ V} \cdot 2 \text{ pF} / 0,2 \text{ mA} \approx 50 \text{ ns}$$

$$t_{pHL} = V_{CC} \cdot C / I(N) \approx 5 \text{ V} \cdot 2 \text{ pF} / 0,5 \text{ mA} \approx 20 \text{ ns}$$

Añadiendo un par de inversores cuya anchura de transistores sea, en cada uno de ellos, 10 veces la del anterior:



Como puede apreciarse en la figura, en cada inversor se produce un «salto» entre su capacidad de carga y su capacidad de entrada de 10; en total, un salto de 1000 que es la relación existente entre los 2 pF y C_i (capacidad de entrada del inversor mínimo).

$$t_{1pLH} \approx 5 \cdot 20 / 0,2 \approx 0,5 \text{ ns} \quad t_{2pLH} \approx 5 \cdot 200 / 2 \approx 0,5 \text{ ns} \quad t_{3pLH} \approx 5 \cdot 2000 / 20 \approx 0,5 \text{ ns}$$

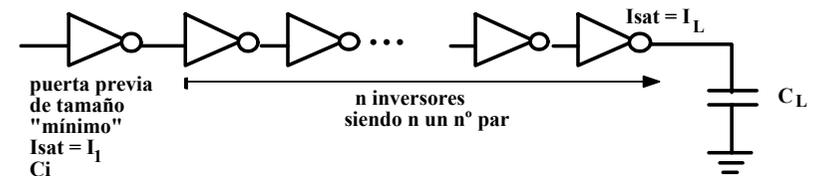
$$t_{1pHL} \approx 5 \cdot 20 / 0,5 \approx 0,2 \text{ ns} \quad t_{2pHL} \approx 5 \cdot 200 / 5 \approx 0,2 \text{ ns} \quad t_{3pHL} \approx 5 \cdot 2000 / 50 \approx 0,2 \text{ ns}$$

$$t_{pLH} \sim t_{1pLH} + t_{2pLH} + t_{3pLH} \approx 1,2 \text{ ns} \quad t_{pHL} \sim t_{1pHL} + t_{2pHL} + t_{3pHL} \approx 0,9 \text{ ns}$$

Ambos tiempos son muy inferiores a los que presenta el primer inversor en solitario: 50 ns y 20 ns, respectivamente.

Este esquema (en que se incluyen parejas de inversores «en escalera», es decir, con anchura de transistores creciente) resulta sumamente útil para configurar adaptadores de intensidad (*buffers*), tanto para los terminales de salida del circuito integrado como para aquellas conexiones internas que han de soportar un alto *fan-out* o alta carga capacitiva y, por tanto, han de proporcionar alta intensidad.

Un adaptador de intensidad puede incluir más de una pareja de inversores «en escalera» (siempre un número par para que no se produzca una inversión suplementaria), de forma que la relación de intensidad entre el último de ellos y la puerta «mínima» inicial se distribuya en múltiples «saltos» pequeños.



Se puede calcular el número óptimo de inversores para conseguir el menor tiempo de propagación global:

- sea **K** el cociente entre la capacidad de carga que debe soportar el conjunto y la correspondiente a una entrada mínima (primera puerta del conjunto): $K = C_L / C_i$
- si el necesario aumento del *fan-out* se distribuye de forma homogénea siendo ∂ el coeficiente incremental $C_{i+1} = \partial \cdot C_i$, será necesario que las anchuras **W** de transistores de dos inversores sucesivos se incrementen en dicho factor ∂ : $W_{i+1} = \partial \cdot W_i$
- la relación de anchuras repercute en proporción directa en la relación entre capacidades de entrada **C_i** y entre intensidades de saturación **I_{sat}**; se verificará que

$$W_{i+1} / W_i = C_{i+1} / C_i = I_{sat,i+1} / I_{sat,i} = \partial$$

- el factor incremental global **K** corresponde a **n+1** incrementos ∂ sucesivos:
 $K = C_L / C_{in} = \partial^{n+1}$, $\partial = \sqrt[n+1]{K}$ y el incremento en intensidad será: $I_L / I_1 = \partial^n$

Los tiempos de propagación de los inversores serán iguales, ya que la relación entre intensidad de saturación y capacidad de carga es constante, y dichos tiempos serán proporcionales a ∂ ya que dicho factor representa la carga relativa a su tamaño que cada inversor soporta. Por tanto, el tiempo de propagación del conjunto será proporcional a $(n+1) \cdot \partial$ y la situación de tiempo mínimo de propagación corresponde a:

$$\frac{d[(n+1) \cdot \partial]}{dn} = 0; \quad \partial = \sqrt[n+1]{K}$$

$$\frac{d((n+1) \cdot \sqrt[n+1]{K})}{dn} = 0; \quad \sqrt[n+1]{K} + (n+1) \cdot \sqrt[n+1]{K} \cdot \ln K \cdot \left[-\frac{1}{(n+1)^2} \right] = 0$$

$$1 - \ln K \cdot \frac{1}{(n+1)} = 0; \quad n+1 = \ln K; \quad K = e^{n+1};$$

$$\partial = \sqrt[n+1]{K} = e = 2,7$$

La adaptación óptima de la «escalera de inversores», para conseguir tiempos mínimos de propagación, se consigue cuando el número de inversores es $n+1 = \ln K$ y, en este caso, el coeficiente incremental ∂ coincide con el número e:

$$W_{i+1} = 2,7 \cdot W_i \quad C_{i+1} = 2,7 \cdot C_i \quad I_{sat,i+1} = 2,7 \cdot I_{sat,i}$$

El tiempo de propagación de cada inversor será $\partial = e = 2,7$ veces el correspondiente a un inversor mínimo conectado a otro inversor análogo.

En el ejemplo desarrollado con la tecnología Ψ :

$$C_L = 2 \text{ pF}; \quad C_{in} = 2 \text{ fF}; \quad K = 1000; \quad n+1 = \ln(1000) \approx 7$$

El mínimo tiempo de propagación se obtiene con 6 inversores intermedios, cuya anchura de transistores sea, en cada uno de ellos, 2,7 veces mayor que en el anterior.

Los tiempos de propagación de cada uno de los inversores son los siguientes:

$$t_{pLH} = 2,7 \cdot 5 \text{ V} \cdot 2 \text{ fF} / 0,2 \text{ mA} \approx 0,15 \text{ ns} \quad t_{pHL} = 2,7 \cdot 5 \text{ V} \cdot 2 \text{ fF} / 0,5 \text{ mA} \approx 0,05 \text{ ns}$$

De forma que, los tiempos de propagación para el conjunto total serán:

$$t_{pLH} \sim 4t_{pLH} + 3t_{pHL} \approx 0,75 \text{ ns} \quad t_{pHL} \sim 4t_{pHL} + 3t_{pLH} \approx 0,65 \text{ ns}$$

Tales son los tiempos mínimos de propagación que pueden conseguirse, claramente inferiores a los que presentaría el primer inversor actuando en solitario sobre la capacidad de carga 2 pF (50 ns y 20 ns, respectivamente) y solamente un poco inferiores a los que se obtienen con una sola pareja de inversores en escalera (1,2 ns y 0,9 ns).

Como puede apreciarse añadir una primera pareja de inversores (adecuadamente dimensionados) tiene un gran efecto respecto a la reducción de tiempos; cuando se añaden más parejas (ajustando en cada caso su anchura **W**), la inclusión de cada una de ellas tiene un efecto menor que la anterior. Por ello, debe valorarse la reducción de tiempos que se desea conseguir en relación con el área de silicio que se va a necesitar.

T7.3. Disipación de potencia

La potencia disipada por el inversor en situación estática es prácticamente nula: no existe ninguna línea de conducción directa entre alimentaciones (uno de los dos transistores de cada entrada se encuentra en corte). Ahora bien, en cada transistor existen uniones PN (difusión-substrato) polarizadas inversamente que conducen una minúscula intensidad inversa, de orden de magnitud del picoamperio ($\text{mA} / 10^6$).

Ahora bien, al conmutar el inversor han de cargarse o descargarse las capacidades propias de sus transistores, en particular, su capacidad de puerta, lo cual determina un consumo dinámico proporcional a la frecuencia de conmutación:

- energía necesaria para cargar o descargar un condensador: $1/2 C \cdot V^2$
- potencia disipada al cargarlo y descargarlo con una onda de frecuencia **f** (en cada período hay dos conmutaciones, habrá que contabilizar una carga y una descarga): $(1/2 C \cdot V^2) \cdot 2f = C \cdot V^2 \cdot f$

Esta potencia (energía por unidad de tiempo) representa un consumo de intensidad desde la alimentación y una disipación de calor en el propio circuito. Son dos aspectos complementarios, relativos al consumo de intensidad, que requieren la correspondiente atención en el diseño y utilización del circuito integrado: una fuente de alimentación con suficiente capacidad de suministro de corriente y una disipación de calor adecuada. Teniendo en cuenta, además, que los tiempos de propagación aumentan con la temperatura y, en consecuencia, disminuye la velocidad de trabajo.

Conforme a la anterior expresión, la potencia consumida y disipada depende de:

- la tecnología, cuyas dimensiones determinan la capacidad **C**; dicha capacidad equivalente de una puerta a efectos de consumo de intensidad tiene varios componentes, entre los cuales predomina la capacidad de entrada (las otras capacidades internas de los transistores son de valores muy inferiores) que es proporcional a la superficie de las zonas de puerta, es decir, al producto **L.W** y disminuye cuadráticamente al disminuir **L** ($W = 1,5L$);
- la tensión de alimentación, que también afecta cuadráticamente **V²**; existe, por ello, una evolución continuada de la microelectrónica hacia tensiones de alimentación más bajas (el paso de 5 V a 3 V reduce el consumo a la tercera parte), siendo cada vez más frecuentes dispositivos de 2,5 V y de 1,8 V;
- la frecuencia de trabajo **f**, aumentando linealmente con ella (lo cual aconseja no trabajar a frecuencia más alta de la estrictamente necesaria para el sistema digital)

Al evaluar el consumo de un circuito integrado, debe tenerse en cuenta que no todas sus puertas conmutan cada vez, sino que solamente lo hace una pequeña parte de ellas; por ejemplo, la escritura de un dato sobre una memoria de 1 Mega solamente afecta a uno de sus 1048576 registros (y al decodificador de direcciones y control de escritura).

Disipación de potencia (tecnología Ψ):

Para transistores de tamaño mínimo el valor de la capacidad de puerta será:
 $C_{ox} = 0,7 \text{ fF} / \mu\text{m}^2$; $C_G = C_{ox} \cdot W \cdot L = 0,7 \text{ fF} / \mu\text{m}^2 \cdot 1,5 \mu\text{m} \cdot 1 \mu\text{m} \approx 1 \text{ fF}$.
 y el de la capacidad de entrada: $C_i = 2 \cdot C_G = \approx 2 \text{ fF}$

Para la tensión de alimentación habitual de 5 V ($V_{CC} = 5 \text{ V}$)

$$P = C \cdot V^2 \cdot f = 2 \text{ fF} \cdot 25 \text{ V}^2 \cdot f = 50 \cdot 10^{-15} \cdot f = 0,05 \cdot f \text{ pW}$$

$$\text{Para } f = 1 \text{ MHz} = 10^6: \quad P = 50 \text{ nW} / \text{MHz}; \quad I_{CC} = 10 \text{ nA} / \text{MHz}$$

La potencia disipada en cada entrada aumenta linealmente con la frecuencia.

Es posible hacer una estimación de la potencia disipada en 1 mm^2 ocupado por puertas CMOS, supuesto que todos los transistores conmuten (situación no habitual pues en cualquier bloque digital sólo conmuta una pequeña fracción de sus puertas).

El área activa ocupada por los transistores en cuanto a su zona de puerta ($L \cdot W$) suele ser inferior al 5 % (ya que la mayor parte de la superficie es necesaria para las zonas de fuente y drenaje, conexiones, separaciones,...):

$$C_{\text{transistores}} = C_{ox} \cdot \text{superficie activa} = 0,7 \text{ fF}/\mu\text{m}^2 \cdot 1 \text{ mm}^2 \cdot 5\% = 35 \text{ pF}$$

$$P = C \cdot V^2 \cdot f = 875 \text{ pW} \cdot f \sim 1 \text{ nW} \cdot f$$

Para $f = 1 \text{ MHz}$, $P \sim 1 \text{ mW} / \text{MHz}$: supuesto que todos los transistores contenidos en 1 mm^2 de silicio conmutasen, disiparían en forma de calor una potencia del orden de $1 \text{ mW}/\text{MHz}$; potencia que no depende de las dimensiones de la tecnología sino del porcentaje de ocupación: superficie activa/superficie total.

Además de la disipación de potencia de tipo capacitivo (principalmente la capacidad de entrada de los transistores MOS) existe otro efecto dinámico debido a que en la conmutación, durante un breve instante de tiempo, conducen ambos transistores PMOS y NMOS, dando lugar a un estrecho «pico de intensidad»; tal efecto resulta despreciable frente al anterior siempre que la conmutación sea adecuadamente rápida (tiempos de conmutación inferiores a 1 ns).

Sin embargo, dichos «picos de intensidad» deben ser tenidos en cuenta en relación con el «ruido», ya que al ser variaciones muy rápidas de intensidad generarán «picos de tensión» sobre las líneas de alimentación (por efecto inductivo).

El orden de magnitud del «pico de intensidad» en la conmutación fue calculado en el apartado anterior, por medio del área de conducción en el *diagrama de Memelink*; en situación de conmutación ($V_i = V_{\text{conm}}$), la intensidad que circula por el inversor, será:

$$I_{\text{conm}} = K_p \cdot \frac{W}{L} \cdot A_N = K_p \cdot \frac{W}{L} \cdot \frac{1}{2} (V_{\text{conm}} - V_{uN})^2.$$

Para la tecnología Ψ:

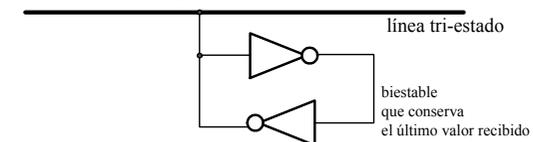
$$I_{\text{conm}} = 40 \frac{\mu\text{A}}{\text{V}^2} \cdot 1,5 \cdot 0,5 \cdot (2,1 - 1)^2 \text{ V}^2 = 36 \mu\text{A}$$

$$P = 180 \mu\text{W}$$

Un inversor, cuya entrada se encuentre en situación de alta independencia (tri-estado) o con un valor de tensión intermedio ($\sim V_{\text{conmutación}}$) genera una disipación de potencia del orden de 0,2 mW y el correspondiente calentamiento del circuito (un valor unas cuatro mil veces superior al generado en la conmutación de dicho inversor a 1MHz).

Por otra parte, existe una situación peligrosa en que este pico de intensidad (debido a la conducción «momentánea» de ambos transistores en la conmutación) puede ser estable y originar un fuerte calentamiento (e incluso la destrucción) del circuito integrado: si una entrada queda en alta impedancia tiende a adoptar un valor de tensión intermedia y ambos transistores conducen.

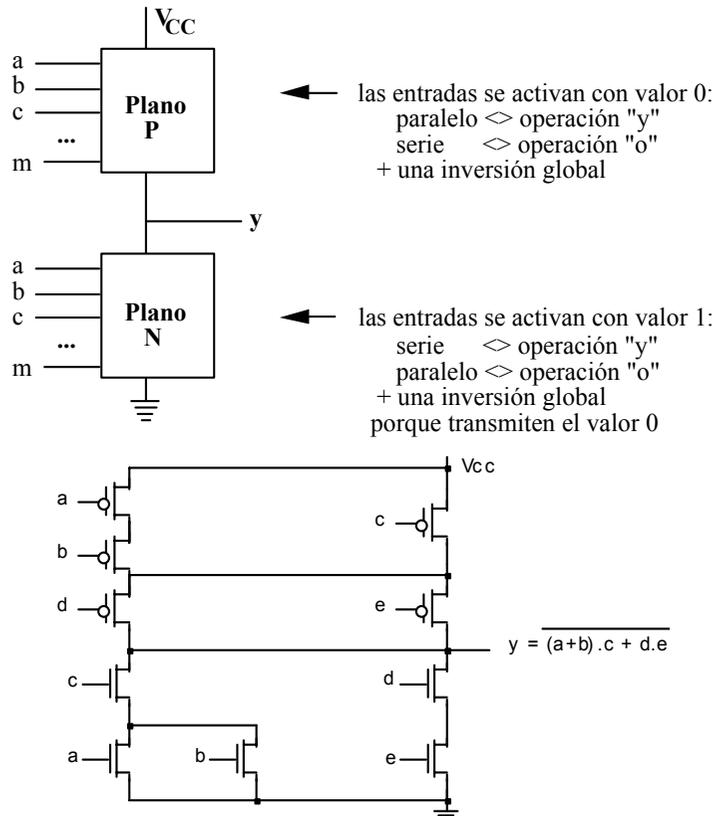
Esta situación puede darse en caso de utilizar buses (o adaptadores tri-estado) internos que puedan quedar en alta impedancia, posibilidad que puede evitarse conectando en cada línea una resistencia de alto valor, que la referencie a 0 V (*pull-down*), o bien un pequeño biestable (un par de inversores) que mantenga el último valor booleano establecido en ella (ver figura).



Conjuntos de circuitos auxiliares de este tipo, cuya finalidad es que las líneas de buses no se queden en estado de alta impedancia, se encuentran también disponibles en los catálogos de circuitos integrados.

T7.4. Puertas CMOS

Haciendo uso del álgebra de conmutadores, las puertas lógicas CMOS se construyen mediante dos planos «duales» de transistores: **plano N** y **plano P**; a cada entrada le corresponden sendos transistores, uno en el plano N y otro en el plano P, conectados de acuerdo con la dualidad serie-paralelo.



- en su plano N, la configuración de transistores corresponde (de acuerdo con el álgebra de conmutadores) a la función $(a + b).c + d.e$, a la cual debe añadirse una negación global ya que dicho plano N transmite el valor 0; de forma que la función conformada por el plano N de la figura anterior es: $y = (a + b).c + d.e$
- en cuanto al plano P, transmite el valor 1 pero sus transistores conducen cuando su entrada es 0, es decir, corresponden a variables negadas (\bar{a} , \bar{b} , \bar{c} , \bar{d} , \bar{e}) y su configuración (de acuerdo con el álgebra de conmutadores) conforma la función $(\bar{a} . \bar{b} + \bar{c}) . (\bar{d} + \bar{e})$; de manera que la función conformada por el plano P de la figura es la misma que la del plano N: $y = (\bar{a} . \bar{b} + \bar{c}) . (\bar{d} + \bar{e}) = (a + b).c + d.e$

Características físicas de las puertas complementarias

Permiten una amplia diversidad de puertas

El juego de conexiones serie-paralelo permite configurar funciones complejas y muy diversas en una misma puerta, con la limitación de que la expresión algebraica de la función ha de presentar una negación global sobre el conjunto de operaciones booleanas (son puertas inversoras) y de que tal negación sea la única que aparezca: cada negación da lugar a una puerta adicional.

Los transistores P y N presentan diferente resistividad (debida a la menor movilidad de los huecos): la conexión PMOS en paralelo, y sus correspondientes NMOS en serie, contribuye a compensar la asimetría, mientras que la conexión de PMOS en serie acentúa dicha asimetría y su efecto resistivo; por ello, son preferibles las puertas "**y-negada**" (*Nand*) a las puertas "**o-negada**" (*Nor*).

Las entradas son de tipo capacitivo

Los transistores presentan una capacidad de puerta, que es preciso cargar o descargar en la conmutación cuando cambia el valor booleano presente en la correspondiente entrada; tal transitorio de carga o descarga:

- da lugar a unos tiempos de conmutación que limitan su velocidad de trabajo;
- limita, asimismo, el *fan-out* de la puerta anterior, es decir, el número de entradas de otras puertas que pueden conectarse sobre una salida (pues tal número condiciona la velocidad de trabajo);
- requiere un aporte puntual de intensidad durante la conmutación, que da lugar a un consumo dinámico proporcional a la frecuencia de conmutaciones;
- genera «ruido» sobre las líneas de alimentación como consecuencia del «pulso» de intensidad necesario para la conmutación.

La salida de cada puerta es resistiva

Cada plano de transistores, cuando conduce, presenta una resistencia relativa a la zona lineal u óhmica de sus transistores; tal resistencia depende de las dimensiones de los transistores (disminuyendo en proporción inversa a la anchura de estos) y afecta a:

- la intensidad suministrable por la puerta;
- los procesos de conmutación (carga y descarga de las capacidades de entrada de las puertas siguientes) y, en consecuencia, los tiempos de propagación y la velocidad;
- la inmunidad frente al «ruido» en términos de potencia.

Se presentan varios transistores en serie

La presencia de transistores en serie supone un aumento de la resistencia de salida de la puerta, que repercute, en concordancia con el apartado anterior, en la intensidad suministrable por la puerta, en los tiempos de propagación y la velocidad de trabajo y en la inmunidad frente al «ruido» en términos de potencia.

En cuanto a intensidades en la salida de puertas con más de una entrada, ha de tenerse en cuenta aquella situación booleana en que conducen varios transistores en serie, en cuyo caso la intensidad suministrable ha de dividirse, al menos, por el número de ellos (ya que se suman sus resistencias equivalentes) o alternativamente ha de aumentarse en igual proporción la anchura de tales transistores.

También el margen de ruido en potencia disminuye por la suma de resistencias de transistores en serie: $\Delta P = (\Delta V)^2 / R_O$.

En los tres aspectos considerados (intensidad suministrable, tiempos de propagación e inmunidad en potencia) interviene la resistencia de salida: $R_O = \Sigma R_O$ (transistores en serie). En una primera aproximación, para n transistores se multiplica por n la resistencia de salida correspondiente a uno sólo de ellos; pero, además, se produce un efecto de «desplazamiento de la tensión de fuente» que aumenta progresivamente la resistencia de los diversos transistores.

Consideremos el caso de transistores NMOS en serie, todos ellos con tensión de entrada $V_G = V(1)$, solamente el inferior de ellos tiene su fuente conectada a 0 V y su tensión $V_{GS} = V(1)$. Mientras que la tensión de salida sea $V_o > 0$ (lo cual ocurre siempre que $I_o \neq 0$ y, en particular, durante la conmutación), la tensión puerta-fuente de los demás transistores es menor $V_{GS} < V(1)$ pues su terminal de fuente no está conectado a 0 V directamente, sino a través de los transistores que se encuentran debajo; por ello, dicha tensión V_{GS} disminuye al ascender en la serie y la resistencia efectiva que presenta cada transistor es mayor cuanto más alejado se encuentra de la conexión a 0 V.

Por ello, puertas con muchas entradas (muchos transistores en serie) presentan malas características funcionales: es altamente aconsejable limitar el número de entradas de las puertas CMOS, de forma que no aparezcan más de 6 transistores en serie.

El consumo estático es siempre nulo

Para cada valor booleano en una de las entradas, uno de sus transistores se encontrará en corte y el otro conducirá: todo camino de conducción entre los dos terminales de alimentación (V_{CC} y 0 V) incluye siempre un transistor en corte, por lo cual el consumo en reposo es nulo. En cambio, sí que hay consumo dinámico originado por la carga o descarga de las diversas capacidades propias de los transistores en la conmutación y dicho consumo es proporcional a la frecuencia de conmutación.

Tiempos de conmutación

Al pasar de un inversor a una puerta booleana de dos o más entradas, los tiempos de propagación aumentan en aquellos procesos (carga o descarga) en que la conducción se produce a través de varios transistores en serie; es un efecto de suma de sus resistencias.

Para la tecnología Ψ :

- NAND de 2 entradas: $t_{pLH} = V_{CC} \cdot C / I(P) \approx 5 \text{ V} \cdot 2 \text{ fF} / 0,2 \text{ mA} \approx 0,05 \text{ ns}$,
 $t_{pHL} = 2 \times V_{CC} \cdot C / I(N) \approx 2 \cdot 5 \text{ V} \cdot 2 \text{ fF} / 0,5 \text{ mA} \approx 0,04 \text{ ns}$
- NOR de 2 entradas: $t_{pLH} = 2 \times V_{CC} \cdot C / I(P) \approx 2 \cdot 5 \text{ V} \cdot 2 \text{ fF} / 0,2 \text{ mA} \approx 0,1 \text{ ns}$,
 $t_{pHL} = V_{CC} \cdot C / I(N) \approx 5 \text{ V} \cdot 2 \text{ fF} / 0,5 \text{ mA} \approx 0,02 \text{ ns}$

Debido a la diferente resistividad de los transistores PMOS y NMOS, interesa más la utilización de puertas "y-negada" (*Nand*) pues en ellas se equilibran un poco los dos tiempos de conmutación (aumenta el menor de ambos), mientras que para las puertas "o-negada" (*Nor*) se aumenta aún más el de subida (que es, ya de por sí, el mayor).

Ahora bien, cuando se desea obtener estimaciones cuantitativas de los tiempos de propagación de una puerta o de un conjunto de puertas booleanas es necesario acudir a la simulación eléctrica (SPICE u otros). Por un lado, los efectos relativos a transistores en serie (comentados en la página anterior) y, de otro, el solapamiento entre los tiempos de conmutación de puertas sucesivas (ya que cada puerta inicia su conmutación antes de haber finalizado la conmutación de la puerta anterior) limita, en gran medida, los resultados obtenidos a través de modelos simplificados o de razonamientos cualitativos.